

Manual de linguística portuguesa

MRL 16

Manuals of Romance Linguistics

**Manuels de linguistique romane
Manuali di linguistica romanza
Manuales de lingüística románica**



Edited by
Günter Holtus and Fernando Sánchez Miret

Volume 16

Manual de linguística portuguesa

Editado por
Ana Maria Martins e Ernestina Carrilho

DE GRUYTER

ISBN 978-3-11-037448-3
e-ISBN (PDF) 978-3-11-036884-0
e-ISBN (EPUB) 978-3-11-039432-0

Library of Congress Cataloging-in-Publication Data

A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>.

© 2016 Walter de Gruyter GmbH, Berlin/Boston

Cover image: © Marco2811/fotolia

Typesetting: jürgen ullrich typosatz, Nördlingen

Printing and binding: CPI books GmbH, Leck

♻️ Printed on acid-free paper

Printed in Germany

www.degruyter.com

Agradecimentos

As editoras do volume expressam o seu especial agradecimento aos editores da coleção MRL, Günter Holtus e Fernando Sánchez Miret, pela proposta inicial, pelo cuidado das suas leituras e pelos comentários, constantes ao longo da composição deste Manual, e a todos os autores, por terem tornado possível, além de agradável e muito estimulante, este projeto conjunto. Agradecem também a Mélanie Pereira o apoio na preparação do índice remissivo, a Wolfgang Konwitschny o cuidado na produção final do livro e a Christine Henschel o constante auxílio, incentivo e otimismo em todas as fases da produção deste Manual.

Manuais de Linguística Românica

A série *Manuais de Linguística Românica (MRL: Manuals of Romance Linguistics)* é uma novidade em termos de manuais internacionais que oferece um panorama abrangente, sistemático e atual do campo da Romanística.¹

Os *MRL* pretendem atualizar e expandir os conteúdos de duas grandes obras de referência anteriores: o *Lexikon der Romanistischen Linguistik (LRL)* (1988–2005, vol. 1–8) e a *Romanische Sprachgeschichte (RSG)* (2003–2008, vol. 1–3). Procura-se igualmente que integrem novas linhas de investigação, bem como alguns tópicos que, até hoje, não tenham sido alvo de uma exploração sistemática.

Uma vez que a revisão completa do *LRL* e da *RSG* não era um empreendimento facilmente executável, ou pelo menos não o seria em tempo útil, os editores dos *MLR* optaram por uma abordagem mais flexível, num formato por módulos.

A série vai incluir 60 volumes, aproximadamente, cada um com uma extensão de 400 a 600 páginas e com 15 a 30 capítulos. Cada volume centrar-se-á nos aspetos mais essenciais do respetivo tópico, adotando um registo claro e estruturado. Enquanto elementos de uma série, os volumes vão cobrir todo o campo da Linguística Românica atual, o que não os impede de funcionarem também individualmente. Dado que a tarefa de produzir cada um dos volumes dos *MRL* é significativamente mais leve do que seria a produção de uma grande obra de referência ao estilo do *LRL*, torna-se também mais fácil considerar aqui os modernos desenvolvimentos e sentidos de investigação em linguística, mesmo os muito recentes.

As línguas de publicação dos *MRL* são o francês, o espanhol, o italiano, o inglês e, em casos excecionais, o português. Cada volume será escrito, de maneira consistente, em uma só dessas línguas, sendo que para cada caso a escolha da língua depende do tópico selecionado. Nesse sentido, o inglês é a língua dos volumes com tópicos de relevância mais geral, que ultrapassam o âmbito da Romanística, como é o caso do *Manual of Language Acquisition* ou do *Manual of Romance Languages in the Media*.

Cada volume tem um de dois focos possíveis: ou (1) uma língua em particular ou (2) um campo específico de investigação. Os volumes que pertencem ao primeiro tipo, *i.e.*, que dizem respeito a cada uma das línguas românicas, incluindo os crioulos de base românica, têm existência individual. Haverá a preocupação em dar destaque às línguas mais pequenas (*linguae minores*), até hoje tratadas só superficialmente por outras obras de referência. Os *MRL* terão, assim, volumes dedicados ao friulano, ao corso, ao galego e ao latim vulgar, entre outras línguas. Haverá também um *Manual of Judaeo-Romance Linguistics and Philology*. Os volumes do segundo tipo reservam-se para a apresentação sistemática de todas as áreas, as tradicionais e as modernas, da Linguística Românica. Quanto aos seus métodos, esses serão discutidos num volume

¹ Os editores da coleção e a editorial De Gruyter agradecem à colega Rita Marquilhas pela tradução desta introdução a partir da versão inglesa.

em separado. As linhas e os campos de investigação mais novos e mais dinâmicos, mais uma vez, revestirão um interesse muito particular, até porque as obras de referência mais antigas não os trataram, ou só o fizeram superficialmente, em contraste com o que acontece na investigação e no ensino da atualidade. Assim, os *MRL* incluirão volumes como estes: *Grammatical Interfaces*, *Youth Language Research*, *Urban Varieties*, *Computational Linguistics*, *Neurolinguistics*, *Sign Languages* ou *Forensic Linguistics*. Cada volume vai conter uma visão panorâmica estruturada, informativa e acessível da história e dos desenvolvimentos recentes de cada uma destas áreas.

Foi-nos muito grato termos podido contar com a colaboração de colegas de craveira internacional, oriundos de uma grande variedade de países de língua românica, e não só, que aceitaram colaborar na série e editar volumes individuais dos *MRL*. A série vai conseguir sintetizar o estado atual do conhecimento em Linguística Românica e apresentar muita informação nova e muitas descobertas recentes precisamente graças à competência dos editores dos volumes individuais, que os conceberam, estruturaram e desenvolveram por meio de convites que decidiram endereçar aos autores mais apropriados.

No seu conjunto, a série dos *MRL* apresentará um panorama simultaneamente abrangente e atual da disciplina, oferecendo a qualquer leitor um leque de informações interessantes e pertinentes, orientações úteis, uma cobertura detalhada de tópicos muito específicos e uma visão de conjunto do que é hoje a Linguística Românica. Acreditamos que a série é capaz de oferecer uma abordagem refrescante e inovadora, ao nível do que tem sido o avanço constante da nossa disciplina.

Günter Holtus (Lohra/Göttingen)

Fernando Sánchez Miret (Salamanca)

Julho 2016

Índice

Ana Maria Martins

- 1 **Introdução: O português numa perspetiva diacrónica e comparativa — 1**

Parte 1: Visões gerais

Tjerk Hagemeijer

- 2 **O português em contacto em África — 43**

Hugo C. Cardoso

- 3 **O português em contacto na Ásia e no Pacífico — 68**

Celeste Rodrigues

- 4 **Variação sociolinguística — 98**

Ana Isabel Mata e Helena Moniz

- 5 **Prosódia, variação e processamento automático — 116**

Maria Antónia Mota

- 6 **Morfologia nas interfaces — 156**

Raquel Amaro e Sara Mendes

- 7 **Lexicologia e linguística computacional — 178**

João Paulo Silvestre

- 8 **Lexicografia — 200**

Amália Mendes

- 9 **Linguística de corpus e outros usos dos corpora em linguística — 224**

Rita Marquilhas e Iris Hendrickx

- 10 **Avanços nas humanidades digitais — 252**

João Costa, Maria João Freitas e Anabela Gonçalves

- 11 **Linguística clínica: alguns dados sobre o português — 278**

Telmo Mória

- 12 **Semântica e pragmática — 308**

Armanda Costa

- 13 Psicolinguística e ciência cognitiva — 336**

Parte 2: Tópicos de sintaxe, semântica e fonologia

Ana Maria Martins e João Costa

- 14 Ordem dos constituintes frásicos: sujeitos invertidos; objetos antepostos — 371**

Ana Maria Martins

- 15 A colocação dos pronomes clíticos em sincronia e diacronia — 401**

João Costa, Alexandra Fiéis e Maria Lobo

- 16 A aquisição dos pronomes clíticos no português L1 — 431**

Inês Duarte, Ana Lúcia Santos e Anabela Gonçalves

- 17 O infinitivo flexionado na gramática do adulto e na aquisição de L1 — 453**

Maria Lobo

- 18 O gerúndio flexionado no português dialetal — 481**

Madalena Colaço

- 19 Especificidades das estruturas de coordenação: padrões de concordância — 502**

Anabela Gonçalves, Ernestina Carrilho e Sandra Pereira

- 20 Predicados complexos numa perspetiva comparativa — 523**

Maria Lobo

- 21 Sujeitos nulos: gramática do adulto, aquisição de L1 e variação dialetal — 558**

Ana Maria Martins

- 22 O sistema responsivo: padrões de resposta a interrogativas polares e a asserções — 581**

Rui Marques

- 23 O modo conjuntivo — 610**

João Veloso

- 24 O sistema vocálico e a redução e neutralização das vogais átonas — 636**

Maria João Freitas

25 A sílaba na gramática do adulto e na aquisição de língua materna — 663

Lista de autores — 689

Índice remissivo — 693

Ana Maria Martins

1 Introdução: O português numa perspetiva diacrónica e comparativa

Abstract: Depois de uma breve apresentação da estrutura do *Manual de Linguística Portuguesa*, identificam-se neste capítulo as principais mudanças linguísticas que, ao longo do tempo, reconfiguraram a língua portuguesa, nos planos da fonologia, da morfologia, da sintaxe e da semântica lexical (ou outras vertentes da evolução do léxico). Em articulação com a perspetiva diacrónica, mostra-se, de forma ilustrativa e com remissões para outros capítulos do *Manual*, que consequências tiveram esses processos de mudança linguística relativamente à variação dialetal e sociolinguística do português europeu, à definição de áreas geolinguísticas em território português e ao afastamento ou aproximação entre o português e outras línguas românicas ou entre o português europeu e o português brasileiro.

Keywords: diacronia, variação dialetal, gramática histórica, fonologia e morfologia, sintaxe e semântica

1 Prólogo

O *Manual de Linguística Portuguesa* oferece uma visão de conjunto, mas também focada em aspetos particulares, do estado atual da investigação em linguística portuguesa, combinando abordagens panorâmicas com análises em profundidade de tópicos selecionados. O manual articula de forma integrada dados sincrónicos e diacrónicos, do português padrão e do português dialetal, da gramática do adulto e da aquisição de língua materna, com o objetivo de permitir um conhecimento abrangente de propriedades gramaticais características do português, usando os métodos, teorias e orientação comparativa da linguística contemporânea. Uma vez que a coleção *Manuals of Romance Linguistics* (MRL) incluirá um volume especificamente dedicado ao português brasileiro (MRL 21), o presente volume (MRL 16) tem por objeto o português europeu, com referências ocasionais ao português brasileiro.¹

Os capítulos panorâmicos do *Manual* cobrem áreas da linguística que registaram um crescimento significativo nas últimas décadas, no âmbito da investigação em linguística portuguesa, tais como o contacto de línguas (particularmente em África e na Ásia), os estudos de prosódia, as abordagens computacionais do léxico, as huma-

¹ A investigação para esta publicação foi suportada pela Fundação para a Ciência e a Tecnologia (FCT), no âmbito do projeto WOChWEL (PTDC/CLE-LIN/121707/2010) e do financiamento UID/LIN/00214/2013 ao Centro de Linguística da Universidade de Lisboa.

nidades digitais, a linguística de corpus, a linguística clínica e a ciência cognitiva, a par de áreas mais tradicionais mas com novos desenvolvimentos, como a morfologia, a semântica e a pragmática. A área da sociolinguística tem sido pouco cultivada na esfera do português europeu (em contraponto flagrante com o português brasileiro – uma interessante questão para a historiografia linguística); o capítulo dedicado à variação sociolinguística revela que passos têm sido dados para alterar esta tendência. Outras áreas da linguística teriam pleno cabimento na parte do *Manual* dedicada às visões panorâmicas (Parte I), tendo em conta o seu dinamismo e impacto na investigação contemporânea em linguística portuguesa. A decisão de não incluir no *Manual* capítulos panorâmicos sobre a aquisição da linguagem e os crioulos de base lexical portuguesa deveu-se ao facto de a coleção incluir volumes especificamente dedicados à aquisição da linguagem (MRL 2) e às línguas crioulas (cf. prefácio). Mas a aquisição do português, como língua materna, tem uma presença forte na segunda parte do *Manual* (Parte II) e a referência aos crioulos de base lexical portuguesa não está ausente dos capítulos sobre o português em contacto (capítulos 2 e 3). Estes capítulos formam um par (África e Ásia/Pacífico) e não um trio (África, Ásia/Pacífico, Brasil) porque as questões de contacto linguístico no Brasil terão, com certeza, amplo espaço de debate no volume dedicado ao português brasileiro (MRL 21). As situações de contacto noutras áreas geográficas ou não têm o mesmo grau de relevância ou não se encontram suficientemente estudadas. Naturalmente, poderiam ter sido feitas escolhas diferentes e, por certo, haverá quem aponte a falta de matérias que teriam cabimento no *Manual*. Mas a situação seria a mesma ainda que as opções fossem outras, tendo em conta que o espaço atribuído a cada um dos volumes da coleção *Manuals of Romance Linguistics* não é ilimitado. A falta de um capítulo dedicado à Fonética Forense, que queríamos ter tido neste Manual, foi um acidente de percurso.

Os capítulos da Parte II do *Manual* cobrem tópicos particulares de sintaxe, semântica, fonologia e suas interfaces com outros domínios da gramática (também aqui uma escolha entre outras possíveis). Estes capítulos apresentam o estado da arte relativamente a traços linguísticos que, pela sua especificidade, são fortemente caracterizadores do português europeu (padrão ou dialetal), tais como a colocação dos pronomes clíticos, o infinitivo flexionado, o gerúndio flexionado, os padrões de resposta a interrogativas polares, o modo conjuntivo e a redução e neutralização das vogais átonas. A par destes tópicos, incluem-se outros que são muito relevantes numa perspetiva românica e comparativa, como a ordem dos constituintes frásicos, a interação entre coordenação e padrões de concordância, os predicados complexos, o parâmetro do sujeito nulo, e a estrutura da sílaba.

Globalmente, o *Manual* dá a conhecer nova evidência empírica e análises fundadas em teorias linguísticas contemporâneas, acolhendo abordagens com inspiração teórica diversa e fundadas em dados não limitados ao português padrão contemporâneo. Os capítulos baseiam-se, por isso, em dados empíricos de diferentes tipos (intuitivos, experimentais, extraídos de corpus, obtidos em trabalho de campo) e

procuram combinar descrições consistentes, e acessíveis a um público que vá para além do universo dos linguistas, com as visões teóricas que emergiram da investigação contemporânea em linguística.

2 Diacronia e variação

Na introdução que se segue será apresentada uma síntese da evolução da língua portuguesa ao longo do tempo, com referência às implicações da mudança linguística para a variação dialetal dentro do português europeu e aos percursos de afastamento ou aproximação entre o português e outras línguas românicas, bem como entre português europeu e português brasileiro. Esta apresentação não será exaustiva, implicando antes a seleção dos traços linguísticos considerados mais relevantes para caracterizar a diacronia do português, nos planos da fonologia, da morfologia, da sintaxe e da semântica lexical em articulação com a sintaxe, incluindo ainda uma breve referência ao léxico. Procurar-se-á assim oferecer uma visão de conjunto das principais mudanças linguísticas do português (europeu) e, ao mesmo tempo, introduzir as referências bibliográficas relevantes para que o leitor possa encontrar informação diacrónica relativamente a tópicos que são tratados nos restantes capítulos do *Manual* numa perspetiva apenas sincrónica (por exemplo: certos aspetos da morfologia e da variação sociolinguística, o infinitivo flexionado, os predicados complexos, o sistema vocálico ou a estrutura da sílaba). A bibliografia específica que irá sendo introduzida, deverá ser completada por bibliografia de âmbito mais geral sobre a história da língua (cf. Brocardo 2014; Brocardo/Lopes 2016a; 2016b; Cardeira 2005; Castro 1991; 2006; Maia 1986; Marquilhas 2000; Martins 2002a; 2003; Silva 1989; 1991; 1994b; 2008a; 2008b; Teyssier 1982; Ali ⁷1971; Câmara 1975; Huber 1986; Nunes ⁸1975; Williams ³1975; ↗15 A colocação dos pronomes clíticos em sincronia e diacronia) ou bibliografia sobre a variação dialetal do português (cf. Cintra 1983; Segura 2013; Álvarez Pérez 2014; 2015; Álvarez Pérez/Saramago 2012; Brissos 2014; 2015; Brissos/Saramago 2014; Carrilho 2008; 2009; Carrilho/Pereira 2011; 2013; Costa/Pereira 2013; Florêncio 2001; Martins 2009; Martins/Saramago 1993; Pereira, Sandra 2003; 2012; 2015; Pereira, Sílvia 2014a; 2014b; 2015; Saramago 1992; Saramago/Segura 2001; Segura 1988; Segura/Saramago 1999; Sória 2013; Vasconcelos 1928; 1985; ↗4 Variação sociolinguística; ↗18 O gerúndio flexionado no português dialetal; ↗20 Predicados complexos numa perspetiva comparativa; ↗21 Sujeitos nulos: gramática do adulto, aquisição de L1 e variação dialetal).

Nos Quadro-síntese 1 a 3 e 5 usa-se, com pequenos ajustes cronológicos, a periodização da história da língua portuguesa de Castro (1999; 2006), que identifica, após o período do português pré-literário, que não será aqui considerado, os períodos designados por português antigo, português médio, português clássico e português moderno (cf. Vasconcelos ⁴1966; Ali ⁷1971; Vázquez Cuesta/Luz 1971; Bechara 1991; Silva 1994a; Maia 1995; Brocardo 2005).

Em cada um dos quadros, identificam-se na coluna correspondente ao português antigo os principais traços gramaticais que caracterizam o português dos séculos XIII e XIV (sendo raros, breves e linguisticamente complexos os textos escritos em português no último quartel do século XII – cf. Castro 2004; Martins 1999; 2007; Souto Cabo 2003). Indica-se depois em qual dos períodos subseqüentes se consolidou a mudança de que resultou o estado de língua do português contemporâneo. O facto de uma mudança particular estar assinalada, por exemplo, na coluna do português médio, não significa que as manifestações iniciais dessa mudança não possam vislumbrar-se já no período anterior nem que atestações de formas conservadoras não possam encontrar-se ainda no período seguinte. Nos casos em que estão assinaladas duas colunas (cf. Quadros 2, 3, 5 e 6), isso indica que uma dada mudança se desenvolveu em duas fases, cuja concretização ocorre em dois períodos diferentes (por exemplo, a elevação das vogais átonas finais precede a elevação das vogais pretónicas, como adiante se exporá). Pode também indicar que está em causa não uma mudança única mas um conjunto de mudanças relacionadas que se desenvolvem durante um intervalo temporal que excede os limites cronológicos de um período, de acordo com a periodização adotada (por exemplo, a evolução das conjunções, apontada no Quadro 3).

2.1 Fonologia

Quadro 1: Fonologia: Consoantes

Principais traços caracterizadores	Cronologia da mudança		
	Português Médio (até meados do séc. XVI)	Português Clássico (até meados do séc. XVIII)	Português Moderno
Sistema de quatro sibilantes, duas fricativas apico-alveolares (/ʃ/, /z/), com origem no S latino, e duas africadas predorsodentais (/ts/, /tʃ/), com origem em C e T latinos palatalizados. Por exemplo: [ʒobrãtseʎaʒ] «sobrancelhas»; [kɔʒer] «cozer» / [koʒer] «coser».	X		
Sibilantes em coda não têm realização palatal. Por exemplo: [treʃ] «três»; [feʃtaʃ] «festas»; [paʒ] «paz».		X	
Existe a oposição fonológica entre a africada palatal /tʃ/ e a fricativa palatal /ʃ/ (enquanto /dʒ/ e /ʒ/ convergem já nesta fase). Por exemplo: [kõʃja] «concha» / [koʃa] «coxa».		X	
As consoantes vibrantes, simples e múltipla, têm ambas articulação anterior (i.e. alveolar). Por exemplo: [rara] «rara».			X

A evolução fonológica do latim levou ao aparecimento de novas consoantes que, emergindo no latim vulgar ocidental, configurarão o sistema consonântico do português antigo. Dois tipos de processos assimilatórios estão na origem da ampliação do primitivo inventário consonântico do latim. Os processos de sonorização em posição intervocálica levarão à expansão da oposição surda/sonora (que no latim clássico estava reservada às oclusivas), primeiro às consoantes fricativas (/f/-/v/, /s̥/-/z/) e depois às novas consoantes africadas (/tʃ/-/dʒ/, /ts/-/dʒ/). As consoantes africadas são o produto de processos de palatalização que originaram, mais extensivamente, um conjunto de consoantes palatais (/ʃ/, /ʒ/, /tʃ/, /dʒ/, /ʎ/, /ɲ/) caracteristicamente românicas, pois o sistema consonântico do latim clássico não fazia uso do ponto de articulação palatal (exceto se /j/ for incluído na classe das consoantes).

Na diacronia do português observa-se uma simplificação do sistema consonântico do português antigo, com a transformação das africadas em fricativas e, concomitantemente, vários processos de convergência fonológica. A neutralização da oposição entre /dʒ/ e /ʒ/ acontece tão cedo que só alguns poucos textos de final do século XII e início do XIII manifestam a distinção ortograficamente (Martins 2007). A oposição fonológica entre o par de consoantes surdas (/ʃ/-/tʃ/), pelo contrário, persiste na língua padrão até ao português clássico e a distinção gráfica entre <x> e <ch>, que a ortografia portuguesa mantém, reflete essa persistência. A diferente cronologia dos processos de convergência fonológica entre africada e fricativa palatal (com a convergência das surdas a ocorrer séculos depois da convergência das sonoras) também explica que a oposição /ʃ/-/tʃ/ se mantenha numa extensa região do Norte de Portugal, sendo um dos traços caracterizadores dos dialetos portugueses setentrionais (Cintra 1971; Segura 2013), enquanto a neutralização da oposição /dʒ/-/ʒ/ não originou variação dialetal, pois estendeu-se a todo o território português. No que diz respeito às sibilantes, foi também a passagem das africadas predorsodentais /ts/-/dʒ/ a fricativas com idêntico ponto de articulação (/s/-/z/) que as tornou suficientemente próximas das fricativas apicoalveolares /s̥/-/z/ para que, em grande parte do território português, viesse a produzir-se a neutralização destas distinções fonológicas. A evolução das sibilantes produziu uma tripartição dialetal (cf. Cintra 1971; Álvarez Pérez 2014; 2015): nos dialetos centro-meridionais houve convergência fonológica a favor das predorsodentais (/s/-/z/); os dialetos setentrionais preservaram as apicoalveolares, mas fizeram-no numa área com convergência fonológica e redução do sistema a duas sibilantes (/s̥/-/z/), paralelamente ao que aconteceu no centro-sul, e noutra com conservação do sistema de quatro sibilantes, mantendo portanto a distinção entre predorsodentais e apicoalveolares (/s/-/s̥/, /z/-/z/).²

2 No castelhano, que perdeu a oposição surda/sonora para as fricativas e africadas (cf. Alonso 1962a; 1962c), não houve convergência fonológica entre /tʃ/-/ʃ/ nem entre /s/-/s̥/ porque a distinção entre estes pares de segmentos consonânticos aprofundou-se com o recuo da fricativa palatal para a zona velar (/ʃ/ > /x/) e a anteriorização da fricativa predorsodental para a zona interdental (/s/ > /θ/). Este tipo de afastamento entre segmentos foneticamente próximos ocorreu no português do século XIX

Só depois da reconfiguração do sistema de sibilantes, ocorreu a mudança que consiste na palatalização do /S/ em coda, que passou a ter uma realização diferente do /S/ em ataque (por exemplo: [sɔʃ], [sɛʃtɐ]). A palatalização parece ser um fenómeno essencialmente associado à área centro-meridional, onde o sistema de sibilantes se reduziu às predorsodontais (Teyssier 1982, 55–56), mas a variação dialetal na realização do /S/ em coda permanece por estudar.

A mais bem documentada das mudanças do consonantismo é talvez a evolução da africada palatal (↗4 Variação sociolinguística; Álvarez Pérez 2014). Ocorrendo numa época para a qual os testemunhos dos gramáticos são já abundantes, é possível acompanhar não só a sua expansão geolinguística, do sul para o centro-litoral, como também a sua progressão sociolinguística, como mostrou Pinto (1981), de onde se extrai a informação que se segue. Em 1671, o lisboeta João Franco Barreto considera a substituição de [tʃ] por [ʃ] um «mao costume» dos «rústicos do sul», referindo-se à região a sul do Tejo. Em 1739, o transmontano João de Morais Madureira Feijó (natural da região que ainda hoje tem /tʃ/) coloca já a mudança em Lisboa, entre a população alfabetizada, sem deixar de a condenar: «nasce da criação das escolas aonde assim aprendem a escrever, e pronunciar; e ficão tão habituados, que depois não há lição que os emende». Pouco depois, em 1746, o estrangeirado Luís António Verney, nascido em Lisboa em 1713, considera, contra Madureira Feijó, que a generalização de /ʃ/ deve ser a norma porque: «Em matéria de pronuncia, sempre se-devem preferir, os que sam mais cultos e falam bem na Estremadura».

A mais tardia das mudanças fonológicas registadas no Quadro 1 é o afastamento entre as vibrantes simples e múltipla, com o recuo desta para a zona uvular. Sendo uma mudança que, pelo menos na região de Lisboa, acontece só no final do século XIX, é possível ter dela um registo do processo em curso por um linguista, o foneticista Gonçalves Viana. Em 1883, na sua obra *Essai de phonétique et de phonologie de la langue portugaise d'après le dialecte actuel de Lisbonne*, Viana toma a realização recuada da vibrante por variante individual de alguns falantes de Lisboa, mas pouco mais tarde, em 1903, já testemunha a sua expansão nos meios urbanos: «On trouvera individuellement des *r* vibrantes uvulaires, même parmi les gens qui prononcent *r* simple comme une linguale» (Viana ³1973, ¹1883, 102); «La prononciation uvulaire de *rr*, mais non pas de *-r* [...] se répand de plus en plus dans les villes. Cependant, on la regarde encore comme vicieuse, le *rr* apical étant toujours préférable au grasseyement du [ʀ], qui individuellement est plus profond qu'en français ou en allemand» (Viana 1903, 19). A oposição entre vibrante simples e vibrante múltipla, de articulação alveolar, mantém-se em dialetos do português europeu mas a sua área não está delimitada e a variação poderá ser mais sociolinguística do que geográfica (↗4 Variação sociolinguística).

relativamente às consoantes vibrantes (/r/- /r/ > /r/-/ʀ/), mas neste caso o castelhano manteve a distinção entre vibrante simples e vibrante múltipla alveolares. As duas línguas divergiram ao longo do tempo neste como noutros aspetos da gramática, embora também manifestem evoluções comuns, como veremos mais adiante.

Quadro 2: Fonologia: Vogais

Principais traços caracterizadores	Cronologia da mudança		
	Português Médio (até meados do séc. XVI)	Português Clássico (até meados do séc. XVIII)	Português Moderno
Alta frequência de hiatos como resultado da síncope de consoantes intervocálicas latinas (mas iniciam-se já nesta fase alguns dos processos que conduzirão à sua eliminação). Por exemplo: [nuu] <nu>; [vīu] <vinho>; [taeʃ] <taís>; [fea] <feia>; ũa <uma>.	X		
Existem as vogais nasais finais <i>ã, õ, ĩ</i> , que virão a originar, por ditongação, os ditongos [ẽw̃] e [ẽj̃]. Por exemplo: [pã] <pão>; [digã] (> [digẽw̃]) <digam>; [nõ] <não>; [leõ] <leão>; [diʒerõ] (> [diʒerẽw̃]) <disseram>; [bẽ] (> [bẽj̃]) > [bẽj̃]) <bem>.	X		
Existem ditongos crescentes átonos finais que virão a ser eliminados por metátese ou síncope da semivogal (a mudança inicia-se neste período). Por exemplo: [koĩbrja] <Coimbra>; [ʃabja] <saiba>.	X		
Ditongos decrescentes [ej], [ow] estáveis (não há ainda processos de monotongação ou dissimilação). Por exemplo: [fejtʊ] <feito>; [powku] <pouco>.	X		
Existe o ditongo [uj], proveniente de [oj], que virá a reduzir-se a [u]. Por exemplo: [lujta] <luta>; [trujta] <truta>; [eʃkujtar] <escuta>; [frujtoʃ] <frutos>; [kujtɛlu] <cutelo>; [tʃujva] <chuva>.	X		
A variação entre [ow] e [oj] tem âmbito limitado. Por exemplo: [kowrela] / [kojrela] <courela>; [dojru] / [dowru] <Douro>	X		
Elevação condicionada das vogais pretónicas, por harmonização vocálica. Por exemplo: [firidu] <ferido>, [viludu] <veludo>, [kurtina] <cortina>, [gurdure] <gordura>.	X		
Não ocorre a elevação generalizada do vocalismo átono. Por exemplo: [temẽte] <temente>; [komer] <comer>; [abril] <abril>.	X	X	
A vogal [e] não sofre dissimilação quando antecede um segmento palatal (semivogal ou consoante). Por exemplo: [veʒu] (> [veʒu]) <vejo>; [feʃu] (> [feʃu]) <fecho>; [teʎa] (> [teʎa]) <telha>; [teɲu] (> [teɲu]) <tenho>; [dej] (> [dej]) <dei>			X

Ao nível do vocalismo, o português manteve o sistema vocálico herdado do latim vulgar ocidental, com as características distinções tímbricas entre /e/-/ɛ/ e /o/-/ɔ/ e sem a chamada ditongação românica, que em graus e formas diversas atingiu a maior parte das línguas românicas. A evolução das vogais /ě/ e /ǒ/ latinas, em posição acentuada, é tradicionalmente usada para traçar a fronteira entre as áreas galega e portuguesa, sem ditongação, a ocidente, e as áreas astur-leonesa e castelhana, com ditongação em [je] e [we] (ou variantes), a oriente. Ao longo do tempo, o sistema vocálico tônico do português manteve-se essencialmente estável. As alterações tímbricas relativamente às vogais herdadas do latim vulgar são sobretudo o resultado de processos de harmonização vocálica, que não reconfiguraram o sistema fonológico. Mudanças profundas produziram-se, no entanto, ao nível do vocalismo átono, tema incontornável (e especialmente polémico) de fonologia histórica do português (cf. Carvalho 1984a; 1984b; 1984c; Hart 1955; 1959; Maia 1986; Marquilhas 2003; 2004; Martins 1985; 2006a; 2007; Naro 1971; Révah 1958; Teyssier 1982; 24 O sistema vocálico e a redução e neutralização das vogais átonas). A redução e neutralização das vogais átonas no português europeu afastam-no do galego, com o qual partilha uma origem comum, e também do português do Brasil. Ao mesmo tempo é um dos traços do português europeu que o aproximam do catalão e do francês (partilhando com este último também a existência de vogais nasais e, nalgumas áreas dialetais particulares, outros traços do vocalismo – cf. Cintra 1971; Segura 2013; Brissos 2014; Brissos/Saramago 2014; 24 Variação sociolinguística).

Os dados geolinguísticos parecem relevantes para reconstituir a história do vocalismo átono, i.e. as diferentes fases do processo de elevação generalizada das vogais átonas que caracteriza o português europeu. A diferença entre vogais finais (com elevação) e vogais pretónicas (sem elevação) observada no português brasileiro sugere que a elevação das vogais átonas finais precede diacronicamente a elevação das vogais pretónicas. A persistência da elevação condicionada das vogais pretónicas /e/ e /o/ no português brasileiro (e.g. [mi'ninu], [kur'tine], [vi'ludu], com assimilação pela pretónica do traço [+alto] da tónica) é compatível com este cenário porque a elevação dependente de harmonização vocálica desaparece quando emerge a elevação generalizada. Por isso, as formas do tipo [fi'ridu], características do português antigo, conservam-se no galego e no português brasileiro, mas não no português europeu. Como o processo fonológico do português europeu envolve elevação e recuo, é possível admitir que o processo de elevação precede cronologicamente o recuo. De novo, o testemunho do português brasileiro é importante: [e] átono final eleva-se para [i], mas não recua para [i] (o que também acontece em dialetos do português europeu). Assim, embora os dados das fontes textuais (e gramaticais) sejam de difícil interpretação, é legítimo admitir que a elevação das átonas finais ocorre durante o período do português médio (iniciando-se talvez antes, se considerarmos que há relação entre elevação das vogais finais e metafonia nominal – cf. Alonso 1958; 1962b; Martins 1985), enquanto a elevação/recuo das pretónicas e o recuo de [i] final para [i] são mudanças do português

clássico. Dada a insuficiência da evidência empírica que pode extrair-se das fontes históricas, um melhor conhecimento do vocalismo átono do português europeu contemporâneo, em particular no que diz respeito às exceções ao processo fonológico de elevação/recoo, poderá contribuir para uma melhor compreensão da evolução diacrónica das vogais átonas (cf. Marquilhas 2003; 2004; Mateus et al. 2003; 74 Variação sociolinguística; 74 O sistema vocálico e a redução e neutralização das vogais átonas).

Globalmente, a evolução do vocalismo átono no português europeu não cria, internamente, variação dialetal. Todos os dialetos do português europeu exibem redução e neutralização do vocalismo átono e, portanto, nenhuma variedade dialetal conservou o processo de harmonização vocálica característico do português antigo, do galego e do português brasileiro (i.e. formas do tipo [mi'ninu], [vi'ludu]). No entanto, é possível encontrar no âmbito do vocalismo átono diversos fenómenos de microvariação, que não se restringem ao [i] átono final, não centralizado, de alguns dialetos meridionais (cf. 74 Variação sociolinguística).

Uma consequência da redução e neutralização do vocalismo átono no português europeu (em particular o apagamento muito comum da vogal [i]) é a tolerância generalizada de núcleos silábicos vazios, criando no plano fonético «grupos consonânticos problemáticos» (75 A sílaba na gramática do adulto e na aquisição de língua materna; Mateus/Andrade 2000). Neste aspeto é grande o contraste com o português brasileiro, ainda que a estrutura da sílaba seja basicamente a mesma nas duas variedades. O português europeu padrão apaga sistematicamente a vogal [i] em posição inicial antes de S+Consoante (embora a palatalização do /s/, como em [ʃ.kɔ.lɐ], indique que ocupa a posição de coda de uma sílaba com núcleo vazio), não introduz vogais epentéticas em palavras como *pneu*, *psicologia*, *adaptar*, *advogado*, *pseudónimo*, nem paragógicas em palavras como *stress*, *snob*, *strip*, *team* (cf. [tʃi.mi] no português brasileiro), nem protéticas em palavras como *stress*, *snob*, em contraste com o português brasileiro. O caso de *stress* e *snob* é particularmente interessante pois nestas palavras o /s/ inicial não palataliza no português europeu, o que mostra que não há um núcleo vazio a preceder o /s/ mas sim, de facto, o grupo consonântico S+Consoante. Ou seja, a inserção de vogal protética que se produziu em idênticas condições na evolução do latim para o português antigo deixou de se produzir no português europeu mas não no português brasileiro. Por exemplo: latim *SCHOLA* > português antigo [eʃkɔla]; inglês *stress* > português europeu [stres], português brasileiro [is.tɾɛ.si], grafado *estresse* (Houaiss/Villar/Franco 2001).

Relativamente à estrutura da sílaba, ocorreram mudanças significativas na evolução do latim para o português. Houve simplificação de rimas, ataques e núcleos ramificados (75 A sílaba na gramática do adulto e na aquisição de língua materna), tendo a simplificação das rimas reduzido drasticamente o número e a natureza das consoantes em coda e conduzido à eliminação das chamadas consoantes geminadas. Por outro lado, o processo de enfraquecimento e síncope de consoantes intervocálicas criou numerosos hiatos, ou seja, teve um efeito contrário à simplificação da estrutura

silábica pois originou sílabas com ataques vazios (e.g. [mala] > [ma.a] > [ma]; [regina] > [re.ĩ.a] > [reĩne]; [êgo] > [ε.o] > [ew]; [una] > [ũ.a] > [ume]). Seguir-se-á, diacronicamente, um processo geral de eliminação dos hiatos que se concretiza através de diversas estratégias (mas não apaga da língua todos os hiatos, mantendo-se os que incluem a vogal *a* e uma vogal tónica alta: *lua*, *tia*, *rainha*, *Raul*). Esse processo estende-se por um período longo que vai do português antigo até ao início do período clássico, pois diferentes tipos de hiatos evoluíram de diferentes formas. Simplificadamente, pode dizer-se que os processos de crase entre vogais idênticas, de inserção da consoante nasal palatal [ɲ] depois de [ĩ] e de semivocalização de vogal átona [-baixa] se iniciam cedo, enquanto a ditongação de [e] e [o] tónicos (a segunda só nos dialetos setentrionais onde se diz [bowe]/[bewe]) acontece mais tarde. A forma [üe], preservada por analogia com [ũ] (em contraste, por exemplo, com *lũa* > *lua*), parece ser a última a ganhar a consoante nasal labial que preenche o ataque vazio da segunda sílaba (*uma*), pelo que *ũa* aparece ainda nos textos clássicos (e está preservada dialetalmente). A evolução dos hiatos originou variação dialetal no território português, quer em relação à eliminação ou conservação de tipos particulares de hiatos quer no que diz respeito à interação entre ditongação de [é]/[ó] em hiato e monotongação dos ditongos [ej]/[ow]. Há informação valiosa sobre esta matéria nas obras de Leite de Vasconcelos (Vasconcelos 1928; 1985; Florêncio 2001), mas está por realizar uma investigação sistemática da variação dialetal decorrente dos processos de evolução dos hiatos do português antigo.

Os ditongos crescentes átonos finais do português antigo foram eliminados por síncope da semivogal ([koĩbrja] > [koĩbra]) ou por metátese ([ʃabja] > [ʃajba]), mas surgiram novos ditongos crescentes, também em posição tónica, como resultado da semivocalização de vogais em hiato, ([koĩbra] > [kwĩbra]; [nomear] > [numjar]). Comparativamente aos ditongos decrescentes, os ditongos crescentes são pouco frequentes no português e de realização instável, o que aponta para que o processo de semivocalização de vogais em hiato seja superficial, restringindo-se ao plano fonético (↗25 A sílaba na gramática do adulto e na aquisição de língua materna).

O português antigo conservava todos os ditongos decrescentes herdados do latim, tanto o ditongo [ow], proveniente de [aw] do latim clássico, como os ditongos [ej], [uj], e outros, surgidos no latim vulgar através dos processos de semivocalização de consoante em coda ([lak.te] > [lajte] > [lejte]; [frük.tu] > [frojtʉ] > [frujtʉ]) ou de metátese ([primariu] > [primarju] > [primajru] > [primejru]; [plüwĩa] > [plovja] > [tʃojva] > [tʃujva]). Neste aspeto, o português antigo contrasta com a maior parte das línguas românicas coetâneas. Na Península Ibérica, a conservação dos ditongos [ow] e [ej] caracterizava o galego, o português e o leonês ocidental face às restantes línguas ibéricas. No decurso da história do português ocorreu a monotongação dos ditongos [ow], [ej] e [uj] (de [oj], com harmonização de altura), o que corresponde sempre a uma inovação do Sul, com maior ou menor força expansiva, e originou variação dialetal no território português. O diferente traçado das isófonas que separam, em cada um dos casos, a área de monotongação da área de conservação do ditongo

explica que a monotongação de [ej] seja um fenómeno dialetal, meridional, enquanto a redução de [ow] a [o] e de [uj] a [u] integram a língua padrão (cf. Cintra 1971; Álvarez Pérez 2014; 2015, no que diz respeito a [ej] e [ow]; cf. Boléo/Silva 1974; Vasconcelos 1901; 1928; 1985; relativamente à persistência de [uj] em dialetos setentrionais). Estes processos de monotongação fizeram o português divergir do galego (que mantém os primitivos ditongos), embora neste como noutros casos não haja separação entre o galego e o português setentrional (é também assim, por exemplo, em relação à inexistência da consoante /v/; ↗4 Variação sociolinguística; Cintra 1971; Álvarez Pérez 2014).

Outras mudanças ocorreram que afetaram os ditongos decrescentes sem lhes alterar a natureza de núcleos ramificados. A variação entre os ditongos [ow] e [oj], que tinha âmbito muito limitado no português antigo, veio a expandir-se através do léxico, exibindo uma evolução cuja base pode ser a dissimilação, a analogia ou ambas (cf. Cintra 1970). Uma mudança muito mais tardia mas claramente dissimilatória é a que afasta entre si os segmentos palatais do ditongo [ej] através do recuo da vogal [e] para [ɛ], originando [eɛ]. Um processo semelhante, mas exclusivamente dialetal e por isso impossível de datar, é o afastamento entre os segmentos labiais do ditongo [ow] através da perda do arredondamento da vogal, que passa de [o] a [ɐ], originando o ditongo [ɛw] característico de dialetos setentrionais. Estes processos de dissimilação podem ser vistos como estratégias de preservação dos ditongos, evitando a tendência para a monotongação.

A transformação de [e] em [ɛ] dá-se não só antes de [j], mas também antes de consoante palatal heterossilábica, sendo esta, parece, a mudança fonológica mais tardia no âmbito do vocalismo. Teyssier (1982, 64–65) situa-a no século XIX e regista, citando José Inácio Roquete, autor de um *Código do Bom Tom* (Paris, 1845), que em meados do século XIX teria ainda um estatuto sociolinguístico desprestigiante: «É muito frequente entre a gente ordinária de Lisboa mudar o *e* em *a* nalgumas palavras: dizem *panha*, *lanha* por *penha*, *lenha*». Mas em 1883, numa carta dirigida a Leite de Vasconcelos, Gonçalves Viana observa que toda a gente em Lisboa diz assim, à exceção de «algum caturra velho» (Teyssier 1982, 65). Ao contrário do que poderia sugerir a sua cronologia tardia, a dissimilação de [e] antes de palatal é um fenómeno com bastante expansão no território português, não sendo estranho ao português setentrional (↗4 Variação sociolinguística).

Por fim, e sem seguir a sequência cronológica das mudanças (os quadros-síntese são aliás, neste aspeto, apenas indicativos), cabe referir a ditongação das vogais nasais finais. Este processo de ditongação, exclusivo do português entre as línguas românicas, originou o ditongo decrescente [ẽw̃] a partir das vogais nasais finais [õ] e [ã] e o ditongo [ẽj̃]/[ẽj̃] a partir da vogal nasal final [ẽ]. O complexo quadro de variação dialetal neste domínio (com as terminações átonas a escaparem frequentemente à ditongação, alguns dialetos a preservarem o ditongo [õw̃] e resultados paralelos para os desenvolvimentos de [ã]/[õ]/[ẽ]) permite confirmar que o mesmo processo fonológico atua sobre as três vogais nasais [-altas], [ã], [õ] e [ẽ], transformando-as em

ditongos decrescentes nasais (Martins 1995; Carvalho 1989). A convergência em *-ão* das terminações do português antigo [ã], [õ] e [ã.u] (um hiato que evolui por semi-vocalização) não tem paralelo nas formas do plural (i.e. [ã.eʃ] > [ẽʃʃ] / [õ.eʃ] > [õʃʃ] / [ã.oʃ] > [ẽũʃ]), do que decorre a irregularidade da formação do plural das palavras terminadas em *-ão* (↗6 Morfologia nas interfaces). Processos vários de regularização analógica, tanto na flexão nominal (e.g. *cidadãos/cidadões; grãos/grães*) como verbal (e.g. *fizerô/andô; fazī/andī*, presente do indicativo), criam variação dialetal e socio-linguística (cf. Vasconcelos 1928; 1985; Florêncio 2001; Martins 1995; Parkinson 2002).

2.2 Morfologia

Quadro 3: Morfologia

Principais traços caracterizadores	Cronologia da mudança		
	Português Médio (até meados do séc. XVI)	Português Clássico (até meados do séc. XVIII)	Português Moderno
O sistema de demonstrativos integra formas simples (<i>este, esse</i>) e formas reforçadas (<i>aqueste, aquesse, aquele</i>). Nas formas neutras, há variação do tipo <i>esto/isto</i> , sendo as variantes com [e] mais comuns.	X		
No sistema de possessivos existem formas fracas do feminino (<i>ma, ta, sa</i>) a par das formas fortes (<i>mãa, tua, sua</i>), mas a simplificação do sistema está já em curso.	X		
Existe o pronome relativo e interrogativo <i>u</i> 'onde' (cognato do francês <i>où</i>), que coexiste com <i>onde</i> 'de onde'.	X		
O sistema pronominal integra os pronomes oblíquos <i>i, en/ende</i> (cognatos do catalão <i>hi, en</i> , do francês <i>y, en</i> , do italiano <i>vi, ne</i>).	X		
O sistema de dêiticos locativos é bipartido: <i>aqui/ali; acá/alá; acó/aló</i> .	X		
As formas verbais de segunda pessoa do plural apresentam a desinência de pessoa/número <i>-des</i> , ficando <i>d</i> em posição intervocálica quando antecedido da vogal temática: <i>-ade(s), -ede(s), -ide(s)</i> .	X		
O particípio passado dos verbos da segunda conjugação tem terminação <i>-udo</i> (<i>-ido</i> ocorre raramente).	X		
Têm pouca expressão os processos de regularização analógica (analogia gramatical), tanto no âmbito da flexão verbal como da flexão nominal.	X		

Principais traços caracterizadores	Cronologia da mudança		
	Português Médio (até meados do séc. XVI)	Português Clássico (até meados do séc. XVIII)	Português Moderno
A classe gramatical das conjunções integra itens que virão a desaparecer (por ex.: <i>ca, pero, mas pero, pero que, ergo, empero (que), macar (que), segundo que/como, sol (que), mentre (que)</i>) ou que evoluirão semanticamente (por ex.: <i>pero</i> e <i>poren</i> , de explicativos para adversativos; <i>pois</i> , de temporal para explicativo).	X	X	
As formas verbais de segunda pessoa do plural e o pronome nominativo <i>vós</i> são plenamente produtivos.			X
O pronome pessoal oblíquo <i>si</i> é sempre reflexivo (nem <i>si</i> nem <i>consigo</i> são usados como formas de tratamento).			X
O clítico dativo <i>lhi/lhe</i> está em variação com <i>lhis/lhes</i> como forma de plural.			X
O pretérito-mais-que perfeito simples, o condicional e o futuro simples são plenamente produtivos.			X

Integravam a gramática do português antigo algumas formas pronominais que vieram a desaparecer. Globalmente, pode dizer-se que o sistema pronominal se simplificou, processo que se iniciou logo no português antigo. Por exemplo, a distinção entre o pronome acusativo *me* e o pronome dativo *mi* não se encontra senão nos primeiros textos, como o Testamento de D. Afonso II, de 1214. Neste mesmo texto regista-se variação, relativamente aos pronomes demonstrativos, entre a forma simples *este* e a forma reforçada *aqueste*. Este tipo de variação estendia-se ao par *esse/aquesse*, enquanto a forma de terceira pessoa era invariavelmente *aquiele*. As formas *aqueste* e *aquesse* (que têm paralelo noutras línguas românicas, tais como o catalão e dialetos do italiano), são já no português antigo menos frequentes do que as formas *este* e *esse*; deixam de ocorrer nas fontes escritas a partir do século XVI, mas dialetos do português contemporâneo guardam testemunho delas (cf. Vasconcelos 1901; 1985; Segura 2013). Tal como o sistema dos demonstrativos, também o dos pronomes possessivos se simplificou, com o desaparecimento das formas átonas *ma, ta, sa*. No português antigo a distinção entre possessivos fortes e possessivos fracos encontrava-se limitada às formas do feminino e as formas fortes já ocorriam em variação com as formas fracas na posição pré-nominal. Cardeira (2005) mostra que no terceiro quartel do século XIV a forma *sua* já era mais frequente do que *sa* em posição adjetival, estabilizando-se o processo de substituição das formas fracas pelas fortes na primeira metade do século XV.

Também no decurso do português médio deixa de ser produtivo o pronome relativo e interrogativo *u* (cognato do francês *où*) embora ainda se ateste esporadica-

mente em textos do século XVI. Sobrevive dialetalmente na forma interrogativa *u-lo* (‹onde está?›), com pouca vitalidade e limitada ao extremo Norte de Portugal (Moreira 1922; Vasconcelos 1928; Silva-Villar 2001). Não chegam ao século XVI os pronomes oblíquos *i* e *en/ende* (o primeiro cognato do francês *y*, do catalão *hi* e do italiano *vi*; o segundo cognato do francês *en*, do catalão *en/ne* e do italiano *ne*). O locativo anafórico *i*, do português antigo, era um substituto de complementos verbais introduzidos pela preposição *em*. O pronome partitivo/locativo *en/ende* era um substituto de complementos verbais introduzidos pela preposição *de*. No século XV o pronome *en/ende* já não ocorre nos textos medievais e o locativo *i* integra-se no sistema dos dêiticos locativos adverbiais, deixando de ser estritamente anafórico e adquirindo a forma *aí* por analogia com o par *aqui/ali* (Teyssier 1981; Muidine 2000). Esta mudança é comum ao português, galego e castelhano. Nestas três línguas, ao contrário do catalão, francês e italiano, os pronomes fracos *i*, *en* desapareceram sem terem chegado a transformar-se em pronomes clíticos (Badia i Margarit 1947; Wanner 1991; 2014; Martins 2014a). A integração de *aí* no sistema de dêiticos locativos transforma o sistema bipartido do português antigo num sistema tripartido, com três graus de proximidade para as formas em *-i* (*aqui*, *aí*, *ali*), enquanto as formas em *-á* (que perdem o *a* inicial e originam *cá*, *lá*) preservam um sistema bipartido e as formas em *-ó* (*acó*, *aló*) desaparecem. Estas últimas mantêm-se no galego, que também tem as formas *cá*, *lá* e apresenta em relação às formas em *-i* a mesma inovação que o português.

Outras modificações do sistema pronominal ocorrerão séculos mais tarde, sem relação com as do português medieval. O desuso do pronome nominativo *vós*, substituído por *vocês*, assim como a emergência do pronome *a gente*, sem substituir *nós*, parecem ser mudanças linguísticas do português moderno, embora não possa descartar-se a hipótese de que tenham emergido mais cedo, sem terem visibilidade na língua padrão (cf. Lopes 2001; 2003). O pronome *vós* continua a ser admitido como forma da língua padrão em registos muito formais, embora seja geralmente excluído pelos falantes dos dialetos portugueses centro-meridionais (que mantêm, contudo, a forma de acusativo/dativo *vos*). A mudança encaixa-se assim na oposição Norte-Sul que é central na paisagem dialetal portuguesa (cf. Segura 2013). A inovação *a gente*, por seu lado, abrange todo o território português embora com implantação mais forte no centro-sul (cf. Sória 2013). Em interação com o desaparecimento do pronome *vós*, várias alterações se produzem no sistema das formas de tratamento, uma das quais é o desenvolvimento de um valor não reflexivo para o pronome *si* (e a forma *consigo*). A fixação de *lhes* como forma de dativo plural (em oposição ao dativo singular *lhe*) é também uma mudança tardia, que regulariza a morfologia dos pronomes clíticos de terceira pessoa no que diz respeito à expressão gramatical da distinção de número.

Fora do sistema pronominal produzem-se alterações importantes na morfologia verbal. No português médio as formas de segunda pessoa do plural perdem a consoante inicial do morfema de pessoa-número *-des* quando esta se encontra em posição intervocálica, originando hiatos que evoluirão de forma regular na primeira e terceira

conjugações e não regular na segunda, com provável influência analógica do modelo flexional da primeira: *-ades* > *-aes* > *-ais*; *-edes* > *-ees* > *-es/-eis*; *-ides* > *-ies* > *-iis* > *-is* (cf. Cardeira 2005; 76 Morfologia nas interfaces). Também no português médio os participípios passados em *-udo* dos verbos da segunda conjugação dão lugar a participípios passados em *-ido*, aprofundando a proximidade entre os modelos flexionais dos verbos da segunda e terceira conjugações. A motivação para esse processo de analogia terá sido a irregularidade morfológica dos participípios em *-udo*, que exibiam antes do morfema de participípio passado *-do* uma vogal diferente da vogal temática, contrariamente aos verbos da primeira e terceira conjugações (cf. Cardeira 2005).

A analogia gramatical não é um tipo de mudança que se encontre limitada a uma época particular. Mas na história do português pode considerar-se que o português médio constitui um período crítico no domínio da regularização analógica, tanto no âmbito da flexão verbal como nominal. Durante este período foi significativamente reduzida a variação exibida ao nível da flexão verbal pelos textos dos séculos XIII e XIV (cf. Colaço/Cardeira 2013; Piel 1944; Williams ³1975; Meier 1948). A regularização analógica manifestou-se de diferentes formas: eliminando alguns participípios fortes (*bento* – *benzido*, *colheito* – *colhido*, *escolheito* – *escolhido*, *defeso* – *defendido*, *roto* – *rompido*; cf. Barros 2002), uniformizando certas formas do futuro a partir do infinitivo (*salrei* – *saírei*, *valrei* – *valerei*, *terrei* – *terei*, *porrei* – *porrei*, *verrei* – *virei*), eliminando alternâncias consonânticas nos radicais verbais (*adugo* – *aduzo*, *bengo* – *benzo*, *senço* – *sinto*, *jaço* – *jazo*), fixando para a terceira conjugação as alternâncias vocálicas [u]/[ɔ] e [i]/[ɛ] (*dormo* – *durmo*, *descobra* – *descubra*, *sento* – *sinto*, *serva* – *sirva*, *mença* – *mintá*; cf. Martins 1988; Barbato 2012), estabelecendo os paradigmas flexionais dos verbos *ser* e *estar*, e reduzindo, em geral, o grau de marcação morfológica de outros verbos irregulares (e.g. *moiro/moira* – *morro/morra*; *fiz/fige*, *fezeste* – *fiz*, *fizeste*; sobre variação dialetal e sociolinguística nos pretéritos fortes, cf. Mota/Rodrigues/Soalheiro 2013; Barbosa/Flores/Bastos-Gee 2016).

Fora da flexão verbal, a analogia teve um papel importante, por exemplo, na criação de formas específicas para o feminino nas palavras terminadas em *-or*, *-ol*, *-ês* e *-nte* (*senhor*, m./f. – *senhor*, m./*senhora*, f.; *espanhol*, m./f. – *espanhol*, m./*espanhola*, f.; *português*, m./f. – *português*, m./*portuguesa*, f.; *infante*, m./f. – *infante*, m./*infanta*, f.) e na mudança de género das palavras terminadas em *-agem* que eram empréstimos lexicais ao francês (cf. Cardeira 2005).

Os processos de regularização analógica constituíram-se numa fonte inesgotável de variação dialetal e sociolinguística, com consequências também para a diferenciação entre as normas portuguesa e brasileira. Sirva de exemplo a oposição [o]/[ɔ] nas formas nominais, criada pela metafonía por [u] final, mas eliminada ou expandida através, respetivamente, de processos de uniformização ou de extensão analógica. O padrão brasileiro tem *s[o]gro/s[o]gros* e *b[o]lso/b[o]lsos*, o padrão português tem *s[o]gro/s[ɔ]gros* e *b[o]lso/b[ɔ]lsos*. *B[o]lsos* (com vogal [o] de [ũ] latino) e *s[ɔ]gros* (com vogal [ɔ] de [õ] latino) são formas resultantes de evolução fonológica regular; *s[o]gros* resulta de uniformização analógica (a forma do plural adota a vogal radical da forma

do singular, como o português europeu padrão faz em *ac[o]rdo/ac[o]rdos*, ainda que muitos portugueses digam *ac[o]rdo/ac[ɔ]rdos*; *b[ɔ]lsos* resulta de extensão analógica da alternância [o]/[ɔ] a palavras em que não é o resultado regular da atuação da metafonia sobre a vogal radical da forma do singular mas não do plural.

Voltando à morfologia verbal, a par da perda do pronome nominativo *vós*, desapareceram do português padrão as formas verbais de segunda pessoa do plural, ainda que os dialetos setentrionais, em geral, as conservem (cf. Segura 2013, 130–131). No português padrão foram substituídas pelas formas de terceira pessoa do plural. Outras mudanças no sistema verbal, como o decréscimo de produtividade do pretérito mais que perfeito simples (afastando o português do galego), do condicional (parcialmente em variação com o imperfeito do indicativo) e do futuro simples (como expressão de tempo mas não de modo), terão ocorrido também nos últimos séculos (cf. Brocardo 2010a; 2010b; Lima 2014; Marques 2013; Oliveira 2013; Segura 2013, 131–132; 712 Semântica e pragmática).

Uma referência final ao domínio complexo dos conectores oracionais e sua evolução ao longo do tempo. Mais uma vez o português médio é um período crítico, ainda que muito continue a acontecer mais tarde, envolvendo renovação lexical, mudança semântica e processos de gramaticalização (cf. Lima 2014; Barreto 2002; Silva 1989; 1994b; Fiéis/Lobo 2008; 2009; sobre processos de gramaticalização de preposições e locuções preposicionais, cf. Poggio 2002a; 2002b). O desaparecido pronome oblíquo *en*, por exemplo, contribuiu para criar o conector adversativo *porém*, que na origem era uma locução conclusiva/explicativa (*por en(de)* = ‘por isso’).

O Quadro 4 dá-nos uma ideia do muito que o inventário de itens gramaticais mudou no decurso do português médio, fazendo com que formas comuns no português do século XIII fossem sentidas como «antigas» no século XVI e, por isso, desadequadas ao padrão literário. Os textos medievais *Demanda do Santo Graal* e *Livro de José de Arimateia* são parte do ciclo arturiano da Post-Vulgata, traduzido para português a partir do francês no século XIII (Castro 1983; 1993). Embora não se conserve o original, a cópia quatrocentista da *Demanda do Santo Graal* é conservadora e reflete bem a língua do original duocentista. A cópia quinhentista do *Livro de José de Arimateia* é, pelo contrário, intencionalmente inovadora, tendo sido executada com o objetivo de limpar o texto de formas desusadas no século XVI, pelo menos ao nível da língua padrão (cf. Martins 2013). Comparando os dois textos fica muito visível quais os itens gramaticais do original que foram, em geral, eliminados pela cópia quinhentista do *Livro de José de Arimateia*. Nas primeiras linhas do Quadro 4 aparecem os itens já referidos no Quadro 3; os restantes aparecem por ordem alfabética.

Quadro 4: Mudança linguística entre o português antigo e o português clássico, de acordo com o testemunho da cópia quinhentista do Livro de José de Arimateia

Itens gramaticais	Nº de ocorrências	
	Demanda do Santo Graal	Livro de José de Arimateia
<i>sa</i>	582	0
<i>ta</i>	33	0
<i>ende</i>	902	1
<i>en</i>	323	2
<i>u/hu</i>	511	2
<i>acá</i>	35	0
<i>alá</i>	49	0
<i>adur</i>	20	0
<i>al</i>	115	12
<i>alhur</i>	20	0
<i>ar/er</i> (cf. Filipe 2007)	146	0
<i>ca</i>	2259 (completivo, comparativo, explicativo/causal)	137 (só explicativo/causal)
<i>certas</i>	222	0
<i>chus/chos</i>	31	0
<i>desi/dessi</i>	102	2
<i>mentre</i>	64	0
<i>rem/ren</i>	278	0
<i>toste</i>	124	0
<i>de consum, de sũu</i>	40	0
<i>sem falha</i>	128	0

2.3 Sintaxe

Vejamos de seguida quais os principais aspetos em que a sintaxe do português mudou ao longo do tempo. A cronologia dos casos de mudança sintática identificados no Quadro 5 parece indicar que, neste domínio da gramática, o período do português médio é menos crítico do que nos domínios da mudança fonológica e morfológica. Na verdade, o século XVI é a época sensível para muitos aspetos da mudança sintática. Isso dificultou, em alguns casos, o preenchimento do Quadro 5 pois havia que escolher entre considerar como ponto de viragem para a mudança o final do período do português médio ou o início do período do português clássico. Escolheu-se, em

geral, a segunda opção, porque mesmo nos casos em que já há atestações da inovação antes de meados do século XVI, a consolidação da mudança parece produzir-se no português clássico, como acontece, por exemplo, com a expansão das estruturas de *se* impessoal, de infinitivo flexionado dependente de verbos causativos e percetivos ou de cliticização ao infinitivo com verbos de reestruturação. Se a opção tivesse sido assinalar a cronologia das primeiras atestações, a marca «X» estaria nestes casos na coluna do português médio. Seguiu-se, afinal, o mesmo critério que para a fonologia e a morfologia, já que muitas das inovações marcadas nos Quadros 1 e 2, acima, na coluna do português médio encontram-se atestadas desde o português antigo. No que diz respeito à variação entre pronomes fortes e pronomes clíticos dativos, as frases do tipo *vendo a vós* deixam de ocorrer nos textos notariais estudados por Martins (1994) no final do século XV (Lisboa) ou início do século XVI (Noroeste). Não quer isto dizer que este ou outro tipo de pronomes fortes usados como objeto direto ou indireto (fora das estruturas de redobro do clítico que encontramos no português contemporâneo), não possam atestar-se, residualmente, em época posterior. Também as orações adverbiais gerundivas introduzidas pela preposição *sem* perdem visibilidade a partir do final do século XV.

Quadro 5: Sintaxe

Principais traços caracterizadores	Cronologia da mudança		
	Português Médio (até meados do séc. XVI)	Português Clássico (até meados do séc. XVIII)	Português Moderno
Podem ocorrer formas fortes dos pronomes pessoais em lugar dos pronomes clíticos (sobretudo dativos: <i>vendo a vós</i>).	X		
O gerúndio ocorre a par do infinitivo em orações completivas e nas adverbiais introduzidas por <i>sem</i> (<i>nom pode ora hy al ffazer ssem oyndo as partes</i>).	X		
A construção de <i>se</i> passivo não está em variação com a construção de <i>se</i> impessoal (que resulta da reanálise de <i>se</i> passivo).		X	
As orações infinitivas dependentes de verbos causativos e percetivos não admitem o infinitivo flexionado.		X	
O infinitivo flexionado ocorre, com valor imperativo, em frases raiz, em variação com o conjuntivo (<i>E fazerem a dita casa e refazerem de todo caso fortuyto</i> / <i>E a faça e refaça de todo caso furtuyto</i>).		X	

Principais traços caracterizadores	Cronologia da mudança		
	Português Médio (até meados do séc. XVI)	Português Clássico (até meados do séc. XVIII)	Português Moderno
É muito rara a cliticização ao infinitivo em estruturas que admitem a subida do clítico. Nas mesmas estruturas não ocorre a negação na oração infinitiva.		X	
Existe interpolação generalizada (i.e. um clítico pré-verbal pode estar separado do verbo por qualquer constituinte sintático).		X	
Existe variação entre ênclise e próclise nos contextos que no português contemporâneo apresentam ênclise obrigatória.		X	
Ocorre a ordem SOV em variação com a ordem SVO em orações subordinadas e em tipos particulares de orações principais (<i>Quando Boorz esto ouvio; se eles aas aventuras do Santo Graal falecerem; já nós este cervo outra vez vimos</i>).		X	
O particípio passado de tempos compostos com <i>haver/ter</i> pode apresentar marcas de concordância com o objeto direto.		X	
O gerúndio é a opção normal na construção progressiva com <i>ser/estar/andar</i> (<i>sija pensando; andava buscando</i>) e nas orações pequenas dependentes de verbos perceptivos (<i>viu-os rezando</i>).			X
Os objetos diretos e indiretos topicalizados são obrigatoriamente retomados por um clítico, i.e., está disponível a construção de «Deslocação à Esquerda Clítica», mas não a construção de «Topicalização».			X
É rara a ocorrência do artigo definido antes de possessivo.			X
Não existe a inversão nominal negativa com <i>algum</i> (i.e. <i>algum</i> é um item bipolar em posição pré- ou pós- nominal).		X	X
Em vários aspetos (extraposição, <i>que/quem, cujo, qual</i>), as estruturas relativas exibem opções que deixaram de ser gramaticais no português (a par das que vieram a fixar-se).		X	X

Nalguns aspetos a sintaxe do português evoluiu como a das outras línguas iberorromânicas. É o caso da ordem SOV, muito característica da sintaxe medieval, que no português se encontra ininterruptamente desde os primeiros textos até, pelo menos, ao final do século XVI, que ocorria tanto em frases finitas como não finitas e que

consistia na anteposição do objeto relativamente ao verbo, mantendo-o numa posição interna à frase. O objeto anteposto podia ser de qualquer natureza, incluindo complementos oracionais, e encontra-se tanto em domínios finitos como não finitos (*sem aventura achar que de contar seja*). A ordem SOV da sintaxe medieval não tem a distribuição da ordem SOV latina, ocorrendo tipicamente em orações subordinadas (finitas ou não finitas), mas também em frases independentes e orações principais que incluam proclisadores (cf. Martins 2002b; 2005, 2011; 715 A colocação dos pronomes clíticos em sincronia e diacronia). (Para uma discussão de outros aspetos da evolução da ordem dos constituintes frásicos, com foco na inversão do sujeito e na chamada «ordem V2», cf. Eide 2006; Fiéis 2003; Galves/Gibraill (aceite); Galves/Kroch 2016; Kaiser 1999; 2002; Martins 2014b; Ribeiro 1995a; 1995b; Rinke 2007; 2009; Salvi 1990; 2000; para uma discussão e análise da interação entre estruturas relativas e extraposição, gerando sintagmas nominais descontínuos, cf. Cardoso 2010; no prelo; aceite; sobre a posição do sujeito em orações gerundivas, numa perspetiva diacrónica e dialetal, cf. Fiéis/Lobo 2010; Lobo/Carrilho 2015).

Outra mudança comum às línguas iberorromânicas é a emergência da construção de *se-impessoal* como resultado da reanálise de *se-passivo* (Naro 1976; Lapesa 1981; 2000), mudança que está na base de outras mais tardias e geradoras de variação. O português brasileiro separa-se do português europeu ao perder a construção de *se-passivo*, embora o padrão escrito ainda a conserve (Nunes 1990; 1991). No português europeu não está estudada a repartição dialetal das duas construções mas *se-nominativo* apresenta nos dialetos do português europeu (com maior incidência no centro-sul e Açores) diferentes graus de defetividade que originam construções não padrão, uma das quais é a construção de duplo sujeito (*A gente chama-se rãs a isto; Chama-se-lhe a gente espigas; Cá nunca se usaram isso; Com licença de caça e tudo, não se podemos caçar*; Martins 2009). Por outro lado, só no final do século XX se atribui às estruturas de *se-impessoal* com verbos transitivos (precisamente o contexto sintático em que *se-impessoal* emergiu como reanálise de *se-passivo*) o estatuto de formas normativas (Peres/Móia 1995, 235), o que sugere que este seria um bom tema para um estudo sociolinguístico.

Noutros casos, um ponto de partida comum, em aspetos essenciais, originou acentuada divergência entre as línguas românicas, como acontece relativamente à colocação dos pronomes clíticos (715 A colocação dos pronomes clíticos em sincronia e diacronia), domínio em que o português brasileiro divergiu radicalmente do português europeu (cf. Carneiro 2005; Lobo 2001; 2002). Certas estruturas infinitivas mostram também evoluções divergentes a partir do que parece ter sido uma relativa unidade inicial (720 Predicados complexos numa perspetiva comparativa; cf. Miller 1992; Silva 2012; Ciutescu 2013a; 2013b). Os domínios infinitivos associados aos verbos de controle e elevação excluía, no português medieval, o marcador de negação predicativa e, em geral, os pronomes clíticos, tornando assim a subida dos clíticos obrigatória. Os verbos causativos e percetivos associavam a estas características a exclusão do infinitivo flexionado. Tomados conjuntamente estes factos mostram que

a estrutura funcional destas orações infinitivas se expandiu ao longo do tempo, tornando-as domínios menos defetivos (cf. Martins 2006b; aceite; sobre os reflexos dialetais desta mudança nas estruturas com verbos causativos e percetivos, cf. Pereira 2012 e 720 Predicados complexos numa perspetiva comparativa). Ao mesmo tempo que se expandia para as estruturas causativas e percetivas, o infinitivo flexionado (originado possivelmente no pretérito imperfeito do conjuntivo latino; cf. Martins 2001; Harris 2013; Scida 2004) deixou de ocorrer nas frases raiz de modalidade imperativa que o permitiam no português medieval, em variação com o conjuntivo (*e se achassem que Moor Eanes siia no plazo com seu marido ualerlj seu plazo e se achassem que non siia no plazo que lhe fezesse o abade plazo*; Martins 1994).

No português medieval, o gerúndio encontrava-se em variação com o infinitivo em orações completivas e algumas orações adverbiais (especialmente as introduzidas pela preposição *sem*), mas veio a ser substituído pelo infinitivo, que sempre foi a opção mais comum nestes contextos (cf. Fiéis/Lobo 2011). Mais tarde o gerúndio voltaria a perder espaço para o infinitivo. Na construção progressiva com auxiliares como *estar*, *andar*, *ficar*, *começar* e nas orações pequenas dependentes de verbos percetivos, o gerúndio foi substituído por *a*+infinitivo nos dialetos setentrionais e do centro-litoral, tendo a língua padrão adotado a inovação. O português brasileiro manteve o gerúndio, contrariamente ao português europeu. A existência de uma fronteira muito nítida entre dialetos com *a*+infinitivo e dialetos com gerúndio no mesmo tipo de contexto sintático mostra que é possível definir áreas dialetais no território português com base em traços sintáticos (Carrilho/Pereira 2011; 2013; Pereira 2014a). As áreas em que o gerúndio não foi substituído por *a*+infinitivo são também aquelas em que emergiu o gerúndio flexionado (718 O gerúndio flexionado no português dialetal), um facto revelador da relevância de incluir a sintaxe na investigação geolinguística (cf. Martins 2009). Assim, o facto de o gerúndio flexionado estar registado no galego mas não nas variedades portuguesas setentrionais deixa de ser surpreendente

Por vezes, o português e as outras línguas iberorromânicas mostram evoluções comuns que as opõem a línguas como o francês e o italiano. É o caso da perda da expressão morfológica de concordância entre participio passado e objeto, um padrão que aliás já não era sistemático no português antigo. Este tipo de oposição (que junta do mesmo lado as línguas iberorromânicas, mas com o catalão a alinhar por vezes com o francês e o italiano) tornar-se-á mais evidente quando comentarmos o Quadro 6.

Algumas inovações do português moderno afastam-no, em maior ou menor grau, da generalidade das línguas românicas. O português desenvolveu a par da construção de Deslocação à Esquerda Clítica, comum à generalidade das línguas românicas, a construção de Topicalização (714 Ordem dos constituintes frásicos: sujeitos invertidos; objetos antepostos), no que coincidem português europeu e português brasileiro, mas o português europeu se afasta dos seus vizinhos mais próximos, ou seja, o galego e o espanhol.

Quanto à generalização do uso do artigo definido antes de possessivo, português europeu, português brasileiro e galego evoluem na mesma direção, mas o português

européu é a língua em que a expansão do uso do artigo é mais radical, tornando-o, em geral, obrigatório, enquanto o português brasileiro apresenta uma larga margem de opcionalidade (cf. Rinke 2010; Silva 1982; 1996; Álvarez/Xove 2002; Silva Domínguez 2002). Alguns dialetos do português europeu, sobretudo insulares, exibem um certo grau de conservadorismo ao permitirem a ausência do artigo antes de possessivo quando este precede um nome de parentesco (Carrilho/Pereira 2011; 2013).

Por fim, o português e o espanhol evoluíram na mesma direção ao associarem a posição pós-nominal do quantificador indefinido *algum* à expressão da polaridade negativa, o que não acontecia no português medieval (*se aqui ficardes em esta furesta, toste vos poderia viir ende mal algũũ*; Martins 2015a). Mas o português levará a evolução mais longe do que o espanhol, reanalisando o item de polaridade negativa constituído por Nome+*algum* como item de polaridade negativa forte, capaz de exprimir a negação sem estar sob o escopo do operador de negação predicativa. Assim, enquanto uma frase como *Não aconteceu coisa alguma* é gramatical tanto no português como no espanhol contemporâneos, só o português admite também *Coisa alguma aconteceu*. Além disso, o português alargou a inversão nominal negativa ao quantificador *nenhum*.

A coincidência cronológica entre um conjunto de mudanças que ocorreram no final do século XVIII (i.e. artigo antes de possessivo; inversão nominal negativa com *algum/nenhum*, com as características que tem no português contemporâneo; desaparecimento do pronome de referência [+humana] *nenhum*, substituído por *ninguém*; restrição da posição de *cujo* à posição pré-nominal e outros aspetos das estruturas relativas) permite sustentar a hipótese de uma mudança paramétrica ao nível da estrutura funcional do sintagma nominal, envolvendo o requisito de dar conteúdo fonológico à categoria Determinante (cf. Martins 2015a; 2015b).

2.4 Semântica lexical e sintaxe

Quadro 6: Semântica lexical e sintaxe

Principais traços caracterizadores	Cronologia da mudança		
	Português Médio (até meados do séc. XVI)	Português Clássico (até meados do séc. XVIII)	Português Moderno
Português Antigo (até ao final do séc. XIV)			
<i>Haver</i> é um verbo existencial e um verbo de posse. Como verbo de posse, começa a atestar-se em variação com <i>ter</i> ainda neste período.	X		

Principais traços caracterizadores	Cronologia da mudança		
	Português Médio (até meados do séc. XVI)	Português Clássico (até meados do séc. XVIII)	Português Moderno
O verbo copulativo <i>ser</i> ocorre tanto com «predicados de indivíduo» como com «predicados de estádio» (<i>ser português</i> a par de <i>ser faminto</i>), mas a oposição <i>ser/estar</i> já começa a delinear-se (i.e. <i>ser português</i> vs. <i>estar faminto</i>).	X		
Os auxiliares dos tempos compostos são <i>haver</i> e <i>ser</i> , mas o sistema já está a mudar (excluindo <i>ser</i> e substituindo <i>haver</i> por <i>ter</i>).	X		
Os itens de polaridade negativa <i>nenhum, nada, ninguém, jamais</i> co-ocorrem com <i>não</i> independentemente da sua posição em relação ao verbo, estabelecendo com o marcador de negação predicativa uma relação de «concordância negativa».	X		
As palavras <i>rem, cousa, homem, al, parte</i> incluem-se entre os itens de polaridade negativa (minimizadores). A sua interpretação polar depende do contexto frásico. Como minimizadores requerem a presença de um marcador de negação.	X		
O item de polaridade <i>algum</i> é bipolar, podendo estabelecer relações de «concordância negativa», e estando assim em variação com <i>nenhum</i> (<i>E foram ally mortos XIII cavallos dos cristãos, pero nõ morreo allgũ dos senhores que em elles herã, e assy se tornou dõ Duarte alegre</i>).	X		
<i>algum</i> e <i>nenhum</i> são comuns como pronomes de referência [+humana], sendo <i>alguém</i> e <i>ninguém</i> infrequentes.		X	X
Os itens de polaridade negativa <i>nenhum, nada, ninguém, jamais</i> podem ocorrer sem interpretação negativa em contextos modais não negativos (reservados a itens de polaridade positiva no português moderno)			X

Duas importantes mudanças do âmbito da semântica lexical, com repercussões na sintaxe, ligam o português, o espanhol e o galego. Nas três línguas reconfigurou-se o espaço semântico e a distribuição sintática dos verbos *haver* e *ter*, por um lado, e *ser* e *estar*, por outro. *Haver* é no português do século XIII tanto um verbo existencial como um verbo de posse, mas como verbo de posse virá a ser substituído por *ter*, mudança que aliás começou a manifestar-se ainda no período do português antigo (Brocardo 2006; Silva 1994b; 2002b; 2002c). O português brasileiro levou a mudança ainda mais

longe, substituindo *haver* por *ter* também como verbo existencial, em contraste com o português europeu. Alguns dialetos portugueses, concretamente os dialetos insulares (Açores e Madeira), exibem o mesmo processo de substituição de *haver* existencial por *ter*, embora sem o grau de consolidação da mudança que se observa no português do Brasil (Avelar/Callou 2007; 2012; Carrilho/Pereira 2011). Quanto aos verbos *ser* e *estar*, produziu-se ao longo do tempo uma especialização de *ser* como cópula de «predicados de indivíduo», enquanto *estar* adquiriu o estatuto de cópula de «predicados de estádio», substituindo neste domínio o verbo *ser* (Silva 1994b; 2002a; 2002c; Marquilhas 2009). Tanto a redistribuição de papéis sintático-semânticos de *ser/estar* como de *haver/ter* separam o português, o galego e o espanhol do francês e italiano que, neste aspeto (como em relação à concordância entre participio passado e objeto e à persistência dos pronomes oblíquos *en/i*), têm propriedades gramaticais semelhantes às do português e espanhol medievais (ou seja, inverte-se aqui o posicionamento das línguas que são tradicionalmente descritas como conservadoras ou inovadoras com base sobretudo em traços fonéticos, o que mostra que a dicotomia língua conservadora/língua inovadora tem pouco interesse numa perspetiva descritiva e conduz a classificações potencialmente falaciosas).

Os verbos *ser*, *haver* e *ter*, como auxiliares de tempos compostos, integram um quadro evolutivo comum e, também neste caso, o francês e o italiano contemporâneos ilustram as propriedades gramaticais que encontramos no português antigo.³ No português do século XIII *ser* ocorre como auxiliar com verbos inacusativos, *haver* é o auxiliar com outros tipos de verbos. Depois *haver* substitui *ser*, deixando o sistema de integrar auxiliares distintos para diferentes verbos, mas virá a ser substituído por *ter* (cf. Ribeiro 1996; Brocardo 2006; Silva 1994b; 2002b; 2002c). Como ambas as evoluções começam cedo e se estendem por um período longo, pode haver sobreposição no tempo entre os dois padrões de substituição, o que torna o processo particularmente complexo e a interpretação dos dados oferecidos pelas fontes textuais nem sempre fácil. A substituição do auxiliar *ser* pelo auxiliar *haver* é uma mudança comum ao português e ao espanhol, mas a substituição de *haver* por *ter* é uma inovação específica do português. (Sobre a presença, marginal, do pretérito perfeito composto no português medieval e clássico e sobre o desenvolvimento do também impropriamente chamado «pretérito perfeito composto» do português moderno, cf. Amaral/Howe 2012; Brocardo 2013; Cardoso/Pereira 2003)

³ Historicamente, o catalão começa por alinhar com o francês e o italiano tanto em relação aos verbos *ser/estar* e *ter/haver* como ao padrão de verbos auxiliares, mas ao longo do tempo desenvolveu processos evolutivos na mesma direção que o espanhol, mantendo contudo características próprias e exibindo nestes aspetos maior variação dialetal que as outras línguas ibéricas. *Ser* mantém-se como auxiliar dos tempos compostos, com verbos inacusativos, no rossellonès, nos dialetos baleares e em alguns dialetos pirenaicos, enquanto os restantes dialetos do catalão usam sempre o auxiliar *haver* (agradeço esta informação, bem como as referências bibliográficas que se seguem, a Ares Llop). Cf. Aramon i Serra (1957); Ramos Alfajarín (2000); Rojas (2004); Batllori (2007); Batllori/Roca (2012).

No domínio da semântica lexical em relação com a sintaxe, a outra grande área de mudança na história do português é a da negação, palavras negativas e itens de polaridade em geral (Martins 1997; 2000; 2015b; Morais 2001; Pinto 2015). Em contraste com o francês e o catalão, mas identicamente ao espanhol, o português perdeu toda uma série de minimizadores indefinidos (*rem, cousa, homem, al, parte*, etc.) que estão bem representados em alguns textos medievais mas virão a desaparecer sem que nenhum deles tenha sido reanalisado como palavra negativa. Assim, por exemplo, *rem*, um dos primeiros minimizadores a perder-se, deixa de se atestar no final do português antigo (cf. Quadro 4 acima). Perde-se também no espanhol, é residual no galego,⁴ mas os seus cognatos francês (*rien*) e catalão (*res*) tornam-se palavras negativas plenamente produtivas.

Perdidos os minimizadores, o português manteve os indefinidos negativos (*nenhum, nada, ninguém*), mas também estes mudaram ao longo do tempo. Começam por ser itens de polaridade negativa fracos, capazes de ocorrer em contextos modais, não negativos (*E ante que outro nenhũu falasse, disse o conde dõ Fernã Gonçalves*) e exigindo a presença do operador de negação predicativa mesmo quando precedem o verbo (*nenhũu nõ scapou*). Mas no português médio, e esporadicamente no português antigo, a co-ocorrência entre os indefinidos negativos e o operador de negação predicativa já se mostra opcional e os indefinidos negativos estão a caminho de se tornarem itens de polaridade negativa fortes (capazes de exprimirem só por si a negação). No entanto, exibirão ainda, até ao português moderno, ambiguidade lexical entre itens de polaridade negativa fortes e itens de polaridade modal. Esta segunda possibilidade acabará por perder-se, mais uma vez afastando o português das outras línguas românicas (e.g. Francês *Pierre est parti avant que personne ait pu faire aucun geste*; Português **O Pedro saiu antes que ninguém tivesse podido fazer nenhum gesto/O Pedro saiu antes que alguém tivesse podido fazer um gesto*; Catalão *Demana-li si en sap res*; Português **Pergunta-lhe se sabe nada/Pergunta-lhe se sabe alguma coisa*). Embora os itens de polaridade negativa mostrem um desenvolvimento na mesma direção na maioria das línguas românicas (excluindo o romeno e o veneziano), o português leva a evolução mais longe. No português, os itens de polaridade negativa especializaram-se exclusivamente na expressão da negação, passando a estar excluídos de qualquer outro tipo de contexto polar.

Paralelamente, os itens de polaridade positiva (como *algum, alguém*) deixaram de poder ocorrer em contextos negativos, ou seja, passaram a ser verdadeiramente itens de polaridade positiva (fracos) e já não itens bipolares. Assim, palavras como *algum* e *nenhum*, que podiam estar em variação nos mesmos contextos sintático-semânticos no português medieval (*des onte ao serão não ouvemos algũa/nenhũa folga*), passaram

4 «A forma xeral para «non humano, non animado» é *nada*, pero tamén se empregan *ren* ou *res*, se ben só como OD, sempre tralo verbo e con frecuencia precedidas de *nin*. *Ren* e *res* son dúas variantes de fala, hoxe de escaso uso, aceptadas ambas na norma actual como opcionais: *Falamos bastante tempo, pero non quixo dicir (nin) ren/res*» (Álvarez/Xove 2002, 487).

a estar em distribuição complementar: *algum* nos contextos afirmativos (assertivos) e modais (também chamados «contextos negativos fracos»), *nenhum* nos contextos estritamente negativos.

Por fim, as mudanças no domínio dos itens de polaridade integram um outro caso de especialização lexical: *nenhum* e *algum* deixam de ser pronomes de referência [+humana], substituídos nesta função pelas palavras *ninguém* e *alguém*, que são raras no português medieval, ganham expressão e crescem em frequência no português clássico e consolidam a sua posição, deixando de estar em variação com *algum* e *nenhum*, no português moderno.

Há marcas residuais nos dialetos portugueses quer da evolução dos auxiliares dos tempos compostos quer da evolução das palavras negativas, mas em nenhum dos casos as formas antigas parecem ter vitalidade significativa (ainda que estes sejam domínios para os quais não há estudos sistemáticos).

2.5 Léxico

Independentemente da história de palavras particulares ou grupos de palavras semanticamente relacionadas, outras dimensões da evolução do léxico de uma língua podem ser consideradas. Uma delas é o contraponto entre inovação e obsolescência, duas tendências tão naturais e constantes na evolução das línguas como outros pares dicotômicos: mudança fonológica/analogia, assimilação/dissimilação, etc. A relação entre os dois termos destes pares não é direta, mas certamente interagem, num jogo complexo de equilíbrios e catástrofes talvez demasiado complexo para que possamos compreendê-lo. No caso do léxico é especialmente assim, dada a natureza extremamente heterogênea do objeto e a multiplicidade de fatores que criam dinâmicas de mudança neste domínio. Por isso obsolescência e inovação lexical são normalmente tratados como aspetos independentes da mudança lexical, exceto nos casos particulares em que é possível encontrar uma relação direta entre uma palavra que morre e a inovação que a substitui (ainda que nem a obsolescência seja necessária para que exista inovação lexical nem a inovação conduza necessariamente à extinção de palavras pré-existentes). Os historiadores da língua portuguesa têm-se interessado sobretudo pela inovação lexical, identificando vagas de entrada de empréstimos na língua e avaliando a sua relevância para a história do léxico português, como exemplificam as passagens de Teyssier (1982) e Castro (2006) abaixo transcritas (cf. Piel 1989, que, além de caracterizar o fundo latino do léxico português, comparativamente a outras línguas românicas, identifica as diferentes origens dos empréstimos que ao longo do tempo se lhe juntaram: pré-latinos; germânicos; árabes; do latim medieval, humanista e científico-erudito; franceses e provençais; «exóticos»; italianos, espanhóis e outros).

«Vimos no capítulo II que formas eruditas e semieruditas, calcadas no latim, penetraram na língua desde as suas origens. Este processo de enriquecimento do vocabulário jamais cessou. Tornou-se, porém, particularmente intenso no século XV, com a prosa didáctica e histórica, e no século XVI, em consequência das tendências gerais do Renascimento humanista. No século XV os latinismos alimentam a prosa de D. Pedro (*Virtuosa Benfeitoria*) e a de D. Duarte, o autor do *Leal Conselheiro*. [...] Com o Renascimento humanista e o prestígio dos estudos latinos, este fenómeno só irá amplificar-se.» (Teyssier 1982, 68–69).

«O vocabulário do português enriqueceu-se, como o de todas as línguas europeias, com um número considerável de termos que designam conceitos e objectos relativos à civilização científica e técnica. Foi-se, por vezes, buscar no léxico existente a palavra própria para denotar o objecto novo (ex.: *comboio*), mas, na maior parte dos casos, recorreu-se como nas outras línguas românicas – e sobre o seu modelo – às raízes greco-latinas; ex.: *automóvel, autocarro, televisão*. A língua continua, assim, a criar termos eruditos como sempre fez, desde as suas origens.» (Teyssier 1982, 73).

«O séc. XV em Portugal assistiu à chegada do Renascimento, enquanto tomada de contacto com a literatura clássica, lida directamente e servindo de inspiração cultural e também linguística. Nesta perspectiva, um resultado quase imediato foi a entrada maciça de empréstimos lexicais tomados ao latim literário (e através dele ao grego) e injectados no léxico português. O latim dos clássicos romanos (que não da Igreja) tornou-se modelo para a renovação do português literário, com efeitos particularmente modificadores a nível da frase e do léxico. Detecta-se aqui algum contraste entre a atitude dos renascentistas do XV, preocupados em explorar as possibilidades de adaptação da língua portuguesa a novas realidades conceptuais e expressivas, e a dos autores do XVI, que não hesitavam em importar latinismos intactos.» (Castro 2006, 167).

Recorrendo de novo ao testemunho do manuscrito quinhentista do Livro de José de Arimateia, comparativamente à cópia quatrocentista, conservadora, da Demanda do Santo Graal, reúne-se no Quadro 7 um conjunto de palavras (verbos, nomes e adjetivos) que eram usuais no português do século XIII, quando os dois textos foram conjuntamente traduzidos do francês (Castro 1983; 1993), mas já não o eram no português do século XVI. A comparação entre o número de ocorrências na Demanda do Santo Graal e o número de ocorrências (muitas vezes 0) no Livro de José de Arimateia torna claro que estamos perante algumas daquelas palavras que mereceram o seguinte apontamento do organizador da cópia quinhentista, Manuel Álvares, no prólogo da obra: «nom mudei senam os vocabulos ininteligiveis, que os que se podem entender na antiguidade daquele tempo os leixei ir».

Quadro 7: Léxico possivelmente pertencente à tradução duocentista da Post-Vulgata arturiana que é residual ou inexistente no Livro José de Arimateia, em contraste com a Demanda do Santo Graal

Léxico (verbos, nomes, adjetivos)	Nº de ocorrências		Possíveis substitutos em Arimateia
	Demanda	Arimateia	
<i>acalçar</i>	52	0	<i>alcançar</i>
<i>aduzer</i>	82	0	<i>trazer</i>

Léxico (verbos, nomes, adjetivos)	Nº de ocorrências		Possíveis substitutos em Arimateia
	Demanda	Arimateia	
<i>assuar / assũar</i>	46	0	<i>ajuntar, juntar</i>
<i>avir</i>	328	29	<i>acontecer</i>
<i>catar</i>	159	15	
<i>coitar</i>	23	1	
<i>chagar</i>	198	14	<i>ferir, dar golpes, maltratar</i>
<i>esmar</i>	20	1	<i>pensar</i>
<i>espertar</i>	33	0	<i>acordar</i>
<i>filhar</i>	178	1	<i>tomar</i>
<i>guarir / gorir</i>	30	0	<i>sarar, guarecer</i>
<i>quitar</i>	28	2	
<i>sacar</i>	51	0	<i>tirar</i>
<i>semelhar</i>	183	4	<i>parecer</i>
<i>pres, presestes, presou, preserom, presera</i> PRENDER, Pret. perf. e mais-que-perf.	19	0	
<i>sia, siam, siia, siiam, seve, severom</i> SER, Pret. Imperf. e Pret. Perf.	42	0	
<i>ařam</i>	26	0	<i>vontade</i>
<i>covardice</i>	23	0	
<i>andança</i>	49	4	
<i>ledice / lidice</i>	100	0	<i>alegria, prazer, sabor</i>
<i>prez</i>	19	1	
<i>seeda / seda</i>	92	1	<i>cadeira</i>
<i>andante</i>	74	0	
<i>arrizado</i>	13	0	
<i>covardo</i>	14	0	
<i>esmorido</i>	20	0	<i>esmorecido</i>
<i>perjurado</i>	21	0	
<i>quite</i>	24	5	<i>livre</i>
<i>sanhudo</i>	26	3	
<i>seestro</i>	19	2	<i>esquerdo</i>

Fica assim exemplificado o reverso da inovação lexical, que é o foco de interesse de Paul Teyssier (1982) e Ivo Castro (2006). Significativamente, não se vê uma relação óbvia entre a «entrada maciça de empréstimos lexicais tomados ao latim literário» e o facto de palavras como as do Quadro 7 serem já desusadas no século XVI. Algumas das palavras que aparecem no Quadro 7 fazem parte da lista de «vocábulos antigos portugueses» de Duarte Nunes de Leão (*Origem da Língua Portuguesa*, 1606), nomeadamente, *adur*, *afã*, *consum*, *falha*, *filhar*, *hu*, *prez*, *toste* e *sanhudo*. Na mesma época que Duarte Nunes de Leão, também Rodrigues Lobo (*Corte na Aldeia*, 1616) se pronuncia sobre as palavras de que «se perdeu o uso»:

«E ao que dizeis das palavras antigas, posto que em algum tempo fôssem boas, não o ficam sendo na parte em que se perdeu o uso d'ellas; pois, como já disse, esse só é o fundamento e razão das palavras: e assim, não diremos *leixou*, *trouve*, *dixe*, *ca*, *sicais*, *acram*, *leidisse*, e outros vocabulos de que usaram auctores gravissimos de cujos escriptos podemos aprender a perfeição da lingua portugueza. E bastou o contrario uso para n'esta parte poderem seguir os que agora escrevem, e falam bem.» (Rodrigues Lobo, *Corte na Aldeia*, 1619).

O testemunho de Rodrigues Lobo mostra como o tema da obsolescência lexical tem uma interessante dimensão sociolinguística, e também geolinguística, já que as palavras desusadas pelos que «escrevem, e falam bem» não são necessariamente palavras inexistentes. Mas o estado da arte da lexicografia portuguesa (↗8 Lexicografia) não permite, satisfatoriamente, seguir o rasto às palavras que o padrão literário foi excluindo e que se terão perdido ou não. O estudo da dinâmica evolutiva do léxico, como a descreve Piel, vê-se assim bastante limitado, face a outros domínios da gramática histórica e da história da língua portuguesa.

«Este longo e laborioso passado explica a falta de homogeneidade, em perspectiva histórica, que caracteriza a sua estrutura. Com efeito, o conjunto vocabular do português nunca se manteve estacionário, antes evoluiu constantemente num ritmo ora mais, ora menos acelerado, evolução que não chegou ainda ao seu termo e que representa o esforço comum de homens procedentes de ambientes geográficos e sociais muito diversos. Entre as formas faladas-populares e as cultas-literárias observa-se uma permuta contínua e fertilizadora, em ambos os sentidos. Ao caudal vocabular do primitivo património afluí, em ondas sucessivas, uma infinidade de elementos estrangeiros, europeus e extra-europeus, entrando em linha de conta praticamente todas as línguas com que os Portugueses, no decorrer da sua história, estiveram, directa e indirectamente, em contacto.» (Piel 1989, 9).

3 Referências

- Ali, Manuel Said (1971, 1931), *Gramática Histórica da Língua Portuguesa*, São Paulo, Melhoramentos.
- Alonso, Dámaso (1958), *Metafonía y neutro de materia en España*, *Zeitschrift für romanische Philologie* 74, 1–24.

- Alonso, Dámaso (1962a), *Ensondecimiento en el Norte peninsular de alveolares y palatales fricativas*, in: *La Fragmentación Fonética Peninsular*, Enciclopedia Lingüística Hispánica, Suplemento I, Madrid, Consejo Superior de Investigaciones Científicas, 85–103.
- Alonso, Dámaso (1962b), *Metafonía, neutro de materia y colonización suditaliana en la Península Hispánica*, in: *La Fragmentación Fonética Peninsular*, Enciclopedia Lingüística Hispánica, Suplemento I, Madrid, Consejo Superior de Investigaciones Científicas, 105–154.
- Alonso, Dámaso (1962c), *B = V en la Península Ibérica*, in: *La Fragmentación Fonética Peninsular*, Enciclopedia Lingüística Hispánica, Suplemento I, Madrid, Consejo Superior de Investigaciones Científicas, 155–209.
- Álvarez, Rosario/Xove, Xosé (2002), *Gramática da Lingua Galega*, Vigo, Galaxia.
- Álvarez Pérez, Xosé Afonso (2014), *European Portuguese dialectal features: a comparison with Cintra's proposal*, *Journal of Portuguese Linguistics* 13, 29–62.
- Álvarez Pérez, Xosé Afonso (2015), *Isoglossas portuguesas nos materiais do Atlas Lingüístico de la Península Ibérica: análise crítica da Nova Proposta de Lindley Cintra*, *Zeitschrift für romanische Philologie* 131, 185–223.
- Álvarez Pérez, Xosé Afonso/Saramago, João (2012), *Áreas lexicais galegas e portuguesas: um novo olhar para a proposta de Cintra*, *Estudis Romànics* 34, 55–97.
- Amaral, Patrícia/Howe, Chad (2012), *Nominal and verbal plurality in the diachrony of the Portuguese Present Perfect*, in: Brenda Laca/Patricia Cabredo-Hofherr (edd.), *Verbal Plurality and Distributivity*, Berlin/Boston, De Gruyter, 25–53.
- Aramon i Serra, Ramon (1957), *Notes sobre alguns calcs sintàctics en l'actual català literari*, in: Günter Reichenkron/Mario Wandruszka/Julius Wilhelm (edd.), *Syntactica und Stilistica: Festschrift für Ernst Gamillscheg zum 70. Geburtstag*, Tübingen, Niemeyer, 1–33.
- Avelar, Juanito/Callou, Dinah (2007), *Gramática e Variação no Português Brasileiro: Considerações sobre «ter-haver» e «de-em»*, in: Maria Lobo/Antónia Coutinho (edd.), *Textos Seleccionados do XXII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística/Colibri, 183–197.
- Avelar, Juanito/Callou, Dinah (2012), *Preservação e mudança na história do português: de possessivo a existencial*, *Matraga* 19:30, 224–235.
- Badia i Margarit, Antoni Maria (1947), *Los complementos pronominalo-adverbiales derivados de IBI e INDE en la Península Ibérica*, *Anejos de la Revista de Filología Española* 38, Madrid, Aguirre.
- Barbato, Marcello (2012), *Origen y evolución de las alomorfias vocálicas radicales*, *Revue de Linguistique Romane* 76, 39–63.
- Barbosa, Pilar/Flores, Cristina/Bastos-Gee, Ana (2016), *Variable Use of Strong Preterites in European Portuguese: A Quantitative and Theoretical Approach*, in: Olinco, *Olomouc Linguistics Colloquium, June 9–11, 2016, Book of Abstracts*, Olomouc, Palacký University, http://olinco.upol.cz/wp-content/uploads/2015/11/Olinco2016_book-of-abstracts.pdf (10.06.2016).
- Barreto, Therezinha Maria Mello (2002), *Observações sobre as conjunções no século XVI*, in: Rosa Virgínia Mattos e Silva/Américo Venâncio Lopes Machado Filho (edd.), *O Português Quinhentista: Estudos Lingüísticos*, Salvador/Feira de Santana, EDUFBA/UEFS, 161–193.
- Barros, Anabela Leal de (2002), *O Particípio Passado: Aspectos da sua morfologia do século XIII ao século XVI*, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Batlloori, Montserrat (2007), *Sincronia i diacronia de «ser» i «estar»*, in: Sadurní Martí et al. (edd.), *Actes del Tretzè Col.loqui Internacional de Llengua i Literatura Catalanes, Universitat de Girona 8–13 de setembre de 2003*, vol. 2, Barcelona, Publicacions de l'Abadia de Montserrat, 43–54.
- Batlloori, Montserrat/Roca, Francesc (2012), *Grammaticalization of «ser» and «estar» in Romance*, in: Dianne Jonas/John Whitman/Andrew Garrett (edd.), *Grammatical Change: Origins, Nature, Outcomes*, Oxford/New York, Oxford University Press, 73–92.

- Bechara, Evanildo (1991), *As fases da língua portuguesa escrita*, in: Dieter Kremer (ed.), *Actes du XVII^e Congrès International de Linguistique et de Philologie Romanes*, vol. 3, Tübingen, Niemeyer, 68–76.
- Boléo, Manuel de Paiva/Silva, Maria Helena Santos (1974, ¹1959), *O Mapa dos dialectos e falares de Portugal Continental*, in: Manuel de Paiva Boléo, *Estudos de Linguística Portuguesa e Românica*, vol. 1: *Dialectologia e História da Língua*, Coimbra, Acta Universitatis Conimbrigensis, 309–352.
- Brissos, Fernando (2014), *New insights into Portuguese central-southern dialects: understanding their present and past forms through acoustic data from stressed vowels*, *Journal of Portuguese Linguistics* 13:1, 63–115.
- Brissos, Fernando (2015), *Dialectos portugueses do Centro-Sul: corpus de fenómenos e revisão do problema da (des)unidade*, *Zeitschrift für romanische Philologie* 131, 999–1041.
- Brissos, Fernando/Saramago, João (2014), *O problema da diversidade dialectal do Centro-Sul português: informação perceptiva versus informação acústica*, *Estudos de Linguística Galega* 5, 53–80.
- Brocardo, Maria Teresa (2005), *Sobre periodização da história do português europeu: Contributo para uma discussão*, *Iberoromania* 62, 97–117.
- Brocardo, Maria Teresa (2006), «Haver» e «ter» em português medieval: *Dados de textos dos séculos XIV e XV*, *Revue de Linguistique Romane* 70, 95–122.
- Brocardo, Maria Teresa (2010a), *Portuguese Plusperfect: Elements for a Diachronic Approach*, *Estudos Linguísticos/Linguistic Studies* 5, 117–130.
- Brocardo, Maria Teresa (2010b), *O «passado do passado» – alguns dados para a história do pretérito mais-que-perfeito em português*, *Verba Hispanica* 20, 33–48.
- Brocardo, Maria Teresa (2013), *O pretérito perfeito – origem e evolução histórica*, in: Jasmina Markič / Clara Nunes Correia (edd.), *Descrições e Contrastes: Tópicos de gramática portuguesa com exemplos contrastivos eslovenos*, Ljubljana, Univerzav Ljubljani, Filozofska fakulteta, 111–117.
- Brocardo, Maria Teresa (2014), *Tópicos de História da Língua Portuguesa*, Lisboa, Colibri.
- Brocardo, Maria Teresa/Lopes, Célia Regina dos Santos Lopes (2016a), *History and Current Setting*, in: Leo Wetzels/João Costa/Sergio Menuzzi (edd.), *Handbook of Portuguese Linguistics*, Hoboken, NJ, Wiley Blackwell, 1–14.
- Brocardo, Maria Teresa/Lopes, Célia Regina dos Santos Lopes (2016b), *Main morphosyntactic changes and grammaticalization processes*, in: Leo Wetzels/João Costa/Sergio Menuzzi (edd.), *Handbook of Portuguese Linguistics*, Hoboken, NJ, Wiley Blackwell, 471–486.
- Câmara, JR., Joaquim Mattoso (1975), *História e Estrutura da Língua Portuguesa*, Rio de Janeiro, Padrão/Prolivro.
- Cardeira, Esperança (2005), *Entre o Português Antigo e o Português Clássico*, Lisboa, Imprensa Nacional – Casa da Moeda.
- Cardoso, Adriana (2010), *Variation and Change in the Syntax of Relative Clauses: New evidence from Portuguese*, tese de Doutoramento, Lisboa, Universidade de Lisboa.
- Cardoso, Adriana (no prelo), *Portuguese Relative Clauses in Synchrony and Diachrony*, Oxford/New York, Oxford University Press.
- Cardoso, Adriana (aceite), *Discontinuous noun phrases and remnant-internal relativization in the diachrony of Portuguese*, in: Ana Maria Martins/Adriana Cardoso (edd.), *Word Order Change*, Oxford/New York, Oxford University Press.
- Cardoso, Adriana/Pereira, Susana (2003), *Contributos para o estudo da emergência do tempo composto em português*, *Revista da ABRALIN* 2, 159–181.
- Carneiro, Zenaide (2005), *Cartas da Bahia: Um estudo linguístico-filológico*, tese de Doutoramento, Campinas, SP, Universidade de Campinas.
- Carrilho, Ernestina (2008), *Beyond doubling: overt expletives in European Portuguese dialects*, in: Sjeff Barbiers et al. (edd.), *Microvariation in Syntactic Doubling*, Bingley, Emerald, 301–323.

- Carrilho, Ernestina (2009), *Sobre o expletivo «ele» em português europeu*, Estudos de Lingüística Galega 1, 7–26.
- Carrilho, Ernestina/Pereira, Sandra (2011), *Sobre a distribuição geográfica de construções sintáticas não-padrão em Português europeu*, in: Armanda Costa/Pilar Barbosa/Isabel Falé (edd.), *Textos Seleccionados do XXVI Encontro Nacional da Associação Portuguesa de Linguística*, CD-ROM, Lisboa, Associação Portuguesa de Linguística, 125–139.
- Carrilho, Ernestina/Pereira, Sandra (2013), *On the areal dimension of non-standard syntax: Evidences from a Portuguese dialect corpus*, in: Alena Barysevich/Alexandra d'Arce/David Heap (edd.), *Proceedings of Methods XIV – Papers from the Fourteenth International Conference on Methods in Dialectology 2011*, Bamberger Beiträge zur englischen Sprachwissenschaft/Bamberger Studies in English Linguistics 57, 69–79.
- Carvalho, Joaquim Brandão de (1989), *L'origine de la terminaison «-ão» du portugais: une approche phonétique nouvelle du problème*, Zeitschrift für romanische Phonologie 105, 148–160.
- Carvalho, José Herculano de (1984a), *Sincronia e diacronia nos sistemas vocálicos do crioulo cabo-verdiano*, in: *Estudos Linguísticos*, vol. 2, Coimbra, Coimbra Editora, 7–31.
- Carvalho, José Herculano de (1984b), *Le vocalisme atone des parlers créoles du cap Vert*, in: *Estudos Linguísticos*, vol. 2, Coimbra, Coimbra Editora, 35–45.
- Carvalho, José Herculano de (1984c), *Nota sobre o vocalismo antigo português: valor dos grafemas «e» e «o» em sílaba átona*, in: *Estudos Linguísticos*, vol. 2, Coimbra, Coimbra Editora, 77–103.
- Castro, Ivo (1983), *Sobre a data da introdução na Península Ibérica do ciclo arturiano da post-vulgata*, Boletim de Filologia 28, 81–98.
- Castro, Ivo (1991), *Curso de História da Língua Portuguesa*, Lisboa, Universidade Aberta.
- Castro, Ivo (1993), *Demanda do Santo Graal; Livro de José de Arimateia; Matéria da Bretanha; Merlim*, in: Julia Lanciani/Giuseppe Tavanì (edd.), *Dicionário de Literatura Medieval Galega e Portuguesa*, Lisboa, Caminho, 203–206, 409–411, 445–450, 456–458.
- Castro, Ivo (1999), *O Português Médio segundo Cintra (nuga bibliográfica)*, in: Isabel Hub Faria (ed.), *Lindley Cintra: Homenagem ao Homem, ao Mestre e ao Cidadão*, Lisboa, Cosmos/Faculdade de Letras da Universidade de Lisboa, 367–370.
- Castro, Ivo (2004), *A primitiva produção escrita em português*, in: *Orígenes de las Lenguas Romanas en el Reino de León, Siglos IX–XII*, León, Centro de Estudios e investigación «San Isidoro»/Caja España de Inversiones/Archivo Histórico Diocesano, 69–97.
- Castro, Ivo (2006), *Introdução à História do Português*, Lisboa, Colibri.
- Cintra, Luís Filipe Lindley (1970), *Os ditongos decrescentes «ou» e «ei»: esquema de um estudo sincrónico e diacrónico*, in: *Anais do Primeiro Simpósio de Filologia Românica (1958)*, Rio de Janeiro, Ministério da Educação e Cultura, 115–134.
- Cintra, Luís Filipe Lindley (1971), *Nova proposta de classificação dos dialectos galego portugueses*, Boletim de Filologia 22, 81–116.
- Cintra, Luís Filipe Lindley (1983), *Estudos de Dialectologia Portuguesa*, Lisboa, Sá da Costa.
- Ciutescu, Elena (2013a), *Micro-parametric variation in Romance causative constructions*, Bucharest Working Papers in Linguistics 15, 45–60.
- Ciutescu, Elena (2013b), *Remarks on the infinitival subject of perception verb complements: evidence for two syntactic configurations*, Revue Roumaine de Linguistique 58, 299–312.
- Colaço, Maria João/Cardeira, Esperança (2013), *«Menço» ou «minto»? Regularização de paradigmas verbais*, Diacrítica 27, 69–94.
- Costa, João/Pereira, Sandra (2013), *«a gente»: pronominal status and agreement revisited*, Linguistic Review 30, 161–184.
- Eide, Kristina (2006), *Word Order Structures and Unaccusative Verbs in Classical and Modern Portuguese: The Reorganisation of Information Structure*, tese de Doutoramento, Oslo, University of Oslo.

- Fiéis, Alexandra (2003), *Ordem de Palavras, Transitividade e Inacusatividade: Reflexão Teórica e Análise do Português dos Séculos XIII a XVI*, tese de Doutoramento, Lisboa, Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas.
- Fiéis, Alexandra/Lobo, Maria (2008), *As orações introduzidas por «des(de) que» na história do português*, in: Sónia Frota/Ana Lúcia Santos (edd.), *Textos Seleccionados do XXIII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística/Colibri, 167–178.
- Fiéis, Alexandra/Lobo, Maria (2009), *Para uma diacronia das orações causais e explicativas do português*, in: Alexandra Fiéis/Maria Antónia Coutinho (edd.), *Textos Seleccionados do XXIV Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística, 265–280.
- Fiéis, Alexandra/Lobo, Maria (2010), *Aspectos da sintaxe das orações gerundivas no Português Medieval e no Português Europeu Contemporâneo*, in: Ana Maria Brito et al. (edd.), *Textos Seleccionados do XXV Encontro Nacional da Associação Portuguesa de Linguística*, Porto, Associação Portuguesa de Linguística, 419–434.
- Fiéis, Alexandra/Lobo, Maria (2011), *Propriedades de gerúndios e de infinitivos em português antigo*, in: Armanda Costa/Pilar Barbosa/Isabel Falé (edd.), *Textos Seleccionados do XXVI Encontro Nacional da Associação Portuguesa de Linguística*, CD-ROM, Lisboa, Associação Portuguesa de Linguística, 256–265.
- Fiéis, Alexandra/Lobo, Maria (2008), *As orações introduzidas por «des(de) que» na história do português*, in: Sónia Frota/Ana Lúcia Santos (edd.), *Textos Seleccionados do XXIII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística/Colibri, 167–178.
- Filipe, Laura Martins (2007), *O caso de «er»/«ar»: um ponto mal esclarecido na história da língua portuguesa*, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Florêncio, Manuela (2001), *Dialecto Alentejano: contributos para o seu estudo*, Lisboa, Colibri/Centro de Estudos do Alentejo.
- Galves, Charlotte/Gibraill, Alba (aceite), *Subject inversion in transitive sentences from Classical to Modern European Portuguese: a corpus-based study*, in: Ana Maria Martins/Adriana Cardoso (edd.), *Word Order Change*, Oxford/New York, Oxford University Press.
- Galves, Charlotte/Kroch, Anthony (2016), *Main Syntactic Changes from a Principle-and-Parameters View*, in: Leo Wetzels/João Costa/Sergio Menuzzi (edd.), *Handbook of Portuguese Linguistics*, Hoboken, NJ, Wiley Blackwell, 471–486.
- Harris, Michael J. (2013), *The origin of the Portuguese inflected infinitive through a corpus analysis*, in: Jennifer Cabrelli Amaro et al. (edd.), *Proceedings of the 16th Hispanic Linguistics Symposium*, Somerville, MA, Cascadilla Proceedings Project, 303–311.
- Hart, Thomas (1955), *Notes on the Sixteenth-Century Portuguese Pronunciation*, *Word* 11, 404–415.
- Hart, Thomas (1959), *The Overseas Dialects as Sources for the History of Portuguese Pronunciation*, in: *Actas do III Colóquio Internacional de Estudos Luso-Brasileiros (1957)*, vol. 1, Lisboa, Instituto de Alta Cultura/Junta de Investigação do Ultramar, 161–272.
- Houaiss, Antônio/Villar, Mauro de Salles/Franco, Francisco Manoel de Mello (2001), *Dicionário Houaiss da Língua Portuguesa*, Rio de Janeiro, Editora Objetiva.
- Huber, Joseph (1986, ¹⁹³³), *Gramática do Português Antigo*, Lisboa, Fundação Calouste Gulbenkian.
- Kaiser, Georg A. (1999), *A ordem das palavras e a posição do verbo finito no português antigo*, in: Ferenc Pál (ed.), *Actas do Congresso Internacional Organizado por Motivo dos Vinte Anos do Português no Ensino Superior*, Budapeste, Departamento de Língua e Literatura Portuguesas da Faculdade de Letras da Universidade Eötvös Loránd, 248–259.
- Kaiser, Georg A. (2002), *Verbstellung und Verbstellungswandel in den romanischen Sprachen*, Tübingen, Niemeyer.

- Lapesa, Rafael (1981), *Historia de la Lengua Española*, Madrid, Gredos.
- Lapesa, Rafael (2000), *Estudios de Morfosintaxis Histórica del Español*, Madrid, Gredos.
- Lima, José Pinto de (2014), *Studies on Grammaticalization and Lexicalization/Estudos de Gramaticalização e Lexicalização*, München, LINCOM Europa.
- Lobo, Maria/Carrilho, Ernestina (2015), *Combining geolinguistic sources in dialect syntax: Three case-studies through ALPI and CORDIAL-SIN*, in: Xulio Sousa/Carlota de Benito/Víctor Lara (edd.), *Syntactic Variation in Western European Languages: From the Noun Phrase to Clause Structure*, Dialectologia, Special Issue V, 141–166.
- Lobo, Tânia (2001), *Para uma sociolinguística histórica do português no Brasil: edição filológica e análise linguística de cartas particulares do Recôncavo da Bahia, século XIX*, tese de Doutorado, São Paulo, Universidade de São Paulo.
- Lobo, Tânia (2002), *A sintaxe dos clíticos: o século XVI, o século XX e a constituição da norma*, in: Rosa Virgínia Mattos e Silva/Américo Venâncio Lopes Machado Filho (edd.), *O Português Quinhentista: Estudos Lingüísticos*, Salvador/Feira de Santana, EDUFBA/UEFS, 83–101.
- Lopes, Célia Regina dos Santos (2001), *O percurso de «a gente» em tempo real de longa duração*, in: Rosa Virgínia Mattos e Silva (ed.), *Para a História do Português Brasileiro*, vol.2.1: *Primeiros Estudos*, São Paulo, Humanitas FFLCH/USP, FAPESP, 127–148.
- Lopes, Célia Regina dos Santos (2003), *A inserção de «a gente» no quadro pronominal do português*, Frankfurt am Main/Madrid, Vervuert/Iberoamericana.
- Maia, Clarinda de Azevedo (1986), *História do Galego-Português: Estado Linguístico da Galiza e do Noroeste de Portugal desde o Século XIII ao Século XVI (com referência à situação do galego moderno)*, Coimbra, Instituto Nacional de Investigação científica.
- Maia, Clarinda de Azevedo (1995), *Sociolinguística histórica e periodização linguística: Algumas reflexões sobre a distinção entre português arcaico e português moderno*, Diacrítica 10, 3–30.
- Marques, Rui (2013), *Modo*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática da Língua Portuguesa*, vol. 1, cap. 19, Lisboa, Fundação Calouste Gulbenkian, 673–693.
- Marquilha, Rita (2000), *A Faculdade das Letras: Leitura e escrita em Portugal no séc. XVIII*, Lisboa, Imprensa Nacional – Casa da Moeda.
- Marquilha, Rita (2003), *Mudança analógica e elevação das vogais pretônicas*, in: Ivo Castro/Inês Duarte (edd.), *Razões e Emoção. Miscelânea de estudos em homenagem a Maria Helena Mira Mateus*, vol. 2, Lisboa, Imprensa Nacional – Casa da Moeda, 7–18.
- Marquilha, Rita (2004), *Traços distintivos, góticos e electrónicos*, in: Rosario Álvarez/Antón Santamarina (edd.), *(Dis) cursos da Escrita: Estudos de filoxía galega ofrecidos en memoria de Fernando R. Tato Plaza*, A Coruña, Fundación Pedro Barrié de la Maza, 475–489.
- Marquilha, Rita (2009), *«Eu ainda sou vivo»: Sobre a edição e análise linguística de cartas de gente vulgar*, Estudos de Linguística Galega 1, 47–65.
- Martins, Ana Maria (1985), *Elementos para um Comentário Linguístico do Testamento de Afonso II (1214)*, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Martins, Ana Maria (1988), *Metafonia verbal no português: Uma abordagem histórica*, in: Dieter Kremer (ed.), *Homenagem ao Prof. Joseph M. Piel por ocasião do seu 85º aniversário*, Tübingen, Niemeyer, 349–366.
- Martins, Ana Maria (1994), *Clíticos na História do Português*, tese de Doutorado, Lisboa, Universidade de Lisboa.
- Martins, Ana Maria (1995), *A evolução das vogais nasais finais «-ã», «-õ», «-ê» no português*, in: Cilene da Cunha Pereira/Paulo R. D. Pereira (edd.), *Miscelânea de Estudos Lingüísticos, Filológicos e Literários In Memoriam Celso Cunha*, Rio de Janeiro, Editora Nova Fronteira, 617–646.
- Martins, Ana Maria (1997), *Aspectos da Negação na História das Línguas Românicas: Da natureza de palavras como «nenhum», «nada», «ninguém»*, in: Ivo Castro (ed.), *Actas do XII Encontro*

- Nacional da Associação Portuguesa de Linguística*, vol. 2, *Linguística Histórica e História da Linguística*, Lisboa, Associação Portuguesa de Linguística, 179–210.
- Martins, Ana Maria (1999), *Os mais antigos textos escritos em português: Documentos de 1175 a 1252*, in: Isabel Hub Faria (ed.), *Lindley Cintra: Homenagem ao Mestre, ao Homem e ao Cidadão*, Lisboa, Edições Cosmos/Faculdade de Letras da Universidade de Lisboa, 491–534.
- Martins, Ana Maria (2000), *Polarity items in Romance: Underspecification and lexical change*, in: Susan Pintzuk/George Tsoulas/Anthony Warner (edd.), *Diachronic Syntax: Models and Mechanisms*, Oxford/New York, Oxford University Press, 191–219.
- Martins, Ana Maria (2001), *On the origin of the Portuguese inflected infinitive: a new perspective on an enduring debate*, in: Laura Brinton (ed.), *Historical Linguistics 1999: Selected Papers from the 14th International Conference on Historical Linguistics*, Amsterdam/Philadelphia, Benjamins, 207–222.
- Martins, Ana Maria (2002a), *Mudança sintáctica e História da Língua Portuguesa*, in: Brian F. Head et al. (edd.), *História da Língua e História da Gramática: Actas do Encontro*, Braga, Universidade do Minho – Centro de Estudos Humanísticos, 251–297.
- Martins, Ana Maria (2002b), *The Loss of IP-scrambling in Portuguese: Clause Structure, Word Order Variation and Change*, in: David Lightfoot (ed.), *Syntactic Effects of Morphological Change*, Oxford/New York, Oxford University Press, 232–248.
- Martins, Ana Maria (2003), *Relatório da cadeira de História da Língua Portuguesa, apresentado a concurso para Professor Associado do 2^o Grupo A (Linguística Geral e Românica) da Faculdade de Letras da Universidade de Lisboa*, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Martins, Ana Maria (2005), *Clitic Placement, VP-ellipsis and scrambling in Romance*, in: Montserrat Batllori et al. (edd.), *Grammaticalization and Parametric Change*, Oxford/New York, Oxford University Press, 175–193.
- Martins, Ana Maria (2006a), *Para a história do vocalismo átono português: a propósito do Testamento de D. Afonso II de 1214*, in: Clarinda de Azevedo Maia/Ana Cristina Macário Lopes/Graça Maria Rio-Torto (edd.), *Miscelânea de Estudos in memoriam José G. Herculano de Carvalho*, Revista Portuguesa de Filologia 35:1 [2003–2006], 295–319.
- Martins, Ana Maria (2006b), *Aspects of infinitival constructions in the history of Portuguese*, in: Randall S. Gess/Deborah Arteaga (edd.), *Historical Romance Linguistics: Retrospective and Perspectives*, Amsterdam/Philadelphia, Benjamins, 327–355.
- Martins, Ana Maria (2007), *O primeiro século do português escrito*, in: Ana Boullón Agrelo (ed.), *Na Nosa Lyngoage Galega: A Emerxencia do Galego como Lingua Escrita na Idade Media*, Santiago de Compostela, Consello da Cultura Galega/Instituto da Lingua Galega, 161–184.
- Martins, Ana Maria (2009), *Subject doubling in European Portuguese dialects: the role of impersonal «se»*, in: Enoch O. Aboh et al. (edd.), *Romance Languages and Linguistic Theory: Selected papers from «Going Romance» Amsterdam 2007*, Amsterdam/Philadelphia, Benjamins, 179–200.
- Martins, Ana Maria (2011), *Scrambling and Information Focus in Old and Contemporary Portuguese*, *Catalan Journal of Linguistics* 10, 1–26.
- Martins, Ana Maria (2013), *Copiar o português duocentista: A Demanda e o José de Arimateia*, in: Rosario Álvarez et al. (edd.), *Ao Sabor do Texto: Estudos dedicados a Ivo Castro*, Santiago de Compostela, Universidade de Santiago de Compostela, Servizo de Publicacións e Intercambio Científico, 383–402.
- Martins, Ana Maria (2014a), *Syntactic Change in Portuguese and Spanish: Divergent and parallel patterns of linguistic splitting*, in: Patrícia Amaral/Ana Maria Carvalho (edd.), *Portuguese-Spanish Interfaces: Diachrony, synchrony, and contact*, Amsterdam/Philadelphia, Benjamins, 35–64.
- Martins, Ana Maria (2014b), *An argument against verb-second in Old Portuguese*, in: Alexandra Fiéis/Maria Lobo/Ana Madeira (edd.), *O Universal e o Particular: Uma vida a comparar, Homenagem a Maria Francisca Xavier*, Lisboa, Colibri, 207–216.

- Martins, Ana Maria (2015a), *Negation and NPI composition inside DP*, in: Theresa Biberauer/George Walkden (edd.), *Syntax over Time: Lexical, Morphological and Information-Structural Interactions*, Oxford/New York, Oxford University Press, 102–122.
- Martins, Ana Maria (2015b), *Ordem de palavras e polaridade: inversão nominal negativa com «algum»/«alguno» e «nenhum»*, *Diacrítica* 29, 401–428.
- Martins, Ana Maria (aceite), *Infinitival complements of causative/perception verbs in a diachronic perspective*, in: Anabela Gonçalves/Ana Lúcia Santos (edd.), *Complement Clauses in Portuguese: Syntax and Acquisition*, Amsterdam/Philadelphia, Benjamins.
- Martins, Ana Maria/Saramago, João (1993), *As sibilantes em português: um estudo de geografia linguística e de fonética experimental*, in: Ramón Lorenzo (ed.), *Actas do XIX Congreso Internacional de Lingüística e Filología Románicas, Universidade de Santiago de Compostela, 1989, Sección IV: Dialectología e Xeografía Lingüística, Sección VIII: Onomástica*, A Coruña, Fundación «Pedro Barrié de la Maza, Conde de Fenosa», 121–142.
- Mateus, Maria Helena/Andrade, Ernesto d' (2000), *The Phonology of Portuguese*, Oxford, Oxford University Press.
- Mateus, Maria Helena Mira, et al. (2003), *Gramática da Língua Portuguesa*, Lisboa, Caminho.
- Meier, Harri (1948), *A evolução dos pretéritos fortes em português*, in: Harri Meier, *Ensaio de Filologia Românica*, Lisboa, Edição da Revista de Portugal, 31–54.
- Miller, D. Gary (1992), *Complex Verb Formation*, Amsterdam/Philadelphia, Benjamins.
- Morais, Maria Aparecida Torres (2001), *Aspectos da história das palavras negativas no português*, in: Rosa Virgínia Mattos e Silva (ed.), *Para a História do Português Brasileiro*, vol. 2.1: *Primeiros Estudos*, São Paulo, Humanitas FFLCH/USP, FAPESP, 149–203.
- Moreira, Júlio (1922), *Estudos de Língua Portuguesa: Subsídios para a Syntaxe Histórica e Popular*, Lisboa, Clássica Editora.
- Mota, Maria Antónia Coelho da/Rodrigues, Maria Celeste/Soalheiro, Elisabete (2003), *Padrões flexionais nos pretéritos fortes em PE falado setentrional*, in: Ivo Castro/Inês Duarte (edd.), *Razões e Emoção: Miscelânea de estudos em homenagem a Maria Helena Mira Mateus*, vol. 2, Lisboa, Imprensa Nacional—Casa da Moeda, 129–155.
- Muidine, Soraya Aboo (2000), *Os pronomes «i» e «en(de)» no português dos séculos XIII a XVI*, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Naro, Anthony J. (1971), *The history of «e» and «o» in Portuguese: a study in linguistic drift*, *Language* 47, 615–645.
- Naro, Anthony J. (1976), *The genesis of the reflexive impersonal in Portuguese: A study in syntactic change as a surface phenomenon*, *Language* 52, 778–810.
- Nunes, Jairo (1990), *O famigerado «se»: Uma análise sincrónica e diacrónica das construções com «se» passivador e indeterminador*, dissertação de Mestrado, Campinas, SP, Universidade de Campinas.
- Nunes, Jairo (1991), *«Se» passivador e «se» indeterminador: o percurso diacrónico no português brasileiro*, *Cadernos de Estudos Lingüísticos* 20, 33–58.
- Nunes, José Joaquim (1975, 1919), *Compêndio de Gramática Histórica Portuguesa: Fonética e Morfologia*, Lisboa, Clássica Editora.
- Oliveira, Fátima (2013), *Tempo Verbal*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 15, Lisboa, Fundação Calouste Gulbenkian, 509–553.
- Parkinson, Stephen (2002), *The Portuguese final nasals: documenting a chronology*, *Santa Barbara Portuguese Studies* 6, 287–306.
- Pereira, Sandra (2003), *Gramática comparada de «a gente»: variação no português europeu*, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Pereira, Sandra (2012), *Protótipo de um Glossário dos Dialectos Portugueses com Anotação Sintática*, tese de Doutoramento, Lisboa, Universidade de Lisboa.

- Pereira, Sandra (2015), *Causative and Perception Constructions in European Portuguese: the dialectal data*, in: Xulio Sousa/Carlota de Benito/Víctor Lara (edd.), *Syntactic Variation in Western European Languages: From the Noun Phrase to Clause Structure*, Dialectologia, Special Issue V, 53–80.
- Pereira, Sílvia Afonso (2014a), *A sintaxe na classificação dos dialetos portugueses*, in: António Moreno et al. (edd.), *Textos Seleccionados do XXIX Encontro Nacional da Associação Portuguesa de Linguística*, Porto, Associação Portuguesa de Linguística, 445–464.
- Pereira, Sílvia Afonso (2014b), *Variação sintática em Portugal e a fronteira com a Galiza*, in: Xulio Sousa/Marta Negro Romero/Rosario Álvarez (edd.), *Lingua e identidade na fronteira galego-portuguesa*, Santiago de Compostela, Consello da Cultura Galega, 201–224.
- Pereira, Sílvia Afonso (2015), *Predicative constructions with gerunds in European Portuguese dialects*, in: Xulio Sousa/Carlota de Benito/Víctor Lara (edd.), *Syntactic Variation in Western European Languages: From the Noun Phrase to Clause Structure*, Dialectologia, Special Issue V, 351–371.
- Peres, João/Móia, Telmo (1995), *Áreas Críticas da Língua Portuguesa*, Lisboa, Caminho.
- Piel, Joseph M. (1944), *A flexão verbal do português (Estudo de morfologia histórica)*, Biblos 20, 359–404.
- Piel, Joseph M. (1989), *Origens e estruturação histórica do léxico português (1976)*, in: *Estudos de Linguística Histórica Galego-Portuguesa*, Lisboa, Imprensa Nacional-Casa da Moeda, 9–16.
- Pinto, Adelina Angélica (1981), *A africada č em português: Estudo sincrónico e diacrónico*, Boletim de Filologia 26, 139–192.
- Pinto, Clara (2015), *Para a história da negação: o minimizador «homem» no português antigo*, Estudos de Linguística Galega 7, 109–123.
- Poggio, Rosauta Maria Galvão Fagundes (2002a), *Processos de Gramaticalização de Preposições do Latim ao Português: Uma abordagem funcionalista*, Salvador, EDUFBA.
- Poggio, Rosauta Maria Galvão Fagundes (2002b), *Comparação entre algumas preposições portuguesas documentadas no século XVI e no século XIV*, in: Rosa Virgínia Mattos e Silva/Américo Venâncio Lopes Machado Filho (edd.), *O Português Quinhentista: Estudos Lingüísticos*, Salvador/Feira de Santana, EDUFBA/UEFS, 217–236.
- Ramos Alfajarín, Joan Rafael (2000), *«Ésser», «estar» i «haver-hi» en català antic: estudi sintàctic i contrastiu*, València/Barcelona, Institut Interuniversitari de Filologia Valenciana/Publicacions de l'Abadia de Montserrat.
- Révah, Israel Salvator (1958), *L'évolution de la prononciation au Portugal et au Brésil du XV^e siècle à nos jours*, in: *Anais do Primeiro Congresso Brasileiro de Língua Falada no Teatro*, Rio de Janeiro, Ministério da Educação e Cultura, 387–399.
- Ribeiro, Ilza (1995a), *A Sintaxe da Ordem no Português Arcaico: O Efeito V2*, tese de Doutorado, Campinas, SP, Universidade de Campinas.
- Ribeiro, Ilza (1995b), *Evidence for a Verb-second Phase in Old Portuguese*, in: Adrian Battye/Ian Roberts (edd.), *Clause Structure and Language Change*, Oxford/New York, Oxford University Press, 110–139.
- Ribeiro, Ilza (1996), *A formação dos tempos compostos: a evolução histórica das formas «ter», «haver» e «ser»*, in: Mary A. Kato/Ian Roberts (edd.), *Português Brasileiro: Uma viagem diacrônica*, Campinas, SP, Editora da UNICAMP, 343–386.
- Rinke, Esther (2007), *Syntaktische Variation aus synchronischer und diachronischer Perspektive: Die Entwicklung der Wortstellung im Portugiesischen*, Frankfurt, Vervuert/Iberoamericana.
- Rinke, Esther (2009), *Verb Placement in Old Portuguese*, in: Andreas Dufter/Daniel Jacob (edd.), *Focus and Background in Romance Languages*, Amsterdam/Philadelphia, Benjamins, 309–332.
- Rinke, Esther (2010), *A combinação de artigo definido e pronome possessivo na história do português*, Estudos de Linguística Galega 2, 121–139.
- Rojas, Eunice (2004), *Análisis Comparativo-Diacrónico de los Usos de «Ser» y «Estar» en Español y en Catalán*, dissertação de Mestrado, Athens, Georgia, University of Georgia.

- Salvi, Giampaolo (1990), *La sopravvivenza della legge di Wackernagel nei dialetti occidentali della Penisola Iberica*, *Medioevo Romanzo* 15, 177–210.
- Salvi, Giampaolo (2000), *La formazione del sistema V2 delle lingue romanze antiche*, *Lingua e Stile* 35, 665–692.
- Saramago, João (1992), *Le parler de l'île de Corvo – Açores*, *Géolinguistique*, Hors série n° 1, Grenoble, Centre de Dialectologie, Université Stendhal, Grenoble III/Centro de Linguística da Universidade de Lisboa, INIC.
- Saramago, João/Segura, Luísa (2001), *Variedades dialectais portuguesas*, in: Maria Helena Mira Mateus (ed.), *Caminhos do Português: Exposição Comemorativa do Ano Europeu das Línguas (Catálogo)*, Lisboa, Biblioteca Nacional, 221–237.
- Scida, Emily (2004), *The Inflected Infinitive in Romance Languages*, New York/London, Routledge.
- Segura, Luísa (1988), *A Fronteira Dialectal do Barlavento do Algarve*, Lisboa, Instituto Nacional de Investigação Científica.
- Segura, Luísa (2013), *Variedades dialectais do Português Europeu*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 5, Lisboa, Fundação Calouste Gulbenkian, 85–142.
- Segura, Luísa/Saramago, João (1999), *Açores e Madeira: autonomia e coesão dialectais*, in: Isabel Hub Faria (ed.), *Lindley Cintra: Homenagem ao Homem, ao Mestre e ao Cidadão*, Lisboa, Cosmos/Faculdade de Letras da Universidade de Lisboa, 707–738.
- Silva, Augusto Soares (2012), *Stages of grammaticalization of causative verbs and constructions in Portuguese, Spanish, French and Italian*, *Folia Linguistica* 42, 513–552.
- Silva, Giselle Machline de Oliveira e (1982), *Estudo da Regularidade na Variação dos Possessivos no Português do Rio de Janeiro*, tese de Doutorado, Rio de Janeiro, Universidade Federal do Rio de Janeiro.
- Silva, Giselle Machline de Oliveira e (1996), *Realização facultativa do artigo definido diante de possessivo e de patronímico*, in: Giselle Machline de Oliveira e Silva/Maria Marta Scherre (edd.), *Padrões Sociolinguísticos*, Rio de Janeiro, tempo Brasileiro, 119–145.
- Silva, Rosa Virgínia Mattos e (1989), *Estruturas Trecentistas: Elementos para uma Gramática do Português Arcaico*, Lisboa, Imprensa Nacional – Casa da Moeda.
- Silva, Rosa Virgínia Mattos e (1991), *O Português Arcaico: Fonologia*, São Paulo, Contexto.
- Silva, Rosa Virgínia Mattos e (1994a), *Para uma caracterização do período arcaico do português*, *D.E.L.T.A.* 10, 247–276.
- Silva, Rosa Virgínia Mattos e (1994b), *O Português Arcaico: Morfologia e Sintaxe*, São Paulo, Contexto.
- Silva, Rosa Virgínia Mattos e (2002a), *A definição da oposição «ser»/«estar» em estruturas atributivas nos meados do século XVI*, in: Rosa Virgínia Mattos e Silva/Américo Venâncio Lopes Machado Filho (edd.), *O Português Quinhentista: Estudos Lingüísticos*, Salvador/Feira de Santana, EDUFBA/UEFS, 103–117.
- Silva, Rosa Virgínia Mattos e (2002b), *Vitórias de «ter» sobre «haver» nos meados do século XVI: usos e teoria em João de Barros*, in: Rosa Virgínia Mattos e Silva/Américo Venâncio Lopes Machado Filho (edd.), *O Português Quinhentista: Estudos Lingüísticos*, Salvador/Feira de Santana, EDUFBA/UEFS, 119–142.
- Silva, Rosa Virgínia Mattos e (2002c), *A variação «ser»/«estar» e «haver»/«ter» nas Cartas de D. João III entre 1540 e 1553: comparação com os usos coetâneos de João de Barros*, in: Rosa Virgínia Mattos e Silva/Américo Venâncio Lopes Machado Filho (edd.), *O Português Quinhentista: Estudos Lingüísticos*, Salvador/Feira de Santana, EDUFBA/UEFS, 143–160.
- Silva, Rosa Virgínia Mattos e (2008a), *O Português Arcaico: Uma Aproximação*, vol.1: *Morfologia e Sintaxe*, Lisboa, Imprensa Nacional – Casa da Moeda.
- Silva, Rosa Virgínia Mattos e (2008b), *O Português Arcaico: Uma Aproximação*, vol. 2: *Sintaxe e Fonologia*, Lisboa, Imprensa Nacional – Casa da Moeda.

- Silva Domínguez, Carme (2002), *Frases Nominais com Posesivo en Galego: Estructura e valores referenciais*, Verba, Anuario Galego de Filoloxía, Anexo 50, Santiago de Compostela, Universidade de Santiago de Compostela.
- Silva-Villar, Luis (2001), *Verbless derivations in historical syntax: A case study of Northwestern Iberian languages*, in: Javier Gutiérrez-Rexach/Luis Silva-Villar (edd.), *Current Issues in Spanish Syntax and Semantics*, Berlin/New York, Mouton de Gruyter, 309–346.
- Sória, Maíra Vasconcellos de Paiva (2013), «Nós», «a gente» e o sujeito nulo de primeira pessoa do plural, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Souto Cabo, José António (2003), *Nas origens da expressão escrita galego-portuguesa: Documentos do séc. XII*, *Diacrítica* 17, 329–385.
- Teyssier, Paul (1981), *Le système des déictiques spatiaux en portugais aux XIV^e, XV^e et XVI^e siècles*, *Cahiers de Linguistique Hispanique Médiévale* 6, 5–39.
- Teyssier, Paul (1982), *História da Língua Portuguesa*, Lisboa, Sá da Costa.
- Vasconcelos, José Leite de (1901), *Esquisse d'une dialectologie portugaise*, Paris/Lisboa, Aillaud.
- Vasconcelos, José Leite de (1928), *Opúsculos II: Dialectologia (Parte I)*, Coimbra, Imprensa da Universidade.
- Vasconcelos, José Leite de (⁴1966, ¹1911), *Lições de Filologia Portuguesa*, Rio de Janeiro, Livros de Portugal.
- Vasconcelos, José Leite de (1985), *Opúsculos VI: Dialectologia (Parte II)*, Lisboa, Imprensa Nacional – Casa da Moeda.
- Vázquez Cuesta, Pilar/Luz, M. Albertina Mendes da (1971), *Gramática Portuguesa*, Madrid, Gredos.
- Viana, Aniceto dos Reis Gonçalves (³1973, ¹1883), *Essai de phonétique et de phonologie de la langue portugaise, d'après le dialecte actuel de Lisbonne*, in: *Estudos de Fonética Portuguesa*, Lisboa, Imprensa Nacional-Casa da Moeda, 83–152.
- Viana, Aniceto dos Reis Gonçalves (1903), *Portugais: phonétique et phonologie, morphologie, textes*, Leipzig, Teubner.
- Wanner, Dieter (1991), *The Tobler-Mussafia Law in Old Spanish*, in: Héctor Campos/Fernando Martínez-Gil (edd.), *Current Studies in Spanish Linguistics*, Washington, DC, Georgetown University Press, 313–378.
- Wanner, Dieter (2014), *The position of Ibero-Romance in the Romania and of Portuguese within Ibero-Romance*, in: Patrícia Amaral/Ana Maria Carvalho (edd.), *Portuguese-Spanish Interfaces: Diachrony, synchrony, and contact*, Amsterdam/Philadelphia, Benjamins, 11–34.
- Williams, Edwin (³1975, ¹1938), *Do Latim ao Português: Fonologia e Morfologia Históricas da Língua Portuguesa*, Brasília, Instituto Nacional do Livro.



Parte 1: Visões gerais

Tjerk Hagemeijer

2 O português em contacto em África

Abstract: Em razão da presença portuguesa em África a partir do século XV e da consequente adaptação do português (europeu) a novas ecologias linguísticas, emergiram neste continente, em dois períodos históricos distintos, dois grupos de línguas crioulas de base lexical portuguesa, os crioulos da Alta Guiné e do Golfo da Guiné, assim como um conjunto de novas variedades do português, faladas nas antigas colónias portuguesas. Neste capítulo, propomo-nos, por um lado, descrever as condições históricas e sociolinguísticas em que estas novas línguas e variedades surgiram e se desenvolveram e, por outro lado, discutir algumas das tendências de variação linguística que têm vindo a ser descritas para o português em África, analisando-as quer à luz do contacto com línguas crioulas e línguas bantu quer à luz de mudança interna.

Keywords: variedades africanas do português, crioulos de base lexical portuguesa em África, línguas bantu, contacto de línguas, variação e mudança linguísticas

1 Introdução

A presença portuguesa na África subsaariana remonta ao século XV, quando se inicia o reconhecimento do continente africano por via marítima. Até ao século XIX, este reconhecimento resulta sobretudo na ocupação dispersa de zonas costeiras e ilhas, através de entrepostos, feitorias e fortalezas, dedicados a atividades económicas, tais como o tráfico de escravos e a produção de açúcar. O século XIX, e especialmente o período que se segue à Conferência de Berlim (1884–1885), marca o início da colonização efetiva de África.

A presença histórica da língua portuguesa em África teve várias consequências linguísticas. Em primeiro lugar, do contacto entre o português e diferentes línguas africanas continentais surgiu, no decorrer dos séculos XV e XVI, um conjunto de línguas crioulas de base lexical portuguesa na Alta Guiné e no Golfo da Guiné. Em segundo lugar, no século XX, assistimos à emergência e gradual consolidação de novas variedades de português, fruto da colonização efetiva dos espaços africanos. A terceira consequência, porventura a menos conhecida e que não será aqui desenvolvida, consiste na integração de léxico de origem portuguesa em inúmeras línguas em África com as quais o português teve, historicamente, contacto, quer línguas faladas nas antigas colónias portuguesas quer outras.

A formação destes crioulos e a emergência destas novas variedades de português resultam de diferentes graus de reestruturação linguística num cenário de aquisição de língua não materna (L2), os quais dependem de fatores sociais e linguísticos

específicos, tais como o grau de acesso e exposição à língua-alvo (o português), aspetos demográficos e a tipologia das línguas em contacto. No caso da criouliização, a necessidade de falantes de diferentes línguas criarem uma plataforma de comunicação comum resulta na emergência de um código L2 que é nativizado. Na maioria dos casos, trata-se de um processo relativamente rápido em que as línguas ancestrais são gradualmente abandonadas em detrimento da nova língua. A situação das novas variedades é distinta, uma vez que as diferentes línguas maternas (L1) continuam a coexistir com o português, a língua promovida por políticas linguísticas oficiais.

Este capítulo segue a seguinte estrutura. A secção 2 apresenta sumariamente os dois grupos de crioulos de base lexical portuguesa falados em África. As secções seguintes incidem sobre o português em África. Na caracterização sociolinguística e histórica contida na secção 3 discute-se o percurso das variedades africanas do português, com especial atenção para o período pós-independência e a questão da transição de L2 para L1 em algumas variedades. A secção 4, que está dividida em diferentes subsecções, aborda, primeiro, aspetos relacionados com o estado da investigação sobre as variedades africanas do português (secção 4.1). Em seguida, nas secções 4.2–4.5, discutimos algumas tendências no domínio da morfossintaxe e da sintaxe que caracterizam as variedades africanas do português e a sua relação com estruturas observadas nas línguas com que estas variedades estão em contacto.

2 Crioulos de base lexical portuguesa em África

O povoamento e o desenvolvimento da Alta Guiné e do Golfo da Guiné, a partir da segunda metade do século XV, apresentam semelhanças significativas do ponto de vista administrativo, económico, social e religioso (e.g. Seibert 2012). A importação maciça de escravos de diferentes origens etnolinguísticas para satisfazer os ciclos económicos da produção de açúcar e do tráfico de escravos resultou na génese de dois grupos de crioulos de base lexical portuguesa de formação independente, os crioulos da Alta Guiné e os crioulos do Golfo da Guiné, fruto do contacto entre o português e diferentes línguas africanas continentais.

Os crioulos da Alta Guiné incluem o kabuverdianu (crioulo de Cabo Verde), o kriol (crioulo da Guiné-Bissau) e o kriyol ou lingu kriston (crioulo de Casamansa) na zona de Ziguinchor, no Senegal. Alguns estudiosos propõem também a inclusão, neste grupo, do papiamentu, falado nas ilhas caribenhas Aruba, Bonaire e Curaçau (e.g. Quint 2000; Jacobs 2010). O debate sobre o berço da proto-língua que deu origem a estes crioulos tem-se centrado em duas hipóteses: a hipótese continental (rios ou praças da Guiné) e a hipótese insular (ilha de Santiago) (ver, por exemplo, a discussão em Jacobs 2010). Em termos linguísticos, para além do português, a língua de superstrato, parecem ter tido um papel preponderante línguas de substrato, como o wolof, da família atlântica, e o mandinga, da família mande. À exceção do crioulo de

Casamansa, cujo número de falantes é baixo e está a diminuir, os crioulos da Alta Guiné apresentam uma grande vitalidade linguística, inclusive nas diásporas.

Os crioulos do Golfo da Guiné, por sua vez, incluem o angolar, o lung'ie (lit. 'língua da ilha' – o crioulo do Príncipe), o santome (também designado forro, crioulo de S. Tomé), todos falados em S. Tomé e Príncipe, e ainda o fa d'ambô (lit. 'fala de Ano Bom'), falado na Guiné Equatorial. Apesar de estas quatro línguas não serem mutuamente inteligíveis, existe abundante evidência histórica e linguística a favor de uma origem partilhada (e.g. Hagemeijer 2011). O proto-crioulo do Golfo da Guiné ter-se-á formado no nordeste da ilha de S. Tomé, tendo-se ramificado no século XVI, devido à dispersão populacional, para as ilhas do Príncipe e de Ano Bom e, no caso do angolar, em consequência da fuga de escravos, para outras zonas de S. Tomé. No que respeita ao substrato dos crioulos do Golfo da Guiné, destaca-se a importância do edo (família edóide, Nigéria), bem como do quicongo e do quimbundo (línguas bantu do Congo e de Angola). Diferentemente dos crioulos da Alta Guiné, a vitalidade dos crioulos do Golfo da Guiné encontra-se atualmente ameaçada em maior ou menor grau. O lung'ie, em particular, terá hoje em dia no máximo algumas centenas de falantes ativos.

Apesar de o léxico predominantemente de origem portuguesa constituir o elo de ligação entre os crioulos da Alta Guiné e os crioulos do Golfo da Guiné, é de realçar que as diferenças linguísticas entre os dois grupos de línguas são substanciais (e.g. Ferraz 1987; Hagemeijer/Alexandre 2012), revelando o impacto de histórias sociais e línguas de substrato distintas no período de formação das protolínguas que surgiram na Alta Guiné e no Golfo da Guiné. Para uma caracterização (socio)linguística, amostras e fontes para estes crioulos, remetemos para Michaelis et al. (2013a; 2013b; 2013c) e Cardoso/Hagemeijer/Alexandre (2015).

3 Caracterização histórica e sociolinguística do português em África

Até às independências dos países africanos de língua oficial portuguesa (1975), o português era essencialmente uma L2 pouco falada por grande parte das respetivas populações. Esta situação veio a alterar-se bastante nas últimas décadas e atualmente pode fazer-se a distinção entre países onde o português não constitui a língua materna da população e países onde o português apresenta um crescente número de falantes nativos.

A primeira situação verifica-se em Cabo Verde, onde o crioulo cabo-verdiano é a língua materna exclusiva da população, e na Guiné-Bissau, onde o crioulo guineense reparte o estatuto de língua materna com cerca de 20 línguas africanas pertencentes às famílias atlântica (e.g. balanta, fula, papel e manjaco) e mande (e.g. mandinga), do Níger-Congo (Couto 2009). De acordo com dados do Instituto Nacional de Estatística da Guiné-Bissau relativos a 2009, o português é falado por 27,1% da população e o

crioulo por 90,4% da população, subentendendo-se que estas percentagens abrangem tanto falantes L1 como L2. Em Cabo Verde e na Guiné-Bissau, os crioulos desempenham o papel de língua de unidade nacional.

A segunda situação caracteriza o contexto linguístico de Angola, Moçambique e S. Tomé e Príncipe, onde o número de falantes L1 e L2 do português tem vindo a crescer rapidamente e onde esta língua desempenha o papel de língua de unidade nacional. De acordo com dados do Instituto Nacional de Estatística de Angola, obtidos no Censo de 2014, num universo de aproximadamente 25,7 milhões de angolanos 71,15% falam português, um número que supera largamente as percentagens das línguas bantu mais faladas, o umbundo, o quicongo e o quimbundo, com percentagens de falantes de, respetivamente, 22,96%, 8,24% e 7,82%. Em relação a Moçambique, o censo moçambicano relativo a 2007 mostra que 10,7% da população falava o português como língua materna e 39,7% como língua não materna (e.g. Gonçalves 2012). De acordo com o mesmo censo, o português é hoje a segunda língua materna mais falada de Moçambique, ficando apenas atrás do macua (falado como L1 por 25,2% da população). No censo anterior, de 1997, no entanto, havia mais três línguas bantu que apresentavam mais falantes L1 do que o português (changana, lomue e sena). Em S. Tomé e Príncipe, o censo de 2011 indica uma percentagem de cerca de 91% de falantes do português, sem distinção entre falantes nativos e não nativos. De acordo com o mesmo censo, o forro, o crioulo maioritário em S. Tomé, era falado, em 2011, apenas por uns 33% da população, corroborando o papel dominante do português neste país.¹

A par de alguns problemas com a origem e a consistência dos dados estatísticos acima referidos, Gonçalves (2012) alerta para o facto de estes representarem uma autoavaliação das competências linguísticas dos recenseados. Ainda assim, fica claro que as independências representam o ponto de viragem: em 1975, apenas 1 a 2% da população angolana falava o português como língua materna e 15 a 20% como língua não materna (Endruschat 1990); em Moçambique, estas percentagens eram de, respetivamente, 1,2% e de 24,4% em 1980 (Gonçalves 2013); e em S. Tomé e Príncipe, cerca de 63% da população falava português (L1/L2) em 1981, o que mostra que a transição para o português neste país já se vinha a consolidar no tempo colonial.

A crescente nativização da antiga língua colonial em Angola, Moçambique e S. Tomé e Príncipe é um fenómeno que não se verifica em relação a outras línguas ex-coloniais em África e pode ser explicado através de fatores comuns e diferenciados. Fatores como o estatuto neutro, de L2, do português face à fragmentação linguística das L1, a massificação do ensino, o aumento da mobilidade social, a timidez das

¹ Numa publicação em linha de 2010, o Observatório da Língua Portuguesa refere as seguintes percentagens de falantes do português: Angola (60%), Cabo Verde (95%), Guiné-Bissau (60%), Moçambique (50%), S. Tomé e Príncipe (95%), referindo que estas percentagens «foram consideradas com base em elementos obtidos junto das diferentes delegações nacionais participantes na IV Conferência Estatística da CPLP realizada em Lisboa a 1 e 2 de Julho de 2010». Embora não seja explicitado, estes números integram falantes L1 e L2.

políticas linguísticas efetivas em prol das línguas nacionais, assim como o prestígio do português enquanto língua de ascensão social, terão sido fundamentais na maior difusão do português.

No caso de Angola, em concreto, a presença portuguesa era muito significativa no período que antecedeu a independência, o que terá contribuído para um maior enraizamento do português como língua franca. Depois da independência, em consequência do longo período do conflito armado (1975–2003), o êxodo de populações falantes de diferentes línguas bantu do interior para o litoral, em particular para a capital, Luanda, reforçou o papel do português como língua franca. A presença portuguesa em Moçambique, por outro lado, foi menor no tempo colonial e o conflito armado que se seguiu à independência foi menos desestruturante do que em Angola, o que poderá ajudar a explicar o crescimento mais lento do número de falantes do português L1/L2 em Moçambique.

Em S. Tomé e Príncipe falavam-se essencialmente as línguas crioulas até meados do século XIX. Contudo, com a abolição da escravatura, o poder colonial instituiu o chamado regime do contrato, que conduziu um grande contingente de mão de obra angolana, cabo-verdiana e moçambicana até às ilhas de S. Tomé e Príncipe para trabalharem nas culturas de café e de cacau. O impacto demográfico foi de tal forma maciço que esta nova população rapidamente ultrapassou em número a população nativa das ilhas. Uma vez que a população nativa, os forros, se mantinha deliberadamente segregada destes contratados e dos seus descendentes, os chamados tongas (Rougé 1992), a população recém-chegada às ilhas começou a adotar o português como L2 e língua franca, em detrimento do forro. Devido ao baixo índice de repatriamentos de cabo-verdianos, o crioulo de Cabo Verde é ainda hoje amplamente falado em S. Tomé e Príncipe, tendo inclusive mais falantes do que o angolar e o lung'ie, dois dos crioulos autóctones, de acordo com o censo de 2011. Tal como em Angola e Moçambique, a consolidação, hegemonia e a nativização do português em S. Tomé e Príncipe é um fenómeno que está ligado ao período pós-independência (Gonçalves/Hagemeyer 2015).

A tendência de subida das percentagens de falantes (nativos) do português no contexto multilingue de Angola, Moçambique e S. Tomé e Príncipe tem necessariamente consequências para as outras línguas faladas nesses espaços. A diminuição acentuada da percentagem de falantes do forro, em S. Tomé, nas últimas décadas, é um exemplo paradigmático. Gonçalves (2012) explica que em Moçambique, em particular no seio da população mais escolarizada das cidades, a partir da entrada para a escola, o português passa frequentemente a ser a língua dominante de populações cuja L1 na primeira infância era uma língua bantu. A exposição reduzida à L1 pode ser considerada uma etapa intermédia num processo coletivo de mudança e erosão linguísticas, o qual está a ocorrer igualmente em Angola e está praticamente concluído em S. Tomé e Príncipe. Para travar este processo, Moçambique, em particular, tem adotado políticas linguísticas efetivas que promovam as línguas nacionais, tais como o projeto de escolarização bilingue, implementado em 1993, que envolve 16 línguas bantu.

Em suma, fatores sociolinguísticos específicos e a democratização da língua portuguesa a seguir às independências tiveram um impacto direto nas relações diglósicas em Angola, Moçambique e S. Tomé e Príncipe, onde a situação de diglossia estável que caracterizava o período colonial deu lugar a uma de diglossia instável que cada vez mais irradia dos centros urbanos para o interior.

4 Caracterização linguística do português em África

Esta secção tem por objetivo descrever alguns traços linguísticos em que as variedades africanas do português se distanciam do português europeu e convergem ou divergem entre si e discutir o papel do contacto de línguas e de processos mais gerais nas estruturas observadas. Focaremos em especial as áreas da morfossintaxe e da sintaxe, com maior atenção para as variedades angolana, moçambicana e são-tomense, onde há um uso mais alargado do português (L1 e L2). Segue-se, primeiro, uma secção que aborda algumas questões gerais relacionadas com o estado do estudo científico do português em África.

4.1 A investigação sobre as variedades africanas do português: questões prévias

A investigação linguística sobre as variedades africanas do português restringe-se essencialmente ao período pós-independência, que marca o início da massificação e consolidação do português. Contudo, existe uma discrepância significativa entre as diferentes variedades africanas do português do ponto de vista da sua descrição. Comparando as duas variedades com mais peso demográfico e geográfico, conclui-se que o estudo científico da variedade moçambicana está mais desenvolvido do que o da variedade angolana, ao passo que as variedades de português de Cabo Verde, de S. Tomé e Príncipe e especialmente da Guiné-Bissau são menos conhecidas, apresentando, de um modo geral, uma produção académica mais recente.² Verifica-se igualmente um desequilíbrio em relação às áreas da gramática investigadas. Têm sido privilegiadas áreas como a sintaxe (e.g. clíticos, duplos objetos, regência verbal) e a morfossintaxe (e.g. concordância de número), ao passo que as áreas da fonética e da fonologia das variedades africanas do português estão praticamente por desbravar.³

² A Cátedra de Português Língua Segunda e Estrangeira da Universidade Eduardo Mondlane (Moçambique) disponibiliza bibliografias atualizadas para cada uma destas variedades (<http://www.catedra-portugues.uem.mz>). Veja-se também Madeira (2008).

³ Apenas algumas descrições mais gerais (e.g. Gonçalves 1997; Mingas 2000) fazem breve referência a questões fonético-fonológicas. Duas exceções a esta regra são as teses de mestrado de Christofolletti

Tendo em conta o amplo espectro de variação linguística que caracteriza cada uma das variedades africanas do português, devido à variabilidade dos fatores sociolinguísticos que tipicamente caracterizam as variedades L2 (ou com um passado recente de L2), é importante determo-nos brevemente na questão dos dados utilizados para a sua descrição e análise. Ao abordar esta questão, Gonçalves (2013, 161) refere que

«[e]sta variabilidade [se] evidencia quer na maneira pouco regular e sistemática com que estes traços se manifestam na produção linguística dos falantes [...], quer na frequente falta de coerência entre as produções dos falantes, orais ou escritas, e os juízos de gramaticalidade que estes emitem quando submetidos a procedimentos experimentais, nomeadamente testes de gramaticalidade».

Por um lado, existem estudos que se baseiam, parcial ou integralmente, em produções escritas, nomeadamente obras literárias. A título ilustrativo, algumas obras de autores como Luandino Vieira ou Jofre Rocha, que procuram reproduzir o português popular de Luanda, têm sido utilizadas em diversos trabalhos sobre o português angolano (e.g. Mingas 2000; Chavagne 2005). Contudo, estes trabalhos literários não constituem fontes seguras para o estudo (representativo) da língua falada (Gonçalves 2013, 162). Nos estudos baseados em textos manuscritos, como cartas ou produções manuscritas em contexto escolar ou académico (e.g. Cabral 2005), será necessário distinguir questões relacionadas com o próprio processo de escrita de características de uma nova gramática. Os estudos baseados em testes (escritos) de produção provocada ou de juízos de gramaticalidade (e.g. Mapasse 2005; Justino 2011), que correspondem tipicamente a um momento de maior consciência linguística, permitem confirmar tendências observadas em dados orais (espontâneos).

Temos, por outro lado, os trabalhos sobre a gramática adulta das variedades africanas do português que privilegiam o uso de dados da oralidade, como por exemplo entrevistas espontâneas ou semiestruturadas, para caracterizar a variação e mudança. Estes dados devem, idealmente, corresponder a processos mais exigentes que conduzam à constituição de corpora pesquisáveis (gravações, transcrições, revisão, anotação, etc.), mas também incluem, com alguma frequência, dados de intuição ou dados que não foram registados em qualquer tipo de suporte. O Panorama do Português Oral de Maputo (Stroud/Gonçalves 1997–2000) e o Corpus África (oral e escrito) do Centro de Linguística da Universidade de Lisboa (e.g. Bacelar do Nascimento et al. 2008) são dois exemplos de projetos que deram um contributo importante para o estudo linguístico das variedades africanas do português. Existem também diversos outros corpora, recolhidos sobretudo no âmbito de teses (e.g. Chavagne 2005; Figueiredo 2010; R. Gonçalves 2010; Inverno 2011). Como as variedades africa-

(2013), sobre a realização de ditongos no português são-tomense, e de Vicente (2009), sobre a consciência fonológica em crianças moçambicanas.

nas do português ainda estão numa fase de mutação mais intensa do que o português europeu e o português brasileiro, reveste-se de especial importância a realização de novas recolhas que possam documentar a sua gradual estabilização.

Um grande número de estudos sobre as variedades africanas do português procura assinalar tendências de divergência da norma do português europeu. Em muitos casos, a robustez destas tendências requer confirmação baseada em estudos quantitativos que também tenham em conta o perfil sociolinguístico dos informantes. O grau de escolarização dos informantes é normalmente apontado como o principal fator de convergência com o português europeu. Gonçalves/Chimbutane (2004, 7) sintetizam esta questão da seguinte forma:

«[a] distribuição e frequência dos traços não-padrão no discurso dos falantes não é idêntica para todos os membros desta comunidade. Com efeito, à semelhança do que acontece com outras línguas ex-coloniais, o português moçambicano apresenta um amplo espectro de variação que inclui desde as subvariedades 'basilectais', mais distantes do padrão europeu, dos falantes com pouco contacto com a língua-alvo, até às subvariedades mais próximas deste padrão, dos falantes mais instruídos».

As línguas com que o português está em contacto em África apresentam tipologias distintas. Em termos de tipologia morfológica geral, os crioulos são línguas isolantes e as línguas bantu aglutinantes. No entanto, é importante realçar que entre os crioulos da Alta Guiné e os crioulos do Golfo da Guiné, por um lado, e entre as diversas línguas bantu, por outro, existem diferenças significativas em domínios gramaticais específicos, o que poderá condicionar a consolidação das gramáticas das variedades africanas do português de forma distinta. A devida exploração destas diferenças (e semelhanças) entre as línguas com que o português está em contacto em África permite compreender melhor o papel do contacto face a processos mais gerais relacionados com a variação e mudança. Nas secções seguintes, é nosso objetivo discutir alguns casos particulares à luz destas questões.

4.2 Morfossintaxe: concordância de número

Uma característica que é frequentemente referida ou aprofundada nos estudos sobre as variedades africanas do português envolve casos em que o uso de morfologia flexional nominal e verbal diverge dos padrões conhecidos no português europeu (e.g. Marques 1983; Gärtner 1989; Gonçalves 1997; Mingas 2000; Chavagne 2005; Figueiredo 2008; 2010; Petter 2008; Inverno 2009; 2011; Lopes/Baxter 2011; Brandão 2011; Jon-And 2011). Os estudos sobre o português angolano e o português moçambicano referem algumas tendências a nível da ausência de concordância sujeito-verbo, tais como o uso do verbo na terceira pessoa do singular em contextos que requerem a terceira pessoa do plural ou a primeira pessoa do singular no português europeu (e.g. Gonçalves 1997; Chavagne 2005), mas é na concordância nominal, e especialmente de

número, que se concentra a maioria dos estudos, sendo o método privilegiado o da sociolinguística variacionista.

Os estudos que analisaram corpora orais permitiram determinar que o cancelamento da marca de número, isto é, tipicamente o apagamento do *-(e)s* do plural, afeta 6,6% dos casos no português são-tomense (Brandão 2011), 13% no português moçambicano e 18% no português cabo-verdiano (Jon-And 2011). Naturalmente, estes resultados estão dependentes do perfil dos informantes: nos referidos estudos, os falantes do português são-tomense eram, em média, mais escolarizados do que os do português moçambicano e do português cabo-verdiano. A relevância da variável escolaridade fica, de resto, patente no referido estudo sobre o português são-tomense, uma vez que a média de 6,6% de cancelamento sobe para 23,5% se considerarmos apenas os informantes com 5 a 8 anos de escolaridade (ensino fundamental) e desce para 1,2% no subgrupo com 12 a 15 anos de escolaridade (ensino superior). Os estudos de Brandão e Jon-And concluíram, sem surpresa, que a variável escolaridade é determinante para os padrões observados, tal como a variável posição linear em relação ao núcleo do sintagma nominal, que prediz que a marcação do plural ocorre tendencialmente mais à esquerda. Dois outros estudos sobre a concordância em comunidades linguísticas específicas de S. Tomé, os tongas de Monte Café (Baxter 2004) e Almoxarife (Figueiredo 2008; 2010), também destacam a importância dessa última variável. No caso do português cabo-verdiano e do português são-tomense, também foi estabelecido que a animacidade do núcleo favorece a marcação de plural, privilegiando nomes com o traço [+humano] em detrimento de nomes com o traço [-humano].

O cancelamento da marca de número é, por vezes, relacionado com a situação de contacto entre o português e outras línguas, onde se encontram, na realidade, estratégias de pluralização bastante diversas. O uso de flexão de número nos crioulos da Alta Guiné é sobretudo condicionado pela animacidade e definitude do núcleo (Baptista 2007). Assim, nomes [+humano] com interpretação definida recebem marcação de plural por sufixação, a não ser que sejam precedidos por elementos flexionáveis (e.g. *omi/omis* ‘homem/homens’; *alguns omi* ‘alguns homens’; *nhas fidju* ‘meus filhos’); o número de nomes [-humano] tende a não ser marcado morfológicamente. Os crioulos do Golfo da Guiné, por sua vez, não apresentam qualquer marcação morfológica de número, recorrendo tipicamente a nomes simples em todas as posições sintáticas. Contudo, para informação que pode ser recuperada discursiva ou contextualmente, estes crioulos apresentam uma estratégia sintática com marca de plural pré-nominal (*inen* (e.g. *inen ome* ‘os homens’), que é obrigatoriamente acompanhada de outros modificadores nominais, em particular do dêitico pós-nominal *se*, quando o nome é [-humano] (e.g. *inen lanza se* ‘as (tais) laranjas’ vs. **inen lanza*) (cf. Alexandre/Hagemeijer 2007).

Nas línguas bantu, por outro lado, a marcação de número segue uma estratégia distinta que consiste na prefixação dos nomes em função de classes nominais que são guiadas, *grosso modo*, por princípios semânticos (e.g. Katamba 2003). Uma língua

bantu típica apresenta em torno de 18 classes nominais: 6 classes emparelhadas, que são compostas por pares singular-plural, e 6 classes não emparelhadas, entre as quais normalmente 3 classes locativas (Katamba 2003, 108). A título exemplificativo, o quimbundo apresenta um total de 18 classes nominais, das quais representámos as primeiras seis em (1). Por exemplo, a classe 1, com plural na classe 2, em (1a), inclui sobretudo nomes com o traço [+humano].

- (1) a. 1 *mú-thu* ‘pessoa’ 2 *á-thú* ‘pessoas’
 b. 3 *mù-xí* ‘árvore’ 4 *mí-xí* ‘árvores’
 c. 5 *dí-lôngà* ‘prato’ 6 *má-lôngà* ‘pratos’ (quimbundo, adapt. de Diarra 1990)

Diversos autores defendem que na aquisição do português angolano (Marques 1983; Gärtner 1989; Mingas 2000; Inverno 2011) e de algumas subvariedades historicamente ligadas a comunidades de contratados bantu em S. Tomé (Baxter 2004; Figueiredo 2008), a marcação de número mais à esquerda, especialmente na posição pré-nuclear adjacente ao núcleo, poderá indicar que há uma relação com sistema de prefixos de classe, dispensando a marcação de número por sufixação no núcleo do sintagma nominal (*os homem*). Em alguns trabalhos mais antigos sobre o português angolano, considerava-se, inclusive, que os artigos do português eram interpretados como prefixos de classe. Contudo, Jon-And (2011), na sua comparação entre o português cabo-verdiano e o português moçambicano, defende que, face à tipologia distinta de marcação do número no crioulo cabo-verdiano e nas línguas bantu, as semelhanças encontradas no domínio da marcação de número nestas duas variedades do português favorecem uma hipótese mais direcionada para processos universais que guiam o processo de aquisição L2.

Ainda em relação às línguas bantu, estas apresentam tipicamente sintagmas nominais de núcleo inicial, como ilustra o seguinte exemplo do quicongo (de São Salvador, no norte de Angola):⁴

- (2) O-ma-tadi ma-ma ma-mpembe ma-mpwena ...
 AUM-8-pedra 8-DEM 8-branco 8-enorme
 ‘Estas pedras brancas enormes...’ (quicongo, adapt. de Bentley 1887, 526)

Este exemplo do quicongo mostra o núcleo do sintagma nominal *tadi* ‘pedra’, com o prefixo da classe 8 (plural). A concordância de classe é marcada nos modificadores pós-nominais, aqui um demonstrativo e dois adjetivos. Como a noção de número é fundamental nestas línguas e a ordem de palavras do sintagma nominal distinta da do português, o cancelamento de número observado no português angolano e no

⁴ As siglas usadas neste capítulo são, por ordem alfabética: 1, 2, 8, 12, etc.=prefixos de classe nas línguas bantu; AUM=aumento (bantu); CONN=conector; CS=concordância de sujeito; DEF=definido; DEM=demonstrativo; FOC=foco; IMP=pronome impessoal; LOC=locativo; NEG=negação; OBJ=objeto; PAS=passivo; PL=plural; PST=passado; REL=relativo; SG=singular; T/A=tempo/aspecto.

português moçambicano dificilmente poderá ser considerado uma consequência direta do contacto.

4.3 Estrutura argumental dos verbos

As alterações às grelhas argumentais dos verbos também estão entre os aspetos linguísticos que mais têm sido realçados nos diversos trabalhos sobre o português em África. Estas alterações afetam frequentemente as preposições selecionadas, como mostram os seguintes exemplos do português moçambicano.⁵

- (3) Eu tinha de ir participar um curso na Suécia. (PM, Gonçalves 1997, 47)
- (4) Os pais escondem os filhos a verdade. (PM, P. Gonçalves 2010, 100)
- (5) Foi na altura que eu separei com os meus pais. (PM, Gonçalves 1997, 49)
- (6) Ensina a criança respeitar aos pais. (PM, Gonçalves 1997, 47)

A ausência de preposição nos exemplos (3–4) tem como consequência a transitivização direta das grelhas argumentais dos verbos *participar* e *esconder*. Em (5), é selecionada uma preposição distinta daquela que é exigida pela norma do português europeu e no exemplo (6) verifica-se um caso de inserção de uma preposição. Os casos de omissão e substituição são mais frequentes nas variedades africanas do português do que os casos de inserção, mas coexistem tendências diferentes. Com base num corpus oral do português são-tomense, R. Gonçalves (2010) conclui que, em contextos desviantes da norma do português europeu, as omissões de preposições representam sensivelmente o dobro dos casos das substituições. Já num corpus de produções espontâneas, escritas, do português angolano Cabral (2005) verifica que há mais casos de substituição do que de omissão, o que se explica, em boa parte, pela substituição generalizada, nesta variedade, de *a* por *em*, que será tratada abaixo.

A reestruturação das grelhas argumentais com argumento dativo, como em (4) acima, também ocorre nas outras variedades africanas do português, como se mostra a seguir:

- (7) Entrega senhor uma cerveja. (PST, R. Gonçalves 2010, 34)
- (8) Ainda temos que pagar nos professores. (PA, Cabral 2005, 132)

⁵ Nos exemplos apresentados ao longo do texto, as siglas indicam a variedade a que o exemplo pertence: CCV=crioulo de Cabo Verde; CST=crioulo de S. Tomé; PA=português de Angola; PM=português de Moçambique; PST=português de S. Tomé e Príncipe.

À semelhança do português moçambicano, o português são-tomense, apresenta uma tendência clara para a construção de duplo objeto, apresentada em (7), que consiste na omissão do marcador dativo *a* que introduz o objeto indireto na norma do português europeu e na reorganização dos argumentos internos: o argumento correspondente ao objeto indireto, tipicamente um beneficiário com o traço [+humano] (os filhos, senhor), precede o objeto direto. Nestes contextos, o português angolano, por outro lado, apresenta uma solução em que o marcador de caso dativo *a* dá lugar à preposição *em* (8). Os estudos sobre o português moçambicano e português são-tomense (e.g. P. Gonçalves 2010; R. Gonçalves 2010) relacionam estas estruturas com o papel do contacto. Os exemplos em (9–10) mostram que as línguas bantu (do sul) de Moçambique, neste exemplo o changana, e o forro de S. Tomé apresentam a construção de duplo objeto.

- (9) Mamani a-nyik-ile [n'wana] [pawa]. (changana, P. Gonçalves 2010, 103)
 1.mamã CS-dar-PST 1.criança 5.pão
 'A mamã deu (algum) pão à criança.'

- (10) Sapatu ka da [ngê] [kalu n'ope]. (CST)
 sapato T/A dar pessoa calo LOC-pé
 'Os sapatos dão calos às pessoas.'

O português moçambicano e o português são-tomense divergem, no entanto, quanto à possibilidade de ocorrência de passivas dativas, uma construção correlacionada com a construção de duplo objeto em que o argumento correspondente ao objeto indireto é promovido a sujeito da frase. No português moçambicano e nas línguas bantu com que este está em contacto, as passivas dativas são legitimadas (cf. 11–12), ao passo que estas estruturas não estão atestadas no português são-tomense e no forro, reforçando a evidência a favor de uma análise de contacto.

- (11) Os filhos são escondidos a verdade. (PM, P. Gonçalves 2010, 100)

- (12) N'wana a-nyik-iw-ile pawa. (changana, P. Gonçalves 2010, 106)
 1.criança CS-dar-PAS-PST 5.pão
 '*A criança foi dada pão.'

Em relação ao português angolano, a generalização da preposição *em* é referida por diversos autores (e.g. Marques 1983; Mingas 2000; Cabral 2005; Chavagne 2005), ocorrendo especialmente em estruturas com argumento dativo (cf. (8)) e verbos de movimento (e.g. *ir, vir, voltar, chegar*).

- (13) Vamos em casa. (PA, Chavagne 2005, 225)

- (14) quando ele veio na nossa terra [...] (PA, Chavagne 2005, 225)

Diversos autores atribuem este uso ao contacto com o quimbundo (Marques 1983; Mingas 2000). Tal como muitas línguas bantu, esta língua apresenta três prefixos de

classe locativos (e.g. Chatelain 1888–1889; Diarra 1990), com funções específicas, que se associam aos nomes. O prefixo locativo *ku-* do quimbundo (ou *ko-* do umbundo), da classe nominal 17, ocorre tipicamente com a ideia de interioridade (dentro de, em), com verbos de movimento e também precede o argumento dativo:

- (15) Mwene w-ala ku-bata (quimbundo, Mingas 2000, 75)
 ele CS-estar LOC-casa
 ‘Ele está em casa.’
- (16) Ndé ku-Palaia. (quimbundo, Chatelain 1894, 30)
 vai LOC praia
 ‘Vai à praia.’
- (17) Uvutuk’ê ku-itadi. (quimbundo, Chatelain 1894, 36)
 ele-voltar LOC-jardim
 ‘Ele volta ao jardim.’
- (18) a. Nga-bana ma-divulu ku-a-thu. (quimbundo, Afonso Miguel [c.p.])
 1SG.PST-dar 6-livro LOC-2-pessoa
 b. Nda-ka a-livulu ko-ma-nu. (umbundo, Félix Chinjengue [c.p.])
 1SG.PST-dar 6-livroLOC-2-pessoas
 ‘Dei os livros às pessoas.’

Em síntese, a reestruturação das grelhas argumentais face à norma do português europeu é um fenómeno que afeta todas as variedades africanas do português. As construções com dois argumentos internos (objeto direto e objeto indireto), em particular, parecem mostrar que o português se aproximou das estruturas gramaticais das línguas com que está em contacto.

4.4 Estratégias de relativização

Tal como as variedades portuguesa e brasileira, as variedades africanas do português apresentam contextos em que as estratégias de relativização se afastam das estratégias canónicas. Diversos trabalhos sobre o português moçambicano (e.g. Chimbutane 1996; Brito 2001) mostram que esta variedade parece privilegiar a estratégia resuntiva em relativas de objeto direto (19), de oblíquo (20) e de genitivo (21). Nestes casos, o antecedente da relativa é normalmente retomado por um pronome (clítico ou forte) cujos traços de género e número concordam com o antecedente.

- (19) Há passagens do texto que podemos vê-las. (PM, Chimbutane 1996, 241)
- (20) Havia rapazes que nós não brincávamos com eles. (PM, Chimbutane 1996, 245)
- (21) Os meios de transporte que vi os nomes deles são... (PM, Brito 2002, 330)

No contexto de relativas de sintagma preposicional nas variedades africanas do português estudadas em Alexandre/Gonçalves/Hagemeijer (2011a; 2011b), verifica-se uma maior incidência da estratégia cortadora – cf. (22)–(23) –, que se caracteriza pelo apagamento da preposição, tanto em contextos de sintagmas preposicionais selecionados (22) como de sintagmas preposicionais adjuntos (23).

(22) Há determinados trabalho que eu me envolvi. (PST, R. Gonçalves 2010, 144)

(23) Depois cheguei um momento que eu vi que era vazio. (PST, R. Gonçalves 2010, 131)

Nos dados do português são-tomense, observou-se uma percentagem de cortadoras mais elevada do que a de realizações canônicas (com arrastamento da preposição), seguindo, de resto, a tendência referida na secção anterior sobre o elevado número de casos de omissão de preposições nas grelhas de subcategorização. Em informantes menos escolarizados do português são-tomense foram ainda encontrados alguns exemplos da estratégia de cópia defetiva (Alexandre 2012), em que o antecedente da relativa é retomado por um pronome invariável que corresponde sempre à terceira pessoa do singular masculino, isto é, não partilha traços de género e número com o antecedente, como ilustra (24).

(24) A própria escola que eu estudei nele. (PST, Alexandre/Gonçalves/Hagemeijer 2011a, 18)

À exceção desta última estratégia, as estratégias cortadora e resuntiva também ocorrem no português europeu e no português brasileiro, sendo a cortadora de longe a mais comum (e.g. Tarallo 1985; Peres/Móia 1995; Alexandre 2000). Tem-se defendido que esta mudança convergente no domínio das relativas nas variedades de português está correlacionada com a generalização do relativizador *que*, em detrimento do uso de pronomes relativos que apresentam traços de concordância (*quem*, *cujo*, *qual*, etc.), pese embora *cujo* estar atestado nas relativas genitivas do português moçambicano (Brito 2001; 2002).

Nas línguas bantu, as relativas que não são de sujeito apresentam diferentes estratégias no que se refere à relação que é estabelecida entre o antecedente da relativa e a estrutura interna da relativa (e.g. Henderson 2007). Há línguas, como o zulu (África do Sul), em que o antecedente da relativa é sistematicamente retomado sob a forma de um prefixo de concordância de objeto (e.g. Cheng/Downing 2010). Embora não obrigatória, esta estratégia é também frequentemente utilizada numa língua com o changana.

(25) A ngwana_i ley_i ni-nga-yi-rhandza yi-f-ile.
 DEF 7.cão 7.DEM 1SG-REL-7-amar 7-morrer-PST
 ‘Este cão que amei morreu.’ (changana, adapt. de Duarte 2011, 89)

Nesta relativa de objeto, o antecedente da relativa *a ngwana ley* ‘este cão’, pertencente à classe nominal 7, é retomada sob a forma de um prefixo de concordância de objeto

da classe 7, *yi*, que precede o verbo (ou, mais precisamente, o radical verbal), resultando numa estrutura equivalente a ‘este cão que eu o amei morreu’. Nas línguas de Angola, nomeadamente no quimbundo e no umbundo, por outro lado, o antecedente da relativa não é retomado em contextos semelhantes.

(26) Ó i-mbwá í-ngà-sùmbù. (quimbundo, adapt. de Diarra 1990, 32)
AUM 9-cão 9-1SG.PST-comprar ‘O cão que comprei.’

(27) Tánga ú-kánda e-kamba ly-á-ngè ly-á-sonefia.
Ler 3-carta 5-amigo 5-CONN-POSS 5-REL-escrever
‘Leia a carta que o meu amigo escreveu.’ (umbundo, adapt. de Schadeberg 1990, 45)

Contudo, de acordo com Diarra (1990, 32), o quimbundo apresenta resunção em relativas de objeto quando o antecedente pertence à classe 1 (tipicamente humanos, no singular), como em (28).

(28) Mù-thú ngá-mú-sàngè.
1-pessoa 1SG.PST-1-encontrar
‘A pessoa que encontrei.’ (quimbundo, adapt. de Diarra 1990, 33)
(isto é, ‘a pessoa que eu a encontrei’)

No que respeita a relativas de sintagma preposicional, verifica-se que as línguas bantu apresentam, de um modo geral, uma estratégia de resunção (Riedel 2010). Apresentamos, em (29), um exemplo do umbundo.

(29) Ava òma-nù ndá-vangw-ílé la-vó.
DEM 2.pessoa 1SG.REL-falar-PST com-3PL
‘Estas são as pessoas com quem falei.’ (umbundo, adapt. de Schadeberg 1990, 46)
(isto é, ‘estas são as pessoas que falei com elas’)

Uma análise detalhada das propriedades dos diferentes tipos de relativas do português angolano e do português moçambicano em contraste com as línguas bantu com que estas variedades estão em contacto permitirá avaliar melhor a importância do contacto. Se efetivamente se verificar uma tendência mais acentuada para pronomes resuntivos no português moçambicano do que no português angolano (e em outras variedades do português), a estrutura das relativas das línguas bantu na zona oriental, onde a estratégia de resunção é generalizada, poderá estar na origem deste contraste.

Em relação aos crioulos portugueses em África, os crioulos da Alta Guiné e do Golfo da Guiné não apresentam resunção em relativas de objeto e a estratégia típica das relativas de sintagma preposicional é a da cópia defetiva (e.g. Alexandre/Hage-meijer 2013), conforme ilustrada para o crioulo de S. Tomé em (30) e (31) respetivamente.

(30) [...] zo xi ê na sa ngê se ku a kônsê fa... (CST)
 então se 3SG NEG ser pessoa DEM REL IMP conhecer NEG ...
 ‘...então se não é uma pessoa que conhecem...’

(31) Inen mwala se ku Zon fla ku ê sa ke. (CST)
 PL mulher DEM REL João falar com 3SG estar casa
 ‘As mulheres com quem o João falou estão em casa.’

No exemplo (31), o antecedente plural da relativa, *inen mwala se*, é retomado pela terceira pessoa do singular *ê*. Note-se que, nestes casos de relativização do sintagma preposicional, o crioulo de Cabo Verde (variantes de sotavento) apresenta variação entre a estratégia da cópia defetiva e a resuntiva (Alexandre 2012).

Em suma, no domínio da relativização, as variedades africanas do português recorrem essencialmente às mesmas estratégias que caracterizam o português europeu e o português brasileiro e à mesma tendência de generalização do relativizador *que*, sendo a variação mais uma questão quantitativa do que qualitativa. Apesar destas tendências convergentes, parece haver espaço para um efeito moderado de contacto, tendo em conta a ocorrência de relativas resuntivas de objeto no português moçambicano e de relativas de cópia defetiva no português são-tomense.

4.5 Objetos pronominais

Nesta secção analisaremos dois aspetos dos objetos pronominais e a sua relação com crioulos e línguas bantu, designadamente a forma do objeto pronominal, que será discutida em 4.5.1, e a sintaxe destes objetos, que será abordada em 4.5.2.

4.5.1 Forma dos objetos pronominais

As descrições do português angolano e do português moçambicano assinalam uma tendência para o uso do clítico dativo *lhe(s)* em contextos acusativos (e.g. Gonçalves 1990; 2004; 2010; Mingas 2000; Miguel 2003; Chavagne 2005; Mapasse 2005), como mostram os seguintes exemplos.

(32) [...] se eu lhe encontrar, eu lhe reconheço. (PA, Chavagne 2005, 227)

(33) Elogiaram-lhe. (PM, P. Gonçalves 2010, 100)

Apesar da semelhança superficial entre estas duas variedades, os mecanismos que estão na origem desta propriedade são distintos. No português moçambicano, a ocorrência do clítico dativo resulta de uma alteração da estrutura argumental de verbos transitivos diretos que consiste na inserção do marcador casual *a* quando o objeto é [+humano].

(34) Eles elogiaram a uma pessoa. (PM, P. Gonçalves 2010, 100)

Esta reestruturação das grelhas argumentais, isto é, a inserção de *a*, não se observa no português angolano, onde o clítico *lhe(s)*, muitas vezes realizado como *le(s)* (e.g. Miguel 2003), funciona como uma forma única para os contextos acusativo e dativo. No português são-tomense, as tendências são menos claras, mas esta variedade, mais do que o português angolano e o português moçambicano, privilegia a ocorrência de pronomes fortes em vez de clíticos.

(35) Porque em muitos momentos não fico psicologicamente preparado para falar o crioulo forro, respondi ele em nosso português. (PST, Gonçalves 2009)

Esta tendência é mais acentuada no caso do complemento direto, visto que há outras estratégias que concorrem para a realização dos complementos indiretos, nomeadamente o uso de *a* ou *para* (e.g. *vendi a/para ele*). Uma consequência direta desta variação observada nas variedades africanas do português traduz-se numa tendência de perda dos clíticos acusativos *o(s)*, *a(s)*.

No domínio dos clíticos reflexos, as variedades africanas do português também apresentam características que as distanciam do português europeu. Em todas elas, observa-se uma tendência para a supressão dos pronomes reflexos, especialmente quando estes são clíticos inerentes (Mendes/Estrela 2008). Esta tendência já tinha sido observada para o português moçambicano (e.g. Gonçalves 1996, 317s.), como em (36), e é particularmente acentuada no português são-tomense (cf. (37)). Embora a supressão do clítico reflexo também esteja documentada para o português angolano (Chavagne 2005, 257; Mendes/Estrela 2008), vários autores referem, para esta variedade, uma generalização do pronome reflexo *se* para as outras pessoas do paradigma, como em (38–39) (e.g. Miguel 2003; Chavagne 2005; Inverno 2011).

(36) A tal namorada foi queixar ao pai. (PM, Gonçalves 1996, 317)

(37) Com o tempo, tornei resistente. (PST)

(38) Nós conseguimos se entender (PA, Inverno 2011, 181)

(39) Eu não se mexi (PA, Miguel 2003, 75)

Nas línguas bantu, os objetos direto e indireto pronominalizados são representados por prefixos idênticos que ocorrem imediatamente à esquerda do radical verbal.

(40) Kàbhúlú ka-mú-bhit-íle. (quimbundo, adapt. de Miguel 2003, 55)
12.lebre 12-1-ultrapassar-PST
'A lebre ultrapassou-o.'

- (41) Nga-mu-bane ma-honjo. (quimbundo, adapt. de Mingas 2000, 71)
 1SG.PST-1-dar.PST 6-banana
 ‘Eu dei-lhe (algumas) bananas.’

Nos casos em que os verbos do português apresentam um pronome reflexo inerente (e.g. *rir-se*, *deitar-se*), as línguas bantu não apresentam morfologia específica. A reflexividade é normalmente marcada com verbos transitivos, isto é verbos que em português licenciariam um pronome reflexo argumental, sob a forma de um prefixo invariável que, à semelhança dos prefixos de objeto, ocorre na posição pré-radical verbal, exemplificado em (42) para as formas do singular.

- (42) a. ngi-di-sukula. ‘Eu lavo-me.’ (quimbundo, adapt. de Miguel 2003, 56)
 b. u-di-sukula. ‘Tu lavas-te.’
 c. wa-di-kwama. ‘Ele(ela) magoa-se.’

Tal como nas línguas bantu, os pronomes de objeto direto e indireto de terceira pessoa do singular nos crioulos apresentam formas idênticas, apresentando-se como formas enclíticas:

- (43) Bu pintxa-l. / Bu da'l dos txabi. (CCV)
 (44) Bô pins'e. / Bô d'e dôsu sabi. (CST)
 2SG empurrar-3SG / 2SG dar-3SG dois chave
 ‘Empurraste-o/a.’ / ‘Deste-lhe duas chaves.’

Nestes crioulos, alguns verbos apresentam uma estratégia reflexiva nominal, que recorre às palavras ‘cabeça’ e ‘corpo’ como elemento de reflexivização. Trata-se, em geral, de verbos que são inerentemente transitivos, isto é, verbos que dispõem de uma posição argumental, como em (45) e (46). Nestes crioulos não existe, de resto, um paradigma reflexivo correspondente aos clíticos inerentes do português – cf. (47)–(48).

- (45) E mata (si) kabesa. (CCV, Quint 2000, 177)
 3SG matar DEM cabeça
 ‘Ele suicidou-se.’
 (46) [...] pa ê na ngana ubwê dê. (CST)
 para 3SG NEG enganar corpo DEM
 ‘...para que ele não se enganasse.’
 (47) Bu xinta. / Bu ri. / Bu deta. (CCV)
 (48) Bô tason. / Bô li. / Bô deta. (CST)
 ‘Tu sentaste-te.’ / ‘Tu riste-te.’ / ‘Tu deitaste-te.’

Conclui-se que a perda dos clíticos acusativos nas variedades africanas do português não pode ser considerada uma consequência direta do contacto, uma vez que as

línguas bantu e os crioulos também apresentam pronomes gramaticalizados, sob a forma de afixos e clíticos, respetivamente. Uma hipótese mais geral de perda dos itens lexicais mais funcionais e, por regra, foneticamente mais fracos em contexto de aquisição de L2 parece adequar-se melhor a este cenário. No entanto, podem pelo menos ser parcialmente atribuídas ao efeito do contacto a tendência para neutralizar a oposição entre a forma dos pronomes de objeto direto e indireto nas variedades africanas do português bem como a perda generalizada dos clíticos inerentes. No caso da generalização do pronome reflexo *se* no português angolano, a hipótese do contacto também deve ser equacionada, conforme mostrámos, mas deverá ter em conta que a mesma tendência também se verifica num subgrupo de dialetos do português europeu (Martins 2009), onde a generalização afeta apenas o paradigma do plural, e no português brasileiro (Brito 2009), onde afeta também o singular, à semelhança do português angolano.

4.5.2 Sintaxe dos objetos pronominais

Nas variedades africanas do português, têm sido identificadas algumas tendências no que diz respeito ao padrão de colocação de clíticos, as quais, em geral, carecem de corroboração através de estudos quantitativos. Na discussão destes padrões é fundamental ter em consideração as diferenças entre contextos verbais simples e os diferentes tipos de sequências verbais, bem como o papel dos atratores de próclise (advérbios, conectores de subordinação, etc.).

Os trabalhos sobre o português de Moçambique (e.g. Mapasse 2005; Gonçalves 2013) destacam o facto de esta variedade privilegiar a colocação enclítica com formas verbais simples, mesmo na presença de atratores de próclise, como em (49), e a posição intermédia em sequências verbais, em (50) e (51), que poderão corresponder a casos de ênclise ao verbo auxiliar ou de próclise ao verbo encaixado (cf. Gonçalves 2013, 172), embora seja necessário analisar se em (51) o clítico é atraído pela preposição *a*.

(49) ..daí que relaciono-me bem com eles. (PM, Mapasse 2005, 70)

(50) Eu sei que podes me ajudar. (PM, Gonçalves 2013, 172)

(51) Este desemprego começou a se fazer sentir. (PM, Mapasse 2005, 68)

Em relação ao português angolano, autores como Chavagne (2005) e Miguel (2003) mencionam a instabilidade no que se refere ao padrão de colocação dos clíticos. Ainda assim, emergem algumas diferenças assistemáticas em relação à norma do português europeu que requerem estudos mais aprofundados, designadamente uma maior preferência por próclise em contextos verbais simples sem a presença de atratores de próclise (52) e mais frequentemente ênclise em contextos com proclisado-

res (53). Nos contextos de sequências verbais do português angolano também se observam casos de clíticos em posição intermédia (54). Tal como no caso do português moçambicano, que poderão ser casos de próclise ao verbo principal, na linha do exemplo (55).

(52) Praticamente, me identifico mais com o povo bantu. (PA, Chavagne 2005, 248)

(53) dum sistema que herdou-se duma colonização. (PA, Chavagne 2005, 248)

(54) E eu até, só uma coisa que vou lhe dizer? (PA, Chavagne 2005, 248)

(55) João, qual é a disciplina que tá a lhe dar mais trabalho aqui?
(PA, Chavagne 2005, An. I, 30)

O português são-tomense, por fim, também parece caracterizar-se por alguma variação em orações principais, em função do tipo de proclisador, e próclise em orações subordinadas finitas (Gonçalves 2009). Globalmente, os padrões de colocação dos clíticos nas variedades africanas do português apresentam mais instabilidade do que as variedades portuguesa e brasileira. Ainda assim, parecem aproximar-se mais dos padrões do português europeu do que do português brasileiro, embora sejam necessários estudos quantitativos baseados em corpora orais para corroborar esta afirmação.

Relativamente ao padrão de colocação de pronomes de objeto, as línguas bantu e os crioulos apresentam estratégias distintas. Nas primeiras, os pronomes são prefixados em contextos verbais simples e prefixados ao verbo encaixado em contextos verbais complexos.

(56) Ndi-va-ipa. (umbundo, Nascimento 1894, 37)
1SG-3PL.OBJ-matar
'Eu mato-os.'

(57) O-fwete kio-baka. (quicongo, Bentley 1887, 282)
2SG-dever 5.OBJ-apanhar
'Deverias tê-lo apanhado.'

Já nos crioulos portugueses em África, os pronomes (clíticos) ocorrem sempre à direita do verbo em contextos verbais simples e à direita do verbo encaixado em contextos verbais complexos, como mostram os exemplos do crioulo de Cabo Verde em (58–59).

(58) Nu ka ta kunpra-l. (CCV)
1PL NEG T/A comprar-3SG
'Não estamos a comprá-lo.'

(59) Djonki pode djuda-m. (CCV)
João FOC poder ajudar-1SG
'João é que pode ajudar-me.'

O contraste entre o português (≥15 A colocação dos pronomes clíticos em sincronia e diacronia), as línguas bantu e os crioulos no domínio da sintaxe dos pronomes de objeto é terreno fértil para estudos comparados. Se o papel das línguas bantu fosse determinante para os padrões de colocação de clíticos, o português angolano e o português moçambicano deveriam apresentar uma tendência para a próclise em contextos verbais simples, hipótese essa que é contrariada pelo menos pela forte tendência para ênclise no português moçambicano. Por outras palavras, a homogeneidade que caracteriza as línguas bantu neste domínio da gramática não se traduz num efeito homogêneo nestas duas variedades. Por outro lado, se se confirmar que a posição intermédia em contextos verbais complexos corresponde, efetivamente, a próclise ao verbo principal nas variedades angolana e moçambicana, estes casos poderiam constituir evidência de contacto com as línguas bantu, especialmente se se verificar que este padrão não ocorre ou é menos comum nos espaços onde o português está em contacto com crioulos, que apresentam ênclise ao verbo encaixado. O recurso a estudos quantitativos baseados em corpora poderá também viabilizar uma análise mais refinada de, por exemplo, o comportamento de clíticos em diferentes tipos de sequências verbais (com verbos «auxiliares» modais, aspetuais, causativos, etc.) ou a perda do efeito de próclise com diferentes tipos de proclisadores.

5 Considerações finais

A presença da língua portuguesa em África resultou, em momentos distintos, na formação de línguas crioulas e de novas variedades do português nas antigas colónias portuguesas. As variedades africanas do português, em particular, começaram a consolidar-se sobretudo a partir das independências, com a democratização do português. Nas últimas décadas, não só o número de falantes do português tem vindo a aumentar em todos os espaços, como também se assiste a um ciclo de nativização desta língua em Angola, Moçambique e S. Tomé e Príncipe.

Apesar de se tratar de uma área de investigação jovem e extremamente complexa do ponto de vista sociolinguístico e linguístico, os estudos efetuados sobre as variedades africanas do português já vão revelando algumas tendências linguísticas mais robustas, designadamente nos domínios da sintaxe e da morfossintaxe, em que se concentra um maior número de estudos. Algumas dessas tendências são observadas, em maior ou menor grau, em todas as variedades do português (e.g. as relativas cortadoras), ao passo que outras são partilhadas por apenas um subconjunto de variedades (e.g. as construções de duplo objeto no português de Moçambique e de S. Tomé). Há ainda traços que são (praticamente) exclusivos de uma só variedade (e.g. as relativas de cópia defetiva no português de S. Tomé ou os objetos indiretos preposicionados por *em* no português de Angola). Esta variação linguística, que seguramente aumentará à medida que forem explorados novos aspetos linguísticos, mostra que o português em África está longe de constituir uma unidade linguística.

A breve descrição de algumas estruturas linguísticas e a comparação com as línguas crioulas e bantu com que estas variedades estão em contacto mostram que a complementaridade entre a mudança induzida por contacto de línguas e a mudança interna guiada por princípios gerais da gramática devem continuar a fazer parte do programa de investigação linguística sobre as variedades africanas do português.

6 Referências

- Alexandre, Nélia (2000), *A estratégia resumptiva em relativas restritivas do português europeu*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Alexandre, Nélia (2012), *The defective copy theory of movement: Evidence from wh-constructions in Cape Verdean Creole*, Amsterdam/Philadelphia, Benjamins.
- Alexandre, Nélia/Gonçalves, Rita/Hagemeyer, Tjerk (2011a), *A formação de frases relativas de PP no português de Cabo Verde e de São Tomé*, in: Armanda Costa/Isabel Falé/Pilar Barbosar (edd.), *Textos selecionados do XXVI Encontro da Associação Portuguesa de Linguística 2010*, Lisboa, APL, 17–34.
- Alexandre, Nélia/Gonçalves, Rita/Hagemeyer, Tjerk (2011b), *PP-relativization in African Varieties of Portuguese: between contact and convergent change*, comunicação apresentada na conferência *Rethinking contact-induced change*, Universidade de Leiden, 9–11 de junho de 2011.
- Alexandre, Nélia/Hagemeyer, Tjerk (2007), *Bare Nouns and the Nominal Domain in Santome*, in: Marlyse Baptista/Jacqueline Guéron (edd.), *Noun phrases in creole languages: a multi-faceted approach*, Amsterdam/Philadelphia, Benjamins, 37–59.
- Alexandre, Nélia/Hagemeyer, Tjerk (2013), *Estratégias de relativização de PPs no mundo luso-atlântico: crioulos de base lexical portuguesa e variedades do português*, in: Maria Denilda Moura/Marcelo Amorim Sibaldo (edd.), *Para a história do português brasileiro, Volume III – Sintaxe comparativa entre o português brasileiro e línguas crioulas de base lexical portuguesa*, Maceió, EDUFAL, 49–71.
- Bacelar do Nascimento, Maria Fernanda, et al. (2008), *Aspetos da unidade e diversidade do Português: as variedades africanas face à variedade europeia*, *Revista Veredas* 9, 35–60.
- Baptista, Marlyse (2007), *On the syntax and semantics of DP in Cape Verdean Creole*, in: Marlyse Baptista/Jacqueline Guéron (edd.), *Noun phrases in creole languages: a multi-faceted approach*, Amsterdam/Philadelphia, Benjamins, 61–105.
- Baxter, Alan N. (2004), *The development of variable NP plural agreement in a restructured African variety of Portuguese*, in: Geneviève Escure/Armin Schwegler (edd.), *Creoles, contact and language change: Linguistics and social implications*, Amsterdam/Philadelphia, Benjamins, 97–126.
- Bentley, William Holman (1887), *Dictionary and grammar of the Kongo language, as spoken at San Salvador, the ancient capital of the old Congo empire, West Africa*, London, Baptist Missionary Society.
- Brandão, Sílvia (2011), *Concordância nominal em duas variedades do português: convergências e divergências*, *Veredas – Revista de Estudos Linguísticos* 15:1, 164–178.
- Brito, Ana Maria (2001), *Relativas de genitivo no português europeu e no português de Moçambique*, in: Clara Nunes Correia/Anabela Gonçalves (edd.), *Actas do XVI Encontro Nacional da APL*, Lisboa, APL, 115–129.
- Brito, Ana Maria (2002), *Relativas de genitivo «estranhas» no português de Moçambique: erros ou sinais de mudança?*, in: *As ciências sociais nos espaços de língua portuguesa: Balanços e desafios: actas 2*, Porto, FLUP, 329–336.

- Brito, Dorothy (2009), *O «se» reflexivo no português brasileiro*, tese de Doutorado, Maceió, Universidade Federal de Alagoas.
- Cabral, Lisender (2005), *Complementos verbais preposicionados do português em Angola*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Cardoso, Hugo/Hagemeyer, Tjerk/Alexandre, Nélia (2015), *Crioulos de base lexical portuguesa*, in: Maria Iliescu/Eugeen Roegiest (edd.), *Manuel des anthologies, corpus et textes romans*, Berlin / Boston, De Gruyter, 670–691.
- Chatelain, Héli (1888–1889), *Grammatica elementar do kimbundu ou lingua de Angola*, Ridgewood, New Jersey, The Gregg Press Incorporated.
- Chatelain, Héli (1894), *Folk-tales of Angola – fifty tales, with Ki-mbundu text literal English translation, introduction and notes*, Honolulu, University Press of the Pacific.
- Chavagne, Jean-Pierre (2005), *La langue portugaise d'Angola: étude des écarts par rapport à la norme européenne du portugais*, tese de Doutorado, Lyon, Université Lumière.
- Cheng, Lisa/Downing, Laura (2010), *Locative relatives in Durban Zulu*, ZAS Papers in Linguistics 53, 33–35.
- Chimbutane, Feliciano (1996), *A estratégia de pronome resumptivo na formação de orações relativas de OD e de OBL do português de Moçambique*, in: *Actas do XI Encontro Nacional da APL*, vol. 3, Lisboa, Colibri, 225–248.
- Christofoletti, Alfredo (2013), *Ditongos no português de São Tomé e Príncipe*, dissertação de Mestrado, São Paulo, Universidade de São Paulo.
- Couto, Hildo Honório do (2009), *Português em contato: o português e o crioulo na Guiné-Bissau*, in: Ana Maria Carvalho (ed.), *Português em contacto*, Madrid/Frankfurt am Main, Iberoamericana / Vervuert, 53–66.
- Diarra, Boubacar (1990), *Gramática kimbundu*, Luanda, Secretaria de Estado da Cultura, Instituto de Línguas Nacionais.
- Duarte, Fábio Bonfim (2011), *Tense encoding, agreement patterns, definiteness and relativization strategies in Changana*, in: Eyamba G. Bokamba et al. (edd.), *Selected proceedings of the 40th Annual Conference on African Linguistics*, Somerville (MA), Cascadilla Proceedings Project, 80–94.
- Endruschat, Annette (1990), *Studien zur portugiesischen Sprache in Angola (unter besonderer Berücksichtigung lexikalischer und soziolinguistischer Aspekte)*, Frankfurt am Main, TFM.
- Ferraz, Luiz Ivens (1987), *Portuguese creoles of West Africa and Asia*, in: Glenn G. Gilbert (ed.), *Pidgin and creole languages: Essays in memory of John E. Reinecke*, Honolulu, University of Hawaii Press, 337–360.
- Figueiredo, Carlos Filipe (2008), *A concordância variável no sintagma nominal plural do português reestruturado de alcoxarife (São Tomé)*, *Papia* 18, 23–43.
- Figueiredo, Carlos Filipe (2010), *A concordância plural variável no sintagma nominal do português reestruturado da comunidade de alcoxarife, São Tomé*, tese de Doutorado, Universidade de Macau.
- Gärtner, Eberhard (1989), *Remarques sur la syntaxe du portugais en Angola e au Mozambique*, in: Jean-Michel Massa/Matthias Perl (edd.), *La Langue Portugaise en Afrique*, Rennes, Université de Haute Bretagne, 29–54.
- Gonçalves, Perpétua (1990), *A construção de uma gramática de português em Moçambique: aspectos da estrutura argumental dos verbos*, tese de Doutorado, Universidade de Lisboa.
- Gonçalves, Perpétua (1996), *Aspectos da sintaxe do português de Moçambique*, in: Isabel Hub Faria et al. (edd.), *Introdução à linguística geral e portuguesa*, Lisboa, Caminho, 313–322.
- Gonçalves, Perpétua (1997), *Tipologia de «erros» do português oral de Maputo: um primeiro diagnóstico*, in: Christopher Stroud/Perpétua Gonçalves (edd.), *Panorama do português oral de Maputo*, vol. 2: *A construção de um banco de «erros»*, Maputo, INDE, 37–70.

- Gonçalves, Perpétua (2004), *Towards a unified vision of classes of language acquisition and change: Arguments from the genesis of Mozambican African Portuguese*, *Journal of Pidgin and Creole Languages* 19:2, 225–259.
- Gonçalves, Perpétua (2010), *A génese do Português de Moçambique*, Lisboa, INCM.
- Gonçalves, Perpétua (2012), *Lusofonia em Moçambique: com ou sem glotofagia?*, comunicação apresentada no *II Congresso Internacional de Linguística Histórica – Homenagem a Ataliba Teixeira de Castilho*, São Paulo, 2012. http://www.catedraportugues.uem.mz/lib/docs/lusofonia_em_mocambique.pdf (04.01.2016)
- Gonçalves, Perpétua (2013), *O português em África*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 7, Lisboa, Fundação Calouste Gulbenkian, 157–178.
- Gonçalves, Perpétua (coord.), *Cátedra de Português, Língua Segunda e Estrangeira* [online], Instituto Camões/Universidade Eduardo Mondlane, Moçambique. http://www.catedraportugues.uem.mz/?__target__=variedades-nao-europeias (30.09.2015)
- Gonçalves, Perpétua/Chimbutane, Feliciano (2004), *O papel das línguas bantu na génese do português de Moçambique: o comportamento sintáctico de constituintes locativos e direcionais*, *Papia* 14, 7–30.
- Gonçalves, Rita (2009), *A colocação dos pronomes clíticos no português oral de S. Tomé: Análise e discussão de contextos numa perspectiva comparativa*, ms. Faculdade de Letras da Universidade de Lisboa.
- Gonçalves, Rita (2010), *Propriedades de subcategorização verbal no português de São Tomé*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Gonçalves, Rita/Hagemeyer, Tjerk (2015), *O português num contexto multilingue: O caso de São Tomé e Príncipe*, *Revista Científica da Universidade Eduardo Mondlane: Série Letras e Ciências Sociais* 1:1, 87–107.
- Gordon, Raymond G. (ed.) (¹⁵2005), *Ethnologue: Languages of the world*, Dallas, SIL International.
- Hagemeyer, Tjerk (2011), *The Gulf of Guinea creoles: Genetic and typological relations*, *Journal of Pidgin and Creole Languages* 26:1, 111–154.
- Hagemeyer, Tjerk/Alexandre, Nélia (2012), *Os crioulos da Alta Guiné e do Golfo da Guiné: Uma comparação sintáctica*, *Papia* 22:2, 207–226.
- Henderson, Brent (2007), *The syntax of agreement in Bantu relatives*, in: Frederick Hoyt et al. (edd.), *Texas Linguistics Society IX: The morphosyntax of underrepresented languages*, Stanford (CA), CSLI Publications, 167–184.
- Inverno, Líliliana (2009), *A transição de Angola para o português vernáculo: estudo morfossintáctico do sintagma nominal*, in: Ana Maria Carvalho (ed.), *Português em contacto*, Madrid/Frankfurt am Main, Iberoamericana/Vervuert, 87–106.
- Inverno, Líliliana (2011), *Contact-induced restructuring of Portuguese morphosyntax in interior Angola*, tese de Doutoramento, Universidade de Coimbra.
- Jacobs, Bart (2010), *Upper Guinea creole: Evidence in favor of a Santiago birth origins of Papia-mentu: Linguistic and historical evidence*, *Journal of Pidgin and Creole Languages* 25:2, 289–343.
- Jon-And, Anna (2011), *Variação, contato e mudança linguística em Moçambique e Cabo Verde: A concordância variável de número em sintagmas nominais do português*, tese de Doutoramento, Universidade de Estocolmo.
- Justino, Víctor (2011), *A distribuição e a expressão gramatical do futuro do conjuntivo no Português de Moçambique*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Katamba, Francis (ed.) (2003), *Bantu nominal morphology*, in: Derek Nurse/Gérard Philippson (edd.), *The Bantu languages*, London/New York, Routledge, 103–120.
- Lopes, Norma/Baxter, Alan (2011), *A concordância verbal variável no português dos tongas*, *Papia* 21:1, 39–50.

- Madeira, Sandra (2008), *Towards an annotated bibliography of restructured Portuguese in Africa*, dissertação de Mestrado, Universidade Coimbra.
- Mapasse, Ermelinda (2005), *Clíticos pronominais em português de Moçambique*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Marques, Irene Guerra (1983), *Algumas considerações sobre a problemática linguística em Angola*, in: Instituto de Língua e Cultura Portuguesa (ed.), *Actas do Congresso sobre a situação actual da língua portuguesa no mundo*, Lisboa, ICALP, 205–223.
- Martins, Ana Maria (2009), *Subject doubling in European Portuguese dialects: The role of impersonal «se»*, in: Enoch Aboh et al. (edd.), *Romance languages and linguistic theory 2007: Selected papers from «Going Romance» Amsterdam 2007*, Amsterdam/Philadelphia, Benjamins, 179–200.
- Mendes, Amália/Estrela, Antónia (2008), *Constructions with SE in African varieties of Portuguese*, *Phrasis* 2008:2, 83–107.
- Michaelis, Susanne Maria, et al. (edd.) (2013a), *The atlas of pidgin and creole language structures*, Oxford, Oxford University Press.
- Michaelis, Susanne Maria, et al. (edd.) (2013b), *The survey of pidgin and creole languages*, vol. 2, Oxford, Oxford University Press.
- Michaelis, Susanne Maria, et al. (edd.) (2013c), *Atlas of Pidgin and Creole Language Structures Online*, Leipzig, Max Planck Institute for Evolutionary Anthropology, <http://apics-online.info> (30.09.2015).
- Miguel, Maria Helena (2003), *Dinâmica da pronominalização no Português de Luanda*, Luanda, Nzila.
- Mingas, Amélia (2000), *Interferência do kimbundu no português falado em Lwanda*, Porto, Campo das Letras.
- Nascimento, José Pereira do (1894), *Grammatica do umbundu ou lingua de Benguella*, Lisboa, Imprensa Nacional.
- Observatório da Língua Portuguesa* [online]. <http://observalinguaportuguesa.org/category/dados-estatisticos> (31.03.2016).
- Peres, João/Móia, Telmo (1995), *Áreas críticas da língua portuguesa*, Lisboa, Caminho.
- Petter, Margarida Maria Taddoni (2008), *Uma hipótese explicativa do contato entre o português e as línguas africanas*, *Papia* 17, 9–19.
- Quint, Nicolas (2000), *Grammaire de la langue cap-verdienne: Étude descriptive et compréhensive du créole afro-portugais des Îles du Cap-Vert*, Paris, L'Harmattan.
- Riedel, Kristina (2010), *Relative clauses in Haya*, *ZAS Papers in Linguistics* 53, 211–225.
- Rougé, Jean-Louis (1992), *Les langues des Tonga*, in: Ernesto d'Andrade/Alain Kihm (edd.), *Actas do colóquio sobre crioulos de base lexical portuguesa*, Lisboa, Colibri, 171–176.
- Schadeberg, Thilo (1990), *A sketch of Umbundu*, Köln, Köppe.
- Seibert, Gerhard (2012), *Creolization and creole communities in the Portuguese Atlantic: São Tomé, Cape Verde and the rivers of Guinea in comparison*, in: Toby Green (ed.), *Brokers of change: Atlantic commerce and cultures in pre-colonial western Africa*, Oxford, Oxford University Press, 29–51.
- Stroud, Christopher/Gonçalves, Perpétua (edd.) (1997–2000), *Panorama do português oral de Maputo*, vol. 1–4, Maputo, INDE.
- Tarallo, Fernando (1985), *The filling of the gap: Pro-drop rules in Brazilian Portuguese*, in: Larry King/Catherine Maley (edd.), *Selected papers from the XIIIth Linguistic Symposium on Romance Languages*, Amsterdam/Philadelphia, Benjamins, 355–375.
- Vicente, Francisco (2009), *Consciência fonológica no ensino básico em Moçambique*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.

Hugo C. Cardoso

3 O português em contacto na Ásia e no Pacífico

Abstract: A partir do século XVI, a língua portuguesa teve grande difusão um pouco por toda a Ásia e o Pacífico, sobretudo em consequência de dinâmicas de expansão colonial mas também por outras vias, tais como a imigração. Este capítulo apresenta e ilustra a diversidade de contextos de contacto linguístico que este facto motivou, bem como os seus principais resultados: os empréstimos mútuos (sobretudo lexicais) entre o português e as línguas da região; a influência duradoura do português nas línguas de contacto posteriormente estabelecidas em contexto colonial; e o desenvolvimento de variedades asiáticas do português, muitas vezes em paralelo e articulação com o de línguas crioulas de base portuguesa, dando relevo às que subsistem, ao panorama atual dos estudos científicos sobre estas variedades e aos domínios de investigação em aberto.

Keywords: novas variedades de português, contacto linguístico, línguas crioulas, Ásia, expansão linguística

1 Introdução

A língua portuguesa tem um envolvimento longo e diversificado com várias partes do continente asiático e do Pacífico. As situações de contacto linguístico que aí se estabeleceram, aliadas a contextos socio-históricos bastante díspares, resultaram num conjunto complexo de variedades linguísticas (L1, L2, pidgins, crioulas), muitas vezes coincidentes no espaço e no tempo. Este capítulo não pode dar conta de toda esta complexidade em pormenor, mas procura apresentar em traços gerais não apenas os contextos sociolinguísticos em que o português se implantou como também os principais resultados do contacto linguístico que aí estabeleceu; para complementar as questões aqui abordadas, recomenda-se a consulta de fontes como Lopes (1936); Barbosa (1969); Paiva Boléo (1974); Loureiro (1992); Thomaz (1990); Baxter (1996); e Tomás (2008).

A secção 2 é dedicada à cronologia da expansão do português pela Ásia e às dinâmicas sociolinguísticas que a suportaram. Na secção 3, aborda-se o impacto do português nas diversas línguas da Ásia, impacto esse que tem nos empréstimos lexicais a sua expressão mais evidente mas, nalguns casos, implicou também influ-

Agradecimentos: O autor expressa o seu agradecimento a Alan N. Baxter, a Susana Afonso e às editoras do volume pelo seu auxílio e comentários.

ência estrutural. Por fim, a secção 4 apresenta uma panorâmica da investigação acerca do papel do contacto linguístico no estabelecimento e desenvolvimento de quatro variedades asiáticas do português: o português pan-asiático, o da Índia (Goa, Damão e Diu), o de Macau e o de Timor.

2 A difusão da língua portuguesa pela Ásia

A difusão da língua portuguesa por diversas zonas da Ásia acompanhou a expansão ultramarina portuguesa nessa região. O ano de 1498, que viu a armada de Vasco da Gama aportar ao sudoeste indiano, pode por isso ser tomado também como o ano 0 desse processo de expansão linguística. Os primeiros estabelecimentos portugueses na Ásia com caráter de permanência foram estabelecidos pouco depois, precisamente na região onde se deu o contacto inicial: a antiga Costa do Malabar, correspondente ao atual estado indiano de Kerala. Em 1505, o Malabar contava já com importantes feitorias portuguesas em Cochim, Cananor e Coulão, e foi a partir da região que, nesta fase, se coordenou a exploração e conquista de outras partes da Ásia. Esta circunstância confere à Costa do Malabar uma importância particular no envolvimento de Portugal com a Ásia e ao contacto entre o português e o malaiala, a língua dravídica aí dominante, um papel fundador no que viria a ser uma história de contacto linguístico vasta e diversificada.

Logo a partir do início do século XVI, a expansão da influência portuguesa em territórios asiáticos foi extremamente rápida. Em meio século, esta estendeu-se a largas porções do litoral sul-asiático (o subcontinente indiano, Ceilão e as regiões em torno da Baía de Bengala), ao Golfo Pérsico, ao sudeste asiático continental e insular (desde o Sião até Timor e as Molucas), ao sul da China e ao Japão. Contudo, esta expansão teve contornos bastante diversos em diferentes regiões da Ásia e ficou a dever-se a atividades também elas diversas. De entre as potenciais vias de difusão do idioma, o historiador Luís Filipe Thomaz (1990, 314) destaca a dominação política, o comércio e a missão, atividades centrais ao envolvimento português com a Ásia mas cujas geografias nem sempre coincidiram. Os primeiros estabelecimentos luso-asiáticos dividiam-se em oficiais e não-oficiais (Subrahmanyam 1993), sendo que os últimos (particularmente numerosos em torno do Golfo de Bengala e sudeste asiático) consistiam normalmente em comunidades de *lançados*, comerciantes por conta própria ou mercenários que desenvolviam as suas atividades fora do controlo da coroa portuguesa. A atividade dos missionários católicos afetos ao *Real Padroado Português*, essa, teve uma extensão muito superior à dos interesses comerciais portugueses, penetrando inclusivamente no interior do continente.

Seria um erro, contudo, imaginar que estas vias de difusão da língua envolveram sempre e apenas falantes nativos de português. Na realidade, é essencial compreender que, no seu período inicial, os empreendimentos coloniais portugueses na Ásia se fizeram com a participação de pessoas de várias proveniências, a começar por outros

européus (espanhóis, italianos, franceses, flamengos, britânicos, etc.) que seguiam não apenas como missionários ou comerciantes mas até como oficiais (Ferreira 2000; Pissarra 2002). Foi ainda extremamente significativo o recurso a mão de obra asiática para vários fins, incluindo o de operar a máquina militar portuguesa, como se percebe pela descrição que o cronista João de Barros faz de uma armada constituída em 1530 para um ataque a Diu:

«Nesta Ilha de Bombaim se fez resenha geral da gente que hia na Armada, e acháram-se tres mil e quinhentos e sessenta e tantos homens de peleja, contando os Capitães, mil e quatrocentos e cincoenta & tantos homens do mar Portuguezes com os Pilotos, e Mestres, dous mil e tantos Malavares, e Canarijs de Goa, oito mil escravos, homens que podiam pelejar, quatro mil marinheiros da terra que remavam, e mais de oitocentos mareantes dos juncos» (Barros 1615, livro 4, cap. XII).

Para além disso, houve ainda um número considerável de africanos transportados para os diversos estabelecimentos luso-asiáticos na condição de escravos (Pinto 1992; Jayasuriya/Pankhurst 2003; Cardoso 2010). Admitindo que a língua portuguesa teria funcionado como meio de comunicação entre uma população linguisticamente tão heterogénea, esta seria então uma L2 para boa parte dos agentes da expansão, um cenário em que cabe um contínuo de competências que vai desde as relativamente avançadas a versões pidginizadas do português.

Em resultado da dominação política, conversão religiosa e/ou processos de mestiçagem, desenvolveram-se no meio deste cadinho linguístico, logo desde o século XVI, comunidades que Baxter (1996, 300) descreve como de «Portuguese cultural orientation», para quem a língua portuguesa (ou variedades relacionadas, ver abaixo) se tornou língua de utilização diária. A partir deste momento, a difusão da língua pela Ásia beneficiou ainda de certos fenómenos de migração que envolveram estas comunidades (veja-se por exemplo Smith 2011). Assim se explica a presença da língua portuguesa em certas zonas onde o envolvimento português foi mínimo ou nulo, como Hong Kong (para onde convergiram muitos macaenses) ou Singapura (que conta com uma significativa população malaquenha). Um caso particularmente interessante é o da cidade de Batávia (atual Jakarta, capital da Indonésia), que serviu entre os sécs. XVII e XX como centro dos domínios coloniais neerlandeses no Oriente. Por várias razões, para aí convergiram, a partir do século XVI, pessoas de outras regiões asiáticas sob domínio neerlandês, incluindo muitas que anteriormente tinham recebido influência portuguesa, como sejam o sul da Índia, a ilha de Ceilão (atual Sri Lanka), Malaca e algumas ilhas do sudeste asiático. Em consequência, formou-se aí uma comunidade de fala «portuguesa», de acordo com descrições dos sécs. XVII a XIX (veja-se Lopes 1936), que se concentrou de forma particular na área de Tugu. No final do século XIX, foi ainda possível recolher um extenso corpus linguístico que veio a ser estudado e publicado pelo linguista alemão Hugo Schuchardt (Schuchardt 1891; veja-se Maurer 2011 para um estudo recente destas fontes).

Com efeito, a língua portuguesa enraizou-se na região asiática ao ponto de se converter em importante língua franca de comércio e diplomacia,¹ sobretudo para comunicação com e entre as demais potências europeias (britânicos, franceses, neerlandeses, dinamarqueses) que se começaram a estabelecer na Ásia a partir de finais do século XVI. Na obra *A Expansão da Língua Portuguesa no Oriente nos Séculos XVI, XVII e XVIII*, de 1936, o filólogo David Lopes faz uma recolha e tradução de referências ao uso da língua portuguesa em fontes estrangeiras que é bastante eloquente a este respeito. As referências que aí encontramos vão até ao século XIX e cobrem grande parte do litoral asiático, desde a Ásia Meridional até Timor e ilhas vizinhas. Contudo, os autores que registam este fenómeno tecem frequentemente comentários que nos levam a crer que, pelo menos em alguns casos, a língua que estava em causa era uma variedade reestruturada do português. A esse propósito, o capitão britânico Alexander Hamilton escrevia:

«[...] along the Sea-coasts, the *Portugueze* have left a Vestige of their Language, tho' much corrupted, yet it is the Language that most *Europeans* learn first, to qualify them for a general Converse with one another, as well as with the different Inhabitants of *India*» (Hamilton 1727).

Pouco mais tarde, o orientalista francês Anquetil-Duperron descrevia também a utilização do português com fins comerciais nos estabelecimentos franceses na Índia e acrescentava:

«Mais il s'en faut bien que ce soit le Portugais pur, appelé dans l'Inde le *Portugais Reinol*. Celui qui s'écrit en approche d'avantage, surtout à la Côte Malabare, où cette Nation a eu de nombreux Etablissements: le Portugais parlé n'est proprement qu'un jargon, consistant en 150 ou 200 mots, presque sans construction» (Anquetil-Duperron 1786, XI).

Descrições como esta apontam para a circulação de um pidgin de base portuguesa, e não necessariamente do português, como meio de comunicação intercomunitária (veja-se também Clements 2000). Para além disso, está bem documentada e estudada a formação de crioulos luso-asiáticos por todo o continente (e.g. Holm 1989; Tomás 1992; Cardoso/Baxter/Pinharanda Nunes 2012), de entre os quais subsistem hoje os de Diu, Damão, Korlai e Cananor (Índia), o de Sri Lanka, o de Malaca e o de Macau. Estas variedades de contacto são, portanto, incontornáveis numa descrição dos destinos da língua portuguesa na Ásia, desde logo porque as mais antigas referências raramente distinguem entre as várias categorias. Na verdade, ao atribuirmos certos fenómenos linguísticos antigos (por exemplo, os empréstimos lexicais a línguas asiáticas, ver

1 No século XVII, que assistiu ao estabelecimento de várias outras nações europeias na Ásia, a negociação política com as autoridades asiáticas foi muitas vezes feita com recurso ao português ou aproximações L2 deste. A título de exemplo, refira-se que foram redigidos nesta língua os tratados entre os dinamarqueses e o rei de Tanjor para estabelecimento de uma feitoria em Tranquebar, em 1620 (Lopes 1936, 36–38), e a correspondência entre os neerlandeses e Rajasinha II de Cândia (Ceilão), em meados do século (Lopes 1907).

secção 3) à influência do «português», é conveniente ressaltar que estas variedades de contacto poderão ter sido os verdadeiros veículos de difusão linguística.

Outra razão pela qual é importante trazer à discussão a questão da criouliização é a coexistência histórica do português com um crioulo luso-asiático em vários locais, com consequências linguísticas que referiremos mais adiante. É interessante notar que esta coexistência está documentada de forma explícita desde a primeira metade do século XVIII, altura em que Nikolaus Dal, missionário Luterano alemão a trabalhar no sul da Índia, descreve a situação linguística das comunidades «portuguesas» em termos de um contínuo ligando três níveis de português: «die rechte» ('o correto'), «die halbverdorbene» ('o semi-corrupto') e «gantz verdorbene» ('inteiramente corrupto') – o último dos quais com características que reconhecemos nos crioulos atuais (Dal 1733, 919; veja-se também Cardoso 2014).

Ficam assim expostas, em linhas muito gerais, as condições da difusão da língua portuguesa pela Ásia. Nas secções seguintes, olharemos para as consequências do contacto linguístico que a acompanhou.

3 O impacto do português em línguas da Ásia

A língua portuguesa teve um impacto sobre as línguas da Ásia, sobretudo aquelas faladas em zonas costeiras, cuja dimensão nem sempre é bem compreendida. A influência do português nota-se principalmente ao nível dos empréstimos lexicais, os quais, sendo em número variável nas várias línguas da Ásia, em algumas contam várias centenas – e, numa delas, ascendem a perto de 2000. Há, porém, duas dimensões desta questão que aqui trataremos separadamente, apesar dos pontos de contacto que estabelecem: o impacto do português nas línguas que o precederam no espaço asiático (secção 3.1) e o contributo do português para as línguas de contacto que se estabeleceram ou desenvolveram após o século XVI como resultado da expansão colonial europeia (secção 3.2).

3.1 Línguas asiáticas que precedem o período colonial

A identificação de palavras de origem portuguesa em línguas da Ásia é antiga e recebeu especial atenção em publicações filológicas de finais do século XIX e início do século XX. A título de exemplo, olhemos para a sequência de estudos que se dedicaram a este tema com referência à mais difundida língua de Ceilão (atual Sri Lanka), o cingalês. Um estudo da autoria de A. E. Buultjens, saído na revista científica ceilonense *The Orientalist* em 1885, faz uma recolha do vocabulário de origem portuguesa em cingalês com o intuito de demonstrar «the extent and the direction of the obligation of the former [i.e., os cingaleses] to the higher civilization of the latter [i.e., os portugueses]» (Buultjens 1885, 214). Buultjens identifica um pouco mais de 100

étimos portugueses para palavras cingalesas, que divide e discute (na sua maioria) de acordo com alguns domínios semânticos particulares: «Dress» (e.g. cing. *bottama* e *lēnsuva*, do ptg. BOTÃO e LENÇO), «The Kitchen» (e.g. cing. *vīnākiri*, do ptg. VINAGRE), «The Bakery» (e.g. cing. *paralu*, do ptg. FARELO), «The Tea-Table» (e.g. cing. *pīrisiya*, do ptg. PIRES), «Fruit-Trees, &c.» (e.g. cing. *pipiñña*, do ptg. PEPINO), «Carpentry, Masonry, &c.» (e.g. cing. *mēse* e *pedarēuvā*, do ptg. MESA e PEDREIRO), «Military» (e.g. cing. *soldāduvā*, do ptg. SOLDADO), «Religion» (e.g. cing. *paskuvā*, do ptg. PÁSCOA), «The Household, &c.» (e.g. cing. *lansōluva*, do ptg. LENÇOL), «Painting» (e.g. cing. *pinture*, do ptg. PINTURA), «Domestic» (e.g. cing. *janēle*, do ptg. JANELA) e «Education» (e.g. cing. *iscōla*, do ptg. ESCOLA).

No número seguinte de *The Orientalist*, saído apenas em 1888/1889, Louis Nell, um distinto jurista e intelectual ceilonense, volta a abordar esta temática no artigo *An explanatory list of Portuguese words adopted by the Sinhalese* (Nell 1888–1889a). A lista é bem mais completa do que a de Buultjens, já que Nell regista e discute com alguma profundidade cerca de 250 étimos portugueses. Num outro artigo no mesmo volume (Nell 1888–1889b), dedicado à gastronomia da comunidade eurasiática ceilonense, Nell aborda ainda os nomes de alguns pratos e ingredientes, tais como *lateria* (do ptg. ALETRIA), *curry-sekku* (cujo segundo elemento deriva do ptg. SECO) ou *bolo*.

Para além destes textos, há vários outros nas revistas filológicas da altura, muitas vezes notas curtas, que discutem a potencial origem portuguesa de palavras cingalesas particulares. Um pouco mais tarde, em 1922, S. G. Perera volta a dedicar um texto ao tópico do impacto lexical do português no cingalês, o que testemunha do interesse que este tópico suscitou entre a elite cultural ceilonense.

O cingalês pertence ao grupo das línguas asiáticas que incorporaram mais palavras de origem portuguesa, mas não se encontra no topo da tabela. Esse lugar pertence ao concani, a língua de Goa, com perto de 2000 palavras de origem portuguesa. Apesar de Adolfo Coelho ter dedicado algumas páginas aos empréstimos lexicais do português em línguas do Oriente (Coelho 1882), a obra que mais contribuiu para uma visão global deste assunto, intitulada *Influência do vocabulário português em línguas asiáticas (abrangendo cerca de cinquenta idiomas)*, é da autoria do padre e filólogo goês Sebastião Rodolfo Dalgado e foi publicada em Coimbra em 1913. Esta obra, organizada num formato muito próximo do de um dicionário, colige informação referente a derivações de étimos portugueses identificadas em exatamente 52 línguas de todo o continente asiático. Apesar das limitações inerentes à sua conceção e de, desde então, ter havido atualizações para línguas específicas – tais como Kim (1976) para o japonês, Jin Guo Ping (1987) para o chinês, Jones/Grijns/de Vries (2007) para o indonésio e o malaio, ou de Lima (2014) para o concani –, a obra de Dalgado permanece a mais abrangente dedicada ao tópico dos empréstimos lexicais portugueses em línguas asiáticas. A Tabela 1 indica as línguas abrangidas pelo estudo e o número de empréstimos lexicais do português que Dalgado identifica para cada uma:

Tabela 1: Número de empréstimos portugueses identificados em Dalgado (1913)

Língua	Nr. de empréstimos	Língua	Nr. de empréstimos
<i>Concani</i>	1768	<i>Dayak</i>	41
<i>Teto</i> (Tétum)	774	<i>Khassi</i> (Khasi)	37
<i>Malaio</i>	431	<i>Sindhi</i>	37
<i>Galóli</i>	429	<i>Siamês</i> (Tailandês)	35
<i>Singalês</i> (Cingalês)	208	<i>Assamês</i>	34
<i>Bengala</i>	173	<i>Malgache</i>	34
<i>Indo-Inglês</i>	173	<i>Panjabi</i>	33
<i>Tamul</i> (Tâmul)	171	<i>Oriya</i>	32
<i>Malayálam</i> (Malaiala)	127	<i>Garó</i>	31
<i>Búgui</i>	122	<i>Batta</i> (Battak)	29
<i>Maratha</i> (Marata)	116	<i>Nicobarês</i>	27
<i>Hindustani</i> (Urdu)	107	<i>Cambojano</i> (Khmer)	25
<i>Gujarati</i> (Guzerate)	105	<i>Persa</i>	22
<i>Javanês</i>	104	<i>Balinês</i>	21
<i>Makassarês</i>	101	<i>Molucano</i>	16
<i>Sundanês</i>	100	<i>Pidgin-English</i> (Pidgin Inglês da China)	15
<i>Japonês</i>	93	<i>Annamita</i> (Vietnamita)	15
<i>Canarês</i>	91	<i>Nepali</i> (Nepalês)	13
<i>Túlu</i>	88	<i>Batávio</i> (Betawi)	10
<i>Laskari-Hindustani</i> (Lasca- rim)	84	<i>Tonkinês</i> (Vietnamita)	8
<i>Telúgu</i>	83	<i>Birmanês</i>	7
<i>Indo-Francês</i>	62	<i>Kashmiri</i> (Caxemiri)	6
<i>Achinês</i>	56	<i>Rabino</i> (Moçárabe?)	4
<i>Hindi</i>	53	<i>Turco</i>	4
<i>Árabe</i>	50	<i>Chinês</i>	3
<i>Madurês</i>	48	<i>Tibetano</i>	3

Os números na Tabela 1 são relativamente conservadores, uma vez que o acesso a informação etimológica deverá ter sido bastante limitado para algumas destas línguas. No topo da tabela, surgem três línguas de territórios à época sob domínio colonial português: o concani (de Goa), o tétum e o galoli (de Timor-Leste). De entre as línguas que se seguem, a maioria (por exemplo, o malaio, cingalês, bengala, tâmul,

malaiala, marata) é falada em regiões costeiras que nalgum momento albergaram importantes estabelecimentos luso-asiáticos. Contudo, encontramos aqui também línguas do interior com as quais o português teve uma coexistência menos intensa ou nula, como são o hindi, o khasi, o sindhi ou o assamês, o que sugere que as palavras de origem portuguesa terão circulado por vias que não envolveram necessariamente o contacto direto.

Em contraponto com a abundância de lexemas de origem portuguesa em línguas asiáticas, a evidência de influência morfossintática é bastante mais escassa. Regra geral, nota-se de maneira clara apenas em línguas que tiveram ou mantêm um contacto intenso e prolongado com o português em contextos (pós-)coloniais. É o caso do tétum falado em Díli, conhecido como *tétum-Díli* ou *tétum-praça*, que adotou alguns morfemas portugueses, como sejam o sufixo derivativo *-dór* ‘-dor’ (veja-se Hajek/Williams-van Klinken 2003) e o sufixo de plural *-s*. Para além disso, as descrições do tétum atribuem ao decalque de construções portuguesas o desenvolvimento de certas construções raras ou social e geograficamente circunscritas. Um caso relatado na literatura refere-se à possibilidade de fazer certos verbos intransitivos preceder o seu sujeito, contrariando a ordem básica de palavras S-V(-O). Na verdade, com o verbo *iha* (que é simultaneamente um verbo existencial e possessivo), a possibilidade da ordem V-S é comum a todas as variedades de tétum: a de Díli, mais influenciada pelo português, mas também o tétum-terik, falado ao longo da costa sul da ilha e com menor contacto com o português. Esta construção é particularmente comum quando o argumento é indefinido ou modificado por uma oração relativa (1), ao passo que a ordem canónica S-V é preferida com argumentos definidos:

- (1) Tétum-Díli (Williams-van Klinken/Hajek/Nordlinger 2002, 55)

Iha distritu lima mak iha ona baunilha.
EXST distrito cinco FOC ter ANT baunilha
‘Há cinco distritos que têm baunilha.’

A potencial influência portuguesa nota-se não nesta construção existencial mas no facto de o tétum-Díli – mas não o tétum-terik – admitir a ordem V-S com mais alguns verbos intransitivos e argumentos indefinidos, nomeadamente *mosu* ‘surgir’, *akontese* ‘acontecer’, *moris* ‘viver’, *falta* ‘faltar’ e *hela* ‘ficar’:

- (2) Tétum Díli (Williams-van Klinken/Hajek/Nordlinger 2002, 57)

a. *Ohin loron mosu violénsia barak tebes iha Timór.*
hoje dia surgir violência muito verdadeiro LOC Timor
‘Nos últimos dias, surgiu/começou muita violência em Timor.’

b. *Falta ida tán karik*
faltar um mais talvez
‘Talvez falte mais um.’

A interpretação de Williams-van Klinken/Hajek/Nordlinger (2002, 56–57) para esta possibilidade assenta no facto de o português também poder colocar os verbos

correspondentes, nestes contextos, antes do sujeito gramatical. Os autores apontam ainda para a circunscrição de estruturas como as de (2) ao registo escrito ou formal, o que sustenta esta interpretação na medida em que, no contexto timorense, são esses os domínios mais associados ao uso da língua portuguesa. Para além das estruturas V-S, Williams van-Klinken/Hajek/Nordlinger propõem ainda a influência do português no desenvolvimento de uma construção passiva (2002, 52) e na utilização do clítico *-an*, prototipicamente um clítico reflexivo, com alguns verbos que em português são pronominais, como no caso de *esforsa-an* ‘esforçar-se’ (2002, 64).

O concani de Goa é outra língua para a qual a literatura aponta certos aportes estruturais do português. Sardessai (1983, 156–157), por exemplo, refere duas construções com respeito às quais o concani literário católico difere dos demais registos de um modo que o aproxima da sintaxe portuguesa:

- a) na presença de algumas frases com a ordem de palavras SVO, ao passo que o concani é, em geral, uma língua estritamente SOV;
- b) na produção de orações relativas pós-nominais e introduzidas por uma partícula relativa, o que contrasta com as relativas participiais mais típicas do concani.

Rocky Miranda dedica um artigo precisamente a identificar a influência sintática do português sobre o concani, que conclui ser um produto do final do século XIX (Miranda 2001, 59) e circunscrito, tal como afirmara Sardessai, ao concani literário da comunidade católica. Uma frase que transcreve para ilustrar o processo, originalmente publicada num jornal goês de 1923 e referida em Chavan (1924), é a seguinte:

- (3) Concani (Miranda 2001, 49)

avoi bapui sabar pauttim khens cortat aplim burguim
 pais muito vezes queixam-se próprios filhos
aiconant vo bexearmi cortat munn aplea sangnneanc
 ouvem.NEG e ignoram assim próprios instruções.para
 ‘Os pais queixam-se muitas vezes de que os seus filhos não ouvem e ignoram as suas instruções.’

Neste caso, a oração subordinada segue-se à oração principal, tal como em português mas contra o que é habitual em concani. Para além disso, a própria estrutura interna da subordinada viola a ordem SOV mais comum no concani, colocando o objeto direto (‘as suas instruções’) após os verbos coordenados (‘não ouvem e ignoram’). Miranda nota ainda outros padrões de influência portuguesa ao nível da ordem de palavras mas afirma que, após a integração de Goa na União Indiana, se deu um progressivo abandono destas estruturas consideradas de influência portuguesa.

3.2 Línguas de contacto do período colonial

A Tabela 1 regista a presença de palavras portuguesas em línguas como o indo-inglês, o indo-francês e o pidgin inglês (da China), facto que remete para o facto de estas variedades asiáticas ou derivações das línguas coloniais europeias, estabelecidas no continente após o português, terem sido grandemente influenciadas pelas variedades linguísticas que à data já circulavam pelas redes coloniais portuguesas. No caso do indo-inglês, este facto torna-se evidente no importante dicionário compilado por Henry Yule e Arthur Burnell no final do século XIX com o título de *Hobson-Jobson*. Logo na introdução, os autores reconhecem que «a large number of our Anglo-Indian colloquialisms, even if eventually traceable to native sources [...], have come to us through a Portuguese medium, and often bear traces of having passed through that alembic» (Yule/Burnell 1886, xviii–xix). Assim, entre as dezenas de exemplos que incluem no glossário contam-se palavras de origem portuguesa (e.g. ingl. *gentoo*, usado em tempos para referir uma pessoa hindu, a partir do ptg. GENTIO) e outras de origem asiática em cuja forma se reconhece um filtro português (e.g. ingl. *jack/jackfruit*, nome de um fruto local, do malaiala CHAKKA através do ptg. JACA), incluindo muitos topónimos (e.g. ingl. *Bassein*, do nome local VASAI através da adaptação portuguesa BAÇAIM).

O pidgin inglês falado em tempos nos portos chineses foi abundantemente documentado a partir de finais do século XVIII. Como comenta Hall (1944, 95), os registos mais antigos desta língua caracterizam-se por uma grande quantidade de palavras de origem portuguesa (e.g. *joss* ‘deus’, do ptg. DEUS, preservado ainda na palavra ingl. *joss-stick* ‘pau de incenso’), algumas das quais foram sendo substituídas por equivalentes de origem inglesa – o que provavelmente explica o número relativamente baixo de lexemas portugueses registados para o «Pidgin-English» na Tabela 1. A este facto não será alheio o facto de o pidgin inglês da China se ter desenvolvido sobretudo no contexto do comércio com o entreposto de Cantão, no qual a cidade de Macau e a ação portuguesa tiveram um papel primordial. É interessante notar que, no relato de uma viagem a Cantão na década de 1740, se conta como marinheiros britânicos foram abordados com uma frase que conjuga elementos lexicais portugueses e ingleses, mas na qual os do primeiro tipo (*carei*, do ptg. QUERER; *grandi*, do ptg. GRANDE; *pickenini*, do ptg. PEQUENINO/A) predominam sobre os do segundo (*hola*, do ingl. WHORE ‘prostituta’):

- (4) Pidgin inglês da China ou Pidgin português da China (Noble 1762, 240)

<i>Carei</i>	<i>grandi</i>	<i>hola,</i>	<i>pickenini</i>	<i>hola?</i>
querer	velho	prostituta	jovem	prostituta

‘Quer uma prostituta velha ou uma prostituta jovem?’

Exemplos como este colocam em questão a fronteira entre um pidgin de base portuguesa e um pidgin de base inglesa (veja-se Baxter 2009). Com efeito, uma proposta que tem suscitado algum debate é a de que a influência portuguesa no pidgin inglês

da China não se fique a dever ao simples contacto entre as duas línguas mas sim a uma relação de ascendência, segundo a qual o pidgin inglês teria derivado por relexificação de um preexistente pidgin de base portuguesa (veja-se Baker/Mühlhäusler 1990; Bolton 2003). A apoiar esta hipótese, algumas análises sustentam que o elemento português ultrapassa o domínio lexical e pode ser observado na própria estrutura do pidgin inglês da China. Matthews/Li (2012), por exemplo, recorrem ao português e às variedades de contacto da Ásia para explicar o uso de *for* como complementador (baseado no *para* português) e de *have* como cópula locativa (baseado nos reflexos de *ter* nos crioulos de base portuguesa da Ásia) no pidgin inglês da China.

O debate acerca da influência do português ou de um pidgin/crioulo português na formação do pidgin inglês da China remete para o da formação do pidgin e do crioulo ingleses do Havai, que, apesar de se referir a uma latitude e a uma diacronia diferente, recupera alguns destes argumentos. A presença de falantes de português no Havai é bastante diferente do que temos descrito até agora, já que se fica a dever a uma onda de migração mais tardia, do século XIX. A indústria do açúcar conheceu grande expansão no arquipélago no final do século XIX e foi neste contexto que para aí convergiram muitos portugueses (sobretudo madeirenses e açorianos) atraídos pelas oportunidades de trabalho nas plantações – mais de 30.000 entre os anos de 1872 e 1930, segundo Reinecke (1969). Esta força demográfica, aliada ao papel de muitos portugueses como supervisores das plantações, levou estudiosos a considerar que, na transição para o século XX, o português (a par de outras línguas, como o chinês e o havaiano) teria tido a oportunidade de influenciar o desenvolvimento do pidgin inglês do Havai e a sua transformação num crioulo falado como língua materna, processo que normalmente situam na década de 1920. Entre os estudos que atribuem certas características do crioulo inglês do Havai ao contributo português, contam-se Reinecke/Tokimasa (1934); Knowlton (1960; 1967); e, mais recentemente, Siegel (2000; 2008). Um caso que desde cedo tem sido apontado como evidência do contributo português é o das funções de *stei/stay*, que pode funcionar como uma cópula com predicados adjetivais que indiquem propriedades não-permanentes ou não-intrínsecas (5a), com valor locativo (5b) e ainda como marcador pré-verbal de aspeto progressivo (5c):

- (5) Crioulo inglês do Havai (Siegel 2000, 229; 212; 218)
- a. *She stei sik / *tall.*
2sf *stei* doente alta
'Ela está doente.' 'Ela está alta.'
 - b. *Everything stay on top one doily.*
tudo *stay* em cima um guardanapo
'Tudo está em cima de um guardanapo.'
 - c. *What you stay eat?*
que 2s *stay* comer
'O que estás a comer?'

Segundo Siegel (2000; 2008), a única das línguas envolvidas na formação do crioulo inglês do Havai (incluindo o inglês, de cujo verbo *stay* deriva a forma em questão) que permite explicar o padrão de multifuncionalidade registado para *stei/stay* é o português, já que, como se observa nas traduções em (5), o verbo *estar* em português tem a mesma distribuição, até mesmo quanto ao tipo de predicados adjetivais com os quais pode ou não ocorrer como cópula. Este debate em torno de *stei/stay* serve apenas para exemplificar a investigação que se debruça sobre o contributo português para a formação do crioulo inglês do Havai; outras características linguísticas que têm sido abordadas nesta perspectiva incluem, por exemplo, certos perfis prosódicos (Knowlton 1967), o complementador *fo/for* – semelhante ao que discutimos para o pidgin inglês da China – (a partir do ptg. *para*; veja-se Siegel 2000, 226), ou o sincretismo possessivo-existencial de (*hab*)*got* (também identificado, como vimos, no caso do pidgin inglês da China e ainda no cantonês, pelo que essas também são possíveis vias de entrada no Havai; veja-se Siegel 2008, 92–95).

4 Variedades asiáticas de português

Para além da formação de línguas pidgin e crioulas a que já aludimos, a expansão da língua portuguesa pela Ásia levou também ao estabelecimento de algumas variantes asiáticas do português que acusam o efeito do contacto linguístico. No contexto asiático, já Leite de Vasconcelos identifica na sua *Esquisse d'une dialectologie portugaise* (1901) vários «dialectos» portugueses. Desses, os que não classifica como «dialectos crioulos» são os de Goa, Macau e Timor, e ainda o «norteiro» (da antiga «província do Norte» do Estado Português da Índia, localizada na região contígua a Bombaim) e o da costa do Coromandel (costa sudeste da Índia, correspondente ao atual estado de Tamil Nadu). Para estes dois últimos, que já não existem, Leite de Vasconcelos baseou-se em corpora escritos que oscilavam entre as variedades crioulas locais e aproximações variáveis do português-padrão, e que levantam algumas dúvidas quanto à sua fidedignidade na representação de uma variante oral real. Os três primeiros subsistem e serão tratados nas secções 4.2, 4.3 e 4.4 respetivamente. Para além disso, desde cedo, temos evidência de ter existido a percepção de um modo tipicamente asiático – ou pan-asiático – de falar português; esse será o tema da secção 4.1.

4.1 O «português oriental»

A noção de que a Ásia produzira variantes particulares do português é antiga, sendo expressa por Contador de Argote, no século XVIII, nos seguintes termos:

«M. Ha mais algum Dialecto?

D. Ha os Dialectos ultramarinos, e conquistas de Portugal, como India, Brasil, &c., os quaes tem muytos termos das linguas barbaras, e muytos vocabulos do Portuguez antigo» (Contador de Argote 1725, 300).

Como aqui se vê, as referências mais antigas a um dialeto asiático do português assentam no reconhecimento de certas especificidades lexicais (a este respeito, veja-se também Tomás 2008, 462–466). Do ponto de vista do contacto, estabeleceu-se assim um contraponto com o que discutimos na secção 3, já que, neste caso, os termos atribuídos a um «português oriental» (Dalgado 1919, xviii) foram em grande medida tomados de línguas asiáticas. Em boa parte, este enorme influxo de léxico local relaciona-se com a necessidade de inserir no discurso em português novas realidades e referentes encontrados na Ásia. As origens etimológicas são diversas, mas encontramos neste grupo de lexemas uma importante parcela de origem malaiala, língua da antiga costa do Malabar, por ter sido a primeira com que o português estabeleceu um contacto estreito na Ásia (veja-se Thomaz 2015). Garcia de Orta, nos seus *Colóquios dos Simples e Drogas da Índia*, é bem explícito ao escrever, com referência ao betel:

«Chamamoslhe *betre*, porque a primeira terra dos Portuguezes conhecida foi o Malavar... Todos os nomes que virdes, que nam sam portuguezes, sam malavares; assi como *betre*, *chuna*, que he cal, *maynato*, que he lavador de roupa, *patamar*, que he caminheiro, e outros muytos» (Orta 1563; transcrito em Dalgado 1919, xxii).

Estas palavras e muitas outras constam daquela que é ainda a obra de referência para o estudo das particularidades lexicais do português asiático, publicada em 1919 por Sebastião Rodolfo Dalgado. O *Glossário Luso-Asiático* é uma grande compilação e estudo etimológico de termos correntes no discurso português na Ásia mas não necessariamente no português de outras regiões. Contudo, a observação das especificidades lexicais do português asiático em obras de carácter lexicográfico precede o *Glossário*. Vários dicionários da língua portuguesa registam palavras que identificam de forma mais ou menos explícita com a esfera asiática. Um dos mais importantes, publicado entre 1728 por Rafael Bluteau, identifica uma grande quantidade de termos com a Ásia através de descrições como «palavra da India», «termo dos portuguezes da India», «termo nautico da India», «entre os gentios da Asia», «he palavra da Asia», «termo da China», etc. São deste tipo centenas de palavras; mas iniciemos a ilustração com uma que já encontrámos: *mainato*. Bluteau define-a do seguinte modo:

«**MAYNATO.** He o homem, que na India tem o officio de lavar a roupa, e assim o Mainato, ou Lavandeiro da Aldea tem o seu Namarsim, ou retalho de varjea, que he assim como officio hereditario na sua familia».

Dalgado (1919), em concordância com Garcia de Orta, clarifica a etimologia dravídica do termo (do malaiala *maṇṇaṭṭan*). Porém, a palavra *mainato* revela também a dispersão que o vocabulário luso-asiático teve em todo o Oriente, já que o termo foi e é corrente em lugares tão distantes como Macau (onde um bairro é conhecido como *Tanque do Mainato*) e Moçambique. É ainda uma das palavras que, segundo Dalgado (1913), foram difundidas por via do português a várias línguas asiáticas, incluindo o malaio e o tétum.

Outro exemplo interessante é o da palavra *bandel* (pl. *bandeis*). Bluteau dá a seguinte definição:

«**BANDEL.** He palavra da Asia, quer dizer o mesmo que Bairro, Destricto de huma nação Estrangeira, tolerada em outro Paiz, donde logra privilegios, e assim os Portuguezes tem Bandel nos Reynos de Bengala, e Siaõ, donde ha Capitaõ môr, nomeado pelo Vice-Rey da India, o qual julga as causas movidas de hum Christaõ a outro, e faz respeitar aos Ministros Evangelicos pelas Justiças da terra».

Dalgado aumenta esta definição e acrescenta que se trata de uma derivação da palavra persa *bandar* ‘porto, cais’, usada também com essa forma na generalidade das línguas da Ásia Meridional, de onde provavelmente teria passado ao português.

Outras palavras originalmente luso-asiáticas acabaram por se impor noutras variedades do português (e.g. *atol*, *biombo*, *canja*, *chá*, *tufão*) e, inclusivamente, passar a outras línguas europeias (veja-se, a este respeito, Cardeira 2010). Um caso interessante e significativo desta última categoria é a palavra *mandarim*, atualmente utilizada em muitas línguas do mundo para designar a variante linguística oficial e mais difundida da China (entre outras aceções). Com origem na palavra sânscrita *mantri* ‘conselheiro, ministro’, as primeiras propostas etimológicas (Yule/Burnell 1886; Dalgado 1919) sugeriam que teria passado ao português por via do malaio *māntāri* e, aplicado inicialmente à classe administrativa chinesa, teria depois passado a referir a língua que falavam. Contudo, Thomaz (2015, 120) nota a ocorrência do termo num documento português de 1499, o que indicia que o empréstimo é mais antigo do que se supunha e deverá ter entrado no português por via do malaiala.

4.2 Goa, Damão e Diu (e Dadrá e Nagar-Haveli)

Goa, Damão e Diu foram colónias portuguesas entre o século XVI e 1961, altura em que uma ação militar as integrou na União Indiana. Tratando-se dos territórios sul-asiáticos com a mais longa presença portuguesa, é aí que se fala ainda o português, a par, nos casos de Damão e Diu, de línguas crioulas. É conveniente notar que a língua

portuguesa se estendeu a muitos outros locais do subcontinente indiano de onde entretanto desapareceu, incluindo largas faixas costeiras do sul, a região de Bombaim e a do Delta do Ganges. Até ao século XX, há referências ao ensino da língua portuguesa em locais como Bombaim, Allepey ou Vepery, e encontramos nos arquivos documentos em língua portuguesa produzidos em muitos outros, como Cochim, Mahé, Baçaim, etc. Estas variantes extintas são muito menos conhecidas, pelo que nos concentraremos aqui nas de Goa, Damão e Diu. Há porém uma diferença radical nas ecologias linguísticas de Goa, por um lado, e de Damão e Diu, por outro, pelo que devemos tratar estes territórios separadamente.

Na transição para o séc. XX, há a registar algumas tentativas de representar as especificidades do português de Goa, desde logo em obras de cariz literário tais como o romance *Jacob e Dulce: Scenas da vida Indiana* (Costa 1896). Os primeiros estudos a debruçarem-se de forma científica sobre as particularidades do português de Goa – neles caracterizado como uma L2 para a maioria dos seus falantes goeses – são da autoria de Dalgado (1900) e Leite de Vasconcelos (1901), sendo o primeiro o mais completo. Nesse texto, Dalgado aponta uma grande quantidade de traços fonéticos, fonológicos e morfossintáticos que distinguem o português de Goa do padrão europeu. Desses, vários são explicitamente atribuídos pelo autor à influência do concani, a língua da generalidade da população.² Estes incluem, no plano fonológico, a presença da consoante aspirada [k^h] e de aspiração com [h] em determinadas palavras de origem concani ou por ela influenciadas (e.g. *khamisa* ‘camisa’, por influência do conc. *khamís*, Dalgado 1900, 66); a generalização de certas formas de tratamento reverenciais («*tio*, se é da idade do pae de quem falla; *irmão*; se é um pouco mais velho que o interlocutor; *mana*, se é mais velha [...]», p. 68), por vezes pospostos ao nome próprio como em concani (e.g. *Paulo-tio*, *Roque-irmão*); ordem de palavras OV ocasional; possibilidade de elipse da cópula (*Toda família boa*, p. 71); colocação do pronome interrogativo em posição não-inicial (e.g. *Pequenos como estão?*, p. 70); ou ainda o emprego da partícula reportativa *hĩ* do concani (e.g. «Elle morreu *hĩ* = elle morreu, dizem», p. 71). Para além das características para as quais Dalgado identifica uma origem no contacto com o concani, muitas outras por ele apontadas podem ser interpretadas nessa linha, como sejam a pronúncia aproximante de <v> («V equivale ao inglês w», p. 67) ou a simples justaposição de dois numerais para indicar uma aproximação ou intervalo numéricos (*Há dois três dias* ‘Há dois ou três dias’; *são dez quinze pessoas* ‘são dez a quinze pessoas’, p. 71) – o que, não sendo impossível em português, é bastante mais regular nas línguas indo-árnicas.

O estudo de Dalgado inclui ainda um glossário de termos característicos do português de Goa, no qual pontuam, como seria de esperar, muitos empréstimos do

² Apesar de o concani ser a origem mais vezes invocada para explicar as peculiaridades do português de Goa, Dalgado (1900, 72) nota também a influência do inglês, que exemplifica com a frase *Os amigos e parentes são pedidos para assistirem*, decalcada do ingl. *The friends and relations are requested to attend*.

concani (e.g. *baí* ‘menina’, *bate/batte* ‘arroz com casca’, *funconias* ‘amuos’, *tape* ‘barrete’) e expressões que resultam de traduções literais do concani (e.g. *toca-bocca* ‘acepipe que se come com o arroz para despertar o apetite’, do conc. *tondák lāvun-chem*); mas também alguns identificados como de origem africana (*batuque*, um instrumento musical), malaiala (*teca*, um tipo de árvore), malaia (*godão* ‘armazém’) ou outras. Muitos dos termos nesta lista pertencem ainda ao conjunto de lexemas tidos como típicos do «português oriental» discutido na secção 4.1 (e.g. *maynato* ‘lavadeiro’, *pera* ‘goiaba’, etc.).

A partir de 1908, José Maria da Costa Álvares dedicou-se a comentar (e contrariar) as peculiaridades do português de Goa através de um consultório linguístico regular no jornal goês *Heraldo*, compilado no ano seguinte num volume (Álvares 1909) com o sugestivo título de *Palestras sobre os vícios de linguagem portuguesa mais comuns em Gôa*. Apesar do seu caráter eminentemente normativo e predominância de questões de ortografia e pronúncia, trata-se de uma fonte importante por identificar uma série de características típicas do português de Goa, para algumas das quais Álvares reconhece também a influência do concani. Um caso deste tipo, descrito como «um dos mais pittorescos e dos mais vulgares indianismos» (ibid., 77), refere-se ao uso da expressão *fazer chave* com o sentido de «fechar à chave», exemplificado com a frase *Fêche porta e faça chave, han?* ‘Feche a porta e tranque-a à chave, sim?’ e explicado por decalque do equivalente concani *tsavi côr* [lit. ‘chave’ + ‘fazer’] (ibid., 78).

Muitos dos traços apontados quer por Dalgado quer por Álvares serão recuperados num artigo bastante mais tardio que procura fazer um levantamento dos «*concanismos* ou *goanismos*» do português de Goa, definidos como «*expressões e modos tomados de concani mas alheios à índole da língua portuguesa*» (Dias 1981, 33). Os «concanismos» descritos são maioritariamente de natureza lexical (por exemplo *babá* ‘menino’) mas não só. No que diz respeito à ordem de palavras, por exemplo, Dias refere não apenas a ocasional ordem OV (já identificada por Dalgado) mas também a possibilidade de a oração relativa preceder o nome, à semelhança do concani, como na frase *ontem que veio rapaz é seu amigo?* ‘o rapaz que veio ontem é seu amigo?’ (ibid., 39).

Outro estudo importante deste período é o de Graciete Batalha, que, em 1982, dedica um livro à presença da língua e cultura portuguesas em Goa. Um dos capítulos faz uma caracterização do português goês com base em inquéritos e conversas com um número limitado de falantes, ressaltando que se tentou obter informação acerca da «fala popular e [...] fala dos jovens» (Batalha 1982, 48). Batalha nota, logo à partida, uma certa continuidade dos traços apontados 82 anos antes por Dalgado, incluindo a supressão do verbo copulativo ou a justaposição de numerais para uma aproximação numérica – e, ao transcrever como *Uasco* a pronúncia de um informante para o topónimo *Vasco (da Gama)*, revela ter identificado ainda a pronúncia aproximante de <v>. Ao nível lexical, Batalha afirma ter notado uma forte tendência para a substituição de termos portugueses pelos seus equivalentes em concani ou inglês. Esta observação coincide com uma investigação desenvolvida aproximadamente na

mesma altura sobre os domínios e frequência de uso do português em Panjim (Wherritt 1985; 1989), que dava conta de uma evidente contração. Quanto ao momento atual, ainda que saibamos que o português continua em uso em Goa, o seu estado concreto, os contextos de contacto que eventualmente estabelece e as suas características linguísticas permanecem um pouco nebulosos.

Ainda no que diz respeito a Goa, existe uma dimensão particular do debate em torno do papel do contacto na formação da variante local. Já nos escritos pioneiros de Dalgado e Leite de Vasconcelos se levanta a hipótese de aí se ter formado, à semelhança da maioria dos territórios luso-asiáticos, uma língua crioula que pudesse, em teoria, explicar certos traços do português de Goa. Apesar de nem um nem outro reconhecerem a sua existência na altura em que escreveram, não deixam de identificar certas semelhanças entre o português de Goa e os crioulos indo-portugueses conhecidos, sobretudo os do litoral a norte de Goa. Com efeito, para a caracterização do português de Goa que faz na sua *Esquisse* (1901, 139–141), Leite de Vasconcelos baseia-se não apenas no trabalho já referido de Dalgado mas também na análise de um romance publicado em 1864 no primeiro volume de um periódico de Margão, *Ilustração Goana*, onde se tenta representar a fala arcaizante de uma mulher «plebeia» de Salcete. Esta representação literária (necessariamente estereotípica) contém, efetivamente, grandes semelhanças com os crioulos indo-portugueses, desde a indicação de tempo e aspeto através de um marcador preposto a uma forma aparentemente invariável do verbo (*cand'eu ta vai par igrêz*) à forma composta do pronome 3pl *ell'outro* (do ptg. ELE + OUTRO). Para além disso, num artigo posterior, Dalgado (1921) comenta um poema claramente crioulo, supostamente recolhido em Goa, propondo que aí tenha vigorado em tempos um «crioulo cafreal», i.e. um crioulo falado pelos escravos africanos que chegaram a ser bastante numerosos (veja-se Cardoso 2010).

A hipótese da influência de um crioulo indo-português no desenvolvimento do português de Goa continua, assim, em aberto. Por outro lado, em Diu e Damão (bem como nos enclaves de Dadrá e Nagar-Haveli, próximos de Damão), a coexistência do português com uma língua crioula ainda se verifica, havendo ainda uma fluência generalizada em guzerate, a língua dominante na região. Na realidade, nestes casos, o interesse dos linguistas tem sido dirigido sobretudo para as línguas crioulas (veja-se, por exemplo, Clements/Koontz-Garboden 2002; Cardoso 2009), pelo que as conhecemos melhor do que ao português local. Sabemos, porém, que nem o uso do português neste território nem o reconhecimento de diferenças em relação ao padrão são recentes; num relato do final do século XIX, referente à população cristã de Diu, lemos que:

«Os usos e costumes dos christãos de Diu são os mesmos ou quasi mesmos, que os dos christãos de Goa e Damão. Fallam o portuguez, mais ou menos correctamente, não sendo todavia raros certos idiotismos, solecismos e barbarismos intoleráveis» (Quadros 1899, 98).

Os estudos recentes sobre Damão e Diu indicam que o português é uma L2 para a maioria (se não a totalidade) dos seus falantes, ao passo que os crioulos têm comuni-

dades importantes de falantes nativos. Torna-se claro, para além disso, que existem nestes territórios diversos tipos de competência em português que revelam diversas vias e motivações para a sua aquisição. No caso de Diu, Cardoso (2007) refere que o conhecimento da língua portuguesa pode ficar a dever-se a um historial de educação nesta língua (para uma faixa etária mais alta), de emigração para um país lusófono (e.g. Portugal ou Moçambique), permanência em Goa, preferências familiares, etc. A língua portuguesa difunde-se hoje em dia à comunidade não apenas pelos canais que lhe são internos mas também através da televisão por satélite, do contacto ocasional com falantes nativos de português (viajantes, jornalistas, investigadores, etc.) e da permanência temporária de emigrantes a residir em países lusófonos como Portugal ou Moçambique.

A relação entre o português e o crioulo não é porém totalmente pacífica, dado o diferencial de prestígio que dá vantagem ao primeiro. Esse facto, associado à circunstância de a competência em português de alguns falantes ser reduzida, produz com frequência fenómenos de hesitação e de hipercorreção na produção desta língua. O contacto entre as duas línguas tem claras consequências linguísticas, que conhecemos melhor na perspectiva do impacto do português sobre os crioulos do que no sentido inverso. Em vários estudos, a influência do português é invocada por Clancy Clements para explicar certas diferenças entre os crioulos de Damão e Diu, por um lado, e os demais crioulos indo-portugueses com menor exposição ao português (Clements 2009b) e até para entender o desenvolvimento de dois socioletos distintos em Damão (Clements 2009b; 2014): um mais urbano, com maior influência do português, e.g. na forma do pronome 3pl *ilz* (do ptg. *ELES*), e outro mais rural, que retém formas crioulas antigas como, no caso em apreço, a forma composta do pronome 3pl *ilot* (do ptg. *ELE + OUTRO*).

Quanto às particularidades do português de Damão e Diu e do papel do contacto linguístico no seu desenvolvimento, esses são campos de pesquisa ainda por explorar.

4.3 Macau (e Hong Kong)

A língua historicamente mais difundida em Macau e na região circundante é o cantonês. Para além disso, Macau é outro dos locais da Ásia onde se desenvolveu um crioulo de base portuguesa, por sinal relativamente bem conhecido e documentado a partir do século XIX (Barreiros 1943–1944). Este crioulo, conhecido como *maquista*, *macaísta* ou *patuá*, tem grandes afinidades com os do sudeste asiático, sobretudo o de Malaca. Como tal, pode depreender-se que terá sido transplantado para este território durante a colonização de Macau no século XVI, e não se terá desenvolvido *ab ovo* num momento posterior.

Este facto é relevante na medida em que, em Macau, se pode dizer que a coexistência entre o português e o crioulo (e o cantonês) foi uma constante desde o primeiro momento, sendo que, para uma boa parte da população colonial, o portu-

guês poderá ter sido, quando muito, uma L2. Durante muito tempo, a variedade do «português» com maior difusão no território parece ter sido uma variedade de contacto (crioulo ou pidgin). Uma obra chinesa de 1745–1746, intitulada *Ou Mun Kei Leok*,³ regista em caracteres chineses alguns excertos da língua em uso na altura em Macau e esta, na realidade, revela-se muito próxima do crioulo maquista (Batalha 1995; Baxter 2009, 287).⁴ Leite de Vasconcelos, um dos primeiros filólogos a tentar uma sistematização do reportório linguístico de Macau, concentra-se sobretudo no crioulo, mas acrescenta:

«A côté du <macaísta> proprement dit, ou dialecte créole de Macao, on emploie un langage à prétentions littéraires, qui rapelle plus ou moins le portugais usuel; il y a en outre un portugais parlé par les Chinois» (Leite de Vasconcelos 1901, 149).

Esta descrição aponta, portanto, para um ecologia linguística que albergava o crioulo, formas de português L2 e uma variedade do português que, sendo próxima do «portugais usuel», não deixava certamente de ter as suas peculiaridades. Esta coexistência do português com o crioulo veio a revelar-se instável, sobretudo a partir da segunda metade do século XIX, com a difusão da imprensa e educação em língua portuguesa. O efeito da pressão exercida pelo português sobre o maquista fica bem expresso numa carta (provavelmente fictícia) de 1865, escrita em crioulo por uma tia macaense à sua sobrinha em Hong Kong, que contém o seguinte *post scriptum*:

«P. S. – Vós lôgo sintí grandi diferença na minha modo di escrivê. Eu já aperfeiçoá bastante neste um pôco tempo. Tudo este escola novo de machu e femia, e aquele gazeta Ta-ssi-yang-kuo já fazê indretá bastante nosso lingu» (reproduzida em Barreiros 1943–1944, 31).

Com efeito, o crioulo entrou então num ciclo de descrioulização que praticamente o aniquilou como língua de uso corrente, sendo substituído entre a comunidade anteriormente crioulófona pelo português. Em consequência deste processo, torna-se necessário considerar, como fazem muitos autores, que as especificidades do atual português de Macau possam resultar, pelo menos em parte, de retenções deste crioulo entretanto abandonado – sem prejuízo de explicações que coloquem em evidência o contacto com outras línguas da ecologia linguística, em particular o cantonês. A própria Graciete Batalha, que a partir da década de 50 do século XX trabalhou o tema do português em Macau, revela muitas dificuldades em traçar a fronteira entre o maquista e uma variedade macaense de português. Numa série de artigos publicados no jornal *Notícias de Macau* em 1958 (posteriormente coligidos em Batalha 1974) condensa toda esta diversidade linguística no termo «Língua de Macau» e descreve a situação do seguinte modo:

³ Para uma tradução recente, veja-se Guangren/Rulin (2009).

⁴ Para mais informação acerca da difusão de uma variedade de contacto do português no contexto do comércio com Cantão, veja-se Matthews/Li (2012), Li/Matthews (2016).

«Quando chegámos a Macau, em 1949, [...] constatámos [...] como essa linguagem, diferente embora da do mesmo nível (popular cidadão) da Metrópole, não era já a mesma que os estudos de Leite de Vasconcelos e seus contemporâneos nos mostravam. De então para cá, devido a mais estreitos contactos com a Metrópole, o dialecto tem evolucionado rapidamente para uma maior aproximação com o português normal, sobretudo no que diz respeito ao vocabulário e à pronúncia. [...] Note-se, no entanto, que a fala das pessoas mais velhas não é ainda idêntica à dos seus filhos, isto é, à dos indivíduos de cerca de 50 anos. A linguagem dos muito idosos (se não são de grandes letras, claro) é quase inteiramente incompreensível para quem chega de fora, ao passo que a da geração média se compreende aos primeiros contactos, embora exigindo de começo certo esforço de atenção» (Batalha 1974, 1958, 18).

Na sua ótica, assistia-se então a uma fase de transição tendente à emergência de uma forma tipicamente macaense de falar português (Batalha 1959; também Diez 1981; Gaião 1999), sendo precisas «mais duas ou três gerações antes que a língua se uniformize em moldes novos» (Batalha 1974, 1958, 19). Apesar da grande instabilidade de formas que encontra, a autora avança algumas características desse novo modo macaense de falar, como sejam a irregularidade na marcação de pluralidade ou na concordância de género e número. Para uma delas, sugere uma explicação assente na influência do cantonês: a substituição da vibrante múltipla do português pela vibrante simples [r] ou pela líquida [l] (Batalha 1974, 1958, 45–46). Outra observação importante no capítulo do contacto é a da mistura do português com o inglês e o cantonês no discurso dos falantes mais jovens. Hoje em dia, este tipo de *code-switching*, motivado por um elevado multilinguismo, é ainda característico da produção da comunidade macaense.

Com a entrega de Macau à China, em 1999, o português manteve o seu estatuto de língua oficial no território, agora a par do chinês. A adaptação da língua portuguesa a esta nova realidade tem vindo a motivar uma nova vaga de investigação linguística, tendente não apenas a observar os efeitos de políticas linguísticas como também o eventual estabelecimento de um «português de Macau» bem definido (veja-se, por exemplo, Correia 1998; Noronha/Chaplin 2012). Em textos recentes dedicados precisamente ao historial de contacto linguístico em Macau e na região, Baxter (1996; 2009) sintetiza algumas características salientes do português macaense, no qual reconhece influências diversas: o efeito da assimilação do anterior crioulo, o chinês, o inglês e o português europeu.

No campo fonológico, Baxter nota a competição entre realizações semelhantes ao português europeu e outras divergentes (Baxter 2009, 296–299). No caso da africada [tʃ] para realizar <ch> (em competição com a fricativa [ʃ] do português-padrão europeu moderno), encontra paralelos tanto nos crioulos de Macau e de Malaca como no cantonês. Já em casos como a realização simples do que seriam vibrantes múltiplas no padrão europeu (já notada por Batalha) ou a monotongação dos ditongos [ɛj] (>[e]) e [ow] (>[o]), por exemplo, encontra uma maior semelhança com os crioulos de Malaca e de Macau, sugestiva de retenção de um antecedente crioulo. Ao contacto com o cantonês atribui, por exemplo, a simplificação de certos grupos consonânticos (e.g. *principal* > *pincipal*).

Ao nível morfossintático, Baxter (1996, 325–326; 2009, 300–306) nota, entre outros fenómenos, a ausência ocasional do artigo definido, variabilidade na concordância de número e género (no sintagma nominal) e de pessoa (entre o verbo e o sujeito) e certas seleções preposicionais peculiares. Também aqui se avançam explicações apoiadas no contacto linguístico ora com o antecedente crioulo, ora com o cantonês, ora ambos – não apenas na perspetiva da transferência direta dessas línguas mas também na dos efeitos que tem para a aquisição do português a *ausência* nessas línguas de determinadas categorias funcionais do português.

Os contributos de todos estes estudos permitem uma caracterização inicial das particularidades do português de Macau, mas resta ainda muito trabalho por fazer, nomeadamente no sentido de apurar a estabilidade ou instabilidade dos traços identificados e as suas fontes mais prováveis. Por outro lado, Macau tornou-se recentemente um pólo de atração para múltiplas populações de todo o mundo, incluindo o mundo lusófono, pelo que será interessante observar de que forma responde o português macaense aos diversos estímulos linguísticos que aí aportam.

A presença do português noutros pontos da costa chinesa em tempos recentes ficou a dever-se, em grande medida, à diáspora macaense. A partir de meados do século XIX, um número considerável de macaenses migraram sobretudo para Hong Kong e Xangai, levando consigo o crioulo maquista mas também o português. A comunidade de Hong Kong, sempre mais numerosa do que a de Xangai, contava cerca de 2000 pessoas no ano de 1900 (Baxter 1996, 326) e conseguira preservar até então o ensino do português através das suas associações comunitárias (Batalha 1990, 294). A comunidade prosperou até 1941, altura em que boa parte dela se refugiou em Macau em fuga da ocupação japonesa de Hong Kong. No final da Guerra, em 1945, apenas uma parte desta população regressou a Hong Kong.

Batalha (1990, 294–295) considera que a permanência em Macau nos anos da Guerra serviu para revitalizar e perpetuar o conhecimento do português entre os macaenses de Hong Kong (o que indicia algum declínio em meados do século XX) mas nota também a presença de numerosos empréstimos ingleses no português desses falantes: por exemplo *oficina* ‘escritório’ (do ingl. OFFICE), *affordar* ‘ter recursos para’ (do ingl. AFFORD) ou *improvar* ‘melhorar’ (do ingl. IMPROVE). Tendo sido a língua colonial em uso em Hong Kong, é compreensível que o inglês tenha tido impacto sobre o português aí falado mas, para além disso, há que considerar, tal como em Macau, o papel do crioulo maquista. Arana-Ward, por exemplo, aborda a questão da descrioulização do maquista entre a comunidade macaense de Hong Kong e nota que, à data da sua recolha de dados, a competência da maioria dos macaenses que contactou se situava «somewhere short of a standard form of Portuguese» (Arana-Ward 1977, 108). A língua portuguesa atualmente falada em Hong Kong carece ainda de pesquisa que defina de forma mais concreta as suas especificidades em relação ao português de Macau e o papel do contacto linguístico no seu desenvolvimento.

4.4 Timor

Quando, na segunda metade do século XIX, se desenvolve o interesse pelas particularidades do português de Timor, assistimos a uma série de equívocos na classificação dos modos de falar timorenses (veja-se Baxter/Cardoso no prelo). Alguns autores admitem a coexistência de um ou mais crioulos com o português L1 e L2, outros negam a existência de um crioulo – sendo que a própria definição de «crioulo» é, nestas fontes, bastante variável. Assim, quando, em 1867, o antigo governador da colónia Afonso de Castro refere um «crioulo», parece estar efetivamente a descrever uma ou mais variedades de português L2:

«Têm os timores tão grande facilidade para aprender o malaio, quanta dificuldade para o portuguez, que rarissimos fallam correctamente, e que uma grande parte dos chefes e dos habitantes de Dilly estropiam, fazendo um crioulo, que nos custa quasi tanto a comprehender, como os dialectos timores» (Castro 1867, 328).

Pouco mais tarde, no prefácio a um catecismo em língua tétum publicado pelo padre Aparício da Silva, o bispo de Macau e Timor, D. António de Medeiros, torna a referir a situação do português timorense nos seguintes termos:

«Aceito o que deseja que aceite, porque é com verdadeira emoção que eu vejo os primeiros fructos d'um trabalho a que V. Rev.^{ma} se votou dedicadamente ha mais de oito annos, que eu reputo de grande valor para a civilisação dos povos malasios da nossa colonia de Timor, sabendo por experiencia, que recebem mais ideias d'uma pratica em lingua do paiz, que de muitas palestras no algaraviado portuguez d'aquelle paiz» (Aparício da Silva 1885, 7).

Neste caso, tanto pode estar em causa uma forma de português L1 como L2, sendo ainda possível que este «algaraviado» fosse adotado por falantes nativos de português (por exemplo, os padres) para comunicação com os timorenses. Mas parece certo que, então tal como agora, o português era falado sobretudo como L2 em Timor. Por outro lado, é agora claro que Timor assistiu efetivamente ao desenvolvimento de pelo menos uma língua crioula com fortes ligações aos crioulos de base portuguesa do sudeste asiático, o crioulo de Bidau (Baxter 1990), e que aí se estabeleceu também uma significativa comunidade macaense que transplantou o crioulo maquista.

Neste contexto altamente complexo, os dados linguísticos recolhidos carecem de uma interpretação que nem sempre é fácil; os espécimes são curtos e fragmentários, falta muitas vezes informação contextual essencial e, em consequência, as fronteiras entre as diversas modalidades e competências linguísticas esbatem-se. Ainda assim, temos dessa época excertos linguísticos altamente divergentes do português-padrão europeu a muitos níveis, nos quais facilmente se reconhece a influência das línguas timorenses. Esta influência era evidente para os observadores da época; nas palavras de Tancredo do Casal Ribeiro, um dos correspondentes de

Hugo Schuchardt em 1882,⁵ «N'este portuguez corrupto que aqui chamam lingua da praça, não fazem mais que dar ao portuguez a construcção da lingua indigena [...]». A título de exemplo, vejamos o que refere um dos textos publicados de que dispomos, saído em 1885:

«Diz-se no *patois* timorense *sol-sube* e *sol-cáe*, correspondendo aquella designação ao lado que o sol nasce e esta ao lado que o sol se põe; por exemplo: nascente chamam *sol-sube*, poente *sol-cáe*» (Vaquinhas 1885, 63).

De acordo com a interpretação de Baxter/Cardoso (no prelo), estes compostos correspondem ao decalque de uma estratégia tipicamente timorense (presente no tétum bem como noutras línguas timorenses) que constrói a referência ao leste com a expressão *lorosae*, literalmente 'sol + sobe', e ao oeste com *loromonu*, literalmente 'sol + cai'.

A difusão da língua portuguesa em Timor, que dependeu em grande medida do aparelho militar e do sistema educativo operado por missionários (Baxter 1996, 312), nunca chegou a ser um fenómeno de massas durante o período colonial. O século XX assistiu a um crescimento tímido da alfabetização em língua portuguesa, intensificado a partir da década de 50 (Hajek 2000; Taylor-Leech 2009), ao mesmo tempo que desapareciam os crioulos falados até então. Posto que instituída ainda no século XIX, a educação teve fraca expressão até à década de 1970, como explica Luís Filipe Thomaz numa obra que inclui ainda uma breve caracterização linguística do português falado pelas classes instruídas de Timor em torno da época da descolonização (Thomaz ²1990). Algumas particularidades aí apontadas são de ordem lexical – incluindo, por sinal, bastantes lexemas típicos do «português oriental» discutido na secção 4.1. Apontam-se algumas derivações semânticas particulares de palavras portuguesas, como sejam o uso de *morador* para referir um «membro da milícia de um reino nativo, de tropas de segunda linha» (Thomaz ²1990, 334–335), mas também empréstimos de origem tétum (e.g. *lúlic* 'tabu, objeto sagrado', p. 336), malaia (e.g. *sarão* 'pano de enrolar em torno da cintura', do mal. SARONG, p. 335) e de outras origens. Ao nível fonético, Thomaz identifica o cancelamento da oposição entre [p] e [f] (e do [b], entre falantes nativos de galoli), a redução do sistema de sibilantes ao segmento surdo [s] e a presença de aspiração – tudo traços atribuídos à transferência de características fonológicas das diversas línguas de Timor. Ao nível morfossintático, registam-se aí fenómenos como a frequente ausência de concordância nominal e ainda a produção de um classificador nominal, demonstrado no exemplo seguinte:

- (6) Português de Timor (Thomaz ²1990, 337)
três pessoas de gente
 'três pessoas'

5 Carta manuscrita. *Hugo Schuchardt Archiv* (Universidade de Graz), nr. 01553.

Thomaz não fornece mais dados acerca da distribuição desta construção mas vê no uso de *peessoas* em (6) um classificador modelado pelo tétum. Com efeito, o tétum tem uma série de classificadores que medeiam alguns nomes e um numeral, sendo que todos estes têm também um valor lexical noutros contextos (por exemplo ‘folha’, no caso do classificador usado com objetos finos e achatados, ou ‘semente’, no caso do que se usa com objetos pequenos e redondos) (Williams-van Klinken/Hajek/Nordlinger 2002, 39). De entre todos estes, interessa-nos para o caso *nain*, que tem o significado primário de ‘senhor, dono’ e na sua utilização como classificador ocorre (opcionalmente) com nomes [+HUM] e numerais maiores que um, como em (7):

(7) Tétum (Williams-van Klinken/Hajek/Nordlinger 2002, 39)

feto nain rua
mulher CLS dois
‘duas mulheres’

É de notar que, neste caso, a proposta influência do tétum sobre o português não implica um empréstimo lexical nem um simples decalque, já que as ordens de palavras de (6) e de (7) diferem consideravelmente (NUM-CLS-N no primeiro caso, N-CLS-NUM no segundo). Trata-se, isso sim, da extensão funcional de um termo português (*peessoa*) com base num elemento funcional tétum (*nain*) mas preservando algumas características sintáticas do português, como sejam a ordem de palavras ou a concordância nominal.

Apesar de estas considerações linguísticas de Thomaz terem sido compostas durante a década de 1980, elas referiam-se ao português falado «nos últimos anos da administração portuguesa» (Thomaz 1990, 313), ou seja, ao período imediatamente anterior a 1975. O período de ocupação indonésia que se lhe seguiu (1975–1999) teve como efeito uma drástica redução do uso da língua portuguesa na sociedade timorense, exceto nos segmentos envolvidos na luta de resistência (Matan Ruak 2001) e em algumas instituições educativas e religiosas. Esta situação inverteu-se quando, em 2002, a nova constituição da nação independente de Timor-Leste consignou o português e o tétum como línguas cooficiais. Em consequência disso, tem-se assistido a uma vaga de difusão da língua portuguesa por via sobretudo do sistema educativo. No censo nacional de 2010, aproximadamente um quarto da população admitia algum conhecimento do português, o que corresponde a um significativo aumento de cerca de 12% em relação ao anterior censo de 2004 (Taylor-Leech 2009; Afonso/Goglia 2015a). Assim, tal como no passado, o português volta a ganhar uma crescente população de falantes L2 sem, contudo, chegar a ocupar o lugar de língua materna de uma grande comunidade.

Este processo de reintrodução do português no domínio público timorense tem sido acompanhado por um grande interesse na sua documentação linguística, em parte porque tal permite observar processos de contacto linguístico em ação. A difusão e apropriação do português não progridem, contudo, de forma equivalente por todo o território e em todos os estratos sociais, ao que acresce o facto de os timorenses serem,

por regra, pelo menos bilingues. Os mais recentes estudos sobre o português de Timor têm, portanto, de lidar com produções extremamente variáveis e flutuantes, refletindo estados diversos de aquisição da língua, o efeito de diferentes estratégias e limitações pedagógicas, etc. Assim, os autores que se dedicam a este assunto tendem a inserir as suas observações no contexto de uma variedade «emergente» (e.g. Albuquerque 2011) ou «em formação» (e.g. Afonso/Goglia 2015a) do português, procurando nas produções linguísticas atuais as sementes do que poderá vir a ser uma variedade nacional estabilizada. Outra proposta (e.g. Albuquerque 2011, 69–71) é a de organizar a variação encontrada em termos de um contínuo de competência linguística que vai do conhecimento mais rudimentar do português como L2 (mais influenciada pelas línguas timorenses) à competência nativa (mais próxima do padrão europeu).

A dimensão do contacto linguístico na perspetiva da aquisição de L2 é portanto inevitável. Com efeito, muitos dos fenómenos linguísticos que têm sido estudados têm a sua raiz precisamente na transferência de características das línguas timorenses faladas pelos aprendentes de português L2 ou nas dificuldades particulares de aquisição motivadas pela ausência de certas características nessas línguas – ainda que este tipo de pesquisa seja por vezes dificultado pela ausência ou insuficiência de descrições para várias das línguas timorenses. Alguns fenómenos que têm sido alvo de investigação nesta perspetiva incluem, entre outros, o uso de *já* pré-verbal como marcador de perfeividade (Batoréo 2010), a desnasalização (Albuquerque 2011), fenómenos de ordem de palavras não-canónica (Albuquerque 2012) ou ainda certas inovações (generalização ou supressão) no uso do clítico *se* (Afonso/Goglia 2015b).

Tal como na transição para o século XX, o debate corrente em torno do papel do contacto tem porém uma outra dimensão que retoma a presença de línguas crioulas no território. Certas características observadas no português timorense atual são reminiscentes também dos crioulos do sudeste asiático, incluindo do que se conhece do crioulo de Bidau. Esta circunstância levanta a hipótese de estarmos em presença de retenções dessas línguas (veja-se, por exemplo, Carvalho 2001), o que entra por vezes em competição com explicações apoiadas na influência sincrónica das línguas timorenses. No caso de Timor, mais até do que noutros locais da Ásia, o campo de investigação sobre o português em contacto permanece aberto e em crescimento.

5 Referências

- Afonso, Susana/Goglia, Francesco (2015a), *Portuguese in East Timor as a non-dominant variety in the making*, in: Rudolf Muhr/Dawn Marley/em colaboração com Anu Bissoonauth-Bedford/Leo Kretzenbacher (edd.), *Pluricentric Languages: New Perspectives in Theory and Description*, Wien, Lang, 193–205.
- Afonso, Susana/Goglia, Francesco (2015b), *Linguistic innovations in the immigration context as initial stages of a partially restructured variety: Evidence from SE constructions in the Portuguese of the East Timorese diaspora in Portugal*, *Studies in Hispanic and Lusophone Linguistics* 8:1, 1–33.

- Albuquerque, Davi Borges (2011), *O Português de Timor-Leste: contribuições para o estudo de uma variedade emergente*, PAPIA 21:1, 65–82.
- Albuquerque, Davi Borges (2012), *Esboço morfossintático do português falado em Timor-Leste*, *Moderna språk* 106:1, 1–10.
- Álvares, José Maria da Costa (1909), *Palestras sobre os vícios de linguagem portuguesa mais comuns em Gôa por A. Castro Alves*, vol. 1, Nova Goa, Casa Luso-Franceza.
- Anquetil-Duperron, Abraham Hyacinthe (1786), *Des recherches historiques et géographiques sur l'Inde*, in: Jean Bernoulli (ed.), *Description historique et géographique de l'Inde*, vol. 2, Berlin, Bourdeaux, XI–XII.
- Aparício da Silva, Sebastião M. (1885), *Catecismo da doutrina cristã em Tétum*, Macau, Tipografia do Seminário.
- Arana-Ward, Marie (1977), *A synchronic and diachronic investigation of Macanese: The Portuguese-based creole of Macao*, dissertação de Mestrado, University of Hong Kong.
- Baker, Philip/Mühlhäusler, Peter (1990), *From business to pidgin*, *Journal of Asian Pacific Communication* 1:1, 87–115.
- Barbosa, Jorge Morais (1969), *A língua portuguesa no mundo*, Lisboa, Agência Geral do Ultramar.
- Barreiros, Leopoldo Danilo (1943–1944), *O dialecto português de Macau*, *Renascimento*, número especial.
- Barros, João de (1615), *Quarta década da Ásia: Dos feitos que os Portugueses fizeram no descobrimento, e conquista dos mares, e terras do Oriente*, Madrid, Impressão Real.
- Batalha, Graciete Nogueira (1959), *Estado actual do dialecto macaense*, *Revista Portuguesa de Filologia* 9, separata.
- Batalha, Graciete Nogueira (1974, 1958), *Língua de Macau: O que foi e o que é*, Macau, Imprensa Nacional de Macau.
- Batalha, Graciete Nogueira (1982), *Língua e cultura portuguesas em Goa: Estado actual*, Macau, Serviços de Educação e Cultura de Macau.
- Batalha, Graciete Nogueira (1990), *Situação e perspectivas do Português e dos Crioulos de origem portuguesa na Ásia Oriental (Macau, Hong Kong, Malaca, Singapura, Indonésia)*, in: *Actas do Congresso sobre a situação actual do português no mundo*, vol. 1, 2.ª ed., Lisboa, Instituto de Cultura e Língua Portuguesa, 287–303.
- Batalha, Graciete Nogueira (1995), *O Português Falado e Escrito pelos Chineses de Macau*, Macau, Instituto Cultural de Macau.
- Batoré, Hanna J. (2010), *Funções do marcador polissémico «já» no Português de Timor-Leste: Importância do conhecimento da(s) língua(s) materna(s) dos aprendentes do Português L2 no processo de aquisição/aprendizagem da língua não-materna*, in: *Textos seleccionados do XXV Encontro Nacional da APL*, Lisboa, Associação Portuguesa de Linguística, 211–224.
- Baxter, Alan N. (1990), *Notes on the Creole Portuguese of Bidau, East Timor*, *Journal of Pidgin and Creole Languages* 5:1, 1–38.
- Baxter, Alan N. (1996), *Portuguese and Creole Portuguese in the Pacific and Western Pacific Rim*, in: Stephen A. Wurm/Peter Mühlhäusler/Darrell T. Tryon (edd.), *Atlas of Languages of Intercultural Communication In the Pacific, Asia and the Americas*, vol. 2.1, Berlin/New York, Mouton de Gruyter, 299–338.
- Baxter, Alan N./Cardoso, Hugo C. (no prelo), *Early notices regarding Creole Portuguese in former Portuguese Timor*, *Journal of Language Contact*.
- Bluteau, Rafael (1728), *Vocabulário Portuguez e Latino*, Coimbra, Colégio das Artes da Companhia de Jesus.
- Bolton, Kingsley (2003), *Chinese Englishes: A sociolinguistic history*, Cambridge, Cambridge University Press.

- Buultjens, Alfred E. (1885), *On some Portuguese words commonly used by the Sinhalese*, *The Orientalist* 2:11–12, 214–218.
- Cardeira, Esperança (2010), *O Português no Oriente e o Oriente no Português*, in: Valeria Tocco (ed.), *L'Oriente nella lingua e nella letteratura portoghese*, Pisa, Edizioni ETS, 81–93.
- Cardoso, Hugo C. (2007), *Linguistic traces of colonial structure*, in: Eric Anchimbe (ed.), *Linguistic identity in postcolonial multilingual spaces*, Newcastle, Cambridge Scholars, 164–181.
- Cardoso, Hugo C. (2009), *The Indo-Portuguese language of Diu*, Utrecht, LOT.
- Cardoso, Hugo C. (2010), *African slave population of Portuguese India: Demographics and impact on Indo-Portuguese*, *Journal of Pidgin and Creole Languages* 25:1, 95–119.
- Cardoso, Hugo C. (2014), *Factoring sociolinguistic variation into the history of Indo-Portuguese*, *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 5, 87–114.
- Cardoso, Hugo C./Baxter, Alan N./Pinharanda Nunes, Mário (edd.) (2012), *Ibero-Asian Creoles: Comparative perspectives*, Amsterdam/Philadelphia, Benjamins.
- Carvalho, Maria José Albarran de (2001), *Timor Lorosa'e, características das línguas crioulas e do português conservado na zona: contribuição para a língua oficial*, *Studies of Languages and Cultures of East Timor* 4, 20–36.
- Castro, Afonso de (1867), *As possessões portuguesas na Oceânia*, Lisboa, Imprensa Nacional.
- Chavan, V.P. (1924), *The Konkan and the Konkani language*, *Journal of the Anthropological Society of Bombay* 12:7, 853–917.
- Clements, J. Clancy (2000), *Evidência para a existência de um pidgin português asiático*, in: Ernesto d'Andrade/Maria Antónia Mota/Dulce Pereira (edd.), *Crioulos de Base Portuguesa*, Braga, Associação Portuguesa de Linguística, 185–200.
- Clements, J. Clancy (2009a), *The linguistic legacy of Spanish and Portuguese: Colonial expansion and language change*, Cambridge, Cambridge University Press.
- Clements, J. Clancy (2009b), *Accounting for some similarities and differences among the Indo-Portuguese creoles*, *Journal of Portuguese Linguistics* 8:2, 23–47.
- Clements, J. Clancy (2014), *Lectal differences in Daman Indo-Portuguese*, *Revista de Crioulos de Base Lexical Portuguesa e Espanhola* 5, 115–156.
- Clements, J. Clancy/Koontz-Garboden, Andrew (2002), *Two Indo-Portuguese Creoles in contrast*, *Journal of Pidgin and Creole Languages* 17, 191–236.
- Coelho, F. Adolfo (1882), *Os dialectos românicos ou neo-latinos na África, Ásia e América; Notas complementares*, *Boletim da Sociedade de Geografia de Lisboa* 3.^a série, vol. 8, 451–478.
- Contador de Argote, Jeronymo (1725), *Regras da lingua portugueza, espelho da lingua latina, ou disposição para facilitar o ensino da lingua latina pelas regras da portugueza*, Lisboa, Officina da Musica.
- Correia, Ana Cristina Rouillé (1998), *A língua portuguesa em Macau: Do passado para o presente: que futuro?*, dissertação de Mestrado, Universidade de Macau.
- Costa, Francisco João da [Gip] (1896), *Jacob e Dulce: Scenas da vida Indiana*, Margão, Tipografia do Ultramar.
- Dal, Nikolaus (1733), *Nachricht von den Portugiesen in Indien*, in: Gotthilf August Francken (publ.), *Drey und dreyßigste CONTINUATION Des Berichts Der Königlichen Dänischen Mißionarien in Ost-Indien/Worin eine Ausfuehrliche Nachricht Von der gantzen Verfassung der Mißion, Wie auch Die Fortsetzung des Tage-Registers vom Jahr 1731. und einige Briefe der Herren Mißionarien enthalten, Nebst einem Anhang Von der Evangelischen Mission zu Madras*, Halle, publ. por Gotthilf August Francken, 917–924.
- Dalgado, Sebastião Rodolfo (1900), *Dialecto Indo-Português de Goa*, *Revista Lusitana* 6, 63–84.
- Dalgado, Sebastião Rodolfo (1913), *Influência do vocabulário português em línguas asiáticas (abrangendo cerca de cinquenta idiomas)*, Coimbra, Imprensa da Universidade.
- Dalgado, Sebastião Rodolfo (1919), *Glossário Luso-Asiático*, Coimbra, Imprensa da Universidade.

- Dalgado, Sebastião Rodolfo (1921), *Berço duma cantiga em Indo-Português*, Revista Lusitana 22, Separata.
- Dias, Filinto Cristo (1981), *Peculiaridades dialectais do português de Goa*, Revista da Academia da Língua e Cultura Portuguesa 1:1, 33–39.
- Diez, Blanca Aurora (1981), *Language change in progress in the Portuguese of Macau*, tese de Doutoramento, Georgetown University.
- Ferreira, Ana Maria P. (2000), *Estrangeiros na Índia no tempo de Afonso de Albuquerque, os anónimos*, Anais de História de Além-Mar 1, 53–58.
- Gaião, Raul Leal (1999), *Aspectos lexicais na obra de autores macaenses*, dissertação de Mestrado, Universidade de Macau.
- Guangren, Yin/Rulin, Zhang (2009), *Breve Monografia de Macau*, edd. Rui Manuel Loureiro/Zhao Chunchen/Jin Guo Ping, Macau, Instituto Cultural.
- Hajek, John (2000), *Towards a language history in East Timor*, Quaderni del Dipartimento di Linguistica – Università di Firenze 10, 213–227.
- Hajek, John/Williams-van Klinken, Catharina (2003), *Um sufixo românico numa língua austronésia: «-dor» em Tétum*, Revue de Linguistique Romane 67, 55–65.
- Hall Jr., Robert A. (1944), *Chinese Pidgin English grammar and texts*, Journal of the American Oriental Society 64, 95–113.
- Hamilton, Alexander (1727), *A new account of the East Indies, being the observations and remarks of Capt. Alexander Hamilton, who spent his time there from the year 1688 to 1723, trading and travelling, by sea and land, to most of the countries and islands of commerce and navigation, between the Cape of Good-Hope, and the Island of Japon*, Edinburgh, Mosman.
- Holm, John A. (1989), *Pidgins and Creoles*, vol. 2, Cambridge, Cambridge University Press.
- Jayasuriya, Shihan S./Pankhurst, Richard (edd.) (2003), *The African Diaspora in the Indian Ocean*, New Jersey, Africa World Press.
- Jin Guo Ping (1987), *Alguns dados sobre léxico chinês de origem portuguesa e lexicografia sino-portuguesa e vice-versa*, in: *Congresso sobre a situação actual da língua portuguesa no mundo – Actas*, vol. 2, Lisboa, Instituto de Cultura e Língua Portuguesa, 361–379.
- Jones, Russell (gen. ed.)/Grijns, C.D./de Vries, Jan W. (edd.) (2007), *Loan-words in Indonesian and Malay*, Leiden, KITLV.
- Kim, Tai Whan (1976), *The Portuguese element in Japanese: a critical survey with glossary*, Coimbra, Faculdade de Letras da Universidade de Coimbra/Instituto de Estudos Românicos.
- Knowlton Jr., Edgar C. (1960), *Portuguese in Hawaii*, Kentucky Foreign Language Quarterly Review 7, 212–218.
- Knowlton Jr., Edgar C. (1967), *Pidgin English and Portuguese*, in: Frederick S. Drake (ed.), *Proceedings of the Symposium on historical, archaeological and linguistic studies on Southern China, Southeast Asia and the Hong Kong Regions*, Hong Kong, University of Hong Kong Press, 228–237.
- Leite de Vasconcelos, José (1901), *Esquisse d'une dialectologie portugaise*, Paris, Aillaud & Cie.
- Li, Michelle/Matthews, Stephen (2016), *An outline of Macau Pidgin Portuguese*, Journal of Pidgin and Creole Languages 31:1, 141–183.
- Lima, Edward de (2014), *Influence of Portuguese vocabulary on Konkani language*, Porvorim, Vikram Publications.
- Lopes, David (1907), *Cartas de Raja Singa rei de Candia aos Holandezes*, Boletim da Sociedade de Geografia de Lisboa, separata.
- Lopes, David (1936), *A Expansão da Língua Portuguesa no Oriente nos Séculos XVI, XVII e XVIII*, Barcelos, Portucalense Editora.
- Loureiro, Rui Manuel (1992), *Ásia*, in: António Luís Ferronha (ed.), *Atlas da língua portuguesa na história e no mundo*, Lisboa, Imprensa Nacional-Casa da Moeda, 91–119.

- Matan Ruak, Taur (2001), *A importância da língua portuguesa na resistência contra a ocupação indonésia*, *Camões – Revista de Letras e Culturas Lusófonas* 14, 40–41.
- Matthews, Stephen/Li, Michelle (2012), *Portuguese pidgin and Chinese Pidgin English in the Canton trade*, in: Hugo C. Cardoso/Alan N. Baxter/Mário Pinharanda Nunes (edd.), *Ibero-Asian Creoles: Comparative perspectives*, Amsterdam/Philadelphia, Benjamins, 263–287.
- Maurer, Philippe (2011), *The former Portuguese Creole of Batavia and Tugu (Indonesia)*, London, Battlebridge.
- Miranda, Rocky V. (2001), *Portuguese influence on Konkani syntax*, in: Anvita Abbi/R.S. Gupta/Ayesha Kidwai (edd.), *Linguistic Structure and Language Dynamics in South Asia*, Delhi, Motilal Banarsidass, 48–61.
- Nell, Louis (1888–1889a), *An explanatory list of Portuguese words adopted by the Sinhalese*, *The Orientalist* 3, 41–56.
- Nell, Louis (1888–1889b), *The archaeology of Ceylon Eurasian gastronomy*, *The Orientalist* 3, 160–167.
- Noble, Charles F. (1762), *A Voyage to the East Indies in 1747 and 1748*, London, Becket/Dehondt.
- Noronha, Manuel António/Chaplin, Ian (2012), *A Survey of Portuguese Language Allegiance among the Macanese and the Macanese Diaspora*, *Portuguese Studies Review* 20:1, 113–135.
- Orta, Garcia de (1563), *Colóquios dos simples e drogas he cousas medicinais da Índia e assi dalgũas frutas achadas nella onde se tratam algũas cousas tocantes a medicina, pratica, e outras cousas boas pera saber*, Goa, João de Endem.
- Paiva Boléo, Manuel de (1974), *O estudo das relações mútuas do português e do espanhol na Europa e na América, e influência destas línguas em territórios da África e da Ásia*, in: *Estudos de Linguística Portuguesa e Românica*, vol. I [Dialectologia e História da Língua], Coimbra, Biblioteca Geral, 355–398.
- Perera, Simon Gregory (1922), *Portuguese influence on Sinhalese speech*, *Ceylon Antiquary and Literary Register* 8, 45–60.
- Pinto, Jeanette (1992), *Slavery in Portuguese India (1510–1842)*, Bombay, Himalaya Publishing House.
- Pissarra, José Virgílio A. (2002), *Chaul e Diu: 1508 e 1509*, Lisboa, Tribuna.
- Quadros, Jeronymo (1899), *Diu, apontamentos para sua historia e chorographia*, Nova Goa, Tipographia Fontainhas.
- Reinecke, John (1969), *Language and dialect in Hawaii: A sociolinguistic history to 1935*, Honolulu, University of Hawaii Press.
- Reinecke, John/Tokimasa, Aiko (1934), *The English dialect of Hawaii*, *American Speech* 9, 48–58; 122–131.
- Sardessai, Manohar L. (1983), *The Portuguese influence on Konkani*, *Journal of South Asian Literature* 18:1, 155–158.
- Schuchardt, Hugo (1891), *Kreolische Studien IX. Über das Malaioportugiesische von Batavia und Tugu*, *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften in Wien, Philosophisch-Historische Klasse* 122, 1–256.
- Siegel, Jeff (2000), *Substrate influence in Hawai'i Creole English*, *Language in Society* 29, 197–236.
- Siegel, Jeff (2008), *The emergence of pidgin and creole languages*, Oxford, Oxford University Press.
- Smith, Stefan Halikowski (2011), *Creolization and diaspora in the Portuguese Indies: The social world of Ayutthaya, 1640–1720*, Leiden, Brill.
- Subrahmanyam, Sanjay (1993), *The Portuguese empire in Asia, 1500–1700: A political and economic history*, New York/London, Longman.
- Taylor-Leech, Kerry (2009), *The language situation in Timor-Leste*, *Current Issues in Language Planning* 10:1, 1–68.

- Thomaz, Luís Filipe F.R. (1990), *A língua portuguesa em Timor*, in: *Actas do Congresso sobre a situação actual do português no mundo*, vol. 1, Lisboa, Instituto de Cultura e Língua Portuguesa, 313–339.
- Thomaz, Luís Filipe F. R. (2015), *De l'autre côté du monde: langues véhiculaires et communication interethnique dans l'Océan Indien à l'époque de la découverte portugaise*, *Mélanges de la Casa de Velázquez. Nouvelle Série* 45:1, 113–131.
- Tomás, Maria Isabel (1992), *Os Crioulos Portugueses do Oriente: Uma Bibliografia*, Macau, Instituto Cultural de Macau.
- Tomás, Maria Isabel (2008), *A viagem das palavras*, in: Mário Ferreira Lages/Artur Teodoro de Matos (edd.), *Portugal, Percursos de Interculturalidade*, vol. 3, Lisboa, ACIDI, 431–485.
- Vaquinhas, José dos Santos (1885) *Timor*, *Boletim da Sociedade de Geographia de Lisboa* 5.ª série, vol. 1, 61–64.
- Wherritt, Irene (1985), *Portuguese language use in Goa, India*, *Anthropological Linguistics* 27:4, 437–451.
- Wherritt, Irene (1989), *Portuguese language shift: About Town in Goa, India*, *Hispania* 72:2, 385–391.
- Williams-van Klinken, Catharina/Hajek, John/Nordlinger, Rachel (2002), *Tetun Dili: A grammar of an East Timorese language*, Canberra, Pacific Linguistics.
- Yule, Henry/Burnell, Arthur C. (1886), *Hobson-Jobson: Being a glossary of Anglo-Indian words and phrases, and of kindred terms; etymological, historical, geographical, and discursive*, London, Murray.

Celeste Rodrigues

4 Variação sociolinguística

Abstract: Múltiplos fatores condicionam a distribuição das variantes de cada fenómeno variável em português europeu: a geografia do território, a estrutura social (definida com base em critérios como a idade, o género, a escolaridade, por exemplo), o estilo de fala, entre outros. Neste capítulo são revisitados alguns dos mais importantes resultados de trabalhos acerca de variáveis fonológicas em português falado no Continente. O texto centra-se nos fenómenos sujeitos a variação relacionada com aspetos sociais da comunidade. Os tópicos estão agrupados em função de grandes categorias de segmentos fonológicos e dos seus contextos de ocorrência. Os processos fonológicos discutidos incluem variação geolinguística, variação em variedades linguísticas específicas (incluindo o dialeto *standard*), variação interna ao falante e entre falantes e variação com potencial impacto na mudança da língua. Reconhece-se contudo que o português europeu carece ainda de muitos trabalhos na área sociolinguística.

Keywords: processos fonológicos variáveis, fatores externos de variação linguística, dialetos, português europeu *standard*

1 Introdução

A variação linguística, um foco especial de atenção dos linguistas contemporâneos, é visível em qualquer língua em função de diversos fatores de importância reconhecida. Para além da geografia do território, a estrutura social, diferindo em função de critérios como a idade, o género, a escolaridade, por exemplo, ou ainda o estilo de fala e a situação pragmática de uso condicionam a frequência das variantes dos fenómenos variáveis. A diferenciação entre as diversas abordagens possíveis é mais uma consequência da perspectiva adotada por cada linguista do que uma diferença do próprio objeto de estudo.

Neste capítulo surgem reunidos resultados de um conjunto de trabalhos que incidem em fenómenos fonológicos e de variação fonética do português europeu falado no Continente, que habitualmente não fazem parte das descrições dialetológicas, sem a pretensão de enquadrar esses trabalhos em áreas específicas de atuação. Esta opção decorre de o português europeu não ter sido objeto de grande número de trabalhos sociolinguísticos de inspiração laboviana (ou outra) que descrevam fenómenos fonológicos, contrariamente, ao português falado no Brasil. Esse facto, porém, não tem relação com a variação existente em português europeu, que, como qualquer outra língua, apresenta um quadro de variação com potencial impacto em fenómenos de mudança linguística. Em contrapartida, os estudos dialetais disfrutaram de uma larga

tradição em Portugal, centrando-se para a definição dos principais contrastes geolinguísticos em traços fonológicos e fonéticos. A razão pela qual o presente capítulo não incide preferencialmente na apresentação das conclusões desses trabalhos é a de estes se reportarem a dados de fala de um pequeno número de falantes por localidade (por vezes, apenas um, embora identificado como representativo, segundo os critérios dialetológicos) e, nessa medida, não nos poderem dar uma imagem da variação sociolinguística existente em cada localidade. Os seus resultados globais estão amplamente divulgados em diversos estudos (nomeadamente, Segura 2013 e suas referências).

Assim, ao longo do capítulo, são apresentados os principais processos fonológicos sujeitos a variação relacionável com aspetos sociais na comunidade portuguesa já estudados. Não se pretende com isso dispensar a leitura dos respetivos trabalhos nos quais se fundamenta, mas contribuir para a reflexão acerca das características globais da variação fonológica e fonética em português. Alguns dos fenómenos surgem no âmbito de estudos que não são entendidos verdadeiramente como sociolinguísticos no sentido laboviano do termo (Weinreich/Labov/Herzog 1968; Labov 1972; 1994; 2001), mas que, por analisarem materiais linguísticos de apenas determinados grupos sociais, contribuem para o conhecimento dos fatores externos que condicionam a variação linguística. Dessa forma, o capítulo enquadra os diferentes trabalhos na área da sociolinguística e introduz cada tópico, centrado em aspetos fonológicos ou fonéticos, de modo a que se possa compreender o seu impacto na variação sincrónica e nas tendências de mudança da língua.

Os tópicos estão agrupados em função de grandes categorias de segmentos (vogais e consoantes) e dos seus contextos de ocorrência (posição relativa nas palavras ou nas sílabas). Assim, as vogais acentuadas serão o alvo da segunda secção, as vogais átonas da terceira e as consoantes em ataque silábico da quarta. Seguem-se algumas reflexões finais.

2 Vogais acentuadas

O português europeu *standard*¹ possui um inventário segmental de sete vogais fonológicas (/i, e, ε, a, ɔ, o u/), opondo as três primeiras unidades produzidas na área anterior à vogal /a/ e a três vogais labiais (/ɔ, o, u/) em posição acentuada (cf. Mateus/Andrade 2000; Mateus et al. 2003; cf., no entanto, ↗24 O sistema vocálico e a redução e neutralização das vogais átonas).²

1 O português europeu *standard* é frequentemente entendido como o português falado pelas classes cultas nativas da região de Coimbra e Lisboa. Esse dialeto possui muitas semelhanças com os dialetos do Centro e Sul de Portugal continental, a que pertence, embora se diferencie deles por algumas características, como se mostrará mais adiante.

2 Alguns autores consideram ser necessário incluir no sistema fonológico do português europeu *standard* a vogal /e/ (Barbosa 1994; Freitas 1997; Veloso 2012). Sendo possível explicar a ocorrência da

Considerando as vogais /e/ e /ɛ/ do sistema do português europeu *standard*, verifica-se que os dialetos apresentam variação em posição acentuada pré-palatal (ou seja antes de /j/, /z/ e, em alguns casos, também de /ɲ/ e /ʎ/). Os dialetos do norte e o português europeu *standard* têm maioritariamente a realização centralizada [e] ou [ɛ] e os do sul não a apresentam, em termos gerais. A variação sociolinguística das vogais acentuadas não está descrita, a não ser no caso destas duas vogais (e no da vogal [u] palatalizada, adiante referido), em trabalhos com dados socialmente estratificados. Barros (1994), o primeiro trabalho a estudar sistematicamente a importância dos processos que atingem estas vogais na fala de Lisboa neste contexto, analisa dados de 24 falantes de ambos os sexos, de duas faixas etárias (rondando os 25 ou os 50 anos), com ensino básico ou com licenciatura. Rodrigues (2003, 128–135), por sua vez, apresenta novos dados de 180 informantes distribuídos por sexo, quatro graus de escolarização e cinco faixas etárias, pertencentes às regiões de Lisboa e de Braga (no norte ocidental), em diferentes tarefas discursivas (incluindo discurso informal, o único registo a que nos referiremos aqui). Com base nestas duas amostras de dados, é possível perceber que:

- (i) Antes de uma consoante fricativa heterossilábica, pode haver centralização de /e/ (e.g. *cereja* [si'reʒɐ]) e de /ɛ/ (e.g. *inveja* [ĩ'veʒɐ]~[ĩ'vezɐ]), inserção de uma glide ([si'rejɐ], [ĩ'vejɐ]) e, em falantes femininas licenciadas de Lisboa, até realização de [a] – [si'raʒɐ], mas não *[ĩ'vaʒɐ] nem *[ĩ'vajɐ]. Esta última variante é pouco produtiva nos dados de Barros (1994) e quase inexistente nos dados de Lisboa de Rodrigues (2003), apesar de em Barros apresentar uma distribuição social que a poderia classificar como variante inovadora de prestígio. A centralização de /ɛ/ é esporádica, enquanto que a de /e/ é muito frequente (mas não categórica). A inserção de glide antes de consoante heterossilábica quase duplicou a sua percentagem de ocorrência entre as duas amostras. Na verdade, era de 17,7% antes de fricativa heterossilábica nos dados de Barros e é de 32,3% nos de Lisboa de Rodrigues (e chega mesmo aos 51,2% em Braga, cidade na qual a inserção também acontece antes de consoante palatal sonante).

vogal [e] sistematicamente, tanto quando é acentuada como quando é átona, consideramos dispensável a inclusão de [ɛ] no inventário fonológico. Recorde-se que vogal [ɛ] constitui o alofone de /a/ átono e, para além desse contexto, só surge em posição acentuada em dois outros: i) antes do morfema de pessoa iniciado por nasal nas formas verbais de 2.^a pessoa do plural dos verbos da 1.^a conjugação no presente do indicativo *amamos* [e'memɐ], onde o seu surgimento depende da nasalização e ii) em consequência da centralização de vogais não recuadas *coelho* ['kwẽlu] do português europeu *standard*. No caso das formas verbais do pretérito nos dialetos setentrionais, a forma verbal foi neutralizada com a respetiva forma do presente, devido ao desaparecimento total do morfema de tempo (-j- segundo Mateus 1982, ¹1975 ou -vi- segundo Andrade 1977), contrariamente ao que sucede no português europeu *standard* e nos dialetos centro-meridionais. Nestes últimos o contraste entre os dois tempos manifesta-se na diferença de nasalização da vogal acentuada e na altura da vogal na 1.^a conjugação e só na da nasalização nas outras duas – já que só as formas do presente podem receber nasalização por antecederem imediatamente o morfema de pessoa começado por nasal -mos.

- (ii) Nos dados de Barros (1994) a inserção de glide também se dá em sílabas com coda fricativa, mas só, em palavras como *sexta* ['sejʃte], ou *texto* ['tejʃtu]. Rodrigues (2003) reforça essa ideia, restringindo a sua aplicação em sílabas com coda fricativa exclusivamente a palavras com /e/ no núcleo seguido de <x> (independentemente do acento), uma vez que as palavras com <s> em coda não permitem a inserção de [j] em português europeu (*cesto* ['seʃtu], *testo* ['teʃtu]).³
- (iii) Paradoxalmente, nos núcleos complexos /eI/,⁴ Barros (1994) reporta a existência, para além de [ej], de formas centralizadas monotongadas antes de não-sonante *beijo* ['beʒu]. A frequência destas formas neste contexto é inferior à do contexto precedente a um /r/ *cadeira* [ke'dere] (contexto não palatal onde mais ocorre). Isso parece indicar que a monotongação está a começar a atingir as vogais em palavras com fricativas heterossilábicas. Todavia, as formas com a glide são consideradas variantes de prestígio, em virtude de surgirem com mais frequência na fala dos informantes mais instruídos da amostra de Barros (1994, 180–184). Isso mesmo foi confirmado por Rodrigues (2003) não apenas nos dados de Lisboa da sua amostra mais ampla, mas também nos de Braga. As percentagens de centralização de /eI/ nos seus dados rondam os 90% para esta estrutura nas duas cidades, sendo as falantes femininas com apenas até ao 9.º ano de escolaridade na faixa etária entre os 26 e os 39 anos as que mais instabilidade apresentam.
- (iv) Nos núcleos /eI/ da amostra de Barros a centralização é muito frequente (80,2%), embora não seja categórica, uma vez que palavras como *ideia* ou *fiéis* podem apresentar [ej] ou [ej]. Nos dados de Lisboa de Rodrigues (2003), a percentagem de ocorrência de centralização é mais baixa, 60% apenas (em oposição evidente à centralização do núcleo /eI/).

Deste modo, em suma, a fala de Lisboa e de Braga assemelham-se no que diz respeito a estas quatro estruturas onde é possível encontrar os processos de centralização, de inserção de glide e de monotongação. No entanto, distinguem-se por só em Braga haver inserção da glide antes de consoante sonante heterossilábica e de só em Lisboa haver o abaixamento da vogal central. A distribuição social das variantes fonéticas não é totalmente dedutível dos dados de Rodrigues, na medida em que há estruturas com poucos itens lexicais frequentes em fala espontânea.

O sistema vocálico do português europeu *standard*, todavia, não é totalmente partilhado por todas as regiões do território linguístico português. Na perspetiva de diversos dialetólogos, devem ser considerados diferentes inventários específicos de algumas localidades ou regiões (Segura 2013; Brissos 2014a; 2014b; 2015; Brissos/

3 Não existem palavras com /ε/ seguido de <x> em coda em português.

4 Segue-se aqui a notação proposta por Mateus/Andrade (2000, 54ss) para as semivogais dos ditongos do português. Estas são representadas por /I/ e por /U/ na representação de base, respetivamente para a semivogal palatal e arredondada, e têm a singularidade de nunca poderem receber acento nem poderem constituir núcleo de sílaba sozinhas.

/Rodrigues 2016). Embora o sistema fonológico do português europeu *standard*, acima referido, se adegue à descrição fonológica da maior parte do território, esses autores identificaram inventários vocálicos específicos para algumas regiões. No Alto Alentejo Interior e na Beira Baixa, existe o sistema /i e ε ə a ɔ o θ u/⁵ (cf. Brissos 2012; 2014a, para uma discussão acerca dessa variedade); o Barlavento algarvio tem o sistema: /i e ε æ ɔ ɔ o y/ (Segura 2013); o litoral Oeste (Minho e Douro Litoral) também diverge em alguns pontos, por ter neutralizado as vogais anteriores /e/ e /ε/ e apresentar para diversas vogais preferencialmente formas ditongadas (Brissos/Rodrigues 2016). De uma forma geral, os inventários referidos incluem mais unidades fonológicas do que o português europeu *standard*, embora partilhem com ele algumas das unidades e, em alguns casos, haja vogais que neles são referidas mas que podem ser explicadas como manifestação fonética apenas (e não como unidades fonológicas). Estes inventários específicos foram detetados por via de análises perceptivas ou acústicas de fala de indivíduos de faixas etárias altas.

Como já referido, um outro trabalho de natureza sociolinguística, o de Santos (2003), analisa numa comunidade do baixo Mondego (Vila Pouca do Campo, a sudoeste de Coimbra) a palatalização de /u/ ([u]). O trabalho destaca-se dos anteriores por incidir numa comunidade rural e por explorar uma metodologia que neles não foi utilizada – o estabelecimento de redes sociais. A palatalização da vogal /u/ encontra-se dispersa por várias zonas do país,⁶ mas não havia sido registada nesta região. Santos baseia-se em dados de 22 falantes dos dois sexos, três faixas etárias, três graus de escolaridade e quatro tipos de ocupação. Na sua proposta de análise, a palatalização consiste num avanço considerável (mas de grau variável) do dorso da língua na produção de /u/, pelo que Santos defende não ser necessário conceber [u] como unidade fonológica em Vila Pouca do Campo. Embora ocorra mais regularmente em sílaba acentuada, a palatalização também pode surgir em sílabas não-acentuadas.⁷

Os dados globais indicam que a palatalização (cf. Santos 2003, 193–225), apesar de estar presente em muitos dos informantes paralelamente a [u], é mais frequente nos das faixas etárias mais altas (83% dos mais jovens – entre os 14 e os 25 anos – praticamente não produz o alofone palatalizado), porém não pode ser totalmente explicada apenas pela interferência desse fator social. As mulheres palatalizam mais

5 Usamos o símbolo /u/ como Brissos para representar a vogal palatalizada que se pode encontrar em várias áreas do território português, porque a vogal, segundo os seus últimos dados acústicos existentes, é uma vogal central. No Barlavento algarvio, no entanto, a sua natureza é mais anterior do que nas outras regiões, razão que pode justificar a sua representação com o símbolo /y/ em Segura da Cruz (1988) e que por isso aqui mantivemos. Dispensamos o símbolo [ü], notação mais conservadora, inexistente no Alfabeto Fonético Internacional (versão de 2005), ainda usada, por exemplo, em Santos (2003).

6 Como diversos trabalhos dialetais haviam já referido (cf. referências em Segura da Cruz 1988) e trabalhos recentes têm corroborado – cf. referências em Brissos (2012; 2014a; 2014b).

7 Pontualmente, a palatalização também pode encontrar-se em palavras com /o/, mesmo em sílaba pré-tónica, como em *folhelho*.

do que os homens (29% vs. 9%), embora isso se correlacione mais com o tipo de ocupação de cada falante e com o respetivo nível de instrução. As mulheres menos escolarizadas (e que, maioritariamente, desempenham só atividades domésticas) apresentam valores mais altos de palatalização do que os homens com o mesmo perfil (86% vs. 57%). Essa diferença não existe, em contrapartida, entre os falantes dos dois sexos, se a sua escolaridade for restringida ou ampla⁸. O tipo de atividade a que se dedicam os informantes correlaciona-se ainda com os valores atingidos pela palatalização, na medida em que os estudantes e, em menor grau, os trabalhadores do setor terciário, apresentam valores mais baixos do que os falantes dos setores primário e secundário. De facto, a autora conclui que só o cruzamento de todos os fatores analisados com as relações que eles estabelecem na rede social local em que participam pode explicar a heterogeneidade dos dados observados em alguns grupos. Se os falantes forem agrupados em função da força do seu relacionamento com a rede local e a vitalidade do fenómeno, verifica-se que quanto mais rural se apresenta a rede pessoal de cada falante, mais expressiva se torna a palatalização (os valores variam entre 6,57% e 64,34% – Santos 2003, 218).

Em resumo, no que respeita à palatalização em Vila Pouca do Campo, Santos (2003) pode comprovar a sua existência expressiva em diversos perfis de falante, em particular nos que apresentam uma teia de relações sociais mais fortes com a comunidade rural local. Esses falantes pertencem preferencialmente ao setor primário, são de faixas etárias elevadas e de grau de escolaridade nulo/mínimo ou restringido. Uma vez que se regista variação interna ao falante e não é possível estabelecer uma relação nítida com o contexto fonético de ocorrência do alofone palatalizado, a autora conclui que a palatalização pode ser descrita com base no segmento fonológico do português europeu *standard* /u/ sincronicamente (embora seja um fenómeno antigo na língua), contrariamente à posição assumida por Segura da Cruz (1988) para a descrição do Barlavento Algarvio.⁹

3 Vogais átonas

Como é sabido, em português europeu o sistema fonológico é simplificado em posição átona em todo o território. No entanto, nem sempre exatamente do mesmo modo. As vogais altas /i/ e /u/ não são alteradas, a não ser por semivocalização junto de outras vogais, e as vogais médias e abertas iniciais podem não ser neutralizadas. A vogal /a/

⁸ A autora utiliza uma distinção de três níveis relacionados com a instrução: 1 – nula/mínima; 2 – restringida; 3 – ampla.

⁹ Esta opção é sustentada pelo facto de no Barlavento a forma palatalizada ser a mais frequente e de constituir o traço característico mais evidente do dialeto (Segura da Cruz 1988, 257), além de acusticamente ser mais anterior do que noutras regiões do país.

âtona é regularmente produzida como [e] e, diferentemente das outras vogais sujeitas a redução, tende a ser sempre preservada.

Geralmente há neutralização de /e/ e /ɛ/ sob a forma de [i] (/e/-[e] *se*lo vs. /e/-[i] *se*lado, /ɛ/-[ɛ] *be*lo, /ɛ/-[i] *be*leza, embora nunca em posição inicial de palavra *[i] *e*levador), neutralização de /o/ e /ɔ/ sob a forma de [u] (/o/-[o] *est*opa vs. /o/-[u] *est*opinha, /ɔ/-[ɔ] *so*rte vs. /ɔ/-[u] *so*rtudo) e elevação de /a/-[e] (/a/-[a] *ca*sa /a/-[e] *ca*sinha) em todas as sílabas átonas da palavra. Apesar disso, estes processos têm exceções, por exemplo, em sílabas com coda sonante /l/ ou /r/ (*so*ldado, *o*lvidar, *tú*nel, *ca*ráter, *he*rbívoro, *se*nior, *al*guidar), em algumas palavras iniciadas por /a/ (*a*tor, *a*ção), em palavras com sufixo derivado do sufixo latino -IONE (*o*bjecção, *pe*rceção, *se*leção) e em verbos como (*in*jetar, *in*fetar). Além disso, há limitação de aplicação da neutralização em posição inicial de palavra no que diz respeito a algumas vogais em determinados dialetos. É o que acontece no português europeu *standard*, onde a neutralização das vogais arredondadas não se aplica em palavras como *o*casão, *o*relha, *o*lhinhos, *o*leado, *o*stentar, *o*rquidário (ou seja não se pronunciam com [u] inicial, como descreveremos adiante a partir do trabalho de Mascarenhas 1996). Situação contrária regista-se em alguns falantes de dialetos setentrionais, por exemplo, de Braga (Rodrigues 2003). Aí os falantes do sexo masculino menos instruídos e das faixas etárias mais elevadas apresentam a forma neutralizada [u], mostrando que a elevação das átonas pode ser menos restritiva do que no português europeu *standard*. Como se trata de falantes com esse perfil sociocultural e essas formas não surgem com grande frequência nas falantes femininas da mesma cidade, essa variante foi considerada em regressão naquele dialeto. Ora no português europeu *standard*, como dissemos, não há produções com elevação das arredondadas em posição inicial de palavra (i.e. [u]), mas Mascarenhas (1996) já salientou que as formas [o] e [ɔ] se podem encontrar em palavras com qualquer uma dessas vogais fonológicas. O seu trabalho comparou falantes de Lisboa e do Porto no que respeita às vogais arredondadas (para além das não recuadas, de que falaremos a seguir), com base em 6 informantes por cidade (3 de cada sexo), com idades entre os 20 e os 24 anos, estudantes de licenciatura nativos de cada uma das cidades, da classe média-alta.¹⁰ Os informantes foram gravados em tarefas de leitura (palavras isoladas e em frase de encaixe), a partir das quais é possível dizer que nos dados de Lisboa há diversas palavras com /o/ inicial que foram lidas ora com [o] ora com [ɔ]. Os falantes do sexo masculino produziram mais [ɔ] do que as falantes do sexo oposto nestas palavras. A forma preferida pelas mulheres é [o]. Entre os homens há um que apresenta quase só [o], ao passo que os restantes utilizam maioritariamente a vogal aberta. Nos dados do Porto as mulheres têm uma taxa de realização da vogal aberta superior à dos homens, que

¹⁰ A classe foi definida em função de um conjunto de circunstâncias, nomeadamente a de os pais dos informantes terem de ser licenciados (quadros superiores de empresas, médicos ou professores, por exemplo).

preferem [o] mas também apresentam [u]. Estes dados sugerem que o processo de abertura seja uma inovação que está a receber um valor desigual na avaliação das falantes femininas nas duas cidades. Se em Lisboa elas não usam a forma inovadora com frequência, contrariamente aos falantes masculinos, é porque não lhe estão a reconhecer prestígio particular. Já na cidade do Porto são justamente as falantes femininas que levam a dianteira nas variantes abertas da vogal, mostrando a habitual tendência de mais rápido acompanhamento dos traços inovadores desse sexo. Rodrigues (2003, 161–167), no que respeita às vogais arredondadas iniciais (para além da pronúncia [u] já acima referida na cidade de Braga), encontrou em Lisboa a pronúncia com [ɔ] mais frequentemente nos falantes das faixas etárias mais baixas do que nos das mais altas e ausência de queda, contrariamente ao que sucede com a vogal /o/ nas restantes posições da palavra. Em Braga, em contrapartida, coexistem os três alofones [ɔ], [o], [u], não se verificando um claro avanço da forma aberta da vogal sobre a média, nas faixas etárias mais baixas. A distribuição das variantes em Braga indica que os falantes mais conservadores preservam [u], e os restantes oscilam entre as duas vogais não altas com preferência das mulheres e dos licenciados dos dois sexos por [o] – certamente por a interpretarem como forma de prestígio, sendo os falantes mais jovens quem mais prefere a vogal aberta.

As vogais não recuadas iniciais estão também em variação em posição inicial nos dialetos portugueses. Mascarenhas (1996) conclui que há elevação das médias (para [i]) em diferentes contextos em Lisboa e no Porto, sendo este o alofone mais frequente. No entanto, parece existir uma preferência por [e] em sílaba com coda /r/ em Lisboa, que não se verifica no Porto (onde as três formas ocorrem em competição). Rodrigues (2003, 204–214), no que se relaciona com as vogais não recuadas iniciais, conclui que em Lisboa há uma clara preferência por [i] em rima não ramificada (92,4%, e.g. em *elétrico*, *emoção*), tal como em Braga (92,2%). A preferência por [i] diminui nos restantes tipos de sílaba nas duas cidades (e.g. *ermida*, *Elvira*), não havendo em nenhum dos casos uma relação entre o perfil dos falantes e a variante ou variantes selecionadas. Desse modo, a forma [e] das sílabas iniciais, registada por Mascarenhas nas sílabas com coda /r/, parece ter perdido terreno face a [i] em Lisboa. As sílabas com coda fricativa (e.g. *expor*, *extraordinário*) admitem, como já foi dito, a realização ditongada e dissimilada [ej], nas cidades de Lisboa e Braga, contrariamente às palavras com núcleos vazios iniciais: *especial*, *estante*, *esgrima*. Nestas últimas, não há especificação do núcleo e por tanto não há geralmente o seu preenchimento, de onde resulta a pronúncia sem vogal inicial. Só em alguns falantes de alguns dialetos setentrionais e no português do Brasil a realização [i] se pode encontrar, resultando nesses casos de um processo de preenchimento do núcleo silábico.

Em conclusão, em posição inicial átona, as vogais altas /i/ e /u/ não variam (apenas admitem semivocalização, se estiverem junto a outras vogais), as vogais médias e baixas estão sujeitas a variação nos diferentes dialetos, havendo fundamentalmente uma preferência por [i] para as não-recuadas e de [ɔ] ou [o], dependendo da idade dos falantes e do prestígio associado à variante [o] em cada dialeto (exceto, se

se tratar de falantes mais conservadores de dialetos setentrionais, como o de Braga, por exemplo). Conclui-se também que as vogais não estão sujeitas a queda em posição inicial (incluindo os casos de palavras como *estante*, *esgrima* por terem núcleos vazios), contrariamente ao que se observa no interior e no fim da palavra em quase todos os dialetos. Conclui-se ainda que em posição inicial de palavra atuam processos diferentes das outras posições da palavra.

Para além da elevação das vogais átonas, o vocalismo átono está sujeito a outros fenómenos de redução, no limite dos quais se encontra a queda de muitas dessas vogais na fala espontânea (Segura da Cruz 1988, 369, por exemplo, refere a apócope de [u] e [i] finais no Barlavento do Algarve). Rodrigues (2003) descreve os dados de duas amostras socialmente estratificadas das cidades de Lisboa e de Braga de todas as vogais átonas não nasalizadas, dos quais se pode concluir que as vogais /i/, /u/ e /a/ átonas não são elididas tão frequentemente como as restantes vogais no interior das palavras. Estas últimas são frequentemente elididas em ambos os dialetos analisados em diferentes estruturas silábicas, de acordo com fatores externos de variação.

- (i) As vogais labiais /o/ e /ɔ/ (neutralizadas sob a forma [u] em *costume*, *fotografia*, etc.) são elididas independentemente da sua origem fonológica, ainda que esta determine a frequência de ocorrência da elisão: Lisboa 14,9% vs. Braga 12,2%.
- (ii) A elisão é parcialmente condicionada pela estrutura silábica nas duas cidades: é mais frequente em sílabas CV (20% em Lisboa vs. 14% em Braga, e.g. *poder*, *boneca*) do que em sílabas CCV ou CVC (abaixo dos 5% na duas cidades, e.g. *prometer*, *comportar*).
- (iii) A elisão está relacionada com o perfil dos falantes. Em Lisboa, os jovens com menos instrução elidem mais do que os jovens mais instruídos (35% vs. 21%); os homens elidem mais do que as mulheres (35% vs. 26%) e os jovens elidem mais do que os menos jovens (36,9% vs. 21%). Em Braga, a elisão relaciona-se com a idade e o nível de instrução, mas não com o género (foram encontrados valores mais altos em falantes jovens com instrução mais elevada – 18,3% vs. 9,3%).
- (iv) Analisando separadamente as sílabas CVC (a maior parte das quais é /CoS/ – *amigos*, *vistos*),¹¹ verifica-se que a vogal é elidida mais frequentemente antes da fricativa vozeada [ʒ] (o que sucede em muitos casos de sândi externo) do que de [ʃ] (71,3 vs. 54,4% em Lisboa e 53,7 vs. 39,7% em Braga); a frequência da elisão é também nesta estrutura silábica mais alta em Lisboa do que em Braga (62,1% vs. 46,1%).

Logo, uma vez que as mulheres do dialeto *standard* elidem menos do que os homens, a taxa de elisão está relacionada com o prestígio (quanto menos elisão, mais prestígio). Não havendo uma clara estigmatização da elisão, ela está a difundir-se rapida-

¹¹ /S/ representa a fricativa em coda. Assume-se aqui uma raiz não especificada quanto a vozeamento e ao ponto de articulação.

mente entre os falantes jovens de Braga, contrariamente ao que sucede entre os mais velhos, sem grandes diferenças de comportamento entre os falantes dos dois géneros.

Em posição final da palavra, /o/-[u] é elidido muitas vezes (*tempo*, *comportamento*): 53,3% em Lisboa e 42,3% em Braga (Rodrigues 2003, 138–153). A variação está dependente do contexto seguinte: é comum antes de uma pausa (na maior parte dos casos uma fronteira de sintagma entoacional, como em *O mau comportamento, doutor*) – (36,8% em Lisboa e 26% in Braga); é muito comum antes de consoantes (65,1% em Lisboa e 55% em Braga, como em *tempo certo*) e ainda mais antes de vogais (81,2% em Lisboa e 81,1% em Braga, como em *campo arranjado, comportamento hostil*). Fatores como a idade interferem na frequência de elisão, já que os mais jovens apresentam mais elisão do que os menos jovens, apesar de em Lisboa esta ser muito frequente em todos os falantes. Em Braga a elisão só é relativamente pouco frequente entre os mais velhos (15% apenas).

No que se refere à vogal átona [i] (alofone de /i/, /e/ ou /ɛ/) em posição medial e final, a probabilidade de elisão está ligada:

- (i) à estrutura da sílaba: afeta as vogais em sílabas CV mais frequentemente do que nas restantes – atingindo em Lisboa em sílabas CV mediais 51,2% e em Braga 43,6%, *embelezar, telefonar* (cf. Rodrigues 2003, 214–224).
- (ii) Para além disso, a frequência da elisão também depende da qualidade fonológica da vogal, sendo aquelas que representam /e/ as que mais são elididas.
- (iii) A frequência da elisão em Lisboa também tem uma correlação com os parâmetros idade e género. São os mais velhos aqueles que menos elidem a vogal, apresentando elisão em menos de 40% dos casos. Nos falantes mais jovens há sempre taxas superiores (em particular, entre os do sexo masculino – 62,5% vs. 55,7% nos do sexo feminino). Em Braga, as percentagens mais altas de elisão registam-se nos falantes da faixa etária intermédia (26–39 anos), sobretudo nos do sexo masculino com menos instrução: 69,2% (até ao 9.º ano) vs. 55,5% (licenciados). As falantes femininas apresentam mais elisão se forem licenciadas (58,7%) do que se tiverem escolaridade só até ao 9.º ano (41,6%), o que mostra que elas são mais sensíveis a um traço muito frequente na variedade *standard*.

Portanto, para além da estrutura da sílaba, o que mais influencia os resultados da elisão é a idade e o género, nos dados já analisados de Lisboa e de Braga.

Um outro fenómeno, para o qual trabalhos recentes voltaram a chamar a atenção, diz respeito à inserção de glide palatal em sândi (cf. Segura 2013; Oliveira et al. submetido). Segundo Segura (2013), que traça a fronteira para a ocorrência do fenómeno com base em materiais do *Atlas Linguístico e Etnográfico de Portugal e da Galiza* (ALEPG), a sua ocorrência regista-se em toda a área dos dialetos setentrionais estendendo-se a norte de uma linha oblíqua que divide o país passando perto de Leiria, num dos extremos, e de Portalegre, no outro. Nos dialetos do sul, em princípio, a inserção não ocorre em falantes típicos dessa região. A inserção da glide palatal ocorre exclusivamente entre as vogais centrais [e##á] de palavras adjacentes, sendo a

primeira sempre átona e a segunda sempre acentuada. Segundo Segura (2013), a inserção que desfaz este hiato é então típica dos dialetos do Norte e do Centro do país.

Como se trata de um processo de fonética sintática, recentemente, o fenómeno tem vindo a ser tratado por investigadores que fazem a análise prosódica de vários fenómenos segmentais. Assim, Oliveira et al. (2014) e Oliveira et al. (submetido) com base em dados do *Interactive Atlas of the Prosody of Portuguese* (InAPoP) – Frota (2012–2015) – averiguaram quais seriam as condições prosódicas em que a inserção ocorre. Com base em dados de falantes femininas de duas localidades do distrito do Porto e outras duas do distrito de Viana do Castelo (rurais e urbanas), Oliveira et al. (2014) observaram que:

- (i) no que respeita ao domínio prosódico, a fronteira de sintagma fonológico bloqueia o fenómeno nas quatro localidades (*A minha [j] alma está em paz vs. Quanto à Maria, qulas às oito da manhã nunca lhe agradaram*);
- (ii) a glide ocorre quando o contexto segmental e acentual se verifica no interior de palavra prosódica (*Nunca tinha ouvido falar da região de Sim^aári Cura, ...*) e também entre clítico e palavra prosódica (*O jornalista perguntou a_{CL} Ana Moura pelo seu cachimbo novo*);
- (iii) o fenómeno atua com maior ou menor frequência em função das localidades, faixas etárias e modalidades discursivas;
- (iv) a presença de proeminência acima do constituinte *palavra prosódica* na segunda vogal condiciona a frequência de aplicação do processo, com especial nitidez se a primeira for um clítico (já que neste caso o acento da segunda vogal potencia a inserção em todas as localidades);
- (v) o fenómeno está presente nas quatro localidades, parecendo haver uma menor tendência para a inserção nos falantes da faixa etária mais baixa considerada (entre os 20–45 anos vs. falantes a partir de 60 anos de idade);
- (vi) está assim criado nesta região um cenário de variação que não implica necessariamente a perda da inserção da glide palatal nem a consequente aproximação à variedade *standard*.

Oliveira et al. (submetido), por seu turno, com base em dados obtidos nas mesmas condições do trabalho de Oliveira et al. (2014) referentes a 42 falantes femininas, mostra que o processo atua nos dialetos do Norte e do Centro, incluindo a região de Nisa, ainda que com uma taxa de aplicação mais baixa. A sua frequência parece estar relacionada com o dialeto de origem dos falantes, assim como a sua faixa etária (foram usadas as mesmas faixas etárias do trabalho anterior). Comparando as tarefas de leitura com as de fala semiespontânea, conclui-se que a inserção da glide não é em nenhuma das regiões analisadas obrigatória, independentemente dos domínios prosódicos considerados, sendo mais frequente na fala semiespontânea. Além disso, em todas as regiões, o fenómeno só ocorre no interior do sintagma entoacional, apesar de em algumas também poder ocorrer no interior de uma palavra prosódica. A frequência da inserção da glide aumenta paralelamente à idade dos falantes, embora as

diferenças não sejam estatisticamente relevantes – o que não é de estranhar tendo em conta a ampla dimensão das faixas etárias analisadas.

Sendo a inserção pouco frequente em algumas das regiões estudadas e a sua distribuição por idade mostrar que os falantes mais jovens raramente a produzem, trata-se certamente de uma mudança em curso, cujo resultado é a perda da inserção. A sua escassa ocorrência no estilo mais controlado (leitura) evidencia o facto de os falantes convergirem com o dialeto padrão em registos mais formais. Logo, estes falantes comportam-se como bi-dialetais, sofrendo em consequência pressões de sentido oposto: do seu dialeto, instigando a inserção e, do dialeto *standard*, pressionando para não haver inserção.

Cruz (2013), por seu turno, no âmbito do seu estudo acerca da variação prosódica, estudou 18 falantes femininas de três áreas urbanas (Albufeira, Castro Verde e Lisboa), de dois escalões etários (20–45 e >59) e com vários graus de escolaridade (igualmente a partir do InAPoP). Um dos fenómenos de que nos dá conta é da presença de vogal final (com a forma [i] ou [i]) em palavras terminadas em consoante (*papel*, *andar*) ou de semivogal em palavras terminadas em vogal acentuada (*café*) nas localidades do Alentejo e do Algarve.¹² A autora observou que nos dados de Albufeira (Algarve) a vogal não surgia, contrariamente ao que se passava em Castro Verde (Alentejo), ainda que após /ε/ também não a tivesse observado em Castro Verde. Este fenómeno já era conhecido (Vasconcelos 1896; ¹1901; Maia 1975; Florêncio 2001), a sua novidade no trabalho de Cruz reside portanto na sua descrição prosódica. A autora mostra que o segmento só ocorre na fronteira direita de sintagma entoacional, e não simplesmente na posição final de palavra como anteriormente era dito. A distribuição por idade em Castro Verde mostra que a presença da vogal é mais frequente entre os falantes da faixa etária mais alta do que na faixa etária mais baixa – após /r/, 38% vs. 13% e após /l/, 13% vs. 2% (Cruz 2013, 41–42). Comparados os dados do Alentejo com os do português europeu *standard*, onde a inserção ocorre também junto à fronteira de sintagma entoacional, a autora conclui que a presença da vogal se deve a razões diferentes (e ainda parcialmente desconhecidas)¹³ da que gera a presença de vogal final no português europeu *standard*.

12 Esta vogal é tida no seu trabalho como paragógica em resultado de um processo pós-lexical de inserção. Pode argumentar-se, em contrapartida, que a vogal no Alentejo se deve à preservação da forma lexical com vogal final nos casos das formas terminadas em sílaba de rima ramificada no português *standard* ou a um alongamento uniformizador do padrão acentual grave nas palavras agudas nesse dialeto, como *café*, explicações que nos parecem mais coerentes com a história destas estruturas.

13 A inserção está restringida a sílabas finais com coda sonante /l/ ou /r/ acentuadas no português *standard* (*papel*, *mar*, *comer*), uma vez que não ocorre em palavras *cadáver*, *Álcacer*. Parece haver portanto um preenchimento da posição de núcleo, perdida na história do dialeto *standard*, mas preservada em dialetos alentejanos (cf. Rodrigues 2015).

O vocalismo átono mostra-se assim um terreno muito produtivo para o estudo de processos variáveis, uns mais controlados por fatores estritamente linguísticos e outros por fatores sociais e estilísticos.

4 Consoantes em ataque silábico

No que ao consonantismo diz respeito, começamos por referir as principais conclusões de Santos (2003) acerca da consoante /tʃ/, em ataque, no Baixo Mondego (Vila Pouca do Campo) e de Aguiar Santos (2008, 73–80) acerca de /tʃ/ no dialeto transmontano da Terra Quente. O primeiro destes trabalhos regista a existência deste traço consonântico, típico dos dialetos setentrionais, junto a Coimbra (numa zona de transição em termos dialetais no que se refere a este traço, como referido em trabalho anterior, de Pinto 1981). A africada /tʃ/ em ataque (*chá*), contrasta fonologicamente com o segmento alvéolo-palatal /ʃ/ (*xarope*) nas duas regiões descritas por estes trabalhos. Santos (2003) conclui, no entanto, que a realização da africada é minoritária na comunidade analisada e parece especializada em termos lexicais. Uma vez que as realizações observadas da africada implicam um relaxamento da sua parte oclusiva, a autora sugere que isso seja reflexo de uma fase de mudança em curso, o que é também sustentado pelo facto de estas se encontrarem sociolinguisticamente restringidas. A frequência das formas africadas é mais elevada nas mulheres do que nos homens (55% vs. 36%). A frequência da realização africada relaciona-se com a idade dos falantes, atingindo nos falantes mais velhos os 80%, e ainda com o seu nível de escolaridade, na medida em que surge mais nos menos instruídos do que nos restantes (57% vs. 25%). No que respeita ao setor de atividade dos falantes, a autora conclui que a africada predomina entre os falantes do setor primário (75%), mas não nos outros dois discriminados no seu trabalho, em que há uma distribuição igual das duas variantes. Um facto interessante reportado por este trabalho é o de dois falantes jovens, estudantes, apresentarem esporadicamente a forma africada. Trata-se de produções justificadas pela emotividade discursiva ou pelo elevado grau de filiação à rede social local, na interpretação de Santos (2003, 181).

Na região transmontana o contraste entre /tʃ/ e /ʃ/ é tido habitualmente como um marcador linguístico, embora ele se estenda à região do Alto Minho e a outras regiões dos dialetos setentrionais e do Centro, como acabámos de dizer. No segundo dos trabalhos referidos acima (Aguiar Santos 2008), foi estudada a variação que o segmento admite em toda a região da Terra Quente (concelhos de Alfândega da Fé, Carrazeda de Ansiães, Mirandela, Macedo de Cavaleiros e Vila Flor). A africada /tʃ/ foi tomada como variável e foram comparados falantes pertencentes a quatro faixas etárias de dois níveis de escolaridade e de ambos os sexos. A análise dos dados permitiu concluir que a idade e a escolaridade são os fatores externos com mais impacto na variação deste segmento, para além do concelho de origem dos falantes.

O segmento admite hoje em dia variação nesta região, apesar de já haver poucas realizações africadas. Nos concelhos em que a africada é mais preservada, Vila Flor (28%) e Macedo de Cavaleiros (26%), a sua frequência é já relativamente baixa, sendo nos restantes muito mais baixa ainda (chegando aos 6% em Carrazeda). Há variação interna a cada falante e entre falantes. Os analfabetos mantêm mais a africada (55%) do que os alfabetizados, embora a variável escolaridade esteja dependente do concelho de origem, já que em Macedo de Cavaleiros a africada é mais realizada por alfabetizados (72%), mas ela é mantida mais frequentemente entre os analfabetos em Alfândega da Fé (55%), Mirandela (88%) e Vila Flor (58%). No concelho de Carrazeda regista-se 50% de realização africada. A distribuição por sexo mostra que as mulheres produzem menos a africada do que os homens (43% vs. 57%), um resultado que é inferior ao referido por Santos (2003) para Vila Pouca do Campo (55%, como referido acima). A maior parte das produções africadas ocorrem nos falantes com mais de 65 anos (94%), em todas os concelhos, embora em alguns, como o de Vila Flor, tenha ocorrido em falantes escolarizados. Há concelhos nos quais a africada só se realiza em falantes dessa faixa etária, estando ausente das produções de falantes com menos de 65 anos em Mirandela e Carrazeda e sendo sempre inferior a 10% nos restantes concelhos. Observando mais em detalhe as unidades lexicais em que a manutenção da africada ocorre, verifica-se que ela se dá preferencialmente em palavras como as formas dos verbos *chegar*, *chamar*, *chorar*, *encher*.

Desta forma, a partir destes dois trabalhos, conclui-se inesperadamente que a realização da africada apresenta uma distribuição social e geográfica com valores baixos na região transmontana estudada e valores ligeiramente mais altos na região de Coimbra (Vila Pouca do Campo). A frequência da africada parece estar condicionada pela idade, instrução, género dos falantes, para além de poder ser restringida a certos itens lexicais.

Outro dos principais traços consonânticos típicos dos dialetos setentrionais consiste na realização dos sons [β] e [b] para realizar a fricativa /v/ do português europeu *standard*. Nesses dialetos, é possível neutralizar foneticamente a oposição fonológica entre os segmentos iniciais de *vala* e *bala*, por exemplo, realizando [b] ou [β] em ambos os segmentos iniciais dos exemplos referidos.

Rodrigues (2003) e Santos (2003) estudaram, respetivamente, o comportamento destas estruturas em falantes de Braga e em falantes da Terra Quente transmontana. No primeiro destes trabalhos, certamente devido a incidir sobre uma comunidade urbana ou em resultado de convergência com a entrevistadora, pode verificar-se que são raros os informantes que não produzem nunca [v] (sendo esta a variante mais produtiva das três registadas), contrariamente ao que se poderia supor com base no facto de se tratar de uma localidade da área dialetal, onde a existência de fusão com /b/ deveria ocorrer regularmente. Por essa razão, a autora considera que o segmento /v/ já deve fazer parte do inventário fonológico dos falantes de Braga, substituindo o

segmento mais antigo /β/¹⁴ progressivamente, em todos os grupos sociais. Os resultados do seu trabalho apontam exatamente nesse sentido: a fricativa [v] surge em discurso informal em 66% das ocorrências possíveis e a bilabial não-estridente [β] em 32,4%, contra apenas 1,3% de [b] quase exclusivamente em posição inicial de frase. Os principais fatores intervenientes na variação, no entanto, são de natureza externa. Se juntarmos as duas variantes bilabiais de /b/ a sua frequência baixa à medida que a escolaridade sobe: passa de 43,8% e 46,2% nos dois níveis mais baixos de escolaridade para apenas 23% nos licenciados. De modo idêntico, a frequência das formas bilabiais aumenta com a idade, apresentando 21,1% nos falantes com menos de 20 anos e atingindo o seu valor máximo na faixa etária dos 40–55 anos. A faixa etária dos >56 apresenta uma taxa de bilabiais de 44%, ligeiramente mais baixa, o que pode ser explicado pela existência na amostra de maior número de falantes entrevistados docentes. As bilabiais são mais frequentes nos falantes masculinos do que nas falantes femininas (46,2% vs. 33,1%). Esta distribuição social das variantes indica que as mulheres, sendo mais rapidamente sensíveis aos fenómenos de mudança linguística, já usam mais do que os homens a inovação no dialeto, [v]. Logo, este consiste numa mudança em curso naquela comunidade urbana (Rodrigues 2003, 281–292), apesar de nos dados de Braga haver só uma informante (licenciada, professora do ensino secundário) sem oscilação na pronúncia de /v/ como [v].

Santos (2003), no seu trabalho acerca de Vila Pouca do Campo, comparou as leituras de um texto onde surgiam dezanove ocorrências de /v/ pelos seus informantes com as produções em *estilo natural motivado por inquérito*. As principais conclusões a que chega são as seguintes:

- (i) A fusão de /v/ e /b/ na localidade estudada é fenómeno corrente, porque só encontra um informante que não apresenta qualquer exemplo que a comprove (é do sexo feminino, jovem e com o nível de instrução mais alto).
- (ii) Homens e mulheres parecem comportar-se de modo diferente no que respeita à fusão: como nesta comunidade elas são muitas vezes menos instruídas do que eles, apresentam taxas de fusão mais elevadas (45% vs. 25%).
- (iii) 66% dos jovens apresenta baixa frequência de fusão.
- (iv) Os falantes com mais instrução apresentam poucas vezes fusão, de um modo geral.
- (v) Os falantes empregados no setor terciário apresentam 60% de fusão, o que mostra que mesmo com relações frequentes com o meio urbano os falantes conservam a pronúncia mais marcada, com a fusão.
- (vi) 57% dos informantes demonstra no entanto ter alguma consciência da marcação da fusão, na medida em que produz casos de hipercorreção (e.g. *ambiente* [ẽ'vjêti]).

¹⁴ Consideramos que, como defendido por Maia (1986) e por Rodrigues (2003), nos dialetos setentrionais o segmento conservador é o fricativo bilabial, não o oclusivo.

Um outro segmento que apresenta variação é /r/ em ataque inicial (*roxo*) e em palavras como *parra*, *Israel*, *palra*, *tenro*. Ainda que o português europeu *standard* apresente quase só [R] e formas fricativas posteriores nestes contextos, é certo que no português europeu *standard* e noutras variedades é possível encontrar falantes que de forma sistemática optam pelo uso de [r] nestes contextos. Os dados de Rodrigues (2003), por exemplo, apontam para uma maior frequência de ocorrência de [r] (a vibrante dentoalveolar múltipla) em Braga do que em Lisboa. Em Lisboa a forma [r] aparece mais em posição inicial do que no interior da palavra (16,4% vs. 12,4%), mas em Braga não há grande diferença nestes dois contextos (inicial 23,6% vs. 21,4%). Estes dados contrariam todos aqueles que dão como categórico no português europeu *standard* o processo de substituição da sonante nestes contextos por formas não anteriores (e.g. Mateus/Andrade 2000, 11; Vigário 2003). Os falantes que produzem as formas anteriores da consoante, no entanto, não pertencem à faixa etária mais jovem da amostra de Lisboa (13–18 anos) nem às duas mais baixas de Braga (13–18 e 20–25 anos). As formas dominantes de realizar o segmento fonológico nas duas cidades são, portanto, as formas posteriores (nomeadamente as formas fricativas). Há falantes nativos das duas cidades que apresentam exclusivamente a forma anterior, outros que preferem as não anteriores e alguns que usam as duas (casos em que tanto oscilam em posição inicial como não inicial). Por esse motivo, a variação parece ser livre, não sendo certo que venha a ocorrer a total substituição do alofone [r] pelos alofones não anteriores, em função da distribuição até agora registada.

5 Conclusões

Por forma a obter uma visão abrangente da variação sociolinguística do português europeu, organizámos o capítulo em função das grandes categorias de segmentos fonológicos. Tendo começado pelas características das vogais acentuadas (secção 2), o trabalho centrou-se na variação associada à vogal /e/ antes de segmentos palatais no português europeu *standard* e nos dialetos setentrionais, que parece ser um fenómeno em ampla expansão em português europeu e na palatalização de /u/, por ser um fenómeno que recebeu diferentes abordagens fonológicas e apresenta uma distribuição geográfica ainda não totalmente conhecida. Na secção 3 detivemo-nos na variação fonológica e fonética associada às vogais átonas (nomeadamente as vogais médias), na medida em que os processos a que o vocalismo átono está sujeito são um contributo importante para a compreensão da natureza rítmica do português europeu. Na secção 4, abordámos alguns traços consonânticos variáveis no território continental em ataque silábico, com base na informação disponível que permite traçar um quadro de variação socialmente condicionada. Sempre que nos pareceu adequado e foi possível, foram tidos em consideração dados de fonologia, de fonética e de prosódia, para além de dados propriamente sociolinguísticos.

Os diversos processos fonológicos do português europeu discutidos mostram variação geolinguística, variação interna a variedades linguísticas específicas (incluindo o português europeu *standard*), variação entre falantes e variação interna a um mesmo falante. Desse modo, procurámos chamar a atenção para os tópicos que mais interesse têm suscitado em português europeu entre os fonólogos, dialetólogos e sociolinguistas e para os assuntos que ainda carecem de investigação mais detalhada. Este breve resumo demonstra, contudo, que a língua carece ainda de muitos trabalhos na área sociolinguística.

6 Referências

- Aguiar Santos, Isabel Joana (2008), *Unidades e processos fonológicos no falar da região da Terra Quente: contributos para a linguística forense*, dissertação de Mestrado, Universidade do Minho.
- Andrade, Ernesto d' (1977), *Aspects de la Phonologie (Générative) du Portugais*, Lisboa, CLUL-INIC.
- Barbosa, João Morais (1994), *Introdução ao Estudo da Fonologia e Morfologia do Português*, Coimbra, Almedina.
- Barros, Rita Queirós de (1994), *Contributo para a uma análise sociolinguística do Português de Lisboa: variantes de /e/ e /ɛ/ em contexto pré-palatal*, dissertação de Mestrado, Faculdade de Letras da Universidade do Minho.
- Brissos, Fernando (2012), *Linguagem do Sueste da Beira no tempo e no espaço*, Lisboa, Centro de Linguística da Universidade de Lisboa.
- Brissos, Fernando (2014a), *New insights into Portuguese central-southern dialects: understanding their present and past forms through acoustic data from stressed vowels*, *Journal of Portuguese Linguistics* 13:1, 63–115.
- Brissos, Fernando (2014b), *A vogal u, os dialectos do Centro-Sul português e a dialectologia acústica*, in: *Textos Seleccionados, XXIX Encontro Nacional da Associação Portuguesa de Linguística*, Porto, APL, 85–102.
- Brissos, Fernando (2015), *Dialectos portugueses do Centro-Sul: corpus de fenómenos e revisão do problema da (des)unidade*, *Zeitschrift für romanische Philologie* 131:4, 999–1041.
- Brissos, Fernando/Rodrigues, Celeste (2016), *Vocalismo acentuado do Noroeste português – descrição acústica, variação dialectal e representação fonológica*, *Revue Romane* 51:1, 1–35.
- Cruz, Marisa (2013), *Prosodic variation in European Portuguese: phrasing, intonation and rhythm in central-southern varieties*, tese de Doutoramento, Universidade de Lisboa.
- Florêncio, Manuela (2001), *Dialecto Alentejano – contributos para o seu estudo*, Lisboa, Colibri.
- Freitas, Maria João (1997), *Aquisição da Estrutura Silábica do Português Europeu*, tese de Doutoramento, Universidade de Lisboa.
- Frota, Sónia (coord.) (2012–2015), *InAPoP-Interactive Atlas of the Prosody of Portuguese* (FCT-PTDC/CLE-LIN/119787/2010, <http://labfon.letras.ulisboa.pt/InAPoP/>) (23.10.2015).
- Labov, William (1972), *Sociolinguistic Patterns*, Philadelphia, University of Pennsylvania Press.
- Labov, William (1994), *Principles of Linguistic Change-Internal factors*, vol. 1, Cambridge/Oxford, Blackwell.
- Labov, William (2001), *Principles of Linguistic Change-Social factors*, vol. 2, Cambridge/Oxford, Blackwell.
- Maia, Clarinda de Azevedo (1975), *Os falares do Algarve (inovação e conservação)*, *Revista Portuguesa de Filologia* 17:1–2, 37–205.

- Maia, Clarinda de Azevedo (1986), *História do Galego-Português*, Coimbra, INIC.
- Mascarenhas, Isabel (1996), *Estudo da variação dialectal entre Lisboa e Porto das vogais [-rec] e [+arred] em contexto inicial*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Mateus, Maria Helena Mira (1982, ¹1975), *Aspectos da Fonologia Portuguesa*, Lisboa, INIC.
- Mateus, Maria Helena/Andrade, Ernesto d' (2000), *The Phonology of Portuguese*, Oxford, Oxford University Press.
- Mateus, M. Helena Mira, et al. (2003) *Gramática da Língua Portuguesa*, Lisboa, Caminho.
- Oliveira, Pedro, et al. (2014), *Onde (ainda [j]) há o fenómeno – Contributo para o estudo da inserção de glide entre vogais centrais*, in: *Textos Seleccionados, XXIX Encontro Nacional da Associação Portuguesa de Linguística*, Porto, APL, 419–436.
- Oliveira, Pedro, et al. (submetido), *Glide insertion to break a hiatus across words in European Portuguese*, in: Pilar Barbosa/Conceição Paiva/Celeste Rodrigues (edd.), *Studies on variation and change in varieties of Portuguese*, Amsterdam/Philadelphia, Benjamins.
- Pinto, Adelina Angélica (1981), *A africada ç em português: estudo sincrónico e diacrónico*, Boletim de Filologia 26, 139–192.
- Rodrigues, Celeste (2003), *Lisboa e Braga: Fonologia e Variação*, Lisboa, FCG/FCT.
- Rodrigues, Celeste (2015), *Evidências de regularização acentual no Litoral Alentejano*, Revista da Abralín 14:1, 463–479.
- Segura, Luísa (2013), *Variedades dialectais do Português Europeu*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 5, Lisboa, Fundação Calouste Gulbenkian, 85–142.
- Segura da Cruz, Luísa (1988), *A Fronteira Dialectal do Barlavento do Algarve*, dissertação para acesso à categoria de investigador auxiliar, Lisboa, CLUL-INIC.
- Santos, Isabel Almeida (2003), *Variação Linguística em Espaço Rural*, Lisboa, Imprensa Nacional Casa da Moeda.
- Vasconcelos, José Leite de (1896), *Dialectos Alentejanos*, Revista Lusitana 4, 13–77.
- Vasconcelos, José Leite de (³1987, ¹1901), *Esquisse d'une Dialectologie Portugaise*, Porto, CLUL-INIC.
- Veloso, João (2012), *Vogais centrais do português europeu contemporâneo: Uma proposta de análise à luz da fonologia dos elementos*, Letras de Hoje 47:3, 234–243.
- Vigário, Marina (2003), *The Prosodic Word in European Portuguese*, Berlin/New York, Mouton de Gruyter.
- Weinreich, Uriel/Labov, William/Herzog, Marvin (1968), *Empirical foundations for a theory of language change*, in: Winfred Lehmann/Yakov Malkiel (edd.), *Directions for historical linguistics*, Austin, University of Texas Press, 87–195.

Ana Isabel Mata e Helena Moniz

5 Prosódia, variação e processamento automático

Abstract: Neste capítulo apresentamos um olhar panorâmico sobre a variação prosódica e sobre a sua interface com a área do processamento automático de fala. Tendo por base essencialmente a investigação que tem sido desenvolvida no português europeu sobre corpora de fala espontânea e preparada, em contextos de exposição e de interação, nomeadamente na variedade padrão falada em Lisboa, analisamos a variação da entoação em contextos declarativos e interrogativos, e abordamos as funções pragmático-discursivas que podem associar-se também a outros parâmetros prosódicos. Partindo de estudos comparativos interestilos (com maior/menor grau de espontaneidade e de planeamento, e natureza mais interativa/expositiva) e interlocutores (espaço geográfico, género, grupo etário/estatuto), destacamos o papel da variação estilística e sociolinguística da prosódia no português europeu. Mostramos também o papel da variação no processamento automático de proeminência prosódica, pontuação, disfluências e emoções.

Keywords: prosódia, entoação, variação estilística, variação social, processamento automático de fala

1 Introdução

São múltiplas as funções desempenhadas pela prosódia. A variação da entoação é um dos mais importantes mecanismos para o agrupamento de constituintes e a demarcação de fronteiras no contínuo de discurso, para a marcação de ênfase e contraste, para a distinção entre intenções comunicativas, e para a expressão de diferenças de atitude, emoção e expressividade. Para além da marcação de natureza melódica, baseada principalmente na variação dos contornos de frequência fundamental (f_0), fazem ainda parte dos mecanismos prosódicos de que os falantes dispõem os parâmetros de marcação temporal (incluindo as pausas) e a variação da energia e da qualidade da voz. Por exemplo, os enunciados declarativos são marcados por uma descida final de f_0 sobre a última sílaba acentuada, por um alongamento do núcleo dessa sílaba, muitas vezes por efeitos de laringalização pré-fronteira, podendo ainda ser seguidos por pausa. Para distinguir terminalidade de continuação, i.e. para sugerir a continuação desses enunciados, pode ser usado um padrão de f_0 diferente, nomeadamente uma subida final de f_0 . Em caso de hesitação, por exemplo antes da introdução de um tópico novo no discurso, a reposição da fluência é marcada nesses enunciados por um contraste prosódico, indicado por diferenças de f_0 (tipo de padrão entoacional, nível e gama de variação mais elevados), mas também por uma maior

energia, por uma pausa de maior duração antes do item a destacar e ainda pela menor duração desse item, por comparação com a sua produção disfluente. O controlo da gama de f_0 e de energia e o da velocidade de elocução podem ainda traduzir-se na nítida impressão de uma maior expressividade dos falantes.

Para além de correlações entre variação prosódica e aspetos de outros níveis de informação linguística, nomeadamente sintáticos, semânticos e pragmático-discursivos, a investigação sobre corpora de fala espontânea tem mostrado que, a par dos índices de variação entoacional, também outros parâmetros prosódicos podem ser usados de forma regular pelos falantes na organização do discurso, contribuindo para a sua fluência comunicativa. A comparação intercorpora (relatos e diálogos espontâneos, apresentações orais escolares, aulas universitárias e noticiários televisivos), intracorpora (relatos espontâneos vs. apresentações preparadas não-lidas) e interfalantes (género, grupo etário, estatuto) tem também mostrado que a variação prosódica configura diferentes estratégias de adaptação dos falantes a estilos de fala distintos e desempenha ainda uma função sociolinguística. Por outro lado, os resultados destes estudos comparativos têm vindo a contribuir para avanços no processamento automático de fala, em especial para a predição de proeminência prosódica, sinais de pontuação, disfluências e marcadores discursivos, ou mesmo para a deteção de eventos paralinguísticos (e.g. emoções).

Neste capítulo incidimos sobre a variação da entoação no português europeu, referindo sempre que oportuno a variação de outros parâmetros, uma vez que não cabe nos limites deste capítulo um tratamento exaustivo das diferentes dimensões prosódicas. Apresentamos também um olhar panorâmico sobre os mais recentes desenvolvimentos na área do processamento automático da prosódia do português europeu. Temos como referência essencialmente os dados acumulados nas últimas décadas a partir da análise de corpora.

Na próxima secção (secção 2), descrevemos o sistema de representação e o modelo teórico adotado para a análise da entoação do português europeu. Ilustramos ainda nesta secção os principais padrões e funções da entoação em contextos declarativos e interrogativos, que têm sido os mais estudados na variedade padrão falada em Lisboa e os mais explorados na área de processamento automático de fala. Na secção 3 introduzimos a variação prosódica numa perspetiva geolinguística, social e estilística. Na secção 4 descrevemos algumas das suas aplicações atuais ao processamento automático da prosódia. Na secção 5 destacamos as principais conclusões e apontamos linhas de trabalho futuro.

2 Sistema de representação e modelo adotado

Para representar a variação dos contornos de f_0 relevante para a distinção quer de padrões quer de constituintes entoacionais no contínuo de fala, há que usar um sistema de transcrição. Preferencialmente, um sistema que: (i) permita especificar

diferentes tipos de informação prosódica, sincronizados com o sinal acústico, em linhas de notação distintas, refletindo uma perspectiva modular acerca das componentes da estrutura prosódica e da sua relação com outros níveis de análise linguística; (ii) tenha sido testado com sucesso no estudo da prosódia a partir de corpora; e (iii) se fundamente num modelo entoacional considerado consistente e de larga aplicação, uma vez que o ato de transcrever a entoação não é completamente neutro em relação a um modelo teórico, e implica uma interpretação linguística da variação fonética.

Neste capítulo seguimos a adaptação feita para o português (Viana et al. 2007) do sistema ToBI – *Tones and Break Indices*, inicialmente proposto para o inglês americano (Silverman et al. 1992; Pitrelli/Beckman/Hirschberg 1994). O ToBI baseia-se num modelo de interpretação das curvas de f_0 como sequências de tons no quadro da fonologia métrica autosssegmental. Desenvolvido essencialmente a partir dos trabalhos de Bruce (1977) para o sueco, de Pierrehumbert (1980) para o inglês americano, de Beckman/Pierrehumbert (1986) para o inglês americano e o japonês, este modelo é, nas últimas décadas, um dos mais influentes no estudo da entoação das línguas (para uma discussão do modelo, cf. Ladd 1996). É neste quadro geral que têm sido realizados estudos comparativos dos padrões entoacionais de línguas diferentes (e.g. Avesani/Hirschberg/Prieto 1995 para o inglês, o espanhol e o italiano; Grice et al. 1995 para o alemão, o italiano e o búlgaro; Grabe 1998 para o alemão e o inglês; Frota et al. 2007 para o português europeu, o italiano, o espanhol e o catalão; Jun 2005 para uma comparação interlínguas mais alargada) e de diferentes variedades da mesma língua (e.g. Grabe et al. 2000, Clopper/Smiljanic 2011 para variedades do inglês britânico e americano, respetivamente; Frota et al. 2015, para variedades do português europeu e do português brasileiro), e ainda muitos dos trabalhos sobre o português europeu nas últimas décadas (e.g. Viana 1987; Falé 1995; Grønnum/Viana 1999; Frota 2000; Vigário 2003; Cabarrão 2013; Moniz 2013; Cruz 2013; Moniz et al. 2014a; Mata et al. 2014a; Mata/Moniz/Batista 2016; Moniz et al. 2016).

A simplicidade do esquema de transcrição e o grau de consistência alcançado entre transcritores em diferentes tipos de fala, segundo algumas avaliações já publicadas (cf. Silverman et al. 1992; Pitrelli/Beckman/Hirschberg 1994, para o inglês; Escudero et al. 2012, para o catalão, e referências aí citadas para outras línguas; Mata et al. 2014b, para o português europeu), são ainda razões para, tal como tem acontecido em outras línguas (como, por exemplo, o G-ToBI para o alemão, ou o Cat-ToBI para o Catalão), se optar, desde a década de noventa, pela adaptação do sistema ToBI na descrição da variação prosódica a partir de corpora de fala do português europeu (cf. projeto CORAL).¹

O ToBI é constituído por 4 fiadas de transcrição paralelas, alinhadas entre si e com o sinal acústico: (i) fiada ortográfica, (ii) tonal, (iii) de índices de rutura e (iv) mis-

¹ https://www.l2f.inesc-id.pt/wiki/index.php/CORAL_-_Labelled_Spoken_Dialogue_Corpus#Publications.

celânea (fiada opcional, usada frequentemente para registar eventos paralinguísticos, como risos, tosse, disfluências, *inter alia*). Outras fiadas podem ser acrescentadas, por exemplo para a notação de categorias sintático-discursivas cuja análise possa ser considerada relevante para uma interpretação da variação entoacional (cf. Mata et al. 2014a). Uma transcrição multilinear ToBI é ilustrada para o português na Figura 1, feita com o programa Praat (Boersma/Weenink 2013). Identifica-se na legenda da figura (entre parêntesis) o corpus de onde foi extraído o exemplo. (O mesmo será feito para os restantes exemplos apresentados neste capítulo.)

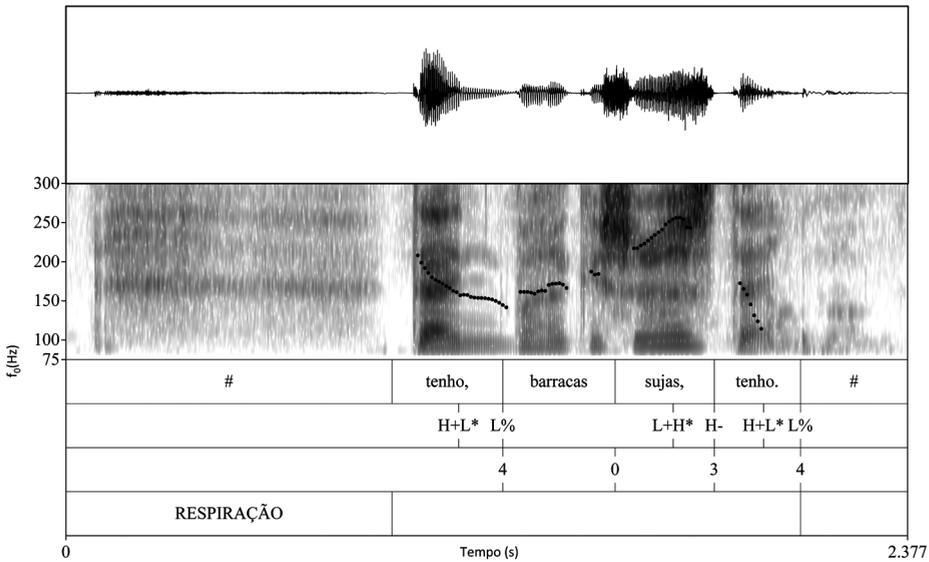


Figura 1: Exemplo de uma transcrição multilinear ToBI para o enunciado *tenho, barracas sujas, tenho.* (CORAL).²

As fiadas tonal e de índices de rutura representam o núcleo de uma análise prosódica. Destinam-se à codificação de dois tipos de informação: os tons que compõem o padrão entoacional e o agrupamento de palavras em constituintes prosódicos, respetivamente. A tipologia de tons do ToBI assume dois níveis básicos, tom alto (H) e tom baixo (L), e duas grandes categorias de descritores: (i) os acentos tonais, que atribuem proeminência aos itens lexicais e estão associados a sílabas metricamente acentuadas no enunciado (marcados com «*»); (ii) os tons associados a fronteiras de constituintes, que demarcam fronteiras de dois níveis de fraseamento entoacional (marcados com «-» e «%/o»).

² O corpus CORAL é constituído por diálogos espontâneos de tipo *map-task* (i.e., diálogos para a indicação de percursos em mapas) entre jovens adultos (Trancoso et al. 1998; ISLRN 499-311-025-331-2).

Tons

No português estão identificados sete acentos tonais ($H+L^*$, H^*+L , L^*+H , $L+H^*$, H^* , L^* , $H+!H^*$) e cinco tons fronteira ($L\%$, $H\%$, $!H\%$, $LH\%$, $HL\%$) – veja-se a representação esquemática de contornos de f_0 na Figura 2. Os acentos tonais podem corresponder a um tom simples (L^* , H^*) ou a eventos bitonais (e.g. $L+H^*$, L^*+H , $H+!H^*$), em que o diacrítico «*» distingue o tom que se encontra alinhado com a sílaba acentuada e o diacrítico «!» é utilizado quando há compressão da gama de variação dos níveis altos (!H). O acento tonal que ocorre em posição final de constituinte é, em regra, o mais importante. Por isso, é identificado como nuclear e em conjunto com o tom fronteira forma o designado contorno nuclear.

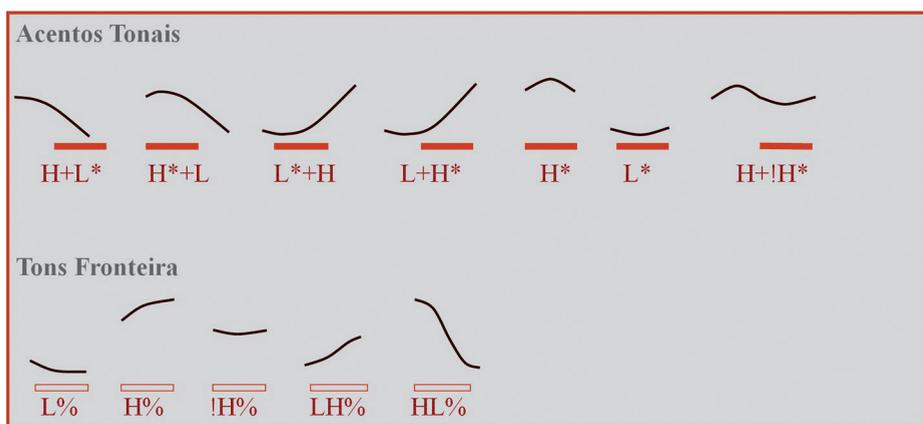


Figura 2: Esquemática de contornos de f_0 para acentos tonais e tons fronteira, seguindo *Towards a P_ToBI* (Viana et al. 2007). As linhas — indicam a sílaba acentuada.

Para o português têm sido considerados dois níveis de fraseamento entoacional, com configurações de fronteira de força diferenciada, tanto em termos de duração e extensão do movimento final de f_0 , como em termos da distribuição de pausas e da ocorrência de vários fenómenos de sândi. Frota (2000) e Viana et al. (2007) associaram os dois níveis a uma unidade do tipo constituinte entoacional (*intonational phrase*), distinguindo um constituinte maior (IP) e um constituinte menor (ip), na sequência de Ladd (1996). Para dar conta das diferenças entre estas fronteiras de fraseamento entoacional, usam-se os diacríticos tonais de fronteira «%» e «-», bem como os correspondentes índices de rutura 4 e 3, como veremos mais adiante. (Veja-se Frota 2014 para uma proposta de uso do diacrítico «%» a representar fonologicamente tanto «-» como «%».) O constituinte entoacional menor (ip), marcado com o diacrítico «-», compreende, em regra, pelo menos um acento tonal e um tom fronteira (L^- , H^- , e $!H^-$, menos frequentemente LH^- e HL^-). O constituinte entoacional maior (IP), por seu turno, é constituído por um ou mais constituintes entoacionais menores e o

seu tom fronteira marcado com o diacrítico «%», usado para descrever o movimento final de f_0 (frequentemente L% e H%, que em geral se associam à expressão de terminalidade/não-terminalidade, respetivamente).

Ilustramos em seguida, em contextos declarativos e interrogativos, os principais contornos nucleares e respetivas funções na variedade padrão falada em Lisboa.

H+L* (modelizado como uma descida pronunciada de f_0 que atinge um alvo tonal baixo dentro da sílaba acentuada) é um acento comum em posição nuclear em português europeu. Seguido de um tom fronteira L%, define o contorno característico das declarativas, bem como o das interrogativas Qu-, como os exemplos das Figuras 3 (*à tua esquerda.*) e 4 (*onde é que está?*) ilustram. Seguido de um tom fronteira LH%, como no exemplo da Figura 5 (*está à tua esquerda?*), define o contorno característico das interrogativas de *sim-não* que funcionam como pedidos de informação (Frota 2002; 2014; Viana et al. 2007; Mata/Santos 2010). Sabe-se que a perceção destas interrogativas depende sobretudo da extensão da subida final de f_0 : as fronteiras ascendentes LH% devem ser superiores a dois semitons (Falé 2005; Falé/Faria 2006) – no que se distinguem das fronteiras LH% que podem surgir em contornos intermédios, em declarativas longas com mais de uma fronteira entoacional. Quanto às interrogativas Qu-, tem sido apontado que estas tendem a distinguir-se do contorno das declarativas por apresentarem, comparativamente, um nível de f_0 inicial mais elevado e uma maior gama de variação (Viana 1987; Falé 2005).

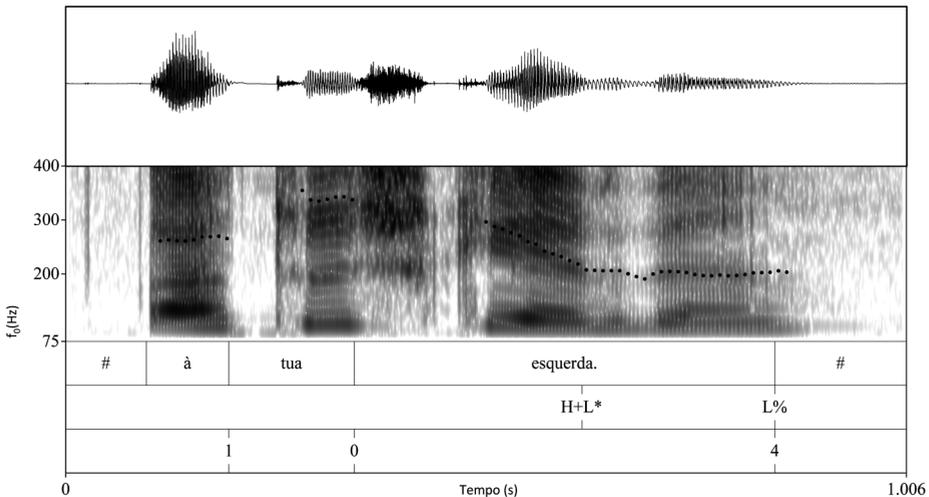


Figura 3: H+L* L% numa declarativa (CORAL).

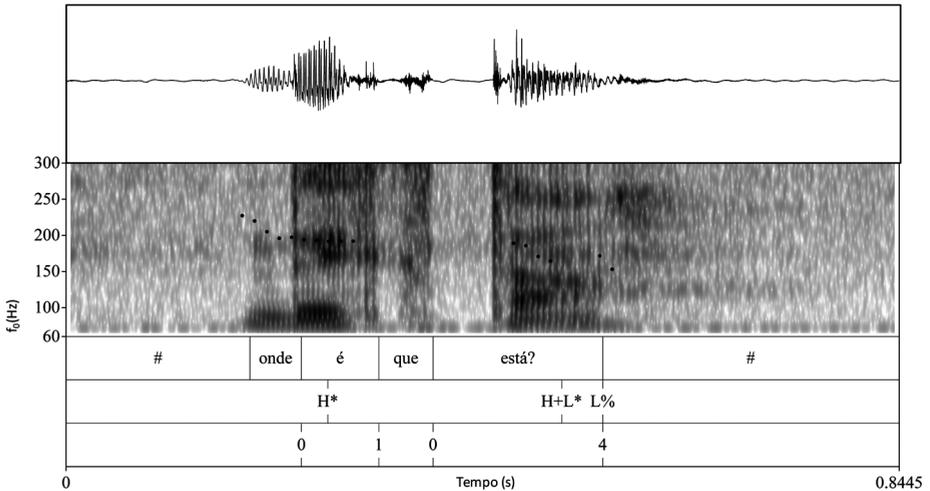


Figura 4: H+L* L% numa interrogativa Qu- (CORAL).

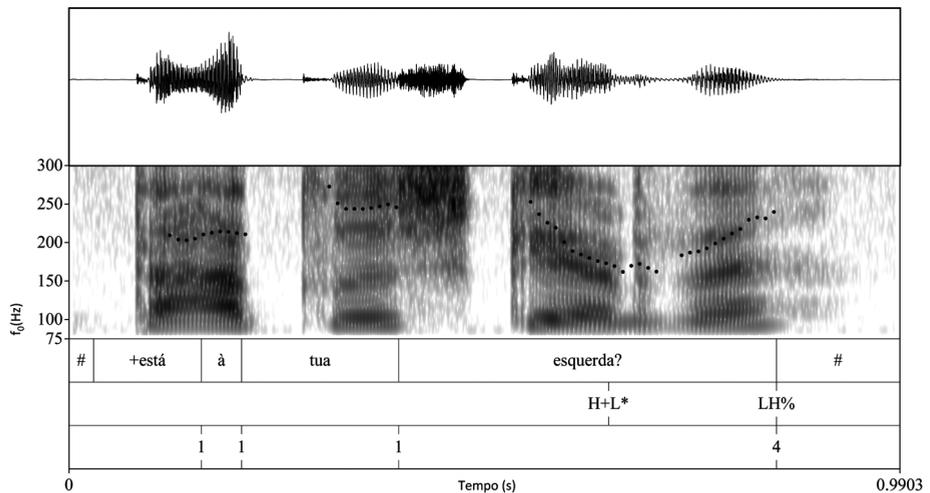


Figura 5: H+L* LH% numa interrogativa de *sim-não* que solicita informação (CORAL).

L* pode ocorrer como acento nuclear em constituintes não finais no interior de frases complexas, em materiais de leitura (Falé 1995). Em diálogos espontâneos, L* L% é, a par de H+L* L%, um padrão de resposta afirmativa comum para a expressão de concordância com o que foi dito, ou de confirmação (Cabarrão 2013), e.g. *sim*, nas Figuras 6 e 7. L* H% é comum para a sugestão de continuação (Viana et al. 2007; Cabarrão 2013), e.g. *sim* na Figura 8, uma resposta afirmativa com função reguladora: o falante indica ao interlocutor que compreendeu a mensagem e quer que este continue a dar-lhe informações/instruções.

A: Que é logo a seguir?

B: Sim.

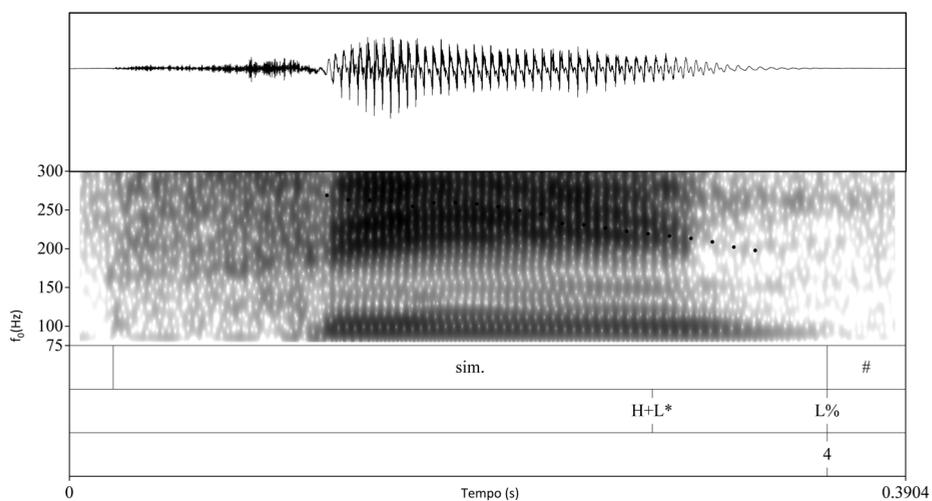


Figura 6: H+L* L% numa resposta afirmativa de confirmação (CORAL).

A: O pasto bravo fica-te à tua esquerda e os poços secos à tua direita.

B: Sim.

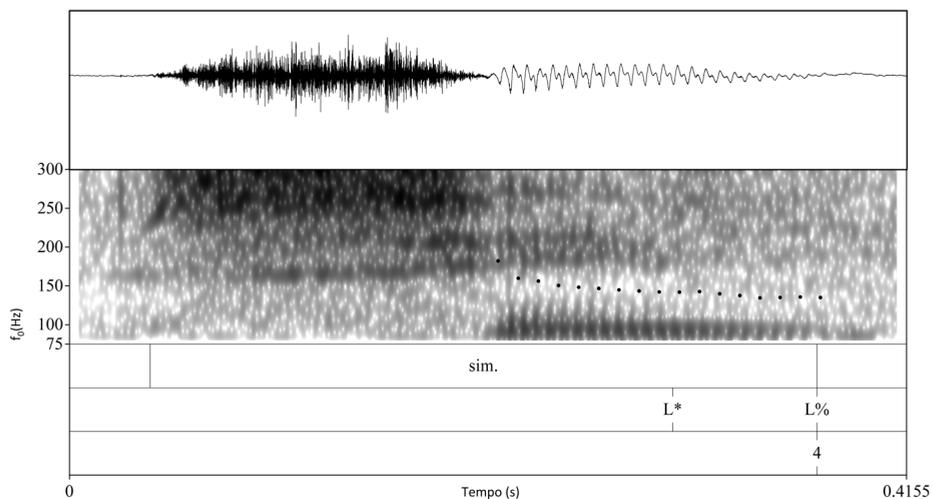


Figura 7: L* L% numa resposta afirmativa de concordância (CORAL).

A: Em cima, mais ou menos ali relativamente perto, ficam-te as barracas sujas.

B: Sim.

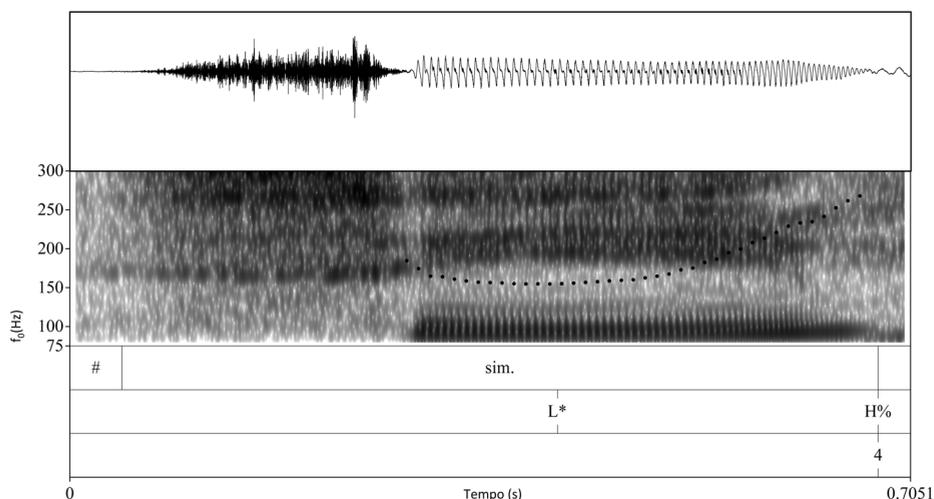


Figura 8: L* H% numa resposta afirmativa com função reguladora (CORAL).

H* e L+H* (modelizado como uma subida pronunciada de f_0 que atinge um alvo tonal alto dentro da sílaba acentuada) são comuns em posição nuclear (Viana et al. 2007), sobretudo seguidos de fronteiras H-/H% (Mata/Moniz/Batista 2016). Em contextos declarativos, H* H- surge associado à expressão de valores pragmáticos de continuação em constituintes entoacionais menores, como no exemplo da Figura 9 (*e então depois, quando fui para casa,*). Como a figura também ilustra, o mesmo pode acontecer em constituintes entoacionais maiores.

Viana et al. (2007) sugeriram que L+H* pode ser usado pelos falantes para assinalar informação nova ou para destacar informação dada que precisa de ser (re)ativada, como no exemplo da Figura 10, onde o falante introduz uma lista das qualidades mais apreciadas num animal (*dócil, elegante, amigo*), de acordo com as respostas a um inquérito que aplicou.

De facto, em contextos interrogativos (Mata/Santos 2010), L+H* e H* seguidos de fronteira tonal não descendente permitem distinguir interrogativas que solicitam confirmação de perceção (e.g. *por baixo do pátio de juizes?*, na Figura 11) vs. interrogativas que funcionam como pedidos de informação (cf. H+L* LH%, em *está à tua esquerda?*, na Figura 5).

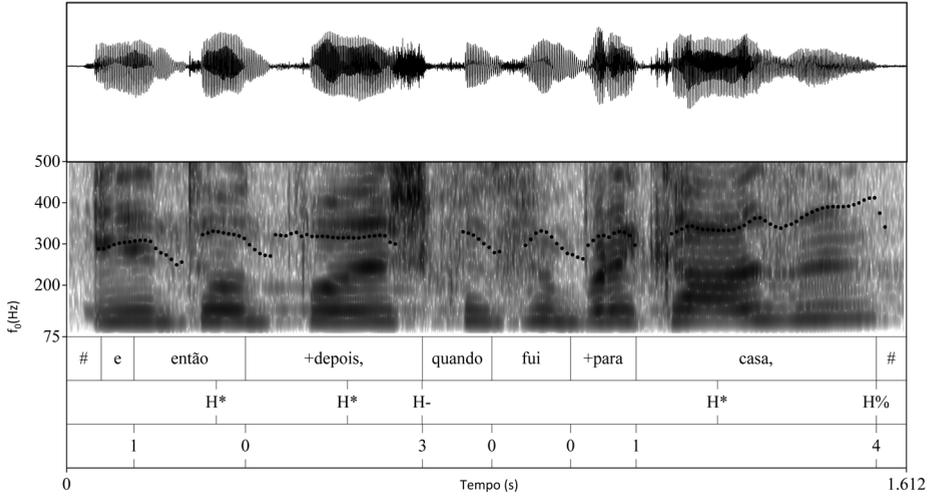


Figura 9: H* H-/H% em contexto declarativo (CPE-FACES).³

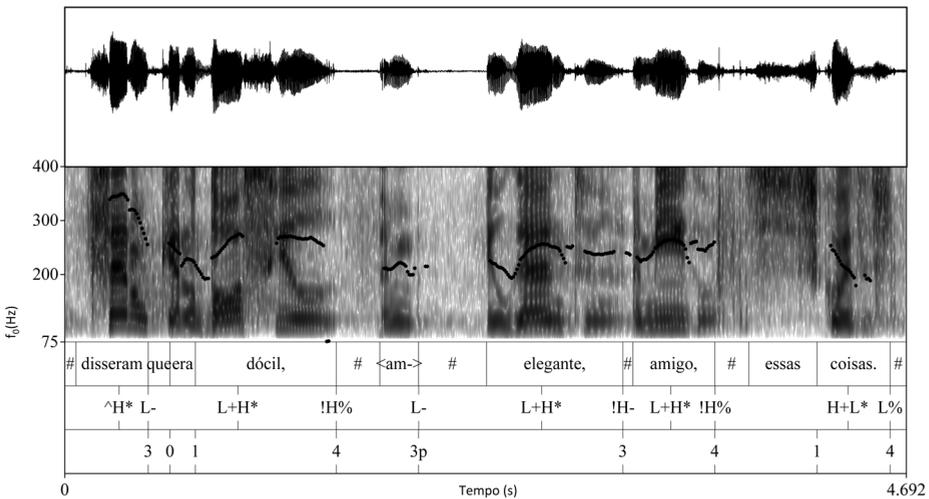


Figura 10: L+H* !H-/!H% em contexto declarativo (CPE-FACES).

Concorre ainda para a desambiguação entre atos de fala o acento L*+H (modelizado como um alvo tonal L dentro da sílaba acentuada seguido de uma subida pronunciada de f_0 que normalmente termina na sílaba seguinte). Em contextos interrogativos,

³ O corpus CPE-FACES é constituído por produções de fala espontânea e preparada (não-lida), de adultos e adolescentes de ambos os sexos, em contexto de sala de aula (Mata 1999; Mata et al. 2014b).

associa-se frequentemente a confirmativas de compreensão não neutras, que sugerem incorreção, incredulidade ou surpresa (Mata/Santos 2010). Veja-se o exemplo da Figura 12 (*vale irrigado?!).* Além disso, comparativamente às confirmativas de percepção, as confirmativas de compreensão são produzidas num nível de f_0 mais baixo (acento nuclear e tom fronteira).

- A: Portanto passas por baixo do pátio dos juízes,
- B: Por baixo do pátio de juízes?

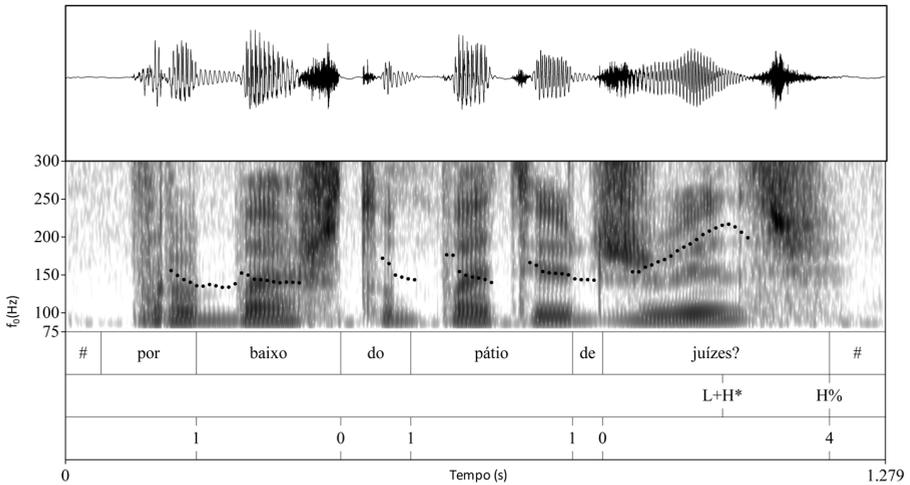


Figura 11: L+H* H% em interrogativa confirmativa de percepção (CORAL).

- A: Passas ao vale irrigado.
- B: Vale irrigado?! (Eu não tenho vale irrigado!)

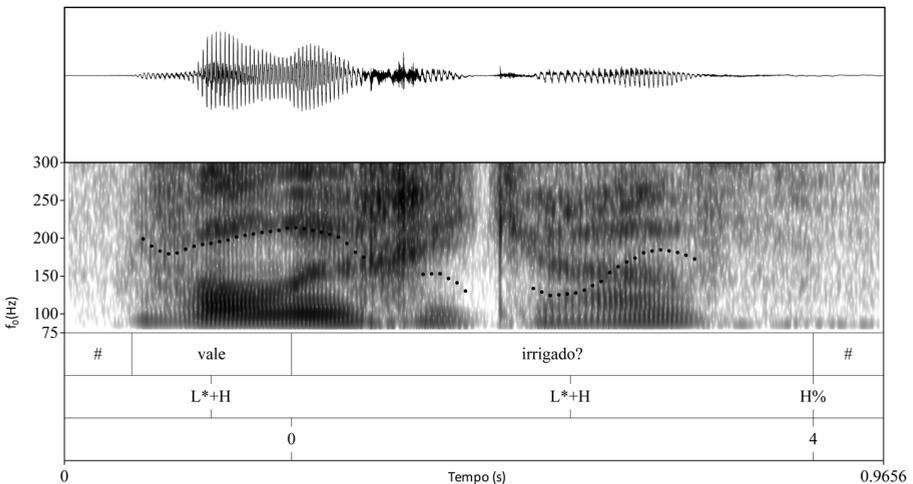


Figura 12: L*+H H% em interrogativa confirmativa de compreensão não neutra (CORAL).

L*+H é ainda comum em interrogativas *tag* (e.g. *não é?*, na Figura 13). Este acento tem também sido associado à marcação de foco prosódico em enunciados interrogativos (Frota 2002).

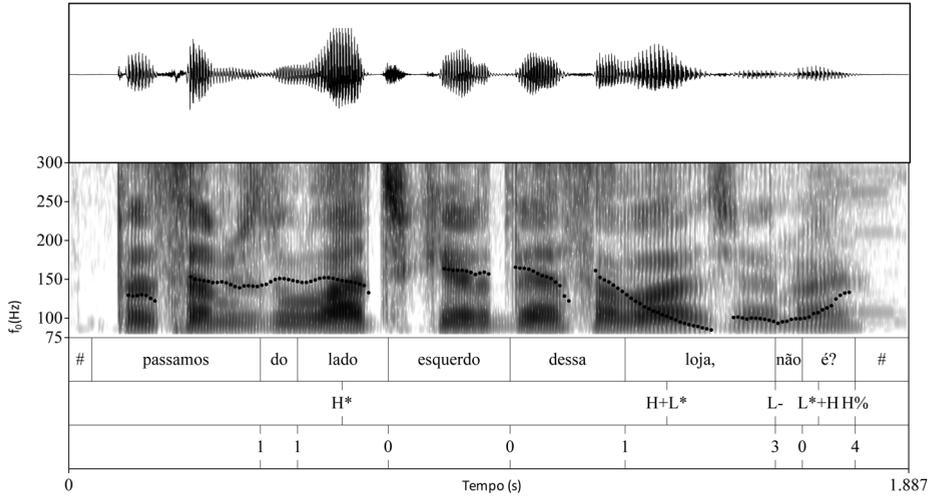


Figura 13: L*+H H% em interrogativa *tag* (CORAL).

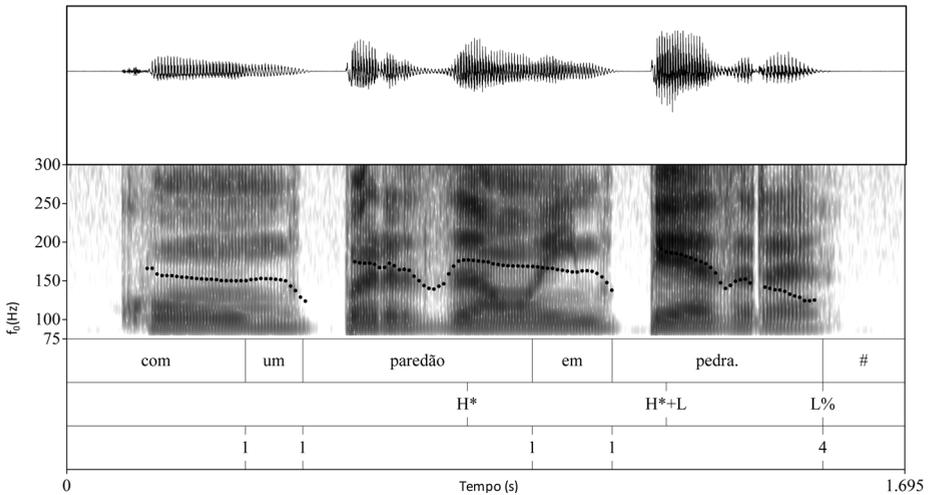


Figura 14: H*+L L% na marcação de foco prosódico (CORAL).

H*+L (modelizado como um alvo tonal H dentro da sílaba acentuada seguido de uma descida pronunciada de f_0 que normalmente termina na sílaba seguinte) tem sido associado a estratégias de demarcação prosódica de foco em enunciados declarativos (Frota 2000; 2014), como no exemplo da Figura 14, onde *em pedra* é acentuado com H*+L para assinalar que se trata de um paredão em pedra e não de um paredão de outro tipo.

Quanto a H+!H* (modelizado como um alvo tonal !H dentro da sílaba acentuada que é precedido por um pico de f_0 na sílaba átona anterior), por ser pouco frequente em posição nuclear, está ainda em larga medida por estudar a sua distribuição e o tipo de valores pragmáticos a que pode associar-se. Ilustra-se a sua ocorrência em contexto de reformulação (e.g. *sabia* vs. *sei*, na Figura 15), um dos contextos em que aparece em fala espontânea.

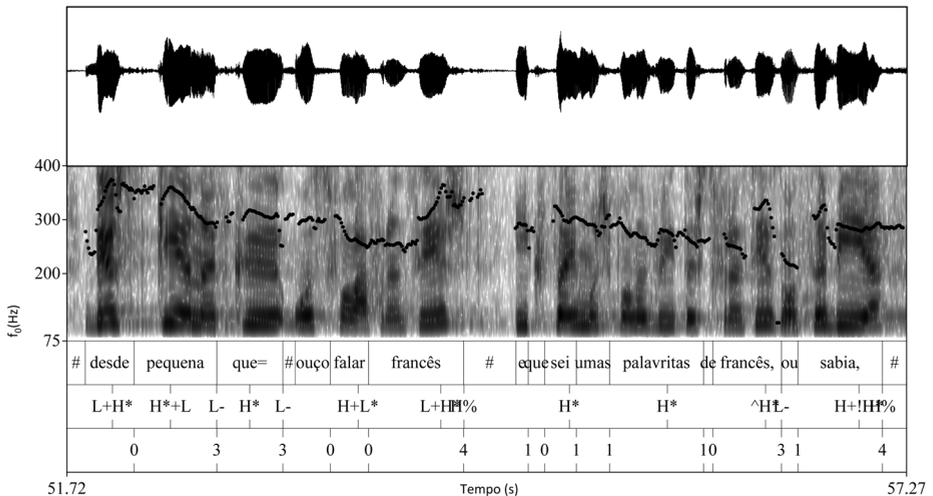


Figura 15: H+!H* H% em contexto de reformulação (CPE-FACES).

Índices de rutura

«Para complementar a informação tonal com uma representação da estrutura rítmica da fala, marca-se na fiada de índices de rutura a força de coerência ou de disjunção entre todas as palavras adjacentes» (Silverman et al. 1992, 868 – tradução nossa). Utiliza-se, para tal, uma escala de índices de rutura, inspirada no trabalho de Price et al. (1991). Esta escala varia entre os níveis 0 e 4. Quanto menor for o valor do índice de rutura, maior será o grau de ligação percebido entre uma palavra e o que lhe está adjacente; quanto maior o valor do índice de rutura, maior o grau de separação percebido nessa fronteira. O índice de rutura 0 indica uma estratégia de conexão máxima entre duas palavras consecutivas, i. e., em português europeu pode corres-

ponder a uma sequência fônica como [’lOg s ’ve],⁴ com a elisão das vogais átonas [u] e [@], em vez de [’lOgu s@ ’ve], correspondente à pronúncia das palavras em forma de citação; ou a [b6’Rak6 ’SuZ6S], com a realização de apenas uma ([S]) das duas fricativas na concatenação das palavras [b6’Rak6S ’suZ6S], como se observa na Figura 1. O índice de rutura 1 é o mais comum entre duas palavras consecutivas no interior de um constituinte entoacional, não implicando omissão de segmentos fonéticos (e.g. [kj ’E], forma de sândi externo predizível, em vez de [k@ ’E]). O índice de rutura 2 é utilizado quando o anotador tem dúvidas em relação ao nível a atribuir, motivadas por: i) perceptivamente considerar que se pode tratar de um nível 1, mas a palavra contém pistas tonais ou alongamentos de segmentos, como os que são comuns em pré-fronteira de constituinte entoacional; ou ii) perceptivamente considerar que se pode tratar de um nível 3 ou 4, com pausa real ou virtual, mas ausência de pistas tonais de fronteira de constituinte. Os índices de rutura 3 e 4 correspondem a fronteiras de constituintes entoacionais menores e maiores, respetivamente. De acordo com estudos anteriores sobre corpora de fala espontânea em português europeu (Mata/Moniz/Batista 2014; 2016; Moniz/Trancoso/Mata 2010), em fronteiras de nível 3 (ip) verifica-se uma frequência mais elevada de tons altos/ascendentes e pausas silenciosas subsequentes mais curtas; em fronteiras de nível 4 (IP), uma frequência mais elevada de tons baixos/descendentes e de efeitos fonéticos pré-fronteira (e.g. efeitos de voz laringalizada), e pausas silenciosas subsequentes mais longas. São também pistas prosódicas para a distinção entre estes dois níveis os declives e desvios padrão de f_0 na última palavra do constituinte: os valores são sempre mais elevados para os constituintes maiores do que para os constituintes menores, tanto no caso de tons altos/ascendentes como no caso de tons baixos/descendentes. Para além disso, em transcrições pontuadas, a quantidade de sinais de pontuação é significativamente mais elevada em fronteiras de nível 4 do que em fronteiras de nível 3. Comparando, por exemplo, a distribuição de pontos e de vírgulas, verifica-se que os pontos são associados sobretudo a fronteiras 4, as vírgulas, embora possam aparecer em ambos os níveis, associam-se maioritariamente a fronteiras de nível 3.

Outros tipos de informação prosódica relevantes para a deteção na fala de estruturas de nível hierárquico superior (como constituintes similares a frase, enunciados entoacionais e parágrafos), e consequentemente para a distinção entre estratégias de fluência comunicativa e estilos de fala, bem como para o processamento automático da estruturação multilinear da fala podem ser recuperados automaticamente, não necessitando de ser marcados numa transcrição manual ToBI. Entre esses tipos de informação destacam-se declives de f_0 e de energia intra e inter constituintes entoacionais, reinicializações de f_0 e de energia pós-fronteira, alongamentos finais, mudanças na qualidade da voz e pausas (sobre a importância destes parâmetros

⁴ Transcrição fonética seguindo o *Speech Assessment Methods Phonetic Alphabet* (SAMPA) (Wells 1997; www.phon.ucl.ac.uk/home/sampa).

numa comparação interlínguas, veja-se Vaissière 1983). É a partir do contraste estabelecido pela variação contínua destes parâmetros, tanto a nível local como global, que se deteta uma estruturação prosódica multilinear.

3 Variação prosódica

Sabe-se que, tal como a variação de natureza segmental (74 Variação sociolinguística), lexical e sintática, também a variação prosódica (melódica e temporal) concorre para a distinção entre grupos de falantes do português e entre situações de comunicação oral. No primeiro caso, esta distinção é habitualmente associada a fatores geolinguísticos e sociolinguísticos, nomeadamente: espaço geográfico de origem, género, idade e estatuto sócio-profissional dos falantes. No segundo, a fatores estilísticos dependentes das circunstâncias da situação de produção que, em geral, remetem para os extremos do contínuo de variação entre fala espontânea e fala preparada.

3.1 Variação geolinguística

Relativamente à variação geolinguística da prosódia, no interior da variedade europeia do português, bem como entre variedades nacionais do português, em particular entre português europeu e português brasileiro, são relativamente recentes os estudos comparativos, tanto os de natureza fonética (cf. projeto Atlas Multimédia Prosódico do Espaço Românico – Língua Portuguesa (AMPER-Por),⁵ Moutinho coord., 2000–), como os desenvolvidos dentro de uma perspetiva fonológica métrica autosssegmental (cf. projeto Atlas Interativo da Prosódia do Português (InAPoP),⁶ Frota coord., 2012–2015). Os estudos publicados incidem sobretudo sobre materiais recolhidos em situação experimental para eliciação de frases, e em situação de leitura de frases, e concentram-se na análise de alguns dos tipos de estruturas frásicas mais estudados no português padrão, embora contemplem múltiplos parâmetros de variação prosódica, quer do ponto de vista da produção quer do da perceção: padrões de f_0 , duração e energia, no caso dos trabalhos no âmbito do AMPER-Por (cf., por exemplo, Moutinho/Coimbra 2014; 2015); aspetos rítmicos, fraseamento e extensão dos constituintes entoacionais, tipo e frequência dos contornos nucleares, distribuição de acentos tonais e realização fonética destes (alinhamento com a cadeia segmental), no caso dos trabalhos no âmbito do InAPoP (cf., por exemplo, Frota et al. 2015, e referências aí citadas; Cruz 2013; Vigário/Frota 2003). Progressivamente, estes estudos têm vindo a incluir variedades setentrionais, centro-meridionais e/ou insulares do português euro-

5 <http://pfonetica.web.ua.pt/AMPER-POR.htm>.

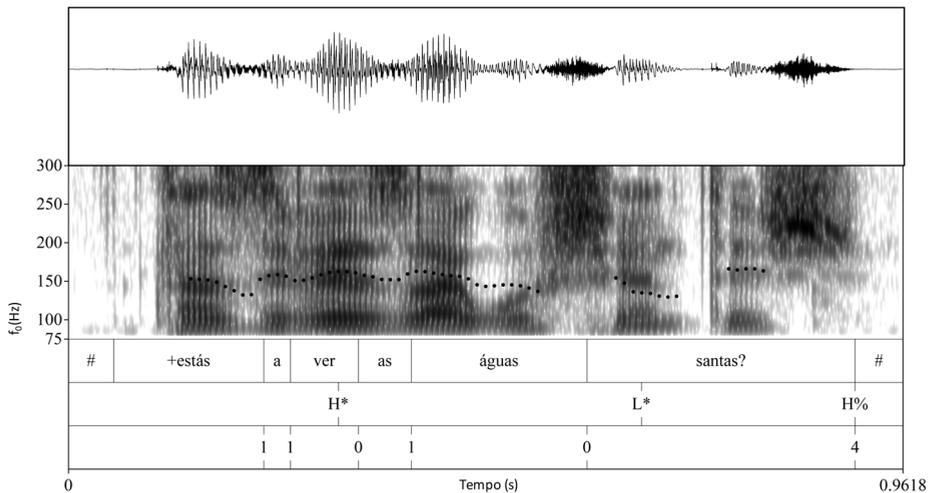
6 <http://labfon.letras.ulisboa.pt/InAPoP/presentation.html>.

peu, bem como variedades do português brasileiro faladas, principalmente, na costa atlântica. Quanto à caracterização prosódica das variedades africanas do português, não existem (quase) estudos publicados (cf. Santos/Fernandes-Svartman 2014, sobre declarativas (lidas) no português da Guiné-Bissau). Neste capítulo, por limite de espaço, concentrar-nos-emos na variação prosódica no interior do português europeu.

Na maioria dos tipos de estruturas que foram comparados entre variedades do português europeu, encontram-se padrões entoacionais idênticos para funções linguísticas idênticas (para declarativas neutras, cf. Moutinho/Coimbra 2014 e referências aí citadas; para declarativas neutras e focalizadas, imperativas com expressão de ordem e de pedido, interrogativas Qu- e vocativos com expressão de chamamento, cf. Frota et al. 2015). Por exemplo, as declarativas caracterizam-se por um padrão entoacional descendente independentemente das variedades estudadas. Este padrão é formado por um tom alto, H, associado à primeira sílaba acentuada e por um tom baixo, L, associado à última sílaba acentuada da frase, embora possa haver variação no alinhamento do alvo tonal L* entre variedades (mais tardio na variedade de Braga, segundo Vigário/Frota 2003), e na frequência de uso de H+L* L% (dominante nas variedades de Lisboa e do Algarve) e L* L% (mais frequente nas variedades do Porto e do Alentejo, segundo Frota et al. 2015). Esta variação entoacional do português europeu é assim compatível com a boa compreensão da entoação que se sabe existir interfalantes no interior da variedade europeia do português, tal como é geralmente reconhecido.

As principais diferenças intervariedades do português europeu são observadas nas interrogativas de *sim-não*: H+L* LH% é usado predominantemente pelos falantes da variedade de Lisboa, L*+H H% por falantes das outras variedades (Frota et al. 2015). Ilustramos para a variedade do Alentejo, na Figura 16, os contornos ascendentes L* H% (cf. Cruz 2013) e L*+H H%.

(a)



(b)

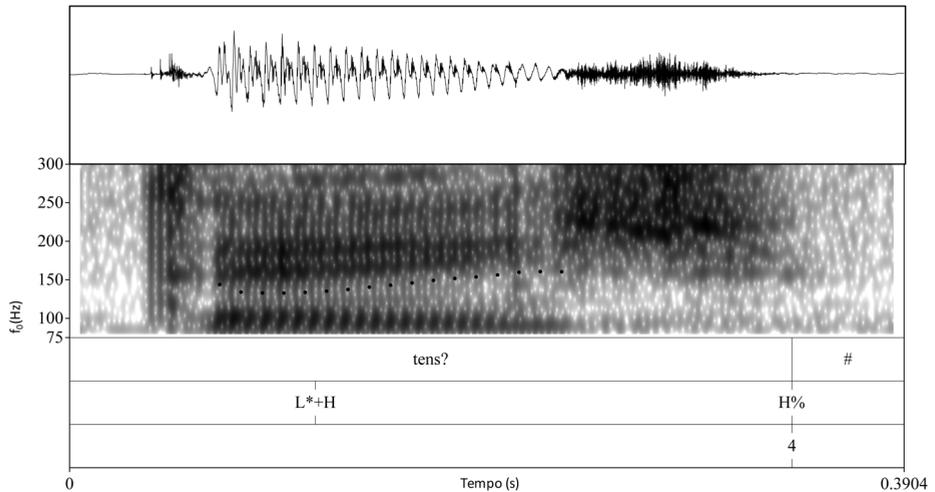


Figura 16: (a) L* H% e (b) L*+H H% em interrogativas de *sim-não* na variedade do Alentejo.

Assim, os resultados destes estudos permitem afirmar que as diferenças prosódicas entre variedades do português não se limitam apenas a questões de realização fonética, sem efeitos significativos na interpretação dos contornos nucleares. Podem incluir contrastes de natureza funcional em alguns tipos de frases, i.e., diferentes padrões podem ter um uso idêntico. Para além disso, um mesmo padrão pode ter usos diferentes em algumas variedades do português europeu (cf. Cruz 2013, sobre a não distinção de perguntas neutras e focalizadas na variedade do Algarve, realizadas com o mesmo padrão L*+H H%). São estas as diferenças que podem concorrer para dificuldades de compreensão interfalantes. No caso das interrogativas de *sim-não*, sabe-se que isso pode acontecer com alguma frequência já que quer em noticiários televisivos quer em fala espontânea as interrogativas de *sim-não* ocorrem muitas vezes com contornos nucleares descendentes (veja-se um exemplo disso na Figura 17), tanto na variedade padrão falada em Lisboa (Mata 1990; Moniz et al. 2011), como em outras variedades, nomeadamente na ilha de São Miguel, nos Açores (cf. Moutinho/Coimbra 2015). Nestes casos a distinção prosódica pode ser associada à gama de valores de f_0 , superior nas interrogativas comparativamente às declarativas, e não a contornos nucleares distintivos.

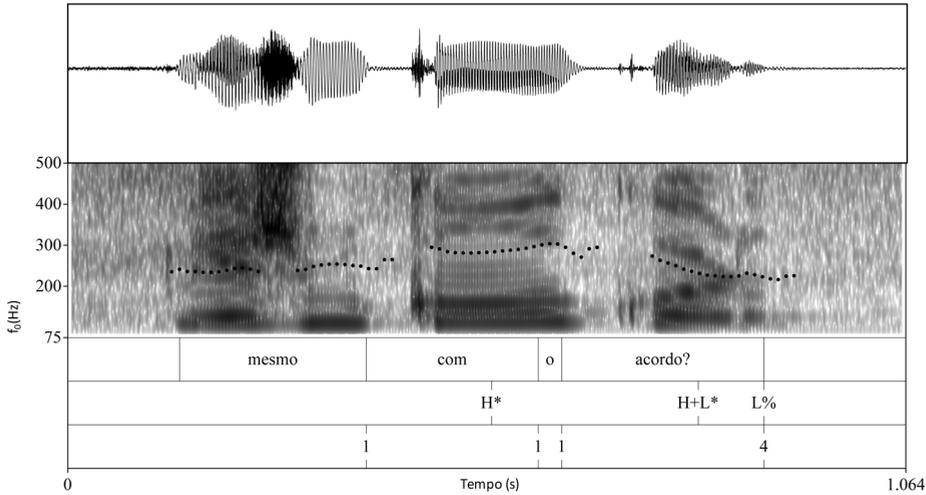


Figura 17: H+L* L% em interrogativa de *sim-não* na variedade de Lisboa (ALERT).⁷

Embora com resultados variáveis entre tipos de frase, os estudos no âmbito do InAPoP (Frota/Cruz 2012–2015) sugerem ainda que a frequência de distribuição de acentos tonais por palavra prosódica em posição pré-nuclear é menor em Lisboa que nas restantes variedades estudadas, tanto nas meridionais (Alentejo, Algarve), como nas setentrionais (Porto, Braga), que em geral são consideradas mais conservadoras que as restantes. Poder-se-á dizer que a maior ou menor ocorrência de configurações ascendentes e o maior ou menor número de acentos tonais em alguns tipos de frase parecem convergir para a sensação da diversidade sonora do país: nem todas as variedades do português europeu terão uma pronúncia seca e um tom cerrado, como habitualmente parece aos ouvidos de falantes de variedades africanas e brasileiras do português, sobretudo no caso da variedade padrão lisboeta.

3.2 Variação social

Comparativamente à variação interfalantes no espaço geográfico, que apresenta um número crescente de publicações, a área da variação social da prosódia tem vindo a receber menos atenção, em português, tal como em muitas outras línguas. É principalmente no âmbito da investigação sobre fala espontânea, e em particular sobre estilos de fala, exclusivamente na variedade padrão falada em Lisboa, que têm vindo a ser sugeridas pistas para uma distinção prosódica interfalantes, por influência de variáveis como género, grupo etário e/ou estatuto sócio-profissional dos falantes.

⁷ O corpus ALERT é constituído por telejornais da RTP (Meinedo/Neto 2003; Neto et al. 2003).

Algumas das características da variação por género observadas na fala espontânea de adultos são a seguir indicadas. Há, em média, sequências interpausas menos longas e menos tempo de pausa silenciosa (menos pausas, mais pausas breves) no discurso feminino em situação de conversa (sem alternância de tomadas de vez). Há, em média, mais tempo de pausa preenchida no discurso masculino (pausas preenchidas em maior número) na mesma situação. As diferenças apontadas são associadas a uma maior rapidez do discurso oral feminino (Freitas 1990). O tom fronteira LH% é em geral mais frequente para as falantes do género feminino em situação de relato espontâneo. Os acentos bitonais tendem a ser mais usados pelo género feminino (e.g. H+L*, L+H*) e L* pelo masculino (Mata/Moniz/Batista 2016), o que é consistente com observações feitas para outras línguas (cf. Clopper/Smiljanic 2011, sobre a maior frequência de tons complexos no discurso feminino).

Também na fala espontânea dos adolescentes (14–15 anos) se observa variação prosódica em função da variável género. Os rapazes distinguem-se das raparigas por uma velocidade de elocução e de articulação⁸ significativamente mais baixa (em média, há respetivamente 4,9/8,1 sílabas por segundo para os rapazes, em contraste com 6,6/9,6 para as raparigas, em constituintes similares a frase sem disfluências). Estas diferenças entre géneros podem ser interpretadas em linha com as diferenças apontadas por Freitas (1990) para adultos. O contraste entre constituintes entoacionais menores e maiores é em geral mais forte para os rapazes do que para as raparigas: estas produzem maior percentagem de ips que os rapazes; estes produzem maior percentagem de IPs que as raparigas (Mata/Moniz/Batista 2016). Os padrões finais descendentes-ascendentes (de tipo H+L* LH%) são mais raros para os rapazes que para as raparigas, e o inverso se passa com os padrões ascendentes-descendentes (Mata 1999), o que é consistente com o observado acima para os adultos.

Relativamente à variação interfalantes por grupo etário/estatuto, até agora apenas as diferenças entre adultos (professores) e adolescentes (alunos de 14–15 anos) foram abordadas numa perspetiva comparativa. As principais diferenças entre adultos e adolescentes em situação de relato espontâneo incluem (Mata et al. 2014b; Mata/Moniz/Batista 2014; 2016): a velocidade de elocução e de articulação, significativamente mais baixa para os adultos (em média, há respetivamente 4,9/6,7 sílabas por segundo para os adultos vs. 5,8/8,8 para os adolescentes, em constituintes similares a frase sem disfluências), e o rácio de fonação,⁹ em geral significativamente mais elevado para os adultos (72,3% vs. 65% para os adolescentes). Incluem também o uso de contornos nucleares de tipo alto/ascendente associados a valores pragmáticos de continuação (principalmente (L+)H* H) em enunciados declarativos, significativa-

⁸ Ambas são calculadas em sílabas por segundo: a velocidade de elocução corresponde ao número de sílabas dividido pela duração total de fala (incluindo silêncios), e a de articulação, ao número de sílabas dividido pela duração de fala, excluindo silêncios.

⁹ O rácio de fonação corresponde a 100% a multiplicar pela duração de fala sem silêncios, dividido pela duração de fala incluindo silêncios (Cucchiari/Strik/Boves 2002).

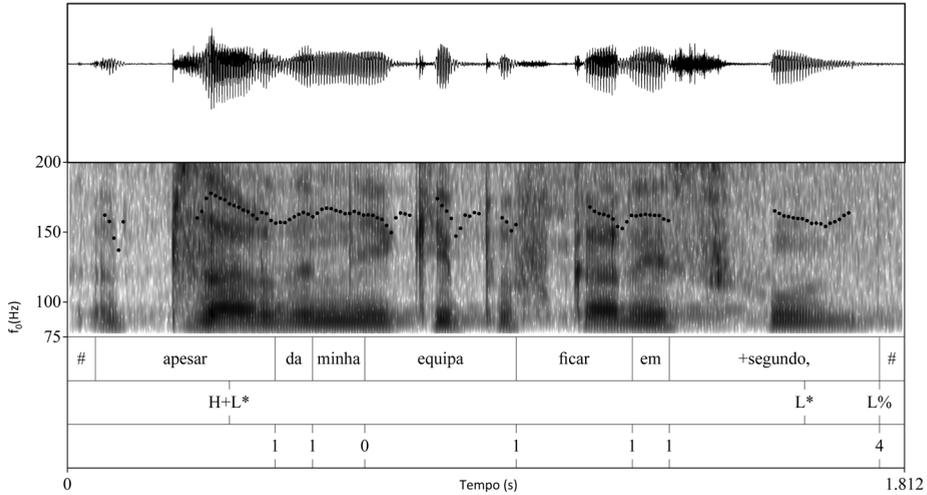


Figura 18: Elisão de vogal pós-tônica em *segundo* (CPE-FACES).

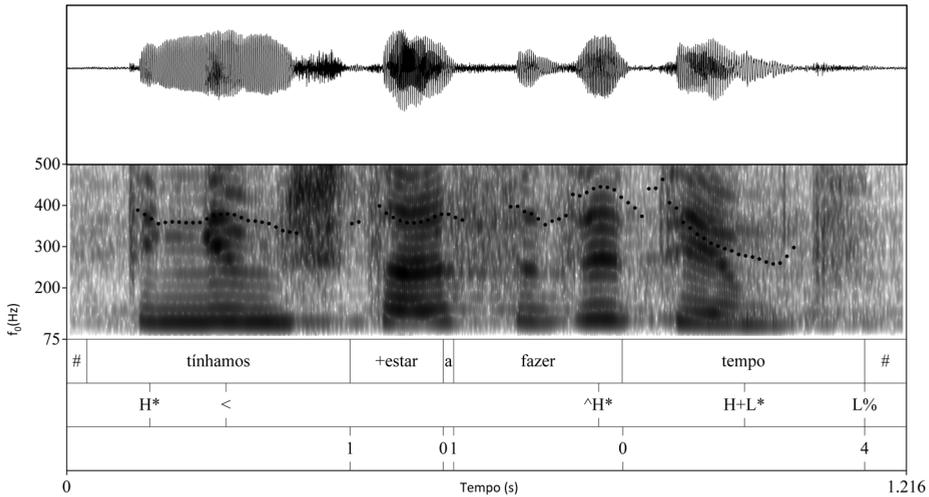


Figura 19: Elisão de vogal postônica e fricativação da última oclusiva em *tempo* (CPE-FACES).

mente mais frequente para os adultos, e a frequência de processos fonético-fonológicos (principalmente, desvozeamento/elisão de material pós-tônico, cf. Figuras 18 e 19; efeitos de laringalização,¹⁰ cf. Figura 20) pré-fronteira de constituinte entoacional, mais elevada para os adolescentes. Estes processos pré-fronteira são difíceis de

¹⁰ O termo «laringalização» corresponde à tradução do inglês *creaky voice* ou *laryngealized voice* (Andrade 1990).

processar automaticamente e, com frequência, originam segmentações automáticas erróneas, sobretudo para a fala dos adolescentes (Mata et al. 2014b).

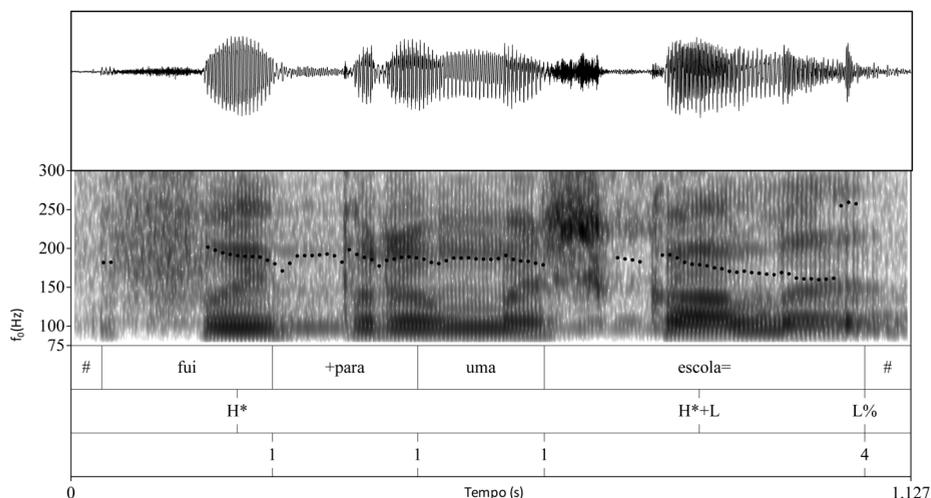


Figura 20: Efeitos de laringalização final em *escola* (CPE-FACES).

Quanto à variação em função do estatuto dos falantes, destaca-se o estudo comparativo de Delgado-Martins/Freitas (1991) sobre índices de marcação temporal em adultas de diferentes grupos sócio-profissionais (professoras universitárias e pivôs de telejornais). Este estudo demonstra que a produção de pausas silenciosas e preenchidas permite distinguir as estratégias usadas por professoras e jornalistas na fala espontânea: as primeiras produzem mais pausas silenciosas (21% vs. 13%), mas muito menos pausas preenchidas do que as segundas (6% vs. 32%). As autoras apontam no sentido de as professoras usarem mais pausas silenciosas com o intuito de clarificar a mensagem e de as jornalistas hesitarem mais no planeamento verbal.

Em conjunto, estes dados sobre a variação de estratégias interfalantes numa mesma situação de comunicação demonstram que a variação prosódica é uma dimensão importante para caracterizar o que é comum à variedade europeia do português e contribui para a codificação da identidade dos falantes de acordo com grupos sociais de pertença (incluindo, género, grupo etário, grupo sócio-profissional e variedade regional).

3.3 Variação estilística

Os dados acumulados nos últimos anos sobre a variação estilística da prosódia no português europeu demonstram que a variação entoacional e temporal permite distinguir entre diferentes estilos de fala, configura diferentes estratégias de fluência, e põe

em evidência diferenças entre gêneros e grupos etários nas estratégias de adaptação dos falantes a atividades de produção oral específicas. É à comparação das principais características da variação prosódica entre estilos de fala (intra e interfalantes) que se dedica o resto desta secção.

O contínuo de variação estilística fala espontânea-fala preparada (designações que remetem para os seus dois extremos) tem sido central na investigação sobre variação prosódica em português europeu, como em muitas outras línguas. Tradicionalmente, a investigação sobre estilos de fala concentra-se nas diferenças entre atividades de produção oral espontânea e de leitura de texto (para uma visão geral e uma discussão, cf. Eskénazi 1993; Barry 1995; Hirschberg 2000). Em português europeu, para além da leitura de texto (Freitas 1990), outros tipos de produção oral com preparação prévia, mas sem recurso à leitura de materiais escritos, têm sido estudados numa perspetiva comparativa. Por exemplo, relativamente à extensão de constituintes similares a frase (do inglês *sentence-like units*, i.e., constituintes delimitados por sinais de pontuação e, na sua maioria, por fronteiras tonais), sabe-se que os tipos de produção que se caracterizam por uma menor extensão do planeamento e uma natureza mais interativa se distinguem tipicamente por terem constituintes similares a frase de menor extensão (e.g. diálogos espontâneos), bem como velocidades de elocução e de articulação mais baixas e um rácio de fonação mais elevado. Com resultados variáveis, o inverso se passa nos tipos de produção que divergem no grau de espontaneidade e de planeamento prévio e têm uma natureza mais expositiva (e.g. aulas universitárias). Comparativamente a um noticiário televisivo, por exemplo, a extensão dos constituintes é menos acentuada numa aula universitária, sem dúvida como reflexo da atenção dada pelo falante à compreensão de conceitos, com paráfrases, explicação e exemplificação detalhadas, em suma, ao tempo que é investido na construção do conhecimento dos participantes. A extensão dos constituintes tende a aumentar como efeito do esforço do falante para apresentar grande quantidade de informação, de forma clara e bem estruturada, num curto intervalo de tempo, muitas vezes com recurso à leitura de materiais escritos (e.g. em noticiário televisivos ou em discursos no parlamento europeu), o que contrasta com o que é habitual numa aula universitária (cf. Moniz 2013, para uma comparação entre diferentes produções orais). Tenderá a diminuir como efeito de fatores de orientação para o ouvinte (e.g. em diálogos dirigidos a crianças). Veja-se o Quadro 1.

Quadro 1: Média de palavras por constituinte similar a frase em diferentes corpora.

Corpora	Média de palavras
Diálogos adulto-criança	3
Diálogos em formato <i>map-task</i>	6
Relato espontâneo de professores de 3.º Ciclo	8
Aula de 3.º Ciclo	17
Aulas universitárias	18
Noticiários televisivos	22
Sessões do parlamento europeu	29

Também a extensão do fraseamento prosódico em relatos espontâneos e exposições escolares (previamente preparadas, mas sem recurso à leitura de materiais escritos), de adultos e adolescentes, surge como um indício do grau de planeamento entre os estilos de fala: há na fala preparada uma divisão em constituintes de maior extensão e regularidade (cf. linhas 3 e 4 do Quadro 1). Surge ainda como marca de uma menor unidade da fala preparada dos alunos: apesar da variação interfalantes, há na fala preparada do professor maior regularidade relativamente à extensão do fraseamento prosódico. Este tende a integrar mais palavras por constituinte do que os alunos, traçando assim blocos mais amplos e mais fortemente marcados do que na fala espontânea. Pode até especular-se sobre o facto de o discurso preparado libertar parcialmente o falante no que diz respeito ao conteúdo, favorecendo uma maior concentração no controlo da forma entoacional (cf. Mata 1999).

No resto desta secção, incluem-se na variação entre estilos de fala as diferenças entre diálogos espontâneos e aulas universitárias (Trancoso et al. 2008; Moniz et al. 2011; Moniz 2013; Moniz et al. 2014b), e as diferenças entre relatos espontâneos e apresentações orais escolares (Mata 1999; Moniz 2006; Moniz et al. 2014a; Mata et al. 2014b; Mata/Moniz/Batista 2014; 2016). Tal como referido acima, estas diferenças estão relacionadas com a quantidade de planeamento prévio envolvido nos tipos de produção, em função da clareza da informação que se pretende transmitir e do grau de interação entre falante e ouvinte(s). Concentramo-nos na forma como se manifestam nos padrões de organização prosódica, em termos de: (a) distribuição de eventos tonais (acentos nucleares e fronteiras tonais); (b) variação da gama de valores de f_0 e de energia; (c) distribuição de pausas silenciosas; (d) distribuição de eventos disfluências e padrões de fluência.

(a) Distribuição de eventos tonais (acentos nucleares e fronteiras tonais)¹¹

A frequência de uso de padrões entoacionais descendentes vs. ascendentes em enunciados declarativos varia entre estilos de fala, com diferenças entre grupos etários. É disso exemplo no português europeu a variação entre relatos espontâneos e exposições preparadas de adultos e adolescentes.

Nos adultos, há mais acentos nucleares de tipo baixo/descendente – principalmente (H+)L* – e fronteiras L% na exposição preparada; os acentos nucleares de tipo alto/ascendente – principalmente (L+)H* – e as fronteiras H% são mais frequentes e usados em proporções mais equilibradas no relato espontâneo. Esta distribuição das declarativas ascendentes é convergente com o que tem sido observado para outras línguas em monólogos espontâneos e narrativas espontâneas (Blaauw 1995; Wennerstrom 2001), e interpretado como efeito quer do uso de estratégias para assinalar a não-terminalidade dos enunciados (em geral correlacionadas com tons finais H e consideradas mais frequentes na fala espontânea) quer do uso de estratégias para envolver os ouvintes, apelando e mantendo a atenção destes (consideradas mais frequentes em atividades de fala de tipo narrativo).

Nos adolescentes, passa-se o inverso: os contornos nucleares de tipo baixo/descendente predominam nos enunciados declarativos do relato espontâneo; os contornos nucleares de tipo alto/ascendente aumentam nos enunciados declarativos da exposição preparada. Esta variação é independente do género dos falantes. Por conseguinte, não se confirma no português europeu, a afirmação corrente para outras línguas de que há mais declarativas ascendentes na fala espontânea dos adolescentes, em particular nas raparigas. A maior frequência de padrões ascendentes na fala espontânea dos adultos é um dos fatores que os distingue dos adolescents.

Refira-se a propósito que as estratégias de distribuição de tons ascendentes vs. descendentes entre situações comunicativas estão correlacionadas com a frequência de efeitos fonético-fonológicos comuns no português europeu em posição pré-fronteira, nomeadamente: inserção de [@], correlacionada com fronteiras H (cf. Figura 21), e efeitos de aspiração final (cf. Figura 22) são muito mais frequentes nas exposições preparadas; desvozeamento/apagamento de vogal postónica (cf. Figuras 18 e 19 acima) e efeitos de laringalização final, correlacionados com fronteiras L (cf. Figura 23), são mais comuns nos relatos espontâneos.

¹¹ Está ainda por concluir uma análise deste tipo de variação em diálogos espontâneos e aulas universitárias. Por essa razão, apresentamos nesta secção as principais diferenças entre relatos espontâneos e exposições preparadas.

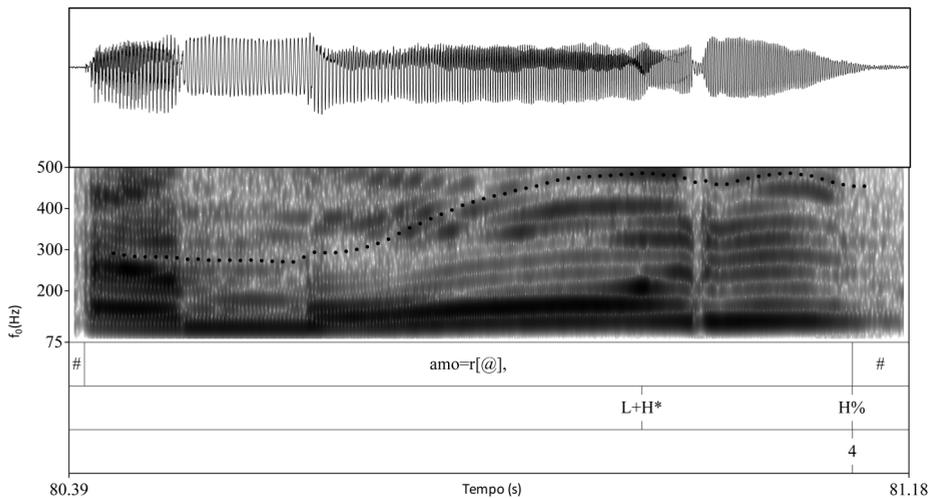


Figura 21: Inserção final de [ʔ] em *amor* (CPE-FACES).

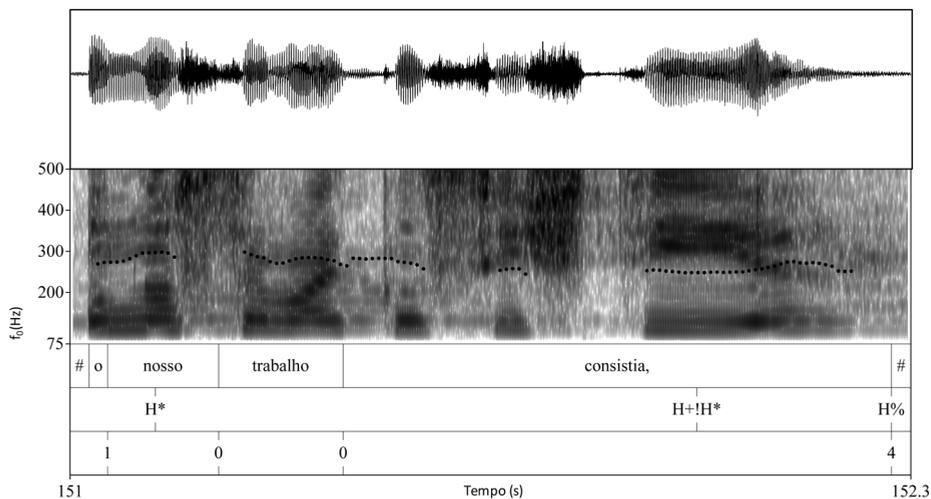


Figura 22: Efeitos de aspiração final em *consistia* (CPE-FACES).

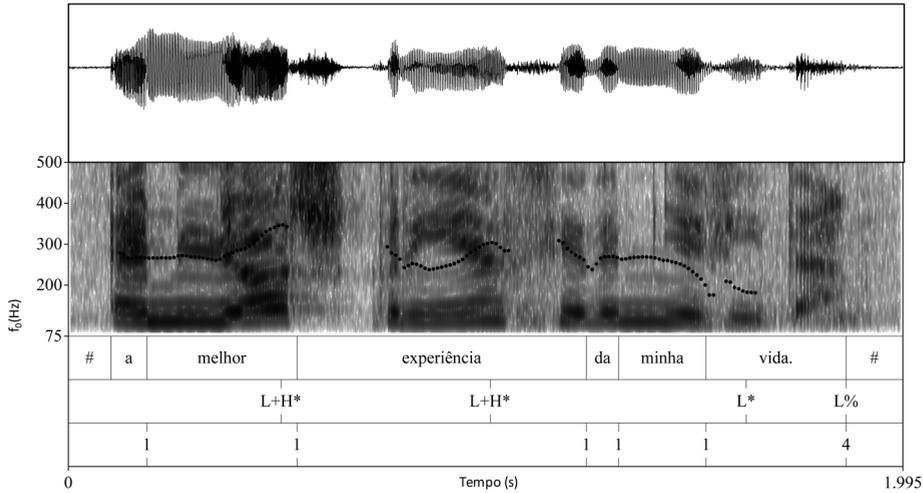


Figura 23: Efeitos de laringalização final em *vida* (CPE-FACES).

(b) Variação da gama de valores de f_0 e de energia

Também a variação da gama de valores de f_0 em que se inscrevem os contornos entoacionais e a variação dos valores de energia são condicionadas pelas circunstâncias das situações de comunicação. Declives de f_0 , valores de f_0 máxima, f_0 mínima e desvios padrão, bem como declives de energia encontram-se entre os parâmetros que mais concorrem para esse efeito. Os diálogos apresentam valores mais elevados de energia; as aulas universitárias valores mais elevados de f_0 . O mesmo tipo de contraste se observa entre relatos espontâneos e exposições preparadas de adultos.

Sabe-se igualmente que os falantes podem aplicar diferentes estratégias no uso destes parâmetros em diferentes estilos de fala, tanto em contextos fluentes como disfluentes, o que se associa a uma maior expressividade e carisma dos falantes. Por exemplo, nas aulas universitárias, o falante avaliado como mais expressivo (S6 no Quadro 2) é o que conjuga com mestria declives de f_0 e de energia mais elevados na reposição da fluência e maior controlo de diversas medidas temporais (velocidades de elocução e de articulação mais elevadas) que contribuem para a dinâmica da interação com os alunos, correlatos que têm sido associados a estratégias de fluência e de competência linguística dos falantes.

Quadro 2: Principais diferenças de f_0 e de energia por constituinte similar a frase em aulas universitárias (LECTRA).¹²

Parâmetros	Média para o falante S6	Média para os restantes falantes masculinos
\neq de f_0 da disfluência para a palavra seguinte (ST)	0,81	0,74
Declive de energia da palavra seguinte à disfluência	1,35	0,77
Média de energia em constituintes similares a frase (dB)	54,2	44,2

(c) Distribuição de pausas silenciosas

Na leitura, as pausas silenciosas ocorrem sobretudo em posições sintáticas hierarquicamente superiores, ou seja, em fronteiras de frase e oração; na fala espontânea, sobretudo entre e/ou dentro de constituintes sintáticos (Freitas 1990). A distribuição de pausas silenciosas não permite apenas distinguir leitura de fala espontânea. Nas aulas universitárias, há mais pausas silenciosas do que nos diálogos, o que se correlaciona, por um lado, com a dinâmica de um diálogo e, por outro, com a multifuncionalidade das pausas silenciosas nas aulas, e.g. dar a palavra, enfatizar informação, levar os alunos a refletirem sobre o tópico em análise antes da introdução de um novo tópico, mudar de tópico. Por outro lado ainda, a presença/ausência de interlocutores também afeta a distribuição de pausas silenciosas: numa aula via internet (cf. falante S5 na Figura 24) a ocorrência de pausas silenciosas tende a ser reduzida; o inverso acontece se o professor estruturar a sua aula em interação constante com os alunos, numa dinâmica similar à de um diálogo (cf. falante S6 na Figura 24).

O mesmo tipo de contraste se verifica quando se comparam exposições preparadas e relatos espontâneos produzidos quer por adultos quer pela maioria dos adolescentes. A maior frequência de pausas silenciosas nas exposições preparadas, em conjunto com a variação de outros parâmetros melódicos e temporais, pode ser interpretada como mais um indicador das diferenças de inteligibilidade entre estilos de fala.

¹² O corpus LECTRA é constituído por aulas universitárias (Trancoso et al. 2008; ISLRN 298-379-572-530-5).

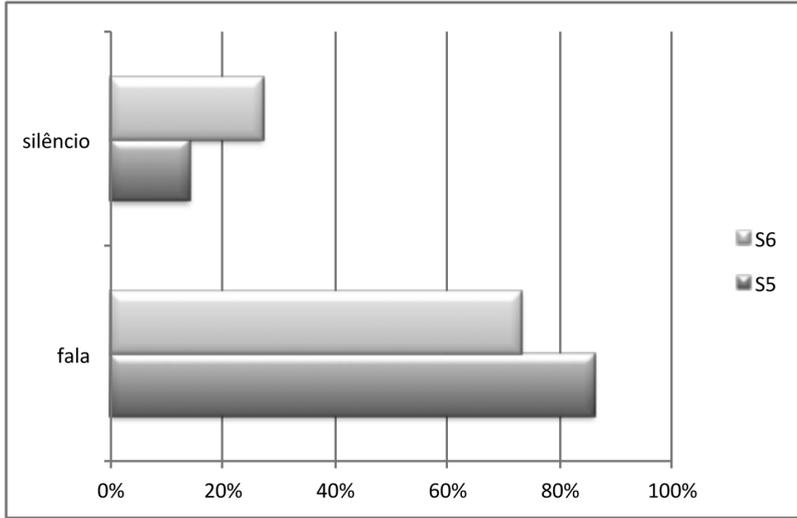


Figura 24: Percentagem de tempo de silêncio e de fala de dois professores universitários (S5 = aula online; S6 = aula presencial) (LECTRA).

(d) Distribuição de eventos disfluentes e padrões de fluência

Sabe-se para o português europeu (Moniz 2006; 2013) que os fenômenos ditos disfluentes, como pausas preenchidas, alongamentos, repetições, fragmentos, substituições, apagamentos, marcadores de edição e sequências complexas (cf. exemplos (1)–(8), a seguir), consoante as propriedades prosódicas que exibem, podem ser perceptivamente julgados como fluentes ou disfluentes.

- (1) <aa> Houve um marco fundamental na evolução da indústria informática (pausa preenchida, sem conteúdo lexical, essencialmente uma vogal central alongada)
- (2) regularização de existências e= o que é que aconteceu agora? (alongamento de material linguístico)
- (3) é preciso codificar <não é> não é direto (repetição de material linguístico)
- (4) porque não há <direç-> direcionalidade na relação (fragmento ou truncação de um item lexical)
- (5) que, aliás, <sai> saiu na vossa ficha (substituição de material linguístico)
- (6) <significa isto que se nós considerarmos nós temos aqui> Ah, e no fim, e no fim, diz aí que vocês tinham ainda um stock de cento e cinquenta traves (apagamento ou abandono de material linguístico e início de uma nova ideia)
- (7) esta cobra <que une que vai que não é que une> que vai depois (marcador de edição ou expressão de reformulação explícita)
- (8) O ano passado houve uns colegas vossos da matemática que <fizeram o projeto só qua-> queriam fazer o projeto quase só com strings. (sequência complexa, i.e., combinatórias de mais do que um tipo de disfluência)

Numa tarefa de classificação de disfluências como mecanismos de fluência/disfluência (com recurso a Árvores de Regressão e Classificação), as propriedades prosódicas são as pistas linguísticas mais salientes – nomeadamente, o fraseamento prosódico e os contornos de f_0 . Os eventos que são produzidos após uma fronteira de constituinte prosódico 3 ou 4 e que se associam a contornos *plateau* ou ascendentes são expressivamente classificados como fluentes (cf. Figura 25), por oposição aos que são produzidos no interior de um constituinte ou formam por eles próprios um constituinte, exibindo contornos descendentes ou efeitos de glotalização (cf. Figura 26). Poder-se-á reafirmar, tal como para outras línguas, que os falantes controlam diferentes aspetos segmentais e suprasegmentais e que parecem fazê-lo de uma forma cirúrgica, i.e., ajustando-os aos contextos adjacentes. Os resultados suportam a hipótese de que existem diferentes graus de mestria no uso das disfluências e sugerem a possibilidade de estes graus poderem revelar, ou estar associados, a diferentes níveis de conhecimento explícito da língua.

A comparação intracorpora (relatos espontâneos vs. exposições preparadas não-lidas) aponta para uma maior frequência de eventos disfluentes na fala espontânea dos adultos, em contraste com uma maior frequência de eventos disfluentes na fala preparada dos adolescentes, o que pode ser associado a graus de adaptação a uma tarefa especificamente escolar. Como na exposição preparada o esforço de monitorização e planeamento por parte dos falantes é significativamente maior, é nesta situação que são produzidas mais disfluências. Os professores, porém, demonstram ter um comportamento diferente, produzindo mais pausas preenchidas na situação de relato espontâneo e com uma percentagem de tempo disfluyente superior nesta situação comunicativa, o que sugere que estes falantes evidenciam maior habituação e treino de exposição de conteúdos do que propriamente no relato de experiências pessoais.

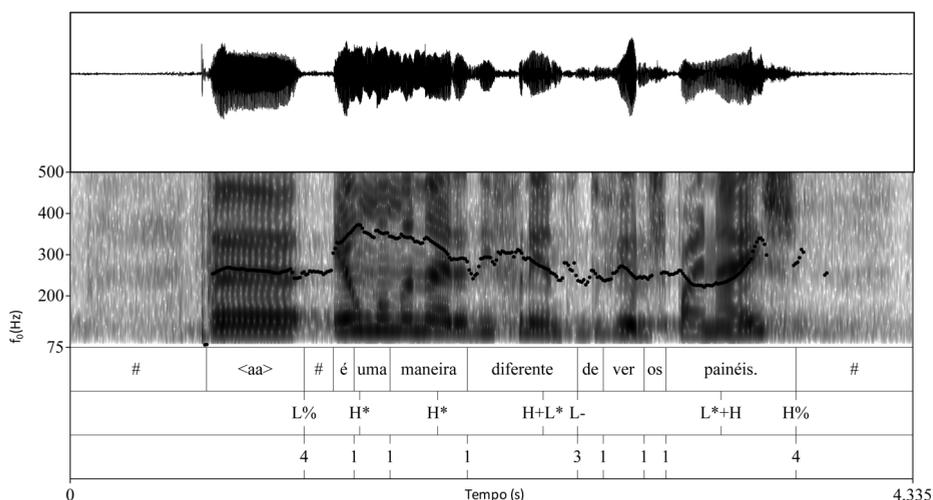


Figura 25: Exemplo classificado perceptivamente como fluente (CPE-FACES).

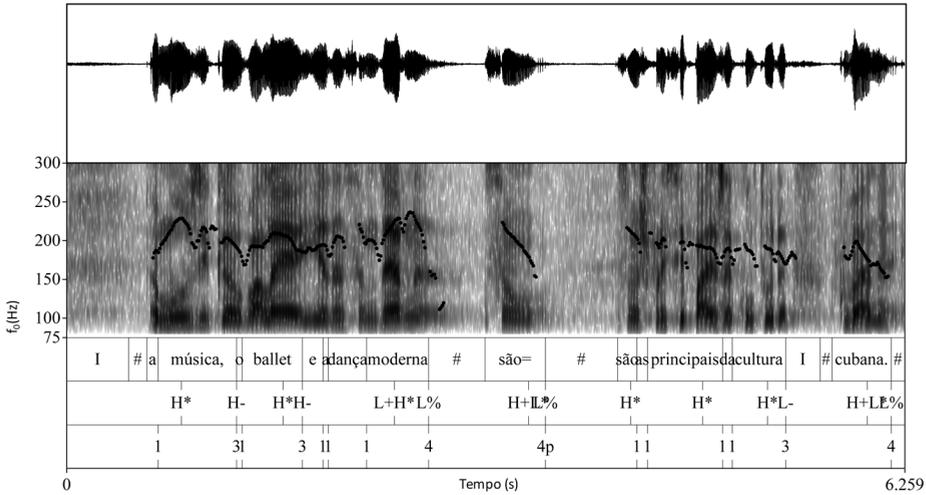


Figura 26: Exemplo classificado perceptivamente como disfluido (CPE-FACES).

A comparação intercorpora aponta efeitos de estilo nos padrões temporais e mesmo na marcação prosódica por contraste através de um aumento dos valores de energia e de f_0 da disfluência para a reposição da fluência na palavra seguinte, entre aulas universitárias e diálogos. Embora as pausas preenchidas sejam a categoria mais representativa em ambos os corpora, as restantes categorias apresentam uma distribuição distinta. Nas aulas, as sequências complexas (e.g. repetições e substituições em conjunto, utilizadas para procura/precisão lexical) são mais frequentes do que as repetições por si só, enquanto nos diálogos ambas têm distribuições similares. Nos diálogos, os fragmentos correspondem a mais do dobro dos fragmentos produzidos nas aulas e os apagamentos são residuais. Estas diferenças na distribuição das categorias disfluentes podem ser interpretadas em função da natureza dos diálogos em análise, nomeadamente das restrições temporais a que os falantes estão sujeitos em diálogos para indicação de percursos em mapas, com recurso mais frequente a categorias como repetições e fragmentos e menos a sequências complexas e apagamentos. Os padrões temporais também apontam para a natureza mais dinâmica dos diálogos por comparação com as aulas, com produção de menos palavras, tanto em frases fluentes como em frases que contêm disfluências. O encadeamento das interações comunicativas num diálogo está sujeito a restrições temporais, evidentes também na duração dos silêncios, na disfluência e nos próprios contextos adjacentes. Uma vez mais, todas as unidades referidas são mais breves nos diálogos do que nas aulas. Tanto os diálogos como as aulas universitárias exibem aumentos de f_0 e de energia (marcação prosódica por contraste) entre a disfluência e a reposição da fluência para a maioria das categorias disfluentes, embora com diferentes graus de contraste. Porém, mesmo a referida estratégia de marcação prosódica por contraste da disfluência para a reposição da fluência está sujeita a variação intercorpora, sendo

esta marcação mais forte nas aulas que nos diálogos. Nas aulas, pistas de f_0 e de energia são produzidas por todos os falantes, para a maioria das categorias, tanto para as disfluências como para os contextos adjacentes. O conjunto de padrões apresentado é um contributo para a diferenciação entre estilos de fala, nomeadamente entre fala espontânea e fala preparada não-lida.

4 Prosódia e processamento automático

No que respeita à interface com o processamento automático de fala, uma das áreas mais produtivas nos últimos anos, a transversalidade da análise prosódica tem tido efeito no renovar de metodologias diversas de processamento da informação prosódica, contribuindo, por um lado, para uma sistematização do ponto de vista linguístico e, por outro, para o melhoramento dos sistemas de síntese e de reconhecimento de fala.

O enriquecimento automático de transcrições de fala com um conjunto de eventos designados no inglês *structural metadata events* (Liu et al. 2006; Ostendorf et al. 2008) – i.e., sinais de pontuação e maiúsculas em fronteira de frase, bem como disfluências e outros marcadores discursivos – contribui significativamente para a legibilidade de uma sequência de palavras obtida através de um reconhecedor automático de fala. Este processo é crucial para diversas aplicações (e.g. legendagem automática de noticiários televisivos, estruturação de tópicos e sumário de notícias para pesquisa, exercícios para a aprendizagem de L1/L2 mediada por computador) e para a própria arquitetura do reconhecedor. Frequentemente, o reconhecedor de fala está integrado numa sequência de módulos que incluem segmentar áudio, grafar a maiúsculas e minúsculas, identificar zonas de disfluência, sumarizar áudio. Por conseguinte, o enriquecimento de transcrições de fala pode ser entendido como uma tarefa de estruturação de uma sequência de palavras em diferentes unidades linguísticas, ou seja, como uma estruturação multilinear, abrangendo módulos da gramática distintos.

Para este processo tem sido utilizado um conjunto alargado de pistas, reflexo da estruturação multilinear acima referida. Estas pistas estão muito para além das lexicais (n-gramas de palavras, ou seja uma ou mais palavras) extraídas da saída do reconhecedor, ou das acústicas extraídas do módulo de processamento de áudio (identificação de segmentos correspondentes a fala vs. outros tipos de segmentos nos quais se incluem silêncios, música, *inter alia*; identificação de falantes, etc.). No trabalho desenvolvido para o português, à semelhança do que acontece para outras línguas, tem sido dado especial destaque às pistas prosódicas.

O enriquecimento de transcrições automáticas depende de: i) transcrições manuais, ii) transcrições produzidas pelo reconhecedor automático e iii) análise do sinal acústico (para mais informações sobre este processo, veja-se Batista et al. 2012a e 2012b; Moniz 2013; Moniz et al. 2014b; Cabarrão et al. 2015). A anotação manual é

complexa sobretudo no caso da anotação de corpora de fala espontânea, uma vez que implica pontuar fala, por um lado, e identificar disfluências e marcadores discursivos, por outro – tarefas que, como se sabe, afetam a concordância interanotadores em distintos corpora do português europeu (cf. Batista 2011 e Cabarrão et al. 2014 para um corpus de notícias televisivas; Moniz 2013 para um de aulas universitárias). A saída do reconhecedor já identifica e delimita temporalmente fones, sílabas, palavras e silêncios. Com base nesta e na informação contida nas transcrições manuais, as sequências são organizadas numa estruturação hierárquica que contempla ainda unidades similares a frases e atos de fala. A partir desta estruturação, são extraídas pistas acústicas segmentais e suprasegmentais, a saber, valores mínimos, máximos, médias, medianas, desvios padrão e declives de f_0 e de energia, bem como a duração de cada uma das unidades. Os valores absolutos dos declives de f_0 e de energia são convertidos em 3 categorias básicas (ascendente, descendente e *plateau*) e em 6 combinatórias (e.g. ascendente-descendente, ascendente-*plateau*). Rácios de duração de unidades e de silêncios adjacentes são também calculados com base em três valores: maior, igual a, menor.

As experiências já realizadas para o português europeu demonstram que sinais de pontuação, tipos de disfluência, reposição de fluência e marcadores discursivos, em geral, podem ser preditos e caracterizados com recurso apenas à informação prosódica.

Quanto aos sinais de pontuação, na predição de pontos finais destacam-se as seguintes pistas: i) contorno descendente na palavra antes de um ponto final; ii) nível estacionário de energia na mesma palavra; iii) duração relativa entre essa palavra e a seguinte; e iv) grau superior de confiança em relação à palavra seguinte. Este conjunto de pistas é ilustrativo do comportamento de uma declarativa neutra no português europeu (cf. secção 2). Os pontos de interrogação, por sua vez, são caracterizados por dois padrões diferenciados: i) contorno de f_0 ascendente na palavra antes de um ponto de interrogação e declive de energia ascendente nessa e na palavra seguinte; ii) contorno de f_0 estacionário na palavra antes de um ponto de interrogação e declive de energia descendente nessa mesma palavra. Os pontos de interrogação surgem, assim, como uma classe difícil de detetar (e.g. Moniz et al. 2011; Margolis 2011), por poderem corresponder a distintos padrões prosódicos em função dos tipos de interrogativas, como se viu na secção 2. Quanto às vírgulas, nas experiências até agora realizadas para o português europeu, tal como para outras línguas, elas são o evento mais difícil de predizer prosodicamente. Tal como expetável, são sobretudo classificadas com base em pistas morfossintáticas, não sendo claramente desambiguadas por meio de pistas prosódicas.

As pistas que mais contribuem para a predição da reposição de fluência a seguir a uma sequência disfluente integram: i) duas palavras contíguas idênticas (única informação lexical, passível de ser identificada acusticamente); ii) subida dos níveis de f_0 e de energia na palavra que inicia uma reposição de fluência e um contorno estacionário na palavra anterior; iii) grau de confiança dado pelo reco-

nhecedor sobre a palavra que inicia a reposição superior ao da disfluência propriamente dita.

Experiências mais recentes (Cabarrão et al. 2015) sobre disfluência e marcadores discursivos em início de enunciado mostram que aquelas são mais bem previstas do que estes. Para isso concorre a diversidade de padrões prosódicos associados a tipos específicos de marcadores discursivos, nomeadamente, desacentuação vs. acentuação, com variação de nível e gama de f_0 (tons H escalonados acima de outros tons H subsequentes; contornos *plateau* em espaço tonal intermédio relativamente aos constituintes subsequentes). Por exemplo, o marcador conversacional «agora» é habitualmente associado a um acento de tipo H+L e fraseado separadamente, formando por si só um constituinte entoacional, provavelmente para indicar uma mudança do tópico discursivo, por oposição a «portanto», que é geralmente desacentuado. Está em curso uma análise mais aprofundada da relação dos padrões prosódicos com a estrutura do discurso e o estatuto informacional destes marcadores.

Para além do enriquecimento de transcrições automáticas, outras linhas de investigação têm vindo a ser desenvolvidas: a predição de proeminência prosódica e a deteção de emoções. A primeira com o objetivo de adicionar ao módulo prosódico informação relativa a proeminência prosódica (acentuação/desacentuação) com base no sistema ToBI. A segunda para um melhor conhecimento dos eventos paralinguísticos de modo a aferir o conteúdo emocional das interações comunicativas e a sua variação social.

Para treinar um modelo de proeminência prosódica do português, foi usado um corpus de fala espontânea e preparada (não lida) de adultos e adolescentes. Adicionalmente, foram utilizados modelos do inglês americano para aferir da validade de identificação de proeminência prosódica entre línguas. Os modelos do inglês permitem prever proeminência prosódica com 71% de acerto nos dados do português e os modelos treinados especificamente com dados do português uma taxa de 77% de acerto, o que representa uma subida de 6% absolutos. Estes resultados demonstram que os eventos proeminentes são detetados com uma percentagem considerável de acerto de uma língua para a outra, como se exemplifica na Figura 27. Demonstram ainda um efeito de idade/estatuto dos falantes, uma vez que a predição de proeminência é sempre melhor nos adultos do que nos adolescentes. Esta linha de investigação sobre predição de proeminência pode, assim, contribuir para a discussão de correlatos acústico-prosódicos universais vs. específicos de uma língua ou para avaliar a variação dependente do domínio de fala e de variáveis sociolinguísticas.

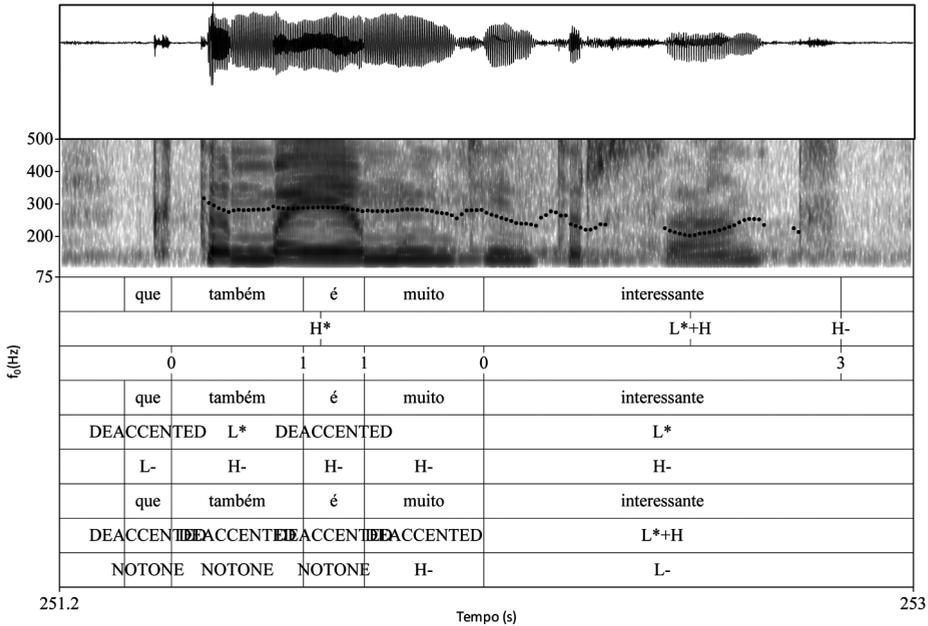


Figura 27: Exemplo de detecção automática de proeminência com o AuToBI. As primeiras 3 fiadas correspondem à transcrição manual; as 3 seguintes à aplicação do modelo do inglês ao português; as 3 últimas à aplicação do modelo treinado especificamente para o português.

No âmbito da detecção de emoções para o português europeu, o stress é o que tem merecido maior atenção, uma vez que o medo de falar em público (glossofobia) é uma das fobias mais comuns nos adultos (Miller/Stone 2009). O principal objetivo é o desenvolvimento de aplicações para o treino de fala em público, com monitorização dos níveis de stress.

Para a predição de níveis de stress foi usado um corpus de fala com ocorrência natural, constituído por apresentações de teses de mestrado em provas públicas (Aguiar et al. 2014). Concorre para essa predição um vasto conjunto de pistas acústico-prosódicas. Entre as pistas prosódicas mais informativas destacam-se as seguintes: parâmetros de qualidade da voz associados a voz laringalizada, nomeadamente *jitter* (perturbações de f_0) e *shimmer* (perturbações de energia) ao longo dos constituintes similares a frase, e movimentos finais de f_0 sistematicamente mais baixos do que o exetável, constituinte após constituinte (cf. Julião et al. 2015). Sabe-se que também em exposições escolares de adolescentes, o stress ou a insegurança se associam a um nível de registo substancialmente mais baixo e a uma redução da gama de f_0 , o que se traduz numa nítida impressão de falta de controlo da situação (Mata 1999).

A detecção de emoções é uma área de especial interesse para interações comunicativas com agentes virtuais ou robôs, na área da robótica social, ou para sistemas de diálogo no domínio da *Intelligent Voice Response* (IVR), ou para a síntese de fala, ou

ainda para o desenvolvimento de modelos de compreensão no âmbito do reconhecimento de fala. Os detetores de emoções desenvolvidos para o português europeu têm sido integrados numa plataforma web designada *SPeech Analytics* (SPA), desenvolvida pelo INESC-ID (Batista et al. 2016).

5 Conclusão

Neste capítulo procurámos integrar algumas das formas e funções da variação da entoação no português europeu, considerando ainda as funções de outros recursos prosódicos em diferentes situações de comunicação. Partindo de funções clássicas da variação entoacional (e.g. declarativas vs. interrogativas) que têm sido objeto de maior número de estudos, destacámos os recursos prosódicos que podem desempenhar funções pragmático-discursivas (e.g. em diferentes tipos de perguntas e respostas; mecanismos de disfluência e reposição de fluência; marcadores discursivos; expressão de emoções), estilísticas (fala espontânea vs. preparada e diferenças entre tipos de fala espontânea) e sociolinguísticas (e.g. adultos vs. adolescentes; falantes masculinos vs. femininos) no português europeu, com base em excertos de corpora de natureza mais/menos expositiva ou dialógica.

Para o português europeu, são escassos os estudos de prosódia em contexto de interação. São também escassos os estudos comparativos de prosódica interestilos e interfalantes, em função de fatores como idade e género. Em trabalho futuro, deverá alargar-se a análise da variação prosódica a mais situações de interação espontânea (humano-humano, humano-agente virtual e humano-robô), comparando diferentes grupos etários (crianças, adolescentes, adultos e idosos) e abrangendo mais fatores de variação (paralinguísticos, sociais e estilísticos). Esta área, embora complexa, é crucial para um conhecimento mais aprofundado das propriedades de estruturação do diálogo e para a integração destas em sistemas de diálogo, por exemplo, contribuindo para modelos de compreensão no processamento automático de fala.

6 Referências

- Aguiar, Ana, et al. (2014), *VOCE Corpus: Ecologically Collected Speech Annotated with Physiological and Psychological Stress Assessments*, in: Nicoletta Calzolari et al. (edd.), *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, Reykjavik, European Language Resources Association (ELRA), 1568–1574.
- Andrade, Amália (1990), *Aspectos fisiológicos e aerodinâmicos da produção da fala*, in: Maria Helena Mira Mateus et al. (edd.), *Fonética, Fonologia e Morfologia do Português*, Lisboa, Universidade Aberta, 57–96.
- Avesani, Cinzia/Hirschberg, Julia/Prieto, Pilar (1995), *The intonational disambiguation of potentially ambiguous utterances in English, Italian, and Spanish*, in: Kjell Elenius/Peter Branderud (edd.),

- Proceedings of the 13th International Conference of Phonetic Sciences 1995*, vol. 1, Stockholm, International Phonetic Association, 174–177, <https://www.internationalphoneticassociation.org/icphs/icphs1995> (28.03.2016).
- Barry, William (1995), *Phonetics and phonology of speaking styles*, in: Kjell Elenius/Peter Branderud (edd.), *Proceedings of the 13th International Conference of Phonetic Sciences 1995*, vol. 2, Stockholm, International Phonetic Association, 4–10, <https://www.internationalphoneticassociation.org/icphs/icphs1995> (28.03.2016).
- Batista, Fernando (2011), *Recovering capitalization and punctuation marks in speech transcripts*, tese de Doutorado, Instituto Superior Técnico.
- Batista, Fernando, et al. (2012a), *Bilingual Experiments on Automatic Recovery of Capitalization and Punctuation of Automatic Speech Transcripts*, *IEEE Transactions on Audio, Speech, and Language Processing* 20:2, 474–485, <http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=6099652&punumber=10376> (28.03.2016).
- Batista, Fernando, et al. (2012b), *Extending automatic transcripts in a unified data representation towards a prosodic-based metadata annotation and evaluation*, *Journal of Speech Sciences* 2, 115–138.
- Batista, Fernando, et al. (2016), *SPA: Web-based platform for easy access to speech processing modules*, in: Nicoletta Calzolari et al. (edd.), *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, Portorož, European Language Resources Association (ELRA), 3886–3892.
- Beckman, Mary/Pierrehumbert, Janet (1986), *Intonational structure in Japanese and English*, *Phonology Yearbook* 3, 255–309.
- Blaauw, Eleanor (1995), *On the perceptual classification of spontaneous and read speech*, Utrecht, LEU.
- Boersma, Paul/Weenink, David (2013), *Praat: doing phonetics by computer* [Computer program], version 5.3.56, <http://www.praat.org/> (15.09.2013).
- Bruce, Gösta (1977), *Swedish word accents in sentence perspective*, Lund, Gleerup.
- Cabarrão, Vera (2013), *Respostas afirmativas em diálogos espontâneos em português europeu (interface prosódia-sintaxe-discurso)*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Cabarrão, Vera, et al. (2014), *Revising the Annotation of a Broadcast News Corpus: a Linguistic Approach*, in: Nicoletta Calzolari et al. (edd.), *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, Reykjavik, European Language Resources Association (ELRA), 3908–3913.
- Cabarrão, Vera, et al. (2015), *Prosodic classification of discourse markers*, in: The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the International Congress of Phonetic Sciences 2015*, Glasgow UK, the University of Glasgow, artigo 634, 1–5, <http://www.icphs2015.info/pdfs/Papers/ICPHS0634.pdf> (05.04.2016).
- Clopper, Cynthia/Smiljanic, Rajka (2011), *Effects of gender and regional dialect on prosodic patterns in American English*, *Journal of Phonetics* 39, 237–245.
- Cruz, Marisa (2013), *Prosodic variation in EP: phrasing, intonation and rhythm in Central-Southern varieties*, tese de Doutorado, Universidade de Lisboa.
- Cucchiari, Catia/Strik, Helmer/Boves, Lou (2002), *Quantitative assessment of second language learner's fluency: comparisons between read and spontaneous speech*, *Journal of the Acoustical Society of America* 111, 2862–2873.
- Delgado-Martins, Maria Raquel/Freitas, Maria João (1991), *Temporal structures of speech: reading news on TV*, in: *Proceedings of the ESCA Workshop on Phonetics and Phonology of Speaking Styles: Reduction and Elaboration in Speech*, Barcelona, European Speech Communication Association (ESCA), artigo 19, 1–5.

- Escudero, David, et al. (2012), *Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system*, *Speech Communication* 54, 566–582.
- Eskénazi, Maxine (1993), *Trends in Speaking Styles Research*, in: *Proceedings of the 3rd European Conference on Speech Communication and Technology (Eurospeech '93)*, Berlin, ISCA Archive, 501–505, http://www.isca-speech.org/archive/eurospeech_1993 (28.03.2016).
- Falé, Isabel (1995), *Fragmento da Prosódia do Português Europeu: as Estruturas Coordenadas*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Falé, Isabel (2005), *Percepção e reconhecimento da informação entoacional em Português Europeu*, tese de Doutoramento, Universidade de Lisboa.
- Falé, Isabel/Faria, Isabel (2006), *Categorical perception of intonational contrasts in European Portuguese*, in: Ralph Hoffman/Hans Mixdorff (edd.), *Proceedings of Speech Prosody*, Dresden, TUDpress Verlag der Wissenschaften, 69–72.
- Falé, Isabel/Faria, Isabel (2007), *Imperatives, orders and requests in European Portuguese intonation*, in: *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Universität des Saarlandes, artigo 1452, 1041–1044, <http://www.icphs2007.de/> (28.03.2016).
- Freitas, Maria João (1990), *Estratégias de organização temporal do discurso em português*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Frota, Sónia (2000), *Prosody and Focus in European Portuguese, Phonological Phrasing and Intonation*, New York, Garland Publishing.
- Frota, Sónia (2002), *Nuclear falls and rises in European Portuguese: a phonological analysis of declarative and question intonation*, *Probus* 14, 113–146.
- Frota, Sónia (2014), *The intonational phonology of European Portuguese*, in: Sun-Ah Jun (ed.), *Prosodic typology II*, Oxford, Oxford University Press, 6–42.
- Frota, Sónia/Cruz, Marisa (coord.) (2012-2015), *Interactive Atlas of the Prosody of Portuguese*, <http://labfon.letras.ulisboa.pt/InAPoP/> (18.04.2016).
- Frota, Sónia, et al. (2007), *The phonetics and phonology of intonational phrasing in Romance*, in: Pilar Prieto/Joan Mascaró/Maria-Josep Solé (edd.), *Prosodic and Segmental Issues in (Romance) Phonology*, Amsterdam/Philadelphia, Benjamins, 131–153.
- Frota, Sónia, et al. (2015), *Intonational variation in Portuguese: European and Brazilian varieties*, in: Sónia Frota/Pilar Prieto (edd.), *Intonation in Romance*, Oxford, Oxford University Press, 235–283.
- Grabe, Esther (1998), *Comparative Intonational Phonology: English and German*, tese de Doutoramento, Universidade de Nijmegen (publ. MPI).
- Grabe, Esther, et al. (2000), *Pitch accent realisation in four varieties of British English*, *Journal of Phonetics* 28, 161–185.
- Grice, Martine, et al. (1995), *The intonation of queries and checks across languages: data from map tasks dialogues*, in: Kjell Elenius/Peter Branderud (edd.), *Proceedings of the 13th International Conference of Phonetic Sciences 1995*, vol. 3, Stockholm, International Phonetic Association, 648–651, <https://www.internationalphoneticassociation.org/icphs/icphs1995> (28.03.2016).
- Grønnum, Nina/Viana, Maria do Céu (1999), *Aspects of European Portuguese intonation*, in: John J. Ohala et al. (edd.), *Proceedings of the 14th International Conference of Phonetic Sciences 1999*, San Francisco, University of California, 1997–2000, <http://www.internationalphoneticassociation.org/icphs/icphs1999> (28.03.2016).
- Hirschberg, Julia (2000), *A corpus-based approach to the study of speaking style*, in: Merle Home (ed.), *Prosody: theory and experiment, Festschrift in honor of Gösta Bruce*, Dordrecht, Kluwer, 335–350.
- Julião, Mariana, et al. (2015), *Speech features for discriminating stress using branch and bound wrapper search*, in: José-Luis Sierra-Rodríguez/José Paulo Leal/Alberto Simões (edd.), *Languages, Applications and Technologies, 4th International Symposium (SLATE'15), Madrid*, Berlin, Springer, 3–14.

- Jun, Sun-Ah (2005), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford, Oxford University Press.
- Ladd, Robert (1996), *Intonational Phonology*, Cambridge, Cambridge University Press.
- Liu, Yi, et al. (2006), *Enriching speech recognition with automatic detection of sentence boundaries and disfluencies*, IEEE Transactions on Audio, Speech, and Language Processing 14:5, 1526–1540, <http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=35293&pnumber=10376> (28.03.2016).
- Margolis, Anna (2011), *Automatic annotation of spoken language using out-of-domain resources and domain adaptation*, tese de Doutorado, Universidade de Washington.
- Mata, Ana Isabel (1990), *Questões de entoação e interrogação no Português. Isso é uma pergunta?*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Mata, Ana Isabel (1999), *Para o estudo da entoação em fala espontânea e preparada no Português Europeu: metodologia, resultados e implicações didáticas*, tese de Doutorado, Universidade de Lisboa.
- Mata, Ana Isabel/Moniz, Helena/Batista, Fernando (2014), *Comparing phrase-final patterns across speech styles and age groups in European Portuguese*, Nouveaux Cahiers de Linguistique Française 31, 171–176.
- Mata, Ana Isabel/Moniz, Helena/Batista, Fernando (2016), *Stylistic variation in the intonation of European Portuguese teenagers and adults*, in: Meghan Armstrong/Nicholas Henriksen/Maria del Mar Vanrell (edd.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields*, Amsterdam/Philadelphia, Benjamins, 45–68.
- Mata, Ana Isabel/Santos, Ana Lúcia (2010), *On the intonation of confirmation seeking requests in child-directed speech*, in: Mark Hasegawa-Johnson (ed.), *Proceedings of Speech Prosody 2010*, artigo 100118, Chicago, 1–4, <http://speechprosody2010.illinois.edu/papers/100118.pdf> (28.03.2016).
- Mata, Ana Isabel, et al. (2014a), *Prosodic, syntactic, semantic guidelines for topic structures across domains and corpora*, in: Nicoletta Calzolari et al. (edd.), *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, Reykjavik, European Language Resources Association (ELRA), 1188–1193.
- Mata, Ana Isabel, et al. (2014b), *Teenage and adult speech in school context: building and processing a corpus of European Portuguese*, in: Nicoletta Calzolari et al. (edd.), *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, Reykjavik, European Language Resources Association (ELRA), 3914–3919.
- Meinedo, Hugo/Neto, João Paulo (2003), *Audio segmentation, classification and clustering in a broadcast news task*, in: *Proceedings ICASSP '03: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, II-5–8, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1202280&isnumber=27072> (28.03.2016).
- Miller, Tim/Stone, Dan (2009), *Public speaking apprehension (psa), motivation, and affect among accounting majors: A proof-of-concept intervention*, Issues in Accounting Education 24, 265–298.
- Moniz, Helena (2006), *Contributo para a caracterização dos mecanismos de (dis)fluência em português europeu*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Moniz, Helena (2013), *Processing Disfluencies in European Portuguese*, tese de Doutorado, Universidade de Lisboa.
- Moniz, Helena/Trancoso, Isabel/Mata, Ana Isabel (2010), *Disfluencies and the perspective of prosodic fluency*, in: Anna Esposito et al. (edd.), *Development of multimodal interfaces: active listening and synchrony*, Berlin/Heidelberg, Springer, 382–396.
- Moniz, Helena, et al. (2011), *Analysis of interrogatives in different domains*, in: Anna Esposito et al. (edd.), *Towards Autonomous, Adaptive, and Context-Aware Multimodal Interfaces: Theoretical and Practical Issues*, Berlin/Heidelberg, Springer, 136–148.

- Moniz, Helena, et al. (2014a), *Extending AuToBI to prominence detection in European Portuguese*, in: Nick Campbell/David Gibbon/Daniel Hirst (edd.), *Proceedings of Speech Prosody 7, 2014*, Dublin, Trinity College, 280–284, <http://www.speechprosody2014.org/> (28.03.2016).
- Moniz, Helena, et al. (2014b), *Speaking style effects in the production of disfluencies*, *Speech Communication* 65, 20–35.
- Moniz, Helena, et al. (2016), *Towards automatic language processing and intonational labeling in European Portuguese*, in: Meghan Armstrong/Nicholas Henriksen/Maria del Mar Vanrell (edd.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields*, Amsterdam/Philadelphia, Benjamins, 227–248.
- Moutinho, Lurdes/Coimbra, Rosa (2014), *Variation prosodique dans les interrogatives totales du Portugais Européen continental*, in: Yolanda Congosto/Maria Luísa Montero/Antonio Salvador (edd.), *Fonética Experimental, Educación Superior e Investigación*, vol. 3, Madrid, Editorial Arco/Libros, 153–170.
- Moutinho, Lurdes/Coimbra, Rosa (2015), *As ondas da fala através do Atlântico. Um estudo acústico*, in: António Manuel Ferreira/Maria Fernanda Brasete (edd.), *Pelos Mares da Língua Portuguesa 2*, Aveiro, Universidade de Aveiro, 529–537.
- Neto, João Paulo, et al. (2003), *A system for selective dissemination of multimedia information resulting from the ALERT project*, in: *Proceedings of the 2003 ISCA Workshop on Multilingual Spoken Document Retrieval (MSDR 2003)*, Hong Kong, ISCA Archive, 25–30, http://www.isca-speech.org/archive_open/msdr2003 (28.03.2016).
- Ostendorf, Mari, et al. (2008), *Speech segmentation and spoken document processing*, *IEEE Signal Processing Magazine* 25:3, 59–69, <http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=79> (28.03.2016).
- Pierrehumbert, Janet (1980), *The Phonology and Phonetics of English Intonation*, tese de Doutoramento, Massachusetts Institute of Technology.
- Pitrelli, John/Beckman, Mari/Hirschberg, Julia (1994), *Evaluation of prosodic transcription labeling reliability in the ToBI framework*, in: *Proceedings of the 3rd International Conference on Spoken Language Processing (ICSLP 94)*, vol. 1, Yokohama, ISCA Archive, 123–126, http://www.isca-speech.org/archive/icslp_1994 (28.03.2016).
- Price, Patti, et al. (1991), *The use of prosody in syntactic disambiguation*, *Journal of the Acoustical Society of America* 90:6, 2956–2970.
- Santos, Vinícius/Fernandes-Svartman, Flaviane (2014), *O padrão entoacional neutro do português de Guiné-Bissau: uma comparação preliminar com o português brasileiro*, *Estudos Linguísticos* 43:1, 48–63.
- Silverman, Kim, et al. (1992), *TOBI: a standard for labeling English prosody*, in: *Proceedings of the Second International Conference on Spoken Language Processing (ICSLP 92)*, Banff, Univ. of Alberta/ISCA Archive, 867–870, http://www.isca-speech.org/archive/icslp_1992/index.html (28.03.2016).
- Trancoso, Isabel, et al. (1998), *Corpus de Diálogo CORAL*, in: Vera Lúcia Strobe de Lima (ed.), *PROPOR'98, Actas do III Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada*, Porto Alegre, Brasil, 99–108, <http://www.speech.inesc.pt/documents/papers/Trancoso98b.pdf> (28.03.2016).
- Trancoso, Isabel, et al. (2008), *The LECTRA Corpus – Classroom Lecture Transcriptions in European Portuguese*, in: Nicoletta Calzolari et al. (edd.), *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, European Language Resources Association (ELRA), 1416–1420, <http://www.lrec-conf.org/proceedings/lrec2008/> (28.03.2016).
- Vaissièrre, Jacqueline (1983), *Language-independent prosodic features*, in: Anne Cutler/D. Robert Ladd (edd.), *Prosody: Models and Measurements*, Berlin, Springer, 53–66.

- Viana, Maria do Céu (1987), *Para a Síntese da Entoação do Português*, dissertação da Carreira de Investigação, Universidade de Lisboa.
- Viana, Maria do Céu, et al. (2007), *Towards a P_ToBI*, apresentação no *Workshop on the Transcription of Intonation in Ibero-Romance*, Universidade do Minho, Portugal, <http://labfon.lettras.ulisboa.pt/SonseMelodias/PaPI2007ToBIWorkshop/> (28.03.2016).
- Vigário, Marina (2003), *The Prosodic Word in European Portuguese*, Berlin/New York, Mouton De Gruyter.
- Vigário, Marina/Frota, Sónia (2003), *The intonation of Standard and Northern European Portuguese: a comparative intonational phonology approach*, *Journal of Portuguese Linguistics* 2:2 (special issue on Portuguese Phonology edited by W. Leo Wetzels), 115–137.
- Wells, John C. (1997), *SAMPA computer readable phonetic alphabet*, in: Dafydd Gibbon/Roger Moore/Richard Winski (edd.), *Handbook of Standards and Resources for Spoken Language Systems*, Part IV, section B, Berlin/New York, Mouton de Gruyter, 684–732.
- Wennerstrom, Ann (2001), *The music of everyday speech: prosody and discourse analysis*, New York, Oxford University Press.

Maria Antónia Mota

6 Morfologia nas interfaces

Abstract: Neste texto, tomam-se como objeto da análise morfológica o lexema, palavra abstrata, a que está associada uma matriz de traços que o caracteriza, e o(s) paradigma(s) de palavras morfossintáticas subsumidas pelo lexema. Apresentam-se exemplos que visam evidenciar que, para a análise das palavras do português, interessa considerar as interfaces que se estabelecem entre o módulo morfológico e os restantes módulos autónomos da gramática. Trata-se de interfaces complexas que, partindo do próprio conteúdo das entradas lexicais, se estabelecem por exemplo entre léxico e morfologia, entre morfologia flexional, fonologia e sintaxe, entre morfologia, semântica e sintaxe, evidenciando a atuação de determinados módulos da gramática sobre outros, por vezes de forma cumulativa. A existência de interfaces pressupõe a modularidade do conhecimento linguístico e, para se definir com clareza o componente morfológico da gramática e estabelecer as suas fronteiras, defende-se ser necessário considerar a interseção de componentes.

Keywords: morfologia, categorias morfológicas, interfaces, palavra abstrata, palavra morfossintática

1 Introdução

A morfologia tem como objeto central de análise o lexema – entendido como unidade abstrata do léxico, com representação morfológica, representação fonológica básica e significação própria – e as operações morfológicas que podem incidir sobre o lexema, ou sobre a sua base radical ou temática, na formação de outros lexemas (por exemplo, *ARCADA*, a partir de *ARC-*, radical de *ARCO*, por derivação sufixal). A cada unidade abstrata está associada uma matriz de traços que a identifica de forma clara quanto a um conjunto de propriedades de diferentes domínios da gramática.¹ Essa matriz assegura que as palavras flexionadas associadas ao lexema – que constituem o seu paradigma (e.g. Aronoff 1994, 11; Booij 2005, 122), como *arco*, *arcos* relativamente ao lexema *ARCO* e *arcada*, *arcadas* relativamente ao lexema *ARCADA* – possuam as especificações necessárias e típicas da sua classe lexical, sejam bem formadas, estejam aptas a serem usadas em enunciados. Logo, é possível conceber o lexema como uma abstração a partir das formas das palavras por ele subsumidas

¹ Existe consenso quanto a o lexema ser membro de uma das classes abertas. Aos afixos é dado um estatuto variável consoante as abordagens; Zwicky (1990, 230) integra-os numa categoria de *lexemas-partículas*.

(Bauer/Lieber/Plag 2013, 8), com conteúdo semântico e morfossintático associado a uma forma fonológica (Anderson 2014, 281). Operar com este tipo de entradas lexicais (lexema e matriz de traços) «ricas» corresponde a considerar o léxico como um módulo da gramática em que se estabelecem associações de informação linguística assegurada pelos módulos (autónomos) morfológico, fonológico, sintático e semântico, e considerar também, obviamente, relações formais e semânticas entre lexemas.² A confluência de informações de diferente tipo e as relações que estas estabelecem no domínio do lexema é veiculada através de interfaces, «canais estreitos de comunicação» entre módulos (Bermúdez-Otero 2012, 43, trad.), modos de as propriedades de diferente tipo se relacionarem, «se verem» mutuamente (Booij 2005, 155).

Do ponto de vista da morfologia, a análise da estrutura interna dos lexemas e dos processos de formação de lexemas combina-se com a análise dos paradigmas de palavras associados a cada lexema; tal implica que se lide com a unidade «palavra abstrata» correspondente a lexema (Stump 2005, 50) e com a unidade «palavra construída» («gramaticalizada»,³ morfossintática, flexionada»), membro do paradigma do lexema.

Tomando como exemplo os paradigmas de lexemas verbais, é do domínio da morfologia que, em português como em muitas outras línguas, haja sincretismo morfológico parcial de marcadores de pessoa-número (PN) – em português, entre a primeira e a terceira células dos paradigmas, na grande maioria dos verbos e dos paradigmas:

(1) Formas idênticas de primeira e terceira pessoa do singular

1. ^a s.g. e 3. ^a s.g.	contara	trouxera	– <i>mais que perfeito</i>
	contava	trazia	– <i>imperfeito, indicativo</i>
	contaria	traria	– <i>condicional</i>
	conte	traga	– <i>presente, conjuntivo</i>
	contasse	trouxe	– <i>imperfeito, conjuntivo</i>
	contar	trouzer	– <i>futuro, conjuntivo</i>
	contar	trazer	– <i>infinitivo flexionado</i>
	–	trouxe	– <i>pretérito perfeito</i>

² O tipo e o número de informações/propriedades contidas numa entrada lexical variam na literatura, mas há consenso quanto a um conjunto básico que associa ao lexema significação própria, forma morfológica e fonológica, classe lexical (e categorias lexicais, sintáticas, etc., inerentes à classe e mais ou menos expandidas).

³ Tradução de *grammatical word*, em Aronoff (1994), correspondendo a *word in sense 3*, em Mathews (1991, ¹1974, 24).

A ausência de sincretismo entre células de um mesmo paradigma é o padrão de flexão considerado canónico, em línguas flexionais (cf., e.g., Corbett 2000; Baerman/Brown/Corbett 2009; Aalberse/Don 2011), mas está fracamente representado em português, prevalecendo nesta língua o sincretismo parcial intraparadigmático. Esta propriedade estritamente morfológica não se aplica ao paradigma do pretérito perfeito do indicativo dos verbos ditos regulares (como *contar*, por oposição a *trazer*, em (1)), em que existem seis expoentes diferenciados (-i, -ste, -u, -mos, -stes, -ram), correspondentes aos seis valores cumulativos de pessoa e de número. Este caso configura o «ideal» supletivismo de marcadores.

Se a morfologia se ocupa das propriedades morfológicas das palavras, das categorias morfológicas que as tipificam e dos constituintes que as realizam, como no caso anterior, e ainda das regras de boa formação morfológica, ocupa-se também, na perspetiva aqui adotada, das interseções de traços morfológicos com traços de outros domínios, frequentemente indispensáveis à explicação das formas das palavras associadas a cada lexema.

A interface entre morfologia e léxico é óbvia: os lexemas, existentes ou possíveis, possuem uma estrutura morfológica e o léxico é constituído por lexemas (Galani/Hicks/Tsoulas 2011, 5). Algumas propriedades (ditas inerentemente lexicais), apesar de relevantes para a morfologia, estão totalmente especificadas no léxico. É o caso, entre outros, de lexemas verbais que apresentam radicais ou temas supletivos; por exemplo, o verbo *trazer* (cf. (1)) ou o verbo *ir*, que apresenta vários radicais, entre os quais VA- e FO-, o mesmo acontecendo com o verbo *ser*, que tem associados, entre outros, SEJ-, SO- e FO-, este último homónimo do do verbo precedente. Os radicais supletivos de cada um destes lexemas verbais não são fonologicamente relacionáveis entre si e nenhum deles decorre de operações morfológicas sobre outro ou outros dos radicais: todos eles estão listados no léxico, associados à entrada do verbo em causa (cf., e.g., Aronoff 1994; 2013; Mateus/Andrade 2000, para o português). Mas se o objeto de análise se deslocar do lexema para os paradigmas do lexema, a associação de cada radical a um dado conjunto de paradigmas flexionais tem relevância para a morfologia, na interface com o léxico. A morfologia deverá, por exemplo, dar conta da relação formal entre paradigmas que partilham a mesma base radical ou ainda, fazendo a ponte entre sincronia e diacronia, do facto de paradigmas associados a dois lexemas serem isomórficos, como no caso do pretérito perfeito de *ir* e *ser*.

Por seu lado, a relação entre as formas que preenchem as diferentes células de um paradigma flexional tem relevância para a fonologia, que deverá explicar, por exemplo, a alomorfia dos constituintes de tempo-modo-aspeto (TMA) -va-/-a-, no imperfeito do indicativo (cf. *mandavas~sabias, fugias*, por exemplo), ou a alternância vocálica em v[a]is~v[e]mos e os fatores fonológicos que a condicionam (note-se que a posição do acento de palavra verbal é morfológicamente determinada por cada paradigma, salvaguardadas as restrições fonológicas que caracterizam o português, nesse domínio, cf. Mateus/Andrade 2000). Ou seja, a fonologia é requisitada para explicar a alomorfia dos constituintes de TMA, em função da classe temática dos

verbos (tema morfológico em *-a*, *-e*, *-i*), no primeiro caso, e a do radical, no paradigma do presente do indicativo.

O supletivismo de bases acima referido é igualmente importante para a semântica aspetual e modal (ou seja, respetivamente, para a temporalidade interna e intrínseca ao tipo de situação descrita pelo verbo e para os valores de modalidade, associados à atitude do enunciador em relação ao descrito). Assim, independentemente da grelha aspetual associada a lexemas como *ir* ou *ser*, cada uma das bases morfológicas destes verbos integra paradigmas que possuem, de forma inerente, valores básicos e diferenciados entre si quanto às categorias aspeto e modo (para além de tempo), os quais se sobrepõem à informação contida nos lexemas e que com eles devem ser compatíveis. Caso contrário, obter-se-ão enunciados agramaticais ou apenas aceitáveis em condições muito particulares. Por exemplo, considerando o valor aspetual iterativo do imperfeito do indicativo, um enunciado como *ele morria de dois em dois dias* só é aceitável com este verbo de culminação (i.e., remetendo para uma situação não durativa) se *morrer* for usado metaforicamente. Mas considerando que o ponto de culminação pode ser precedido por um processo preparatório, *ele morria lentamente e eu não podia fazer nada* é perfeitamente aceitável e compatível com a duratividade também associada ao imperfeito do indicativo: a presença do advérbio, combinada com o imperfeito, leva a que se foque a duratividade (Cunha 2013). Neste caso, está em jogo a interface entre morfologia e semântica (básica) no domínio dos paradigmas flexionais e entre aquele complexo morfossemântico e a semântica dos lexemas verbais.

Da estrutura argumental dos verbos, na sua entrada lexical, decorrem por outro lado restrições quanto às construções que admitem, na interface entre propriedades semânticas e propriedades sintáticas do lexema. Por exemplo, o traço [+copulativo] de *ser* e *estar* implica que estes verbos não admitam a construção passiva (*o rapaz é/está nervoso* vs. **o rapaz é sido/estado nervoso*), o mesmo acontecendo com verbos que estão marcados como selecionando um objeto indireto ou oblíquo (*o rapaz escapou à polícia* vs. **o rapaz foi escapado (à polícia)*, *o homem viveu em Havana* vs. **o homem foi vivido em Havana*).

Desencadeiam-se assim interfaces complexas que, partindo do próprio conteúdo das entradas lexicais, se estabelecem de forma mais específica por exemplo entre léxico e morfologia, entre morfologia flexional, fonologia e sintaxe, entre morfologia, semântica e sintaxe, evidenciando a atuação de determinados módulos da gramática sobre outros, por vezes de forma cumulativa. A existência de interfaces pressupõe a modularidade do conhecimento linguístico e, para se definir com clareza o componente morfológico da gramática e estabelecer as suas fronteiras, é necessário considerar a interseção de componentes. Partindo da perspetiva da morfologia, nas secções seguintes são dados alguns exemplos que evidenciam a relevância das interfaces para a descrição do português e a complexidade do cruzamento de informação entre componentes (a interface entre dois componentes da gramática, apenas, não constitui o caso mais comum e produtivo). A opção de inscrever os casos descritos adiante sob

determinados tipos de interfaces é motivada pelo entendimento possível do direcionamento de influência de um módulo sobre o(s) outro(s).

2 Interface morfologia-fonologia

Os lexemas possuem informação fonológica básica sujeita a modificações motivadas, sobretudo, pela flexão morfológica (Goldrick/Rapp 2007). Assim, por exemplo, relativamente a verbos de tema em *-e* e *-i* que apresentam alternância consonântica no radical – como [d]~[s] em *podes, pode*, etc. ~ *posso* ou *pedes, pede*, etc. ~ *peço* –, é sem dúvida possível considerar-se que se trata de uma idiosincrasia lexical, marcada nas entradas lexicais de um conjunto fechado de verbos. Mas só a fonologia pode explicar que esta alternância ocorra apenas se a vogal do radical for [-alt, -/+rec, -/+arr] (i.e., /o/, em POD-/POS-; /e/, em PED-/PES-)⁴ e na 1.ª p. sg. (cf. Mateus 1982, 1975, 125ss.). Interessa contudo à morfologia a relação formal estabelecida entre esta célula do indicativo e todo o paradigma do presente do conjuntivo (*possa, possas*, etc.; *peça, peças*, etc.), extensível a todos os verbos do português (ver (1), acima), assim como o facto de a vogal temática ser apagada na 1.ª p. sg. do presente do indicativo (não ter realização na célula em causa), originando paradigmas flexionais semidefetivos quanto à VT e de essa defetividade ser total, no presente do conjuntivo. Comparem-se as estruturas morfológicas em (2a-b-d) e (2c), notando-se que «-» em (2a-b) representa a supressão da VT neste contexto (Mateus 1982, 1975) ou simplesmente a sua não presença em forma de palavra flexionada:

- | | | | | |
|--------|------------------------------------|----|------------------------------------|---------------|
| (2) a. | acedo, abro, chamo | c. | acedes, abres, chamas | <i>indic.</i> |
| | ...]RADICAL -]VT -]TMA Ø]1.ª sg. | | ...]RADICAL ə]VT -]TMA S]2.ª sg. | |
| b. | aceda, abra, chame | d. | acedas, abras, chames | <i>conj.</i> |
| | ...]RADICAL -]VT a/e]TMA Ø]1.ª sg. | | ...]RADICAL -]VT a/e]TMA S]2.ª sg. | |

A flexão de nomes contáveis (lexicalmente marcados como tal) revela também a interface entre morfologia e fonologia. Nos nomes temáticos,⁵ a flexão dá origem a paradigmas de duas células diferenciadas quanto à categoria morfológica do número, explicitamente marcada em termos fonológicos e fonéticos: por exemplo, *tenda, tendas* ou *lente, lentes*, em que os valores +sg. vs. -sg. são marcados por Ø~/S/ (TENDAØ ~TENDAS). Mas nos nomes atemáticos como *lápis* ou *pires* verifica-se que

4 Neste capítulo, a forma do radical é grafada em letra maiúscula. A forma do tema e do lexema, em versaletes.

5 Um nome temático apresenta na sua estrutura, em posição final, uma vogal átona (cf. *manta, módulo, lente*) realizada respetivamente como [e], [u], [i]. Diferentemente, os nomes atemáticos são caracterizados por não possuírem uma vogal com essas características ou por não apresentarem qualquer vogal nessa posição e por o lexema e os seus constituintes morfológicos, tema e radical, serem isomórficos entre si (por exemplo, lexemas CAFE OU LAPIS).

esta operação morfológica desencadeia isomorfismo nos membros do paradigma, por razões fonológicas (como se verifica em sintaxe – *o(s) lápis verde(s)*, *o(s) pires verde(s)*): ocorre a simplificação da consoante *-/S/* no final da estrutura morfológica-fonológica do lexema e de *-/S/* marcador morfológico de plural, fonologicamente idênticos. Adjetivos como *fresco*, *fresca* ou *leve* (e *verde*) apresentam o mesmo contraste com adjetivos atemáticos em *-/S/*, como *simples* ou *piegas*: vejam-se os contrastes *o(s) casaco(s) fresco(s)*, *a(s) blusa(s) fresca(s)*, {*o(s) casaco(s)*, *a(s) blusa(s)*} *leve(s)* vs. *a(s) frase(s) simples*, *a(s) frase(s) piegas*, por exemplo). Ou seja, a alomorfa $\emptyset\sim/S/$ que caracteriza a marcação de número é foneticamente encoberta nos nomes e adjetivos atemáticos cuja estrutura apresenta *-/S/* final, devendo a morfologia lidar com essa informação, na análise da flexão em número.

Um outro caso que releva da interface morfologia-fonologia é o do contraste entre um tipo de nomes e adjetivos que, em português, se caracteriza por apresentar estabilidade do lexema (equivalente ao tema), no paradigma flexional (cf. (3a)), e um segundo tipo em que as células do paradigma exibem alomorfa temática (cf. (3b)):

- | | | |
|--------|---|--|
| (3) a. | LEME] _{LEXEMA – TEMA ÚNICO} | – paradigma <i>leme</i>
<i>lemes</i> |
| b. | POMAR] _{LEXEMA-TEMA1 = RADICAL}
POMARE] _{LEXEMA-TEMA2} | – paradigma <i>pomar</i>
<i>pomares</i> |

O comportamento em (3b), em que tema e radical são formalmente idênticos na palavra singular mas não na palavra plural, pode conduzir a diferentes tipos de análise:

- (i) a adunção de *-/S/* marcador de plural desencadeia o fenómeno fonológico de epêntese da vogal por defeito, */e/*, dada a impossibilidade da sequência **/rS/* como coda silábica (**pomars*), havendo, assim, restrições impostas pelas regras de boa formação silábica (e.g. Bisol 1998; Alcântara 2003; ↗25 A sílaba na gramática do adulto e na aquisição de língua materna);
- (ii) o lexema tem a estrutura *POMARE*, sendo que a vogal final sofre apagamento no singular por estar em posição final (e átona) e ser possível que um nome terminado em coda *-/r/* seja bem formado, em português (e.g. Veloso 2007; Ferreira 2009).

Nestas duas hipóteses, considera-se haver cruzamento de instruções entre morfologia e fonologia. A terceira alternativa, acima avançada (Mota no prelo a), pressupõe que o tipo de lexemas em (3b) apresenta alomorfa lexical de temas (cf. Ralli 2003; 2007), integrando-se o lexema nominal numa subclasse temática; ou seja, numa subclasse em que o lexema possui, na forma morfológica, uma posição para o constituinte temático, à direita do radical. Essa posição é ocupada por */e/*, na forma de palavra plural, e por um segmento nulo (\emptyset), na forma de palavra singular, alomórfico do anterior; o conceito de «nulo» supõe, de facto, que exista um outro constituinte, com conteúdo morfológico (e fonológico) que a ele se opõe (cf. Aronoff/Fudeman 2005,

16s.). Assim, ao tema morfológico caracterizado pela presença da vogal e ao tema alomórfico caracterizado por conter um constituinte nulo correspondem, respetivamente, as formas temáticas fonológicas /xe/ e /xØ/, em que *x* representa o radical. Correspondem, também, as estruturas silábicas admitidas pelo português e ilustradas por *po.ma.res*_{-sg} (*pomares*) e *po.mar*_{+sg} (*pomar*), respetivamente.⁶ Importa referir que se parte do pressuposto, geralmente aceite, de que restrições fonológicas não motivam a forma morfológica dos lexemas, mas sim que a morfologia é o *input* de operações fonológicas (Paster 2006); não havendo justificação fonológica para que a vogal, marcador temático, não ocorra na forma de palavra singular, como adiante comentado, estamos perante uma subclasse morfológica de nomes e de adjetivos que pode ser caracterizada como uma subclasse temática Ø~e (POMARØ~POMARE). Um dos argumentos a favor desta opção consiste em ser necessário distinguir o comportamento flexional de POMAR – *pomar*, *pomares* e de ARVORE – *árvore*, *árvores*. De facto, um nome como *árvore*, ou *mármore*, apresenta /r/ precedendo /e/, devendo então justificar-se a presença desta vogal, já que a referida consoante poderia funcionar como coda final de palavra. A questão não se coloca, portanto, em termos exclusivos de boa formação da estrutura silábica. Esta está associada a determinado padrão temático em que o nome se inscreve, que é uma categoria morfológica especificada nas entradas lexicais. Um nome com padrão temático em -e, como *árvore*, ou em -o ou -a nunca é oxítono, já que qualquer destas vogais com função de marcador temático é inerentemente átona e ocorre sistematicamente nas duas palavras do paradigma do lexema (*árvore*, *árvores* e *módulo*, *módulos* ou *tenda*, *tendas*). Se *pomar* fosse um nome temático de tema em -e, apresentaria a forma singular **pomare*, paroxítona, a par de *pomares*, não existindo qualquer restrição fonológica aparente para excluir essa possibilidade – não sendo necessário prever o apagamento do marcador temático na palavra singular. A *pomar* ser atemático (com inserção de /e/ na estrutura do plural), estar-se-ia perante a necessidade de reconhecer a existência de um número significativo de nomes (e adjetivos), terminados em consoante na forma singular do paradigma do lexema e com uma vogal epentética na forma do plural, que não seriam nem conformes ao padrão geral, típico, do português (que consiste em os lexemas nominais e adjetivais serem formalmente idênticos ao seu tema morfológico) nem ao padrão, muito menos representado nesta língua, dos nomes e adjetivos claramente atemáticos (em que na palavra plural do paradigma não ocorre uma vogal epentética, cf. {lápis, simples}_{+/-sg}, {café, gagá}_{+sg} ~ {cafés, gagás}_{-sg} ou *atum*_{+sg} ~ *atuns*_{-sg}, entre outros). A única particularidade de nomes como *pomar*, *colar*, *mentor*, *poder* ou de adjetivos como *circular*, *polar*, *secular*, *esmoler* é a de os respetivos lexemas terem associados dois temas, permitindo paradigmas de palavras bem formados, de todos os pontos de vista. Este mesmo comportamento se constata em muitos nomes e adjetivos cujo radical termina em /S/ (*eirós*, *eiroses*; *rapaz*, *rapazes*, por exemplo). A

6 Os pontos, em *po.ma.res* e *po.mar*, indicam fronteira de sílaba.

alomorfia de temas não constitui um fenómeno excepcional. Não é incomum existirem nas línguas dois ou mais temas (ou radicais) associados a um mesmo lexema, em relação de alomorfia ou de supletivismo (cf., e.g., Carstairs-McCarthy 1987; Booij 1997; 2002; 2012; 2015; Ralli 2007; Spencer 2010), como comprovam vários verbos do português (Mateus/Andrade 2000). Outros casos ilustram de forma clara a interface entre morfologia e fonologia, como acontece com os constituintes de pessoa-número (PN) associados à quinta célula dos paradigmas verbais. Assim, *-des* e *-is* revelam alomorfia condicionada fonologicamente; o primeiro ocorre se a unidade morfológica de base, a que se liga, terminar em consoante /r/ (*contardes*, *trouxerdes*), em segmento nasal (*tendes*, *pondes*) ou ainda em vogal, se a referida unidade for monossilábica e apresentar uma mesma vogal no radical e como vogal temática (*ides*, *credes*, *ledes*, que correspondem, respetivamente, a I]_{RAD} i]_{VT} (*ir*), CRE]_{RAD} e]_{VT} (*crer*) e LE]_{RAD} e]_{VT} (*ler*)). A segunda forma de PN ocorre nos restantes casos (*contais*, *trazeis*) (ver Mateus 1982, 1975, 106–107 para desenvolvimento). Note-se que a vogal em *-is* é a representação ortográfica do glide [j] decorrente da alteração da vogal *-i* em contacto com a vogal precedente (a VT) com traços diferentes, originando-se um ditongo pesado (*cont[a]js*, *traz[e]js*). Neste caso, a vogal *-i* fica em contacto com a VT após a truncção da consoante inicial de *-des* e corresponde à elevação da vogal média átona presente nesse PN.⁷ Em *fugis*, dada a identidade da VT e de *-i*, ocorre a simplificação das duas vogais. Esta alomorfia configura uma situação de heteróclise interparadigmática de marcadores de PN com o mesmo valor. Contudo, em português, a alomorfia no domínio deste tipo de marcadores morfológicos é excepcional.

A interface entre morfologia e fonologia é muito relevante quando se trata, entre outros, dos seguintes casos:

- (i) a harmonização vocálica entre a última vogal do radical, se esta não for alta (/i/, /u/) nem baixa (/a/), e a vogal temática. O contexto que favorece este fenómeno é o da 1.^a p. sg. do presente do indicativo e de todo o paradigma do presente do conjuntivo e a vogal temática /e/ ou /i/: ou seja, formas acentuadas no radical e com uma vogal à direita da vogal temática, na estrutura do lexema (por exemplo, CED]_{RAD} e]_{VT} o]_{PN}]_{1.ªsg} e CED]_{RAD} e]_{VT} a]_{TMA}]_{1.ªsg}, ...). A vogal temática é truncada nos contextos acima referidos, em forma abstrata, mas deixa o traço de altura na estrutura (um autossegmento flutuante, cf. Mateus/Andrade 2000, 75), que influencia a vogal do radical, havendo uma «projeção da altura da vogal temática sobre a vogal do radical» (Mateus 2003, 295; ver ainda Andrade 1994, 95; Mateus/Andrade 2000, 81–85). No caso de *ferir*: é um verbo com uma vogal média no radical e com vogal temática alta, /i/, no tema FER/i/; por efeito da regra de harmonização vocálica, a vogal do radical torna-se alta, dando origem a *firo*, que contrasta com *feres*, *ferre*, *ferem*, sujeitas a uma regra de abaixamento da vogal /e/ do radical para [ɛ] (cf. Mateus 2003, 298–300);

⁷ Em Andrade (1994, 96), é proposta a forma /dis/ para este marcador de PN.

- (ii) a elevação e/ou centralização da vogal do radical [-alta], relacionada com o padrão acentual de cada paradigma flexional: por exemplo, *f[i]rimos* e *f[i]ris* em sílaba átona vs. *f[ɛ]res*, *f[ɛ]re*, *f[ɛ]rem* em sílaba tónica;
- (iii) a inserção fonética, em verbos com radical terminado em /e/, de glide homorgânica, em formas acentuadas no radical e que apresentam na sua estrutura de base um hiato entre a vogal do radical e a vogal à sua direita: em *rodear*, por exemplo, as formas apresentam um ditongo leve realizado [ej] (*rodeio*, *rodeias*, *rodeia*, *rodeiam*), quando ao radical terminado em vogal se acrescenta um outro constituinte morfológico vocálico. Em *rod[j]amos*, *rod[j]ais*, a última vogal do radical, /e/, em posição átona, realizar-se-ia previsivelmente como [i], por efeito da regra da elevação e centralização das vogais átonas. Contudo, dado o contexto de hiato com a VT, a elevação redundava em [j] (i.e., e sem entrar em explicação pormenorizada, não se obtém *rod[iɛ]mos/-is* mas sim o resultado indicado acima).
- (iv) a vocalização de /L/, na forma plural de nomes e adjetivos em que esta consoante lateral constitui a coda final do radical e da forma singular de palavra (*canal*, *canais*; *anel*, *anéis*; *lençol*, *lençóis*; *paul*, *pauis*, por exemplo). Esta consoante tem a peculiaridade de poder realizar-se como glide, em contexto apropriado como é o de o acento de palavra recair sobre a sílaba fechada pela consoante em causa e ser seguida do constituinte de plural.⁸ O processo morfológico de formação do plural, por adjunção de /S/ e despromoção de /L/ como segmento final, está na base de um processo fonológico que leva esta consoante a ancorar-se ao núcleo silábico, o que origina, por sua vez, um ditongo fonético. Dito de outro modo, a consoante deixa de ocupar a posição de coda, associa-se ao núcleo da sílaba que fecha, no singular, dando origem a um ditongo decrescente seguido do marcador

⁸ É sabido que em português do Brasil esta consoante em coda pode realizar-se [w] em posição interna (*ca[w]çado*, por exemplo) e em muito maior grau em posição final, como nas formas de palavra singulares como *norma[w]* ou *ane[w]* (o mesmo pode acontecer em variedades do português europeu não standard, da região norte). Tal ocorre sobretudo quando - /L/ se encontra depois de vogal [+bx] mas também de vogal [-alt,-bx], raramente depois de vogal [+alt] (Callou/Moraes/Leite 2013). O facto de não se obterem palavras plurais com as formas **norma[w]s*, **ane[w]s* tem previsivelmente a ver com o facto de, no plural, estar presente - /S/, posicionado à direita de /L/, o qual deixa de estar em posição final. Em Callou/Moraes/Leite (2013, 175s.), seguindo uma proposta de modelo não linear da fonologia, é sugerido que [w] supõe um [ɰ] velarizado em fase anterior, com um nó consonântico e um nó vocálico, sendo as etapas «(i) a inserção de um nó vocálico, com ponto de articulação dorsal, que corresponde à representação da lateral velarizada; (ii) o cancelamento do nó consonantal, levando ao glide posterior arredondado [w]» (Callou/Moraes/Leite 2013, 176). Em português europeu standard, o singular destas palavras apresenta a consoante velarizada [ɰ] (*norma[ɰ]*, *ane[ɰ]*). Por seu lado, e também no plural, a agramaticalidade de palavras como **normals* ou **anels*, quer no português europeu como no do Brasil, decorre de a sequência **-/LS/* em coda silábica (complexa) não ser canónica (ver Andrade no prelo). Resolve-se essa restrição pela vocalização da consoante lateral: associando-se à vogal precedente, redundando foneticamente no glide velar [j], originando um ditongo (cf. *normais*, *anéis* ou *caracóis*), quando seguido da consoante coronal marcadora de plural. Remete-se para Hora/Pedrosa/Cardoso (2010) para questões relacionadas com as consoantes em coda em português do Brasil.

de plural ([aj]), em *canais*, por exemplo). Esta é a análise considerada mais simples para explicar o fenómeno (como em Mateus/Andrade 2000, 71), apesar de existirem outras. Este fenómeno fonológico apenas afeta lexemas terminados em *-l/* mas não tem efeito se o nome é oxítono e apresenta *-i/* como vogal final do radical (cf. *canil*, *canis* ou *senil*, *senis*, por exemplo): nesse caso, dá-se a fusão dos dois segmentos com traços idênticos, [+alt, -rec]. Se o requisito da acentuação da sílaba final não existir, tem lugar a dissimilação da vogal /i/ da última sílaba do radical em contacto com [j], originando-se um ditongo [ej] átono, como em FRAGIL e *frágil*, *frágeis* OU FACIL e *fácil*, *fáceis*.

Este comportamento fonológico da consoante *-l/* na formação do plural é consentâneo com considerar-se estes nomes e adjetivos como pertencendo à classe atemática: não há qualquer evidência da presença de uma vogal final no lexema, mesmo que apenas no plural, como se pode propor relativamente a outras estruturas nominais e adjetivais com lexema e palavra singular terminada em *-s/* ou *-r/* (ver secção 3).

3 Interface morfologia(-fonologia)-sintaxe

A interface morfologia-sintaxe decorre de os paradigmas dos lexemas disponibilizarem informação relevante para as operações especificamente sintáticas. Se a componente sintática não interfere na capacidade de variação que as palavras apresentam, na formação dos paradigmas, ela beneficia das suas propriedades flexionais. Por exemplo, a categoria semântica «contável» (marcada nas entradas lexicais dos nomes) tem contrapartida na categoria morfológica «número» (na interface morfologia-semântica), realizada por constituintes sufixais se o valor daquele traço semântico for positivo (*CAPA* N[+cont(ável)] e *capa*, *capas* vs. *AREIA* N[-cont] e *areia*, **areias*, na aceção de massa ou matéria (se o lexema *AREIA* remeter para tipos, cores, espessuras, etc. da matéria ‘areia’, deixará de ser um nome massivo e terá o comportamento de [+cont], podendo pluralizar, como em *usaram-se areias de três tipos*). Por seu lado, um dos dois valores de número possíveis para nomes e adjetivos só é ativado em função da referência, em sintaxe.

A categoria «género», por seu lado, com um valor lexicalmente determinado nos nomes (inerente, não flexional) mas variável nos adjetivos, permite que a relação de concordância de género entre nome e adjetivo, quando requerida pela sintaxe, encontre na morfologia a possibilidade de marcar explicitamente essa relação, na maioria dos casos, como em *capa*_{-masc} *longa*, *xaile*_{+masc} *longo*, mas não sistematicamente, como se observa em *capa/xaile quente/confortável*. Neste segundo caso, um lexema adjetival de tema em *-e* ou atemático não dispõe senão de uma forma no seu paradigma de palavras morfossintáticas. O valor de género por ele veiculado é apenas sintaticamente marcado, em função do valor de género do nome com que se associa. A concordância N-Adj será assim explícita ou não, em termos morfológicos («overt»

ou «covert», cf. Corbett/Baerman 2006), dependendo da morfologia do lexema, concretamente da sua classe temática (ela própria fixada no léxico para cada lexema). A concordância é assim o fenómeno que desencadeia de forma mais evidente a interface morfologia-sintaxe. No caso anterior, pondo em relevo a distinção entre género como categoria lexical sem contrapartida morfológica e como categoria de concordância, flexional; no caso da concordância entre sujeito e verbo, levantando a questão da estruturação interna do paradigma verbal selecionado em sintaxe. Por exemplo, paradigmas com sincretismo de células (como *mandava*, *mandaria* e *mandasse*, que preenchem a primeira e a terceira células do singular, ambas sem expoente morfológico visível de pessoa-número) podem tornar necessária a explicitação do sujeito, na frase, a fim de desambiguar a sua interpretação, caso o contexto não seja suficiente.

As categorias de flexão são morfológicas, estando associadas à existência de paradigmas dos lexemas, mas são, como acima referido, importantes para a sintaxe, podendo assim considerar-se, de forma mais rigorosa, serem categorias morfossintáticas, do mesmo modo que os constituintes que as realizam.

Um outro caso que ilustra a interface entre morfologia e sintaxe é o da conversão de radicais, processo com implicações também fonológicas. Por exemplo, o radical nominal FABRIC- (lexema *FABRICA* e nomes do seu paradigma *fábrica* e *fábricas*) ou o radical adjetival FIRM- (lexema *FIRME* e *firme*, *firmes*), ao sofrerem conversão em radical verbal (processo morfológico que não implica intervenção de afixos ou de outros radicais que alterem a forma do radical de base, cf. Villalva no prelo a), vão integrar-se numa nova classe de palavras, adquirindo propriedades sintáticas típicas dessa classe. Assim, os lexemas correspondentes a *fabricar* e *firmar* passam a constituir novos itens lexicais com uma matriz de traços adequada à classe Verbo, de entre os quais os traços sintáticos [+transitivo] (*fabricar sapatos*, *firmar um acordo*) e a seleção de sujeito [+Agente]. A alteração da classe dos radicais de base implica que, a nível morfológico, tenham associados paradigmas flexionais típicos do Verbo, respondendo assim às exigências da sintaxe. A nível fonológico, implica que os novos lexemas estejam sujeitos a padrões acentuais específicos: o nome *fábrica* é proparoxítono e as formas do verbo são proparoxítonas (*fabricá*ssemos), paroxítonas (*fábrica*) ou oxítonas (*fabricar*, *fabricará*). A posição do acento de palavra, nos nomes como nos verbos, está especificada a partir da entrada lexical, em função da respetiva classe e de outras propriedades da estrutura morfológica e fonológica dos lexemas. Nos nomes, aqueles que apresentam um padrão de tema em *-a*, *-o* ou *-e* são inerentemente não oxítonos, já que o marcador de padrão temático (índice temático – IT) é por definição átono e esta é uma propriedade do domínio da fonologia. Estes nomes só podem, assim, ser paroxítonos ou proparoxítonos, pelo condicionamento imposto pelo IT. Nos verbos, essa posição está marcada em função de cada paradigma flexional – em geral, o acento recai ou na vogal do radical ou na vogal temática, mas cada paradigma apresenta padrões acentuais fixos. Por exemplo, o padrão acentual no presente do indicativo corresponde a acento (i) na última sílaba do radical nas células

1 *fábrico*, 2 *fábricas*, 3 *fábrica*, 6 *fábricam* e (ii) na VT nas células 4 *fábricamos* e 5 *fábricais*; no futuro do indicativo e no condicional, o acento recai sistematicamente na sílaba que integra o TMA: *fábricarás*, *fábricarias* (Mateus/Andrade 2000, 109–128, sobre acento de palavra em português). Em suma, a conversão tem implicações no léxico (formação de novos lexemas) e decorre de um processo morfológico (embora não típico, diferentemente da derivação – ver Rio-Torto no prelo – e da composição); a flexão é do domínio morfológico, com motivação e repercussão sintática; a possibilidade de um lexema apresentar flexão marcada morfológicamente está especificada no léxico, maioritariamente de acordo com a forma morfológica e fonológica do lexema;⁹ a morfologia e a fonologia estabelecem relações estreitas no domínio dos lexemas e dos paradigmas dos lexemas; a classe de palavras está marcada no léxico como uma propriedade sintática e determina as possibilidades de flexão e o seu tipo, entre outras propriedades. Todo este complexo configura múltiplas interfaces, sendo que aqui se destacou a relativa a morfologia-fonologia-sintaxe.

Ainda relativamente à flexão, uma das questões complexas do português é a da classe temática dos nomes que, no singular, terminam no ditongo *-ão* mas que apresentam dois padrões de flexão em número: *-ãos* (*irmãos*, *cidadãos*) e *-ões*, *-ães* (*limões*, *pães*). Como acontece com os nomes que terminam numa das codas possíveis do português (*flor*, *ananás*, *mal*), também relativamente àqueles nomes existem diferentes hipóteses explicativas da relação morfológica entre os dois membros do paradigma do lexema, um marcado para singular e outro para plural. A estrutura dos lexemas correspondentes às três palavras acima referidas, com ditongo nasal no singular, contém, na forma do seu constituinte radical, um segmento final /N/, denominado autosegmento flutuante. Esse segmento fonológico não tem realização fonética como consoante nasal, nos nomes simples,¹⁰ mas nasaliza a vogal anterior à qual está associado, na estrutura silábica (Mateus/Andrade 2000, 131–134). Relativamente ao primeiro caso, há consenso quanto a analisar o lexema correspondente a *irmão* como temático (de tema em *-o*), sendo que a adjunção de *-/S/* de plural não altera a forma do lexema. Assim, a partir da estrutura abstrata $IRMA^N]_{RAD} O]_{IT}$, obtém-se uma forma do radical com a vogal final nasalizada; resolvida a influência de /N/ sobre essa vogal, o índice temático *-/u/* encontra-se agora em hiato com a vogal do radical, semivocaliza, associa-se ao núcleo e recebe nasalidade da mesma vogal.

⁹ Os nomes *pluralia tanta* (semanticamente não contáveis, apenas possuem o valor de número morfológico [-sg], como *exéquias* ou *núpcias*, por exemplo) e *singularia tanta* (semanticamente não contáveis, apenas possuem o valor de número morfológico [+sg], como *água*, *areia* ou *lama*) são excepcionais, do ponto de vista da variação dos valores de «número».

¹⁰ Esse segmento não se ancora ao esqueleto silábico. Apenas ocorre como consoante nasal, ancorada, em palavras derivadas, ocupando a posição de ataque da sílaba seguinte àquela a que, nos nomes simples, está associado: por exemplo, *irmanar*, *panificar*, *leonino*. Nunca ocorre como coda, como se observa em *cetim*($CTI/N/$), por exemplo, cuja forma fonética apresenta vogal nasal final, *cet[ĩ]*, e não consoante nasal.

Estes processos redundam no ditongo $[-\text{ẽw̃}]$, na palavra singular; na palavra plural, o percurso de formação é idêntico ao acima descrito, com a adunção de $/S/$ de plural e o resultado $[-\text{ẽw̃}j]$. No que respeita ao segundo padrão, há diferentes análises disponíveis, das quais, de forma aqui muito simplificada, se podem dar três exemplos: o lexema tem a forma *LEONE* e *PANE*, de tema em $-e$ (o processo de nasalização da vogal do radical é o acima descrito); a vogal marcadora de classe temática mantém-se no plural, ligando-se a ela $-/S/$, mas é apagada no singular para dar origem às formas $le[\text{ẽw̃}]$ e $p[\text{ẽw̃}]$, com desenvolvimento de um ditongo fonético leve (Abaurre-Gnerre 1983; Veloso 2005). Dado que *leão* tem a vogal $/o/$ no radical (cf. *leônimo*), o ditongo esperável seria $[-\text{õw̃}]$, no singular, mas o mesmo não tem ocorrência em português europeu standard, devendo prever-se a dissimilação que origina $[\text{ẽw̃}]$. Numa outra abordagem, os lexemas cujas palavras plurais correspondentes terminam em $-\text{ões}$, $-\text{ães}$ são considerados de tema em $-o$, o que pressupõe as formas de lexema *LEONO* e *PANEO*. Esse marcador temático está na base do glide que ocorre no singular (*leão*, *pão*) mas está também presente na forma abstrata do plural; aí, está sujeito à regra fonológica de dissimilação (passando a $/e/$), motivada pela adunção do marcador de plural (Bisol 1998). Outra análise, ainda, consiste em defender que o lexema tem a forma *LEON* e *PAN*, a vogal do radical é nasalizada pelo segmento $/N/$ e, por se encontrar em posição final de palavra e sob o acento, desenvolve o ditongo $[\text{ẽw̃}]$, na forma singular; no plural, sofre a epêntese do glide $[j]$, desencadeada pela adunção de $-/S/$ ($le[\text{õj}j]$). A seleção do glide, inesperada, tendo em consideração os traços das vogais $/a/$ e $/o/$ no radical, é remetida para condicionamentos diacrónicos (Mateus/Andrade 2000, 72s.). Uma outra hipótese é a de que este tipo de nomes se inscreve na classe temática $\emptyset\sim e$ (cf. secção 2), com alomorfa lexical de temas. Nesta análise, o tema 1 tem a estrutura *LEON* \emptyset e *PAN* \emptyset (presente na palavra singular do paradigma) e o tema 2 apresenta a estrutura *LEONE* e *PANE* (presente na palavra plural). Esta possibilidade, que aqui se defende, explora a existência de bases alomórficas morfológica e lexicalmente motivadas em português, como também acontece noutras línguas (Ralli 2007); evita marcações excepcionais nas entradas lexicais e, sobretudo, que, como noutros tipos de abordagem, sejam necessárias explicações menos económicas para explicar os *outputs* fonéticos adequados (*limão* e *pão*, e não $*leõe$ e $*pãe$; *leões* e *pães* e não $*leões$ e $*pãos$).

O autosegmento flutuante $/N/$ está implicado também na estrutura da 3.^a pessoa do plural, nos paradigmas verbais (por exemplo, *ANDA*/ $N/$, *DEVE*/ $N/$, *FOGE*/ $N/$), sendo o resultado da sequência fonológica $/V[+\text{segmento nasal}]/$ um ditongo nasal fonético, um ditongo leve, com glide homorgânico ($[w]$ em *cercam*, $[j]$ em *devem*, *fogem*); ver nota 10. A vogal deste ditongo é sempre a vogal final do constituinte mais à direita, na estrutura: a vogal temática, como nos exemplos precedentes, a vogal de PN (*cercar \underline{a} m*, pretérito perfeito, PN $-ra/N/$) ou a vogal que integra o TMA (*cercar \underline{a} m*, mais que perfeito, por exemplo). Embora o marcador morfológico de pessoa-número seja assumido por $/N/$, encontra-se assim sob a forma de uma amálgama com outros constituintes, por condicionamentos fonológicos.

A relevância desta marcação é importante para a concordância sintática. Seguindo Corbett/Baerman (2006), na «concordância canónica» o controlador e o alvo da concordância partilham valores dos traços relevantes para a relação sintática, havendo total correspondência (*matching*). Mas os expoentes morfológicos verbais podem não ter a resolução esperada e encobrir a partilha de valores que a concordância pressupõe, não pondo contudo em causa a sua existência: ela será apenas não explicitada por um marcador visível, será *covert*. Os marcadores de PN são não lexicais (não integram o lexema), as suas propriedades são menos nucleares do que as do TMA, sendo periféricos na estrutura da palavra. Associados a estas propriedades lexicais e morfológicas, estão condicionamentos fonológicos, relativos ao comportamento de /N/. Por exemplo, em variedades não standard do português, este segmento pode nasalizar a vogal anterior sem que seja desencadeado um ditongo (ou este pode ser posteriormente reduzido, como em *comer*[u], cf. Martins 1995; Rodrigues 2012, 221) ou então simplesmente não atua, originando-se formas de 3.^a p. pl. como *cerc*[e], *ced*[i], *fog*[i] (e, em nomes com /vogal do radical [+nasal]/, formas como *vag*[i], *vagem*, ou *hom*[i], *homem*; ver Bisol 2001, para uma análise deste fenómeno). A sequência /vogalN/ encontra-se em posição átona final e esse facto é determinante para a possibilidade de vir a anular-se a oposição com /vogal/ átona final de 3.^a p. sg., em forma fonética, não dispondo as formas de plural de um expoente diferenciado. Como consequência, a concordância sintática não é, nestes casos, canonicamente explicitada. Este facto é verificável, em português, em frases simples (por exemplo, *eles foge para lá – corpus ALEPG, Válega*). Paralelamente, condicionamentos sintáticos podem desencadear a seleção da forma singular (*portanto, continua os tempos a evoluir cada vez mais; vai vir montes de charters com politólogos; bonequinhos [isso] era figuras de acção*);¹¹ ver Carrilho 2003; Cardoso/Carrilho/Pereira 2011). Mas a ocorrência de sincretismo fonético e da consequente concordância «covert» ocorre predominantemente quando está em causa a oposição fonológica /vogalN/~vogal/; de facto, em português europeu, não há casos, de acordo com pesquisas recentes (Mota/Miguel/Mendes 2012, por exemplo), de concordância do tipo *tu anda* ou *nós anda*, como acontece em variedades não standard do Brasil (e também de África), concluindo vários estudos que nelas existe de facto sincretismo interparadigmático, o qual pode redundar na existência de até duas formas apenas, em casos extremos: 1.^a sg. vs. todas as restantes (Duarte 2000, por exemplo). Nestes casos, pode considerar-se existir uma concordância *singleton* (Costa/Silva 2006): a uma concordância canónica [[eles]_{+PL} [andam]_{+PL}]]_{+PL} corresponde a variante [[eles]_{+PL} anda]_{+PL/não marcado para PN}]]_{+PL}.

A flexão morfológica responde, assim, previsivelmente, à informação sintática requerida pela concordância, mas podem existir condicionamentos, na interface

11 Dados do *corpus Lisboa* do projeto *Estudo comparado dos padrões de concordância em variedades africanas, brasileiras e europeias*, www.concordancia.letas.ufrj.br.

morfologia-fonologia-sintaxe, que encubram a partilha de traços, entre os quais os que dizem respeito ao comportamento do segmento /N/.

Uma relação interessante entre morfologia e sintaxe interna do lexema pode ser ilustrada pelos afixos descontínuos. No processo de formação de verbos pelo processo de parassíntese, em português, é ativado um afixo que se caracteriza por possuir uma parte prefixal, em anteposição ao radical, e uma parte sufixal, em posposição ao radical, existindo uma dezena de combinatórias possíveis. O processo opera sobre radicais nominais ou adjetivais e os radicais verbais assim derivados originam lexemas de tema em *-e*, se a parte sufixal tiver a forma *-ec-* ou *-esc-* (*apodrecer*, *enflorescer*, por exemplo), e de tema em *-a*, nos restantes casos:

- | | | | |
|--------|--|----|--|
| (4) a. | aCOBRe] _{RAD} , <i>acobrear</i>
afixo <i>a...e</i> | c. | esQUARTej] _{RAD} , <i>esquartejar</i>
afixo <i>es...ej</i> |
| b. | enSURDec] _{RAD} , <i>ensurdecer</i>
afixo <i>en...ec</i> | d. | aPODReC] _{RAD} , <i>apodrecer</i>
afixo <i>a...ec</i> |

A parte sufixal deste afixo pode ser \emptyset , sendo a existência desse elemento nulo indispensável à formação de lexemas verbais a partir de bases não verbais, como em *acetinar*, *aquietar*, *desossar*, *enlatar*, *refinar*, que, em forma final, revelam apenas um constituinte pré-radical de base. Não é possível considerar apenas a parte pré-verbal ou a pós-verbal como ilustram **podre~~c~~er*, **podre~~j~~ar*, **podri~~f~~icar*, **podri~~z~~ar* ou **apo~~d~~rar* vs. *apodre~~c~~er*, pelo que *acetinar*, etc. são formados com o afixo *a/des/en/re... \emptyset* . Com efeito, e considerando-se que os prefixos não têm a capacidade de alterar a classe de palavras da base a que se ligam (Rio-Torto no prelo), a presença de *a* apenas, a ser considerado um prefixo, não bastaria para originar uma base verbal a partir de uma base adjetival ou nominal. Os afixos são elementos listados no léxico que apresentam restrições quanto à classe lexical da base a que se ligam, quanto à posição que podem ocupar, na estrutura interna da palavra, determinando frequentemente a classe temática dos lexemas, como acima referido. A parassíntese ilustra, assim, um caso de interface múltipla, que implica propriedades lexicais, morfológicas e da sintaxe interna das palavras.¹²

No domínio da formação de palavras, as interfaces são contudo menos determinantes do que no domínio da flexão. Por exemplo, não é necessário extrapolar do domínio da morfologia para explicar o contraste formal entre o radical AMAVEL, no adjetivo derivado *amável*, e o radical AMABIL-, no derivado nominal *amabilidade*. Trata-se de um caso de supletivismo lexical de radicais, associados ao lexema AMAVEL. O lexema derivado (AMA]_{TEMA VERBAL} vel]_{SUF. ADJETIVALIZADOR}) é caracterizado por ser atemático, o que significa que a forma do lexema e a do seu radical morfológico coincidem. Mas visto existir incompatibilidade morfológica entre radicais já sufixados

¹² A parassíntese é concebida por alguns linguistas como resultado da conversão de radicais nominais e adjetivais em verbais e da adjunção de afixos.

com *-vel* e o sufixo *-idade* (**contabilidade*, **movelidade*), ao lexema está associada, na sua entrada lexical, uma base supletiva, um radical preso especializado para a derivação sufixal (*amabilidade*, como *contabilidade* ou *mobilidade*) e sem estatuto de lexema (cf. adjetivo **amabile/o/a*).

4 Interface morfologia(-fonologia)-semântica

A interface entre morfologia e semântica pode constatar-se na formação de lexemas verbais que apresentam o sufixo *-it-*, veiculando valor iterativo, ou *-isc-*, com valor avaliativo (*comiscar*, *neviscar*, *mordiscar*), acrescentados a bases já verbais. Por exemplo, *dormitar* remete para sequências breves do evento ‘dormir’ e *saltitar*, da ação ‘saltar’. Os sufixos avaliativos e z-avaliativos nominais e adjetivais acrescentados a radicais ou a palavras, respetivamente, demonstram o mesmo tipo de interface, como em *florinha*, *fraquito*, *montão* ou *levezinho*, por exemplo.

A interface entre estes dois módulos é também clara no grupo de nomes que, sendo semanticamente marcados com o traço [+anim(ado)], têm um comportamento morfológico específico: *mestre*, *mestra*, *primo* e *prima* ou *gato* e *gata* correspondem a lexemas em que o marcador de classe temática *-e/-o* e *-a* acumula informação de género, com os valores respetivos de [+masc] e [-masc]. Os nomes [-anim] nunca apresentam este fenómeno, e, dentro do grupo dos [+anim], ele restringe-se a um subgrupo limitado em que existem dois nomes com radical idêntico e fortemente aparentados, semanticamente. Na quase generalidade dos nomes, o valor de género é arbitrário, está especificado nas entradas lexicais nominais e não tem relação com a classe temática e seus expoentes. Facilmente se verificam estas propriedades observando *creme*_{+masc} e *serpente*_{-masc} (IT *-e*) ou *mapa*_{+masc} e *cabra*_{-masc} (IT *-a*), por exemplo. Por outro lado, a associação de determinado índice temático a um valor de género externamente motivado pelo sexo biológico das entidades também não pode ser generalizado, em português, como mostram *serpente* (macho ou fêmea, sem alteração formal e com concordância de género feminina) ou *modelo* (homem ou mulher), que se distingue do anterior por desencadear concordância masculina ou feminina (*o/a modelo*), apenas em sintaxe. O facto de os sufixos derivacionais determinarem, na sua maioria, o género gramatical dos nomes (*fritura*_{-masc}, *alarmismo*_{+masc}) e de, nos nomes compostos (morfo)sintáticos, ser o primeiro nome que determina o valor de género do composto – [[homem]_{+masc} [rã]_{-masc}]_{+masc}, [[carruagem]_{-masc} [restaurante]_{+masc}]_{-masc} – mostra também que a categoria «género» é basicamente lexical e morfológica (Rio-Torto no prelo, Villalva no prelo b). Assim, e apesar de a relação entre o comportamento morfológico dos nomes *mestre*, *mestra*, *primo* e *prima*, etc. e o traço semântico [+anim] não ser generalizável, em português, este grupo de nomes mostra que a natureza morfológica de *-e/-o* e *-a* pode extrapolar a função de marcador temático e concentrar a informação relativa a duas categorias, classe temática e género. Nesta situação, o índice temático caracteriza-se como um constituinte morfo-

lógico *portmanteau*, típico das línguas tipologicamente flexionais. Este facto implica tomar decisões quanto a considerar-se que *MESTRE* e *MESTRA*, por exemplo, são dois lexemas independentemente listados no léxico ou, alternativamente, que existe um grupo de nomes que admitem flexão de género, como defendido na gramática tradicional e em alguns trabalhos sobre a morfologia do género. Porque se considera que «género» é uma categoria cujo valor está especificado na entrada lexical da grande maioria dos nomes, não tem relação constante com a forma do índice temático (propriedade que os distingue dos adjetivos do tipo *profundo~profunda*) e, sobretudo, porque aos nomes temáticos corresponde um lexema com uma estrutura equivalente ao tema, a primeira hipótese é aquela que parece mais adequada, permitindo uma descrição englobante e dando conta de comportamentos com alto grau de generalização em português (Villalva 2000; Mota no prelo a). Em suma, o traço [+anim] deve ser considerado na análise morfológica, dada a particularidade de certos nomes com esse traço apresentarem expoentes cumulativos, mesmo que para se concluir que a sua interferência na morfologia não é sobremaneira determinante.

O marcador morfológico de tempo-modo-aspeto (TMA), na flexão verbal, é outro exemplo de expoente cumulativo que implica relação entre morfologia e semântica (Mota no prelo b). Através de constituintes flexionais (morfofossintáticos e, logo, relevantes também para a sintaxe), situados numa estrutura verbal à esquerda da flexão de pessoa-número, a morfologia do verbo disponibiliza componentes apropriados para veicular informação sobre (i) a localização de uma situação, relativamente a outra, no tempo cronológico, (ii) a perspetivação dessa situação, (iii) o grau de possibilidade que o falante lhe atribui (ou seja, a modalidade) e ainda (iv) o aspeto, tempo interno da situação. Relativamente ao primeiro tipo de informação, anterioridade, posterioridade e sobreposição são relações entre situações, descritas pelo verbo, associadas à categoria semântica do tempo. Quanto ao segundo caso, ainda relacionado com tempo, a perspetivação da situação pode colocá-la num tempo coincidente com o tempo da enunciação, o presente, ou no passado ou no futuro. Os valores modais (associados ao modo morfológico indicativo e conjuntivo) e aspetuais – de duratividade ou homogeneidade, por exemplo –, também semânticos, que o verbo pode veicular correspondem aos dois últimos tipos de informação acima focados – cf. 723 O modo conjuntivo.

Numa língua fusional, como o português, as três categorias que correspondem a tempo, modalidade e aspeto estão morfológicamente amalgamadas. Os sufixos de TMA veiculam tipicamente as três em simultâneo, configurando assim constituintes *portmanteau*. Noutras línguas, de outra tipologia, essas categorias podem ser asseguradas por dois ou mais marcadores, adjacentes e independentes (Stump 2001, 139). O TMA codifica assim três categorias distintas mas fortemente associadas. Por exemplo, *-a*, em *os malabaristas construíam a pirâmide quando um deles caiu*, permite que a situação seja dada como localizada num tempo passado relativamente ao tempo da enunciação, com início anterior e parcialmente coincidente com outra situação (a

queda de um dos malabaristas), que seja considerada como possível ou realmente ocorrida e como implicando duratividade.

As formas de TMA que as estruturas verbais apresentam identificam e distinguem entre si os geralmente chamados «tempos verbais» dos modos indicativo e conjuntivo. Trata-se, assim, de um constituinte muito relevante, do ponto de vista formal e do ponto de vista das interfaces que a morfologia estabelece. Contudo, e pelo que acaba de se ilustrar, a designação «tempo» é inadequada, porquanto não traduz a complexidade de categorias implicadas nas formas verbais (Mota no prelo b). Acresce que nem todos os paradigmas têm igual grau de informação temporal. Por exemplo, o paradigma do pretérito perfeito admite pouca variação quanto ao tempo das situações que veicula; mas os paradigmas do imperfeito ou do condicional, por terem uma forte componente modal e/ou aspetual, são mais flexíveis quanto à localização temporal das situações. Basta considerar casos como *chegámos a tempo; uma hora mais e o murete caía* (situação localizada no passado mas projetada no futuro, relativamente ao tempo da oração anterior, e equivalendo a *teria caído* – o murete efetivamente não caiu), *o murete caía cada vez que chovia* (situação regular e habitual ocorrida no passado) ou *o murete caía, se cavasses a terra* (situações correlacionadas, imperfeito na oração antecedente de uma oração condicional, equivalendo a *cairia*), por exemplo (Oliveira 2013).

Note-se, por fim, nesta breve abordagem da interface morfologia-semântica, no verbo, que nem todos os paradigmas apresentam um constituinte morfológico de TMA, como no caso do pretérito perfeito (*formei, formaste*, por exemplo, correspondem a $FORM]_{RAD} a]_{VT} i]_{PN} 1.^{a}sg$ e a $FORM]_{RAD} a]_{VT} ste]_{PN} 2.^{a}sg$, respetivamente), diferentemente do imperfeito ou do condicional (vejam-se *formava, formavas*, correspondendo a $FORM]_{RAD} a]_{VT} va]_{TMA} \emptyset]_{PN} 1.^{a}sg$ e $FORM]_{RAD} a]_{VT} va]_{TMA} s]_{PN} 2.^{a}sg$, respetivamente, e *formaria, formarias*, correspondendo a $FORM]_{RAD} a]_{VT} ria]_{TMA} \emptyset]_{PN} 1.^{a}sg$ e $FORM]_{RAD} a]_{VT} ria]_{TMA} s]_{PN} 2.^{a}sg$, respetivamente). Remete-se para Cunha (2013), Marques (2013), Oliveira/Mendes (2013), para mais informação sobre questões de aspeto, modo e modalidade, respetivamente.

No domínio da formação de palavras, as estruturas de composição levantam também questões interessantes, relativamente às interfaces que aí se cruzam. Nos compostos morfológicos, estão envolvidos dois ou mais radicais morfológicos unidos por uma vogal de ligação, foneticamente [ɔ] ou [i]:

- (5) a. $CARDI]_{RAD1} \emptyset]_{VASCULAR}]_{RAD2}]_{RAD} COMPOSTO] \emptyset]_{IT}$
cardiovascular
- b. $FRANC]_{RAD1} \emptyset]_{LUS}]_{RAD2} \emptyset]_{MALAI}]_{RAD3}]_{RAD} COMPOSTO] \emptyset]_{IT}$
franco-luso-malaio
- c. $SILV]_{RAD1} i]_{CULTUR}]_{RAD2}]_{RAD} COMPOSTO] a]_{IT}$
silvicultura

Cada radical é portador de acento próprio e a vogal de ligação é tipicamente átona, o que conduz a que um composto de dois ou três radicais possua um acento principal e

um ou dois acentos secundários; nos exemplos cardiovascular, silvicultura, franco-luso-malaio, o acento secundário está assinalado com sublinhado. A vogal de ligação é um recurso fonológico que permite que, do ponto de vista morfológico, uma palavra possa conter na sua estrutura dois ou três radicais associados de maneira a formar uma base única, o radical composto do lexema. Uma vez obtida a forma final daquele radical, pode ocorrer na periferia direita um índice temático, se o lexema o requerer (cf. (5b–c)). Do ponto de vista semântico, estruturas com estas propriedades morfológicas e fonológicas têm um núcleo semântico (cf. (5c)), ou mais de um (cf. (5a–b)), e uma leitura composicional, mais ou menos transparente; ou seja, a interpretação do composto decorre da significação dos elementos lexicais que o constituem e, frequentemente, a interpretação dos compostos morfológicos revela uma relação semântica de hiperonímia-hiponímia. Por exemplo, (5c) pode contrastar-se com *agricultura*, *piscicultura* ou *suinicultura*; em todos os casos, a leitura adequada é ‘cultura de x’, sendo que em cada uma destas palavras o núcleo semântico (o hiperónimo) é o segundo membro (Villalva no prelo b). Apesar de aqui sumariamente apresentada, a interface morfologia-semântica (com intervenção da fonologia, em alguns casos) é sem dúvida importante no domínio do processo de composição.

5 Comentários finais

Neste capítulo, visou-se contribuir para uma abordagem da morfologia nas suas interfaces com outros módulos da gramática. Focaram-se alguns casos que mostram que a consideração das interfaces morfologia-fonologia, morfologia(-fonologia)-sintaxe e morfologia(-fonologia)-semântica contribui para uma descrição da palavra, objeto da morfologia, consentânea com a conceção de léxico adotada: nas entradas lexicais, articulam-se e «dialogam» traços que relevam de diferentes módulos da gramática. Numa língua de morfologia «rica», como o português, o módulo morfológico tem a possibilidade de codificar, maioritariamente de forma explícita, informação não só estritamente morfológica como relevando de outros domínios da gramática. Essa possibilidade, que o distingue de outras línguas, mesmo dentro da família românica, revela precisamente a confluência de informações de diferente tipo e as conexões que estas estabelecem no domínio do lexema, veiculadas através dos «canais estreitos de comunicação» entre módulos.

6 Referências

- Aalberse, Suzanne/Don, Jan (2011), *Person and number syncretisms in Dutch*, *Morphology* 21, 327–350.
 Abaurre-Gnerre, Maria Bernadete (1983), *Alguns Casos de Formação de Plural em Português: uma Abordagem Natural*, *Cadernos de Estudos Lingüísticos* 5, 127–156.

- Alcântara, Cíntia da C. (2003), *As classes formais do português e sua constituição: um estudo à luz da teoria da morfologia distribuída*, tese de Doutorado, PUC do Rio Grande do Sul, Porto Alegre.
- ALEPG, *Atlas Linguístico-Etnográfico de Portugal e da Galiza*, Centro de Linguística da Universidade de Lisboa, <http://www.clul.ul.pt/pt/investigacao/205-linguistic-and-ethnographic-atlas-of-portugal-and-galicia-alepg>
- Anderson, Stephen R. (2014), *Morphological Change*, in: Claire Bower/Bethwyn Evans (edd.), *The Routledge Handbook of Historical Linguistics*, London/New York, Routledge, 264–285.
- Anderson, Stephen R./Lightfoot, David (2002), *The Language Organ: Linguistics as Cognitive Physiology*, Cambridge, Cambridge University Press.
- Andrade, Amália (no prelo), *Sílaba*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 3, Lisboa, Fundação Calouste Gulbenkian.
- Andrade, Ernesto d' (1994), *Temas de fonologia*, Lisboa, Colibri.
- Aronoff, Mark (1994), *Morphology by itself*, Cambridge MA, MIT Press.
- Aronoff, Mark (2013), *Varieties of morphological defaults and exceptions*, ReVEL special issue 7, 84–97, www.revel.inf.br/eng
- Aronoff, Mark/Fudeman, Kirsten (2005), *What is Morphology?*, Oxford, Blackwell.
- Baerman, Matthew/Brown, Dunstan/Corbett, Greville G. (2009), *The Syntax-Morphology Interface. A Study of Syncretism*, Cambridge, Cambridge University Press.
- Bauer, Laurie/Lieber, Rochelle/Plag, Ingo (2013), *The Oxford Reference Guide to English Morphology*, Oxford, Oxford University Press.
- Bermúdez-Otero, Ricardo (2012), *The architecture of grammar and the division of labour in exponence*, in: Jochen Trommer (ed.), *The morphology and phonology of exponence*, Oxford, Oxford University Press, 8–83.
- Bisol, Leda (1998), *A nasalidade, um velho tema*, D.E.L.T.A. 14 (n. esp.), 27–46.
- Bisol, Leda (2001), *Ditongo Nasal: uma hipótese e duas leituras*, in: Ivo Castro/Inês Duarte (edd.), *Razões e Emoção: Miscelânea de estudos em homenagem a Maria Helena Mira Mateus*, vol. 1, Lisboa, INCM, <http://www.fl.ul.pt/dlgr/arquivo/mateus/15.ps> (30.09.2015).
- Booij, Gert (1997), *Autonomous Morphology and Paradigmatic Relations*, in: Gert Booij/Jaap van Marle (edd.), *Yearbook of Morphology*, Dordrecht, Kluwer, 35–54.
- Booij, Gert (2002), *The Morphology of Dutch*, Oxford, Oxford University Press.
- Booij, Gert (2005), *The Grammar of Words*, Oxford, Oxford University Press.
- Booij, Gert (2012), *Allomorphy and the architecture of grammar*, in: Bert Botma/Roland Noske (edd.), *Phonological Explorations. Empirical, Theoretical and Diachronic Issues*, Berlin/Boston, De Gruyter, 9–24.
- Booij, Gert (2015, ¹2010), *Morphological Analysis*, in: Bernd Heine/Heiko Narrog (edd.), *The Oxford Handbook of Linguistic Analysis*, Oxford, Oxford University Press, 449–471.
- Callou, Dinah/Moraes, João/Leite, Yonne (2013), *Consoantes em coda silábica*, in: M. Bernadete M. Abaurre (org.), *A construção fonológica da palavra*, vol. 3, in: Ataliba T. de Castilho (org.), *Gramática do português culto falado no Brasil*, São Paulo, Contexto, 167–194.
- Cardoso, Adriana/Carrilho, Ernestina/Pereira, Sandra (2011), *On verbal agreement variation in European Portuguese: syntactic conditions for the 3SG/3PL alternation*, *Diacrítica* 25:1, 135–158.
- Carrilho, Ernestina (2003), *Ainda a «unidade e diversidade da língua portuguesa»: a sintaxe*, in: Ivo Castro/Inês Duarte (edd.), *Razões e Emoção: Miscelânea de estudos em homenagem a Maria Helena Mira Mateus*, Lisboa, Imprensa Nacional – Casa da Moeda, 163–178.
- Carstairs-McCarthy, Andrew (1987), *Allomorphy in Inflection*, London, Croom Helm.
- Corbett, Greville G. (2000), *Number*, Cambridge, Cambridge University Press.
- Corbett, Greville G./Baerman, Matthew (2006), *Prolegomena to a typology of morphological features*, *Morphology* 16:2, 231–246.

- Costa, João/Silva, M. Cristina F. (2006), *Notas sobre a concordância verbal e nominal em português*, Estudos Linguísticos 35, 95–109.
- Cunha, Luis Filipe (2013), *Aspeto*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 17, Lisboa, Fundação Calouste Gulbenkian, 585–619.
- Duarte, Maria Eugênia L. (2000), *The loss of the Avoid Pronoun Principle in Brazilian Portuguese*, in: Mary A. Kato/Esmeralda V. Negrão (edd.), *Brazilian Portuguese and the Null Subject*, Frankfurt am Main, Vervuert, 17–36.
- Ferreira, Ana (2009), *Flexão de Número dos Nomes Terminados em Ditongo Nasal à luz da Fonologia Lexical*, dissertação de Mestrado, Universidade do Porto, <http://repositorio-aberto.up.pt/bitstream/10216/23174/2/tesemestanaferreira000093456.pdf> (30.09.2015).
- Galani, Alexandra/Hicks, Glyn/Tsoulas, George (edd.) (2011), *Morphology and Its Interfaces*, Amsterdam/Philadelphia, Benjamins.
- Goldrick, Matthew/Rapp, Brenda (2007), *Lexical and post-lexical phonological planning to articulatory processes: Evidence from tongue twisters*, Language and Cognitive Processes 21, 649–683.
- Hora, Demerval/Pedrosa, Juliene/Cardoso, Walcir (2010), *Status da consoante pós-vocálica no português brasileiro: coda ou onset com núcleo não preenchido foneticamente?*, Letras de Hoje 45:1, 71–79.
- Marques, Rui (2013), *Modo*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 19, Lisboa, Fundação Calouste Gulbenkian, 673–693.
- Martins, Ana Maria (1995), *A evolução das vogais nasais finais «-ã», «-õ», «-ê» no português*, in: Cilene da C. Pereira/Paulo R.D. Pereira (edd.), *Miscelânea de Estudos Lingüísticos, Filológicos e Literários In Memoriam Celso Cunha*, Rio de Janeiro, Editora Nova Fronteira, 617–646.
- Mateus, Maria Helena Mira (1982, ¹⁹⁷⁵) *Aspectos da Fonologia Portuguesa*, Lisboa, INIC.
- Mateus, Maria Helena Mira (2003), *A harmonização vocálica e o abaixamento de vogais nos verbos do português*, in: Fernanda Irene Fonseca et al. (edd.), *Língua portuguesa: estruturas, usos e contrastes*, Porto, Centro de Linguística da UP, 289–302, <http://ler.letras.up.pt/uploads/ficheiros/7314.pdf> (30.09.2015).
- Mateus, Maria Helena/Andrade, Ernesto d' (2000), *The Phonology of Portuguese*, Oxford, Oxford University Press.
- Mathews, Peter H. (1991, ¹⁹⁷⁴) *Morphology*, Cambridge, Cambridge University Press.
- Mota, Maria Antónia C. da (no prelo a), *Categorias nominais e adjetivais*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 3, Lisboa, Fundação Calouste Gulbenkian.
- Mota, Maria Antónia C. da (no prelo b), *Categorias verbais*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 3, Lisboa, Fundação Calouste Gulbenkian.
- Mota, Maria Antónia/Miguel, Matilde/Mendes, Amália (2012), *A concordância de P6 em português falado. Os traços pronominais e os traços de concordância*, *Papia* 22:1, 161–187.
- Oliveira, Fátima (2013), *Tempo verbal*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 15, Lisboa, Fundação Calouste Gulbenkian, 509–553.
- Oliveira, Fátima/Mendes, Amália (2013), *Modalidade*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 19, Lisboa, Fundação Calouste Gulbenkian, 623–669.
- Paster, Mary Elizabeth (2006), *Phonological Conditions on Affixation*, tese de Doutorado, University of California, Berkeley, http://www.linguistics.berkeley.edu/dissertations/Paster_dissertation_2006.pdf (30.09.2015).
- Ralli, Angela (2003), *Morphology in Greek Linguistics: The State-of-the-art*, *Journal of Greek Linguistics* 4, 77–130.
- Ralli, Angela (2007), *On the Role of Allomorphy in Inflectional Morphology: Evidence from Dialectal Variation*, in: Giandomenico Sica (ed.), *Open Problems in Linguistics and Lexicography*, Milano, Polimétrica, 123–147.

- Rio-Torto, Graça M. (no prelo), *Derivação*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 3, Lisboa, Fundação Calouste Gulbenkian.
- Rodrigues, Celeste (2012), *Variantes não-standard e tipo de discurso: (des)encontro de resultados*, in: Inês Duarte/Armada Costa (edd.), *Nada na linguagem lhe é estranho. Estudos em homenagem a Isabel Hub Faria*, Porto, Afrontamento, 215–228.
- Spencer, Andrew (2010), *Lexical Relatedness and the Lexical Entry – a Formal Unification*, in: Stefan Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Paris 7*, Stanford, CA, CSLI Publications, 322–340.
- Stump, Gregory T. (2001), *Inflectional morphology: a theory of paradigm structure*, Cambridge, Cambridge University Press.
- Stump, Gregory T. (2005), *Word-formation vs Inflectional Morphology*, Linguistics Faculty Publications, http://works.bepress.com/gregory_stump/22 (30.09.2015).
- Veloso, João (2005), *Estrutura interna e flexão de número dos nomes terminados em «-ão»: onde reside a «irregularidade»?*, in: Graça Rio-Torto et al. (edd.), *Estudos em homenagem ao Professor Doutor Mário Vilela*, vol. 1, Porto, FLUP, 325–338.
- Veloso, João (2007), *Schwa in European Portuguese: The Phonological Status of [i]*, in: Olivier Crouzet/Jean-Pierre Angoujard (edd.), *Actes des JEL'2007. Schwa(s). 5èmes Journées d'Études Linguistiques*, Nantes, Université de Nantes, 55–60.
- Villalva, Alina (2000), *Estruturas Morfológicas. Unidades e Hierarquias nas Palavras do Português*, Lisboa, Fundação Calouste Gulbenkian/FCT.
- Villalva, Alina (no prelo a), *Conversão*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 3, Lisboa, Fundação Calouste Gulbenkian.
- Villalva, Alina (no prelo b), *Composição*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 3, Lisboa, Fundação Calouste Gulbenkian.
- Zwicky, Arnold M. (1990), *Inflectional morphology as a (sub)component of grammar*, in: Wolfgang U. Dressler et al. (edd.), *Contemporary Morphology*, Berlin/New York, Mouton de Gruyter, 217–236.

Raquel Amaro e Sara Mendes

7 Lexicologia e linguística computacional

Abstract: Tendo por objetivo o processamento automático do significado das unidades lexicais, o cruzamento da lexicologia e da linguística computacional é uma área de investigação fundamental no domínio do processamento das línguas naturais. Neste contexto, a especificação lexical, i.e., a determinação das características e do nível de granularidade pertinentes para uma adequada representação do conhecimento lexical, é um aspeto essencial a definir, já que disso depende em grande medida a modelização desse conhecimento. Partindo de investigação desenvolvida para o português, o presente capítulo apresenta uma descrição de propriedades e fenómenos lexicais essenciais para a linguística computacional (e.g. relação entre propriedades semânticas e comportamento sintático, ambiguidade e criação de significado em contexto), uma breve discussão de dados empíricos que evidencia a adequação de modelos de representação do léxico e a integração dos resultados da investigação no quadro destes modelos na WordNet.PT.

Keywords: conhecimento lexical, modelização, *wordnet*, léxico generativo, português

1 A natureza do conhecimento lexical

O estudo do léxico desempenhou, desde sempre, um papel central na descoberta progressiva do funcionamento do sistema da língua e no aprofundamento do conhecimento linguístico. A lexicologia, área da linguística dedicada por excelência ao estudo do léxico, inclui, na definição que consideramos neste trabalho, tópicos como a etimologia, a morfologia, a variação do significado e a semântica lexical, i.e., o estudo dos fenómenos sincrónicos relacionados com o significado dos itens lexicais. Atualmente, em lexicologia, o estudo do léxico estende-se muito além da classificação cronológica dos significados no âmbito da lexicografia e da descrição das leis que regem a mudança semântica no âmbito da etimologia. De facto, hoje em dia, as regularidades identificadas são postas em relação com a utilização real da língua, por um lado, e com a natureza da mente humana e a conceptualização do conhecimento, por outro, associando-se assim também ao domínio da psicologia (Geeraerts 2010, 9).

Na verdade, o léxico é uma componente essencial das línguas e tem um papel central nas teorias linguísticas da atualidade e, naturalmente, na conceção e na teoria dos sistemas para o processamento das línguas naturais desde o seu aparecimento (Nirenburg/Raskin/Tucker 1987; Hobbs et al. 1987; Guthrie et al. 1996; etc.). O cruzamento da lexicologia e da linguística computacional é assim uma área de investigação fundamental no domínio do processamento das línguas naturais, já que tem como objetivo geral o processamento automático do significado das unidades lexicais (via-

bilizando, por conseguinte, a partir da combinação destas unidades básicas das línguas, também o processamento do significado das frases em que estas ocorrem), considerando a determinação do significado nuclear das unidades lexicais, a interação entre as suas propriedades semânticas e as suas propriedades sintáticas e a representação formal desse conhecimento, numa perspetiva sincrónica das línguas. Assim, o desenvolvimento de um conhecimento cada vez mais aprofundado e detalhado acerca das unidades lexicais e da forma como estas se relacionam entre si dentro do sistema que é o léxico, bem como o modo como é estabelecida a interface entre este módulo da gramática e os outros, é uma área de investigação em linguística com grande impacto, em particular no desempenho dos sistemas de processamento das línguas naturais.

1.1 Unidades lexicais: unidades complexas e multifacetadas

O léxico é por excelência o repositório da informação que permite dar conta de fenómenos complexos como a criação de significado em contexto ou os diferentes tipos de ambiguidade lexical. Dar conta de fenómenos como estes, bem como do facto de o léxico ser um dos módulos da gramática mais dinâmicos – é o que mais alterações regista em períodos de tempo muito curtos, facto que se deve em grande parte à estreita ligação entre o léxico e a representação da realidade que se pretende comunicar e que está em constante evolução –, implica uma conceção de léxico fundamentalmente distinta da que está por detrás da construção de recursos lexicais tradicionais. Estes caracterizam-se tipicamente por uma conceção estática do léxico, de natureza enumerativa, ou seja, em que o conjunto dos sentidos possíveis de cada unidade lexical, bem como os seus contextos de uso, é determinado de antemão, não existindo quaisquer mecanismos que permitam expandi-lo.

A observação de que a construção de significado em contexto apresenta regularidades, e que estas dependem de propriedades particulares que são comuns a determinados conjuntos de itens lexicais, faz com que conceções alternativas de léxico descrevam as unidades lexicais já não como unidades atômicas e indivisíveis, mas antes como unidades caracterizadas por diferentes aspetos de significado, ou facetas, por propriedades diversas e relacionadas com diferentes níveis de representação, que podem ser descritas individual e independentemente umas das outras, embora frequentemente se relacionem e influenciem entre si.

Conceber as unidades lexicais como unidades complexas e multifacetadas, que codificam um leque de aspetos representativos do significado lexical, permite por um lado uma descrição mais adequada dos fenómenos lexicais e por outro uma cobertura mais abrangente dos dados observados. Na verdade, neste tipo de conceção de léxico, estes componentes de significado definem simplesmente os limites semânticos do seu uso. Quando consideradas no contexto de outras palavras, no entanto, facetas mutuamente compatíveis na informação semântica de cada unidade lexical tornam-se mais proeminentes, determinando interpretações específicas.

Finalmente, esta concepção de léxico e de unidade lexical permite descrever de forma objetiva aspetos que presidem à constituição de classes lexicais e que determinam regularidades no comportamento sintático-semântico de conjuntos de itens lexicais.

1.2 Ambiguidade e construção de significado em contexto

A organização tradicional dos recursos lexicais assume que o significado das unidades lexicais pode ser definido exaustivamente por um número determinado de sentidos. O resultado desta abordagem é que sempre que a interpretação de um enunciado é confrontada com uma situação de ambiguidade lexical é necessário usar um qualquer mecanismo de desambiguação que procura seleccionar da lista de sentidos enumerados a definição disponível mais adequada, baseando-se no estabelecimento de correspondências entre a caracterização dos sentidos de uma unidade lexical e fatores relacionados com o contexto.

Uma das principais desvantagens deste tipo de abordagem decorre da necessidade de se especificar de antemão os contextos em que uma determinada unidade lexical pode ocorrer, sendo que uma descrição incompleta resulta numa cobertura deficiente do sistema. Para além de tudo o mais, esta concepção dos recursos lexicais revela uma visão estática do léxico, já que a divisão do significado dos itens lexicais em sentidos independentes bloqueia a permeabilidade de sentidos e não permite dar conta do uso criativo de palavras em novos contextos.

Modelos de léxico alternativos, em que a caracterização do significado das unidades lexicais é por natureza aberta, permitem ultrapassar as limitações das concepções tradicionais de léxico referidas acima. Isto é conseguido através de esquemas de codificação explícita do conhecimento lexical em diferentes níveis de generalização, em particular fazendo com que a resolução de ambiguidades lexicais seja uma parte integrante de um procedimento de análise semântica uniforme. Os mecanismos integrados nestes procedimentos de análise operam sobre um conjunto de sentidos básicos, com mais estrutura interna do que o que era tradicionalmente assumido, permitindo alargar o conjunto de sentidos das unidades lexicais individuais quando estas são consideradas em conjunto em expressões mais abrangentes. Desta forma, este tipo de modelo de léxico não pressupõe uma enumeração de sentidos finita para representar o conhecimento lexical e permite dar conta do uso criativo da linguagem.

Tendo em conta os aspetos abordados nesta secção, torna-se evidente que a natureza do conhecimento lexical é, portanto, complexa e dinâmica, envolvendo propriedades conceptuais (modelização mental do conhecimento do mundo), semânticas (estrutura do significado, relações temáticas, restrições combinatórias e relações entre as diferentes unidades que compõem o sistema do léxico) e sintáticas (realização sintática, (inter)dependência e relação entre significado e contexto). Para dar conta destas propriedades do léxico surgiram diversos modelos computacionais para a sua representação, cujas características principais são apresentadas abaixo, neste capítulo.

2 Modelos computacionais de representação do léxico

Os modelos computacionais de representação do léxico têm como objetivo a determinação e a descrição do significado e das propriedades das unidades lexicais de modo a permitir a utilização desta informação no processamento automático das línguas. Partindo de princípios estruturalistas, estes modelos assumem a independência do conhecimento linguístico relativamente ao conhecimento do mundo, e a ideia de que a formalização desse conhecimento é possível, e podem ser genericamente agrupados em dois tipos de abordagens: as abordagens decomposicionais e as abordagens relacionais.

As abordagens ditas decomposicionais (Bierwisch 1971; Wierzbicka 1972; 1996; Talmy 1985; Jackendoff 1990; Pustejovsky 1995) baseiam a descrição do significado das unidades lexicais em unidades mais pequenas, partilhando todas elas o interesse pela interação entre o léxico e a cognição num sentido mais lato, quer pela procura dos fundamentos cognitivos para as descrições do significado, quer pela procura dos fenómenos de interface entre a semântica e a informação contextual e não linguística. As abordagens denominadas relacionais (Mel'čuk 1988; 1998; Miller et al. 1990; Fellbaum 1998), por outro lado, têm como base o significado da unidade lexical como um todo e procedem à descrição desse significado com base nas relações que as unidades lexicais estabelecem entre si ou entre si e unidades de outros níveis, como, por exemplo, as unidades sintagmáticas.¹

De entre os modelos computacionais de representação do léxico, destacam-se o Léxico Generativo (Pustejovsky 1995), pela proposta de uma representação formal do significado das unidades lexicais e pelo claro objetivo computacional que está na base da construção do modelo e dos diferentes mecanismos propostos para a construção do significado das unidades linguísticas, e o modelo da WordNet (Miller et al. 1990; Fellbaum 1998), pela adequação à produção de recursos lexicais computacionais de grande escala, viabilizando a sua usabilidade em tarefas de processamento das línguas naturais.

2.1 Modelos generativos de léxico: geração de significado em contexto a partir das propriedades das unidades lexicais

O modelo do Léxico Generativo (Pustejovsky 1995) vê o léxico como um sistema complexo e dinâmico que permite a geração de significado em contexto a partir da

¹ Para uma visão mais aprofundada da evolução das teorias e modelos de léxico, ver Geeraerts (2010).

informação associada a cada item lexical e de um pequeno conjunto de mecanismos que estabelecem princípios gerais para combinar e selecionar essa informação ou parte dela. Este modelo contrasta, assim, em vários aspetos com os modelos enumerativos do léxico. Em primeiro lugar, distingue-se destes na medida em que a unidade básica do modelo é o item lexical com significado definido, mas subespecificado, podendo integrar diferentes tipos de ambiguidade lexical. Este modelo recupera, assim, a noção de significado base de Bréal (1897). De acordo com este autor, o significado base corresponde ao conhecimento partilhado pela comunidade linguística que permite a utilização eficiente da língua. A partir da definição deste significado base subespecificado, que se concretiza em cada contexto específico, procura-se prever e explicar o comportamento das unidades lexicais em contexto e o seu uso, sempre que tais propriedades advenham diretamente da informação associada a um dado item lexical. Em segundo lugar, e embora, tal como referido acima, a unidade básica do modelo do Léxico Generativo seja de facto o item lexical, esta unidade é concebida como uma unidade complexa, composta por diversos elementos mais pequenos que, em conjunto, formam o significado de cada unidade lexical. Esta forma de conceber os itens lexicais permite, por um lado, explicar a partilha de características entre diferentes unidades lexicais, em particular alternâncias de significado regulares, e, por outro, tornar a codificação lexical mais económica, na medida em que é possível especificar determinados tipos de informação apenas uma vez e depois partilhá-la entre diferentes itens lexicais. Neste sentido, o modelo do Léxico Generativo é um modelo decomposicional, em que a representação das unidades lexicais como estruturas informacionais – cujo conteúdo é decomposto em traços e valores – segue um conjunto definido de regras, permitindo formalizar a criação de significado em contexto e a interface entre sintaxe e semântica, tornando assim evidentes regularidades e princípios comuns ao nível do comportamento de grupos de itens lexicais.

O Léxico Generativo usa uma abordagem declarativa e sistemática para representar as unidades lexicais e dar conta do seu comportamento semântico e sintático. A informação lexical é representada através de matrizes de atributo-valor, cujos valores podem ser atômicos ou, recursivamente, ser constituídos por outras matrizes, estruturadas em 4 níveis de representação: estrutura argumental, estrutura eventiva, estrutura qualia e estrutura de herança lexical. A *estrutura argumental* contempla a definição das propriedades semânticas dos argumentos lógicos de um dado item lexical, bem como informação acerca do seu mapeamento sintático.² A *estrutura eventiva* é constituída pela declaração dos eventos constituin-

2 O Léxico Generativo contempla 4 tipos de atributos distintos: *argumentos próprios* (P-ARG_n) (parâmetros do conteúdo semântico do item lexical que correspondem ao objeto semântico denotado pelo item lexical em causa), *argumentos verdadeiros* (ARG_n) (parâmetros do conteúdo semântico do item lexical cuja omissão sintática só é permitida quando recuperável pelo contexto), *argumentos por defeito* (D-ARG_n) (parâmetros que entram nas expressões lógicas da estrutura qualia de um item lexical, mas que não são necessariamente expressos sintaticamente e que só podem ser expressos através de operações

tes do evento representado, por restrições de ordem temporal entre os subeventos listados e pela especificação do subevento mais proeminente. A *estrutura qualia* é composta por quatro atributos – papéis qualia – e respetivos valores: constitutivo (CONSTR) (que expressa a relação entre um dado objeto e os seus elementos constituintes); formal (FORM) (que estabelece as propriedades estáveis que distinguem um dado objeto dentro do seu domínio semântico); tético (TEL) (que diz respeito à função ou finalidade do objeto ou evento); e agentivo (AG) (que determina a origem ou cadeia causal do objeto ou do evento). Por último, a *estrutura de herança lexical* estabelece as condições para que haja herança lexical, através de vetores qualia, ou seja, de ligações entre os nós na rede de tipos estabelecidas de acordo com os papéis qualia. Finalmente, é importante referir que o modelo permite a representação da partilha de informação entre estruturas através de um mecanismo de unificação.

No entanto, a complexidade e riqueza da informação associada aos itens lexicais no Léxico Generativo não é, por si só, suficiente para dar conta do seu comportamento dinâmico em contexto. Por isso, o Léxico Generativo propõe três mecanismos generativos: *coerção de tipos* (que consiste numa operação semântica que converte um argumento de um dado tipo no tipo esperado pela seleção argumental de um predicado, sem que tal implique alteração da realização sintática do argumento – Pustejovsky 1995; 2001; 2007); *ligação seletiva* (que permite dar conta da relação entre modificadores e entidades modificadas, permitindo ao modificador selecionar um argumento específico de entre o conjunto de valores declarados no conteúdo semântico do objeto modificado); e *cocomposição* (que permite completar o significado de um predicado com valores presentes no conteúdo semântico dos seus argumentos).

O Léxico Generativo propõe também a classificação dos itens lexicais em tipos semânticos, o que, para além de permitir uma estrutura de tipos organizadora do léxico, torna os mecanismos definidos adequados para exprimir as relações entre os objetos semânticos e entre estes e as correspondentes realizações sintáticas.

Tomemos as frases em (1) (adaptadas de Mendes 2009, 125) como exemplo e atentemos nas estruturas (simplificadas) do Léxico Generativo que as representam, em (2). Estes exemplos ilustram de forma simplificada os níveis de representação do Léxico Generativo e a forma como as representações neste modelo permitem dar conta, de forma sistemática e simples, de restrições semânticas que explicam restrições de coocorrência dos itens lexicais.

- (1) a. A Ana escamou o peixe.
 b. *A Ana escamou as escamas (do peixe).
 c. *A Ana escamou a ave.

de subtipificação ou de especificação do discurso) e *argumentos-sombra* (s-ARG_n) (parâmetros incorporados na semântica de um item lexical, que só podem ser expressos através de operações de subtipificação ou de especificação do discurso) (Pustejovsky 1995; Amaro 2009).

$$(2) \left[\begin{array}{l} \textit{peixe} \\ \text{EST-ARG} = [\text{P-ARG}_1 = x: \textit{animal}] \\ \text{QUALIA} = [\text{CONST} = \textit{tem_parte}(x, y: \textit{escama})] \end{array} \right]$$

$$\left[\begin{array}{l} \textit{ave} \\ \text{EST-ARG} = [\text{P-ARG}_1 = x: \textit{animal}] \\ \text{QUALIA} = [\text{CONST} = \textit{tem_parte}(x, y: \textit{pena})] \end{array} \right]$$

$$\left[\begin{array}{l} \textit{escamar} \\ \text{EST-ARG} = \left[\begin{array}{l} \text{ARG}_1 = x: \textit{humano} \\ \text{ARG}_2 = y \\ \text{S-ARG}_1 = z: \textit{escama} \end{array} \right] \\ \text{EST-EV} = \left[\begin{array}{l} \text{E}_1 = e_1: \textit{processo} \\ \dots \end{array} \right] \\ \text{QUALIA} = [\text{FORM} = \textit{retirar}(e_1, x, z, y)] \end{array} \right]$$

A declaração da propriedade «*tem_parte escama*», distintiva de *peixe*, por um lado, e a declaração de um argumento (de tipo sombra, isto é, cujo conteúdo é incorporado na semântica do predicado) de tipo semântico associado ao verbo *escamar*, por outro, prevê corretamente a gramaticalidade da frase em (1a) e a agramaticalidade das frases em (1b) e (1c).

2.2 Modelos relacionais de léxico: o léxico como um conjunto organizado de unidades inter-relacionadas

Como referido anteriormente, o modelo da WordNet (Miller et al. 1990; Fellbaum 1998) é atualmente o modelo computacional do léxico com maior relevo no que respeita à criação de léxicos computacionais e à utilização destes recursos em aplicações no domínio da linguística computacional (Hanks 2003). Para além disso, na medida em que teve origem em investigação acerca da organização do léxico mental, o modelo da WordNet tem também uma forte motivação psicológica.

As *wordnets* são bases de dados eletrónicas estruturadas como uma rede de relações entre os nós que constituem as suas unidades básicas, os *synsets* (designação que denomina conjuntos de sinónimos que lexicalizam um dado conceito, e cujo significado é definido, neste modelo, pelas relações estabelecidas entre cada um deles e os restantes nós da rede). As relações que determinam o significado de um *synset* são relações lexicais (sinonímia), léxico-conceptuais (hiperonímia/hiponímia, meronímia, etc.), de função ou papel temático (relação de agente; relação de instrumento-envolvido em; etc.), relações de oposição semântica (antonímia, quase-antonímia) e relações de causa (é causado por/causa, etc.).

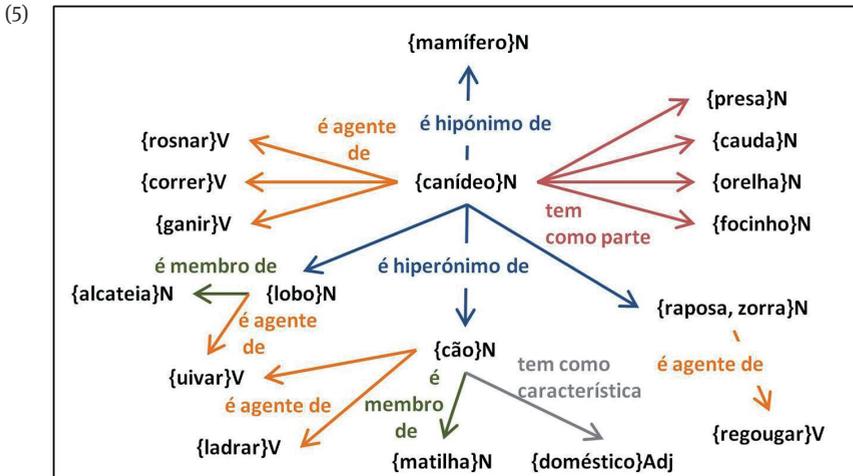
A relação estruturante de maior relevo no modelo da WordNet é a relação de hiperonímia/hiponímia, definida como:

- (3) X é hiperónimo de Y se Y é um tipo de X e X não é um tipo de Y.

A relação de hiperonímia/hiponímia é uma relação léxico-conceitual que contempla simultaneamente conhecimento do mundo, pelo seu carácter ontológico, e conhecimento linguístico, como se demonstra nos exemplos seguintes, em que o hiperónimo é usado para referir uma entidade mais específica (o hipónimo) previamente apresentada (Amaro 2009, 25):

- (4) a. Ele comprou um pastor alemão_{hipónimo}, mas o cão_{hiperónimo} não morde.
 b. Ele rastejou_{hipónimo} pela floresta, movendo-se_{hiperónimo} assim para evitar ser visto.
 c. Os animais marinhos_{hipónimo} são ameaçados pela proximidade de cidades, uma vez que os animais aquáticos_{hiperónimo} são facilmente afetados pela poluição dos esgotos.

A relação de hiperonímia/hiponímia permite organizar os itens lexicais de acordo com o seu tipo, já que cada hipónimo tem todas as propriedades do seu hiperónimo e mais algum aspeto que o caracteriza especificamente e o distingue de todos os seus co-hipónimos.³ Desta forma, a relação de hiperonímia/hiponímia contempla simultaneamente a definição de um mecanismo de herança monotónico – por oposição, por exemplo, ao descrito no modelo do Léxico Gerativo, acima – que permite a descrição dos itens lexicais de um modo muito económico, uma vez que os hipónimos herdam as propriedades conceptuais dos seus hiperónimos, como ilustrado na figura abaixo (adaptado de Marrafa et al. 2006).



³ Por co-hipónimos entendem-se todos os nós de uma rede léxico-conceitual que partilham um mesmo hiperónimo direto.

Através da relação de hiponímia/hiperonímia, na figura (5), os nós {lobo}_N, {cão}_N e {raposa, zorra}_N herdam todas as propriedades estabelecidas para o seu hiperónimo, {canídeo}_N: têm como parte {presa}_N, {cauda}_N, {orelha}_N, {focinho}_N; e são agentes de {rosnar}_V, {correr}_V e {ganir}_V.⁴ Estes três co-hipónimos, por sua vez, distinguem-se entre si e do seu hiperónimo pelas relações que lhes são particulares: por exemplo, {lobo}_N, para além de ser agente de {rosnar}_V, {correr}_V e {ganir}_V, propriedades herdadas do hiperónimo, é o único que é membro de {alcateia}_N; {cão}_N é o único que tem como característica ser {doméstico}_{Adj}, é agente de {ladrar}_V e membro de {matilha}_N; {raposa, zorra}_N é o único agente de {regougar}_V.

Esta breve apresentação das principais características destes dois modelos de léxico torna evidente a grande capacidade expressiva destas abordagens alternativas aos modelos tradicionais de léxico. Nas secções 3, 4 e 5 abaixo discutem-se alguns fenómenos empíricos concretos que levantam problemas em termos de codificação lexical e apresentam-se propostas de modelização no quadro destes dois modelos, demonstrando a capacidade dos mesmos de representar adequadamente fenómenos linguísticos complexos.

3 Modelização e especificação lexical

Como referido na breve descrição das características e relações existentes entre a lexicologia e a linguística computacional feita até aqui, a modelização do conhecimento lexical tem como objetivo a determinação do significado nuclear das unidades lexicais e a representação das propriedades semânticas e sintáticas que lhes estão associadas. Desse modo, a especificação lexical, ou seja, a determinação das características e do nível de granularidade pertinentes para uma adequada representação do conhecimento lexical, é um aspeto essencial e, por sua vez, indissociável dos modelos de representação do léxico.

⁴ Embora, tal como todos os canídeos, biologicamente as raposas rossem e ganam, este tipo de conhecimento, que está associado a um conceito taxonómico – *canídeo* –, é um conhecimento mais científico, nem sempre partilhado por todos os falantes, o que pode fazer com que a associação de *raposa* a estes dois eventos possa provocar alguma estranheza. Esta estranheza é decorrente do facto de, na rede de relações, coexistirem conceitos de léxico comum, que em princípio são partilhados por todos os falantes da língua, e conceitos cujo conteúdo semântico é de alguma forma mais técnico, e por isso não necessariamente dominado por todos os falantes.

3.1 Identificação de facetas de significado e definição de classes lexicais

A divisão do léxico em campos ou domínios semânticos (e.g. *alimentação, seres vivos, saúde, comunicação*, etc.) pode ser vista como um primeiro exemplo da utilização do significado para o estabelecimento de subconjuntos de itens lexicais com propriedades partilhadas. No entanto, a abordagem por domínio semântico (Trier 1931; 1934; Wittgenstein 1965; Vassilyev 1974; Lyons 1977), ou seja, a definição de conjuntos de itens lexicais com significados relacionados, distingue-se da definição de classes lexicais na medida em que na primeira, ao contrário do que acontece com a segunda, pode, objetivamente, não haver nenhuma propriedade, semântica ou sintática, partilhada entre membros de um mesmo domínio. Os casos de *tossir* (expelir violentamente ar pela garganta, com ruído) e *fôrceps* (instrumento em forma de pinça usado para puxar a cabeça do bebé para fora do útero da parturiente)⁵ ilustram este caso, já que estas duas unidades lexicais, embora pertençam ambas ao domínio da saúde, não partilham qualquer tipo de propriedade linguística.

A definição de classes lexicais, na perspetiva considerada aqui, implica uma partilha de propriedades semânticas e sintáticas entre os itens de uma mesma classe que permita obviar o processo da sua representação e a predição de propriedades e comportamentos generalizados.

Com base neste pressuposto, Mendes (2009) apresenta uma análise exaustiva dos adjetivos do português e propõe a definição de classes de adjetivos com base na identificação de facetas de significado. O primeiro aspeto considerado prende-se com o tipo de denotação que os adjetivos podem assumir: *adjetivos atribuidores de propriedades*, que denotam estados e cuja contribuição para a delineação da denotação do sintagma nominal resulta da adição de restrições à denotação, como exemplificado em (6); e *adjetivos não-restritivos*, que denotam uma função semântica (negação, probabilidade, etc.) e cuja contribuição para a denotação do sintagma nominal opera ao nível da intenção do nome modificado, funcionando como operadores modais, como exemplificado em (7).

(6) a maçã luzidia (x é maçã e x é luzidio)

(7) o diamante falso (x não é diamante)

A classe dos adjetivos atribuidores de propriedades contempla a maioria dos adjetivos do português, contendo várias subclasses que refletem propriedades semânticas distintas. Por exemplo, é possível agrupar os adjetivos *descritivos*, adjetivos atribuidores de propriedades denotadores de estados que correspondem a uma única propriedade, que selecionam e modificam um determinado aspeto do significado do

5 Exemplos retirados de Marrafa et al. (2006).

nome, tal como a cor ou o tamanho (ver (8) e (9)). No entanto, ainda dentro da classe dos adjetivos descritivos, é possível distinguir duas subclasses distintas relativamente à forma como a denotação do sintagma nominal é estabelecida: no caso dos adjetivos como *vermelho*, adjetivos *absolutos*, a denotação do sintagma nominal é resultado de uma interseção de conjuntos (*vestido vermelho* = vestido \cap vermelho, i.e., um objeto que pertence simultaneamente ao conjunto dos vestidos e ao conjunto das entidades que são vermelhas); no caso dos adjetivos como *pequeno*, adjetivos *relativos*, a denotação do sintagma nominal consiste num conjunto que está incluído num outro mais abrangente, que funciona, no fundo, como uma classe de comparação (*elefante pequeno* = elefante pequeno \subset elefante, i.e., um objeto que pertence ao conjunto das entidades que são pequenas no universo dos elefantes). Os elementos da classe dos adjetivos relativos partilham, assim, a propriedade de recuperarem do conteúdo do nome modificado o universo ou a classe de comparação que permite a determinação completa da sua denotação.

- (8) o vestido vermelho (o vestido cuja cor é vermelha)
- (9) o elefante pequeno (o elefante cujo tamanho é pequeno (para um elefante))

Finalmente, dentro dos adjetivos atribuidores de propriedade pode ainda considerar-se a subclasse dos adjetivos *relacionais*. Estes distinguem-se dos descritivos pelo facto de não estarem associados a uma propriedade individual, mas sim a conjuntos de propriedades. Tal como acontece com os nomes, os adjetivos relacionais denotam propriedades mais complexas e normalmente estabelecem com o nome modificado relações semânticas mais diversificadas.

- (10) a. o orçamento municipal (x é um orçamento que tem uma relação com o município)
 b. o animal marinho (x é um animal que pertence/é parte do mar)

De facto, os adjetivos relacionais introduzem conjuntos de propriedades que tipicamente correspondem a propriedades que caracterizam um nome (ver exemplos em (10)) e determinam relações semânticas mais complexas e que podem ser de diferente ordem. Assim, em (10a), o adjetivo determina uma relação subespecificada (R_1) entre o nome modificado e *município* (orçamento(x) \wedge $R_1(x, \text{município})$), fortemente dependente do contexto: *orçamento municipal* (disponível para o município), *obras municipais* (realizadas pelo município/financiadas pelo município). Em (10b), a relação determinada pelo adjetivo é uma relação de pertença ou quase-meronímia (animal(x) \wedge x \in mar), que se distingue, por exemplo, do adjetivo relacional quase-sinónimo *marítimo* ($R_1(x, \text{mar})$).

Esta breve descrição empírica do caso dos adjetivos põe, deste modo, em evidência a possibilidade de estabelecer classes lexicais com base na identificação de um pequeno conjunto de propriedades semânticas suscetíveis de serem formalizadas. É importante também não perder de vista o facto de os itens lexicais que partilham estas propriedades evidenciarem um comportamento linguístico análogo,

nomeadamente em termos de distribuição sintática, tal como apresentado em detalhe na secção 3.2.

3.2 Relação entre facetas de significado e comportamento sintático

As facetas de significado, que correspondem a propriedades semânticas, estão também intimamente ligadas ao comportamento sintático das unidades lexicais. As restrições de seleção argumental, por exemplo, são o caso que primeiramente ilustra esta relação: *escrever* seleciona um argumento de tipo humano como agente, por exemplo, o que explica a gramaticalidade das frases em (11a) e a malformação semântica das frases em (11b).

- (11) a. O rapaz/padeiro escreveu a carta.
b. #A lombriga/bicicleta escreveu o testamento.

Retomando os adjetivos descritos na secção anterior, é possível verificar que as propriedades que permitem definir classes lexicais estão diretamente relacionadas com propriedades sintáticas. Por exemplo, os adjetivos não-restritivos, por não denotarem estados que correspondam a propriedades, não ocorrem com modificadores adverbiais, nem em construções comparativas (exemplos em (12) retirados de Mendes 2009, 49). Por seu lado, os adjetivos relacionais distinguem-se dos descritivos por dificilmente ocorrerem em contextos predicativos (13a), não ocorrerem em posição pré-nominal (13b) e não ocorrerem com advérbios de grau (13c) a não ser em casos excepcionais.⁶

- (12) a. *o diamante muito falso
b. *um diamante mais falso do que o outro
- (13) a. *As casas são rurais. *Adoro as casas que são rurais.
b. *Adoro as rurais casas.
c. ?*Adoro as casas muito rurais.

Propriedades semânticas relacionadas com a estrutura interna dos eventos e com facetas de significado incorporadas no conteúdo semântico das unidades lexicais podem, também, explicar diferentes comportamentos de itens lexicais semanticamente relacionados. Tomemos como exemplo os verbos de movimento do português. Os verbos que expressam movimento (i.e., a mudança de localização num dado

⁶ Há uma margem de aceitabilidade para as construções em que adjetivos relacionais coocorrem com advérbios de grau, aceitabilidade que está relacionada com a possibilidade de reinterpretar estes adjetivos como descritivos. Para uma discussão e explicação aprofundadas das condições excepcionais que permitem que os adjetivos relacionais possam ocorrer neste tipo de contextos, ver Mendes (2009).

intervalo de tempo) partilham, ainda que com restrições mais ou menos específicas, a sua estrutura argumental (Amaro 2009). Por exemplo, os verbos *mover-se* (alterar a própria localização) e *regressar* (mover-se de novo para o local de partida) selecionam um argumento de tipo animado. No entanto, não coocorrem necessariamente com os mesmos sintagmas adverbiais:

- (14) a. Ele moveu-se durante meia hora./*Ele moveu-se em meia hora.
 b. ?*Ele regressou durante meia hora./Ele regressou em meia hora.

Este diferente comportamento sintático pode ser diretamente relacionado com a estrutura interna do evento denotado por cada um dos verbos considerados que, por sua vez, pode ser explicada pelo seu conteúdo semântico. Recuperando a noção de incorporação semântica proposta por Talmy (1985), é possível descrever o conteúdo semântico do verbo *regressar* assumindo que este incorpora no seu significado o elemento semântico *DESTINO*, definido como a localização final do objeto que experiencia/participa no evento, descrição que é consistente com a definição intuitiva de *regressar*. Assim, *regressar* incorpora uma localização final, que determina um estado final que, adicionado ao evento denotado por *mover-se*, resulta num evento de tipo transição, com propriedades de *Aktionsart* distintas (Vendler 1967; Mória 2000).⁷ Estas propriedades permitem prever o diferente comportamento sintático destes verbos ilustrado em (14): uma atividade, evento não delimitado temporalmente, não permite a especificação de um intervalo temporal que pressupõe um estado final (*em meia hora*) – ver (14a); pelo contrário, um *accomplishment*, sendo um evento delimitado pelo ponto de culminação e consequente estado final, não permite a especificação de um intervalo temporal que não pressupõe esse mesmo estado final (*durante meia hora*) – ver (14b). Um outro exemplo de relação entre facetas de significado e comportamento sintático pode observar-se nas restrições de coocorrência entre co-hipónimos, aqui ilustrados em (15).

- (15) a. Ele subiu a rua correndo.
 b. *Ele subiu a rua descendo.

Os verbos *subir*, *correr* e *descer* constituem subtipos (hipónimos) do verbo *mover-se*: *subir* é mover-se para cima; *correr* é mover-se no solo, rapidamente, usando os membros; *descer* é mover-se para baixo. A relação de hiponímia entre itens lexicais reflete a lexicalização de diferentes componentes semânticos, o que também distin-

7 São quatro as classes de *Aktionsart*: *estado* – evento não pontual (ou seja, não delimitado temporalmente), homogêneo e simples (e.g. *ser alto*); *atividade* – evento não pontual simples relativamente homogêneo (i.e., com repetição de subeventos a intervalos regulares) (e.g. *correr*); *accomplishment* – evento não pontual, heterogêneo e complexo, composto por um processo preparatório, um ponto de culminação e um estado consequente (que, na tipologia de eventos usada no Léxico Generativo, corresponde *grosso modo* a uma transição) (e.g. *construir*); *achievement* – evento pontual, heterogêneo e complexo, constituído por um ponto de culminação e um estado consequente (e.g. *morrer*).

gue os nós irmãos (Fellbaum 1998). De acordo com Mendes/Chaves (2001) e Amaro (2009), esta distanciação de significado explica a incompatibilidade de alguns co-hipónimos. Assim, os itens co-hipónimos são compatíveis se não lexicalizarem valores distintos para um mesmo aspeto semântico. No exemplo acima, os verbos *subir* e *descer* lexicalizam diferentes valores (para cima e para baixo, respetivamente) para o mesmo aspeto, a DIREÇÃO, não sendo por isso compatíveis. Por seu lado, *correr* lexicaliza um valor para o aspeto MODO, sendo por isso compatível com *subir* e *descer*.

3.3 Codificação lexical e construção de significado em contexto

Para além da relação entre as facetas de significado e a definição de classes, por um lado, e o comportamento sintático das unidades lexicais, por outro, a modelização e a especificação lexicais têm de dar conta das propriedades polimórficas das unidades lexicais, em particular da construção de significado em contexto. Tal implica que os modelos de representação do léxico permitam uma codificação lexical adequada das propriedades dos itens lexicais, tendo também em conta a partilha dessas propriedades, de modo a garantirem a economia do processo de representação do conhecimento lexical e a utilidade e a viabilidade dos recursos lexicais resultantes.

Retomemos o caso dos adjetivos que, por definição, tendem a exibir grande plasticidade de significado, devido à sua subespecificação semântica. Por exemplo, o adjetivo *bom*, dependendo do nome com que ocorre, aparenta ter diferentes significados, como ilustrado em (16) (Mendes 2009):

- (16) a. um bom professor = um professor que ensina bem
 b. uma boa faca = uma faca que corta bem; uma faca bem construída

A codificação lexical das propriedades destes adjetivos no quadro do Léxico Gerativo permite, a partir de uma única entrada lexical, e através do mecanismo de ligação seletiva, dar conta da relação estabelecida entre o modificador e as entidades modificadas. A ligação seletiva, como o nome indica, consiste numa operação semântica que permite ao modificador selecionar um argumento específico de entre o conjunto de valores declarados no conteúdo semântico do objeto modificado. Deste modo, e atentando nas entradas lexicais representadas em (17), é possível verificar que o adjetivo *bom* tem um conteúdo semântico subespecificado (\cong avaliação positiva) cujo significado particular é definido em contexto, a partir da ligação estabelecida com um argumento presente na estrutura qualia (representado pelo índice 1) do nome que modifica. Assim, ao modificar *faca*, *bom* pode avaliar positivamente *construir* ou *cortar*; ao modificar *professor*, *bom* avalia positivamente *ensinar*.

$$(17) \left[\begin{array}{l} \textit{bom} \\ \text{EST-ARG} = \left[\text{ARG}_1 = \left[\begin{array}{l} \dots \\ \text{QUALIA} = \boxed{1} \end{array} \right] \right] \\ \text{EST-EV} = \left[E_1 = e_1 : \textit{estado} \right] \\ \text{QUALIA} = \left[\text{FORM} = \textit{avaliação positiva}(e_1, \boxed{1}) \right] \end{array} \right] \quad \left[\begin{array}{l} \textit{faca} \\ \text{EST-ARG} = \left[\text{P-ARG}_1 = x : \textit{utensilio} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{AG} = \textit{construir}(e_1, y, x) \\ \text{TEL} = \textit{cortar}(e_2, y, z, x) \end{array} \right] \end{array} \right] \\ \left[\begin{array}{l} \textit{professor} \\ \text{EST-ARG} = \left[\text{P-ARG}_1 = x : \textit{humano} \right] \\ \text{QUALIA} = \left[\text{TEL} = \textit{ensinar}(e_1, x, y) \right] \end{array} \right]$$

O caso dos adjetivos relativos, descritos na secção 3.1, constitui um outro exemplo de criação de significado em contexto, na medida em que estes adjetivos necessitam de recuperar do conteúdo do nome modificado a classe de comparação que permite a determinação completa da sua denotação. As matrizes de atributo-valor em (18) (Mendes 2009, 153) ilustram como esta plasticidade é codificada no Léxico Generativo, fazendo uso da estruturação da informação nas entradas lexicais e dos níveis de representação do modelo, por um lado, e do mecanismo de cocomposição,⁸ por outro, para representar adequadamente o significado «real» de *elefante pequeno*.

$$(18) \left[\begin{array}{l} \textit{pequeno} \\ \text{EST-EV} = \left[E_1 = e_1 : \textit{estado} \right] \\ \text{EST-ARG} = \boxed{1} \left[\text{ARG}_1 = x : \textit{entidade} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{CONST} = \textit{relativo_a_classe}(e_1, \boxed{1}) \\ \text{FORM} = \textit{dimensão_reduzida}(e_1, x, \boxed{1}) \end{array} \right] \end{array} \right] \boxed{1} \left[\begin{array}{l} \textit{elefante} \\ \text{EST-ARG} = \left[\text{P-ARG}_1 = x : \textit{paquiderme} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{CONST} = \textit{tem_parte}(x, y : \textit{tromba}) \\ \text{FORM} = \textit{grande}(e_1, x) \end{array} \right] \end{array} \right]$$

4 WordNet.PT: da investigação no quadro de diferentes modelos lexicais computacionais a um léxico computacional para o português

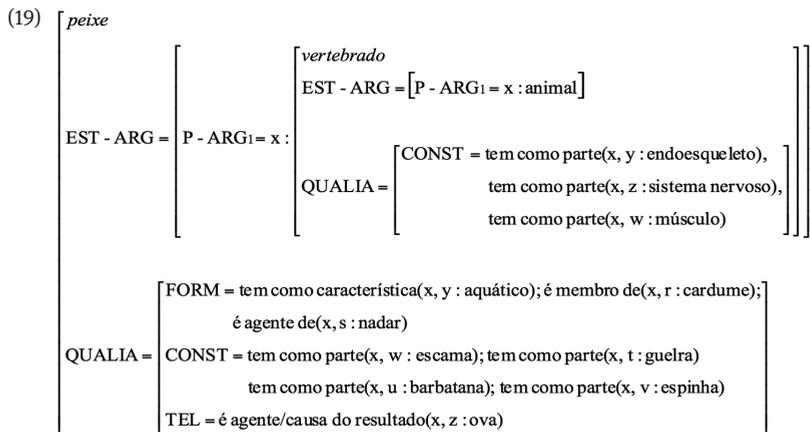
A WordNet.PT (Marrafa 2001; 2002; Marrafa et al. 2006), rede léxico-conceptual do português desenvolvida a partir da abordagem da EuroWordNet (Vossen 2002), espe-

⁸ Mecanismo que permite completar o significado de um predicado com valores presentes no conteúdo semântico dos seus argumentos.

lha um modelo que se distingue das propostas anteriores, WordNet e EuroWordNet, no que respeita à extensão do conjunto de relações usadas e às estratégias de cobertura lexical seguidas. A WordNet.PT foi inicialmente desenvolvida privilegiando a adequação dos resultados em detrimento da dimensão. Este facto, juntamente com a integração de investigação sobre as propriedades dos itens lexicais, motivaram a opção pela seleção, descrição e codificação manuais dos dados na WordNet.PT, resultando esta numa base de dados de menor dimensão, mas de grande fiabilidade e com maior densidade de relações, quando comparada com wordnets construídas automaticamente. O incremento da base de dados seguiu maioritariamente a abordagem por campos ou domínios semânticos, envolvendo a integração de itens lexicais de diferentes categorias gramaticais, o que motivou a necessidade do enriquecimento do modelo com mais informação e com a codificação de mais relações, em particular, relações transcategoriais (Mendes 2009; Amaro 2009; Amaro/Mendes/Marrafa 2010a; 2010b; 2013).

Atualmente, a WordNet.PT contém cerca de 19.000 entradas lexicais, cobrindo as principais categorias gramaticais de diversos domínios semânticos (e.g. alimentação, geografia, saúde, seres vivos), e contempla um conjunto de mais de 50 relações léxico-conceptuais, agrupadas em relações de equivalência, de oposição, geral/específico, todo/parte, categorização, participação em evento e definidoras da estrutura do evento.

Pelas semelhanças de organização conceptual e ontológica, a rede de tipos proposta no modelo do Léxico Generativo pode ser substituída, com vantagens, pela integração das próprias entradas lexicais numa rede léxico-conceptual, permitindo a formalização uniforme da herança lexical (Amaro 2009; Amaro/Mendes/Marrafa 2010a), ou seja, a explicitação das condições de partilha de propriedades semânticas – formalizadas através de ligações estabelecidas na rede – entre estruturas lexicais relacionadas por hiperonímia. Com base nesse pressuposto, e tendo em conta que as relações estabelecidas no modelo da WordNet são relações clara e estritamente definidas, a informação veiculada como valores dos papéis qualia é facilmente transposta para as relações disponíveis na WordNet.PT (Amaro/Mendes/Marrafa 2010a), mantendo a adequação de um mecanismo de herança que permita que apenas as inferências corretas sejam geradas. A WordNet.PT permite esta integração com modelos decomposicionais do léxico na medida em que substitui a herança monotónica genericamente assumida no modelo WordNet por um mecanismo de herança por defeito.



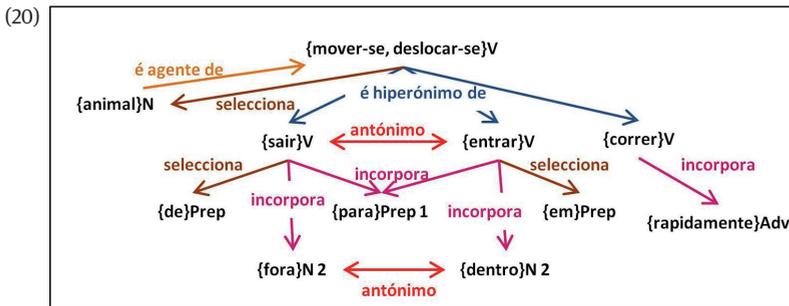
Retomando o exemplo de *peixe*, em (2), a propriedade *tem_parte*(*x*, *y*: *escama*), valor do papel constitutivo, é expressa na WordNet.PT pela relação de meronímia (tem como parte) estabelecida entre os *synsets* {*peixe*}_N e {*escama*}_N. Além disso, a declaração de uma relação de subtipo ‘*x*: *y* → *x* é hipónimo de *y*’ como valor do argumento próprio, na estrutura argumental, permite definir a estrutura de herança através da integração dos itens lexicais numa rede léxico-conceptual. Os argumentos próprios são, assim, ligados ao seu hiperónimo e respetiva estrutura informacional, dele herdando todas as propriedades, com exceção das que caracterizam especificamente o item lexical em questão, como referido na secção 2.2 e como ilustrado em (19).

Para além da motivação linguística deste tipo de representação do léxico, referida anteriormente e que será discutida em maior pormenor nas secções abaixo, é importante referir também as grandes potencialidades deste tipo de recurso lexical em termos dos seus usos computacionais graças ao grande leque de informação codificada e à sua estrutura. Neste contexto, para além da cobertura, a densidade dos léxicos relacionais é particularmente relevante. De facto, as *wordnets* têm vindo a ser usadas para resolver obstáculos de base que se colocam ao funcionamento de aplicações computacionais fiáveis que envolvam processamento de língua natural e, particularmente, o acesso à informação veiculada pela linguagem, tais como sistemas de busca e extração de informação, sistemas de tradução automática, sistemas de sumarização, sistemas de geração em linguagem natural ou aplicações de desambiguação do sentido das palavras, entre outros, já que todos estes diferentes sistemas de processamento de língua natural precisam de informação lexical rica para funcionarem corretamente.

4.1 Decomposição do significado e propriedades dos predicados verbais

O conteúdo semântico dos predicados verbais, pelas suas propriedades intrínsecas, caracteriza-se tipicamente pelas restrições argumentais, que contribuem significativamente para a definição do seu significado. No entanto, as relações disponíveis no modelo da EuroWordNet não permitem a representação adequada dos predicados verbais, nomeadamente a distinção entre propriedades de seleção argumental vs. propriedades de subcategorização – que nem sempre têm correspondência direta –, para além de não contemplarem a representação da incorporação de elementos semânticos, que permitem expressar as facetas de significado que diferenciam diversos co-hipónimos entre si, bem como hipónimos dos seus hiperónimos.

A WordNet.PT, com base em trabalho desenvolvido sobre verbos de movimento do português (Amaro 2009), integra, assim, as relações *seleciona/é selecionado* (argumentos verdadeiros), *incorpora/é incorporado* (argumentos-sombra) e *tem como argumento por defeito/é argumento por defeito* (argumentos por defeito), adaptadas do modelo do Léxico Gerativo, que se distinguem das relações de participação em evento já disponíveis (e.g. *é agente de*) pelo facto de expressarem propriedades de subcategorização (Amaro/Mendes/Marrafa 2010b; 2013). A figura abaixo ilustra esta proposta.



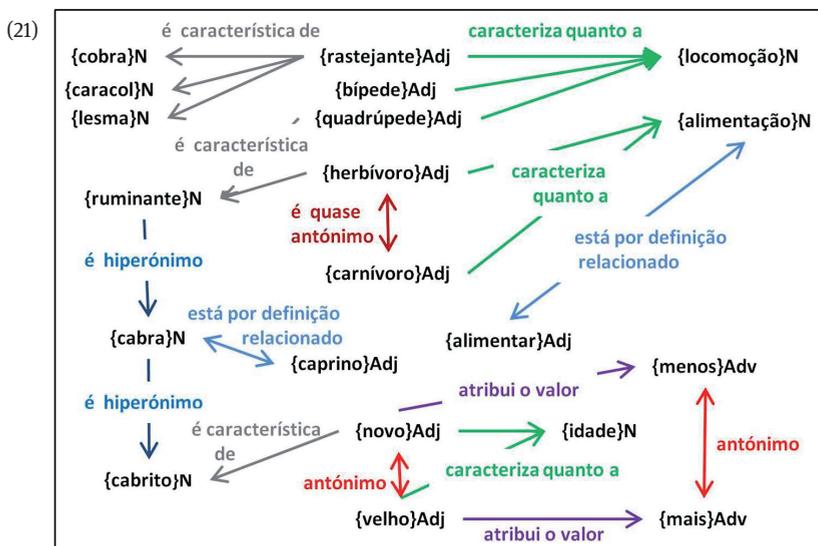
À semelhança da figura em (5), a rede representada em (20), que mostra uma rede protagonizada por *synsets* verbais, representa três hipónimos de {mover-se, deslocar-se}_V. Estes partilham com o hiperónimo a subcategorização de um argumento nominal, {animal}_N, que por sua vez tem a característica distintiva de ser agente de *mover-se*, característica que o diferencia do seu co-hipónimo {planta}_N, por exemplo. Os três co-hipónimos em causa distinguem-se entre si pelas suas propriedades de subcategorização (*sair* subcategoriza um argumento preposicional introduzido por *de*; *entrar* subcategoriza um argumento preposicional introduzido por *em*; *correr* não subcategoriza nenhum outro argumento) e pelos elementos que incorporam (*sair* incorpora o conteúdo semântico dos nós *para* e *fora*; *entrar* incorpora o conteúdo semântico dos nós *para* e *dentro*; *correr*, por seu lado, incorpora o conteúdo semântico do nó

depressa). Esta modelização permite, deste modo, expressar de forma objetiva e integrada a lexicalização de diferentes componentes semânticos, o que, como referido na secção 3.2, tem uma relação direta com restrições combinatórias dos itens lexicais.

4.2 Definição de classes de adjetivos com base em propriedades estruturais e semânticas

A investigação sobre o comportamento sintático-semântico dos adjetivos, parcialmente discutido na secção 3.1, e a sua modelização em léxicos relacionais (cf. Mendes 2009) determinou a definição de um pequeno conjunto de relações léxico-conceituais para representar os elementos desta categoria na WordNet.PT. Nesta secção apresenta-se brevemente a estratégia de modelização definida, que permite dar conta de diversos fenómenos semânticos de forma generalizada e sistemática, simultaneamente refletindo e tornando evidentes propriedades contrastantes das diferentes classes de adjetivos.

Embora a hiperonímia seja a principal relação estruturante no modelo da WordNet, a organização dos adjetivos é consideravelmente diferente. Na WordNet, os adjetivos descritivos organizam-se em *clusters* de *synsets*, refletindo princípios psicológicos de organização desta categoria no léxico mental (Miller 1998). No entanto, para além de esta estratégia levantar problemas ao nível da implementação, tal como discutido em detalhe em Mendes (2009), esta permite apenas representar adequadamente uma das classes de adjetivos, os adjetivos descritivos. Ora, as diferentes classes de adjetivos evidenciam propriedades distintivas muito salientes (cf. secção 3.1), que determinam a necessidade de estratégias de codificação adequadas às suas especificidades.



Neste contexto, na WordNet.PT definiu-se um pequeno conjunto de relações léxico-conceptuais⁹ para modelizar os adjetivos tendo em conta estes diferentes aspetos: *caracteriza quanto a*, que codifica a relação entre os adjetivos e os atributos que estes modificam; *atribui o valor*, que determina o valor do atributo introduzido por um determinado adjetivo; *está por definição relacionado com*, que estabelece a relação entre um adjetivo relacional e o nome que lexicaliza o conjunto de propriedades a que aquele está associado; e *é característica de*, que permite relacionar um adjetivo com um nome de que constitui uma propriedade distintiva, i.e. uma diferença específica relativamente aos seus co-hipónimos. A figura em (21) apresenta as relações na WordNet.PT para um pequeno conjunto de adjetivos de diferentes classes.

A rede representada apresenta adjetivos representativos de diferentes classes, pondo em evidência as diferentes estratégias de modelização usadas para representar adequadamente o seu significado e as suas propriedades semânticas específicas. Assim, um adjetivo descritivo como {novo}_{ADJ} relaciona-se com a lexicalização do atributo que modifica, {idade}_N, através da relação de *caracteriza quanto a*, bem como com o valor que especifica para este mesmo atributo, {menos}_{ADV}, neste caso, através da relação *atribui o valor*. Para além disso, exatamente por lexicalizarem valores, frequentemente opostos, de um mesmo atributo, os adjetivos desta classe evidenciam também relações de oposição semântica, codificadas no exemplo apresentado através da relação de antonímia entre {novo}_{ADJ} e {velho}_{ADJ}. Adicionalmente, ao estarem associados a uma propriedade única, estes adjetivos lexicalizam frequentemente diferenças específicas de outros nós da rede, tal como é o caso de {cabrito}_N no nosso exemplo, expressão que lexicaliza um tipo de cabra que se caracteriza por ter pouca idade, ou seja, por ser nova, aspeto que é expresso na rede através da relação *é característica de* entre {cabrito}_N e {novo}_{ADJ}. Os adjetivos relacionais, por outro lado, apresentam uma rede de relações significativamente menos densa, tipicamente apresentando apenas a ligação ao nome que lexicaliza o conjunto de propriedades a que estão associados (ver relação entre {caprino}_{ADJ} e {cabra}_N). Deste modo, graças à representação na rede da WordNet.PT das propriedades semânticas mais salientes de cada adjetivo, facto que resulta em contrastes significativos em termos de codificação, é possível fazer emergir do conjunto de ligações expressas na rede as diferentes classes de adjetivos.

5 Em suma

O breve itinerário pelos modelos generativos e relacionais do léxico apresentado neste capítulo põe em evidência a sua adequação para modelizar o conhecimento lexical, na medida em que dão conta de fenómenos lexicais complexos, essenciais para o

⁹ Algumas herdadas do modelo da EuroWordNet.

processamento das línguas naturais, e reflete deste modo as estreitas relações atualmente existentes entre a lexicologia e a linguística computacional.

A partir dos resultados de investigação sobre as propriedades distintivas de diferentes categorias morfossintáticas em português, apresentam-se propostas de especificação lexical para os itens lexicais estudados e demonstra-se a sua adequação através da sua integração na WordNet.PT.

6 Referências

- Amaro, Raquel (2009), *Computation of Verbal Predicates in Portuguese: relational network, lexical-conceptual structure and context*, tese de Doutoramento, Universidade de Lisboa.
- Amaro, Raquel/Mendes, Sara/Marrafa, Palmira (2010a), *Lexical-conceptual relations as qualia role encoders*, in: Petr Sojka et al. (edd.), *Text, Speech and Dialogue*, LNAI 6231, Berlin/Heidelberg, Springer, 29–36.
- Amaro, Raquel/Mendes, Sara/Marrafa, Palmira (2010b), *Encoding Event and Argument Structures in Wordnets*, in: Petr Sojka et al. (edd.), *Text, Speech and Dialogue*, LNAI 6231, Berlin/Heidelberg, Springer, 21–28.
- Amaro, Raquel/Mendes, Sara/Marrafa, Palmira (2013), *Increasing Density through New Relations and PoS Encoding in WordNet.PT*, *International Journal of Computational Linguistics and Applications* 4, 1–12.
- Bierwisch, Manfred (1971), *On classifying semantic features*, in: Danny Steinberg/Leon Jakobovits (edd.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, Cambridge, Cambridge University Press, 410–435.
- Bréal, Michel (1897), *Essai de sémantique: science des significations*, Paris, Hachette.
- Fellbaum, Christiane (ed.) (1998), *WordNet: An Electronic Lexical Database*, Cambridge MA, The MIT Press.
- Geeraerts, Dirk (2010), *Theories of Lexical Semantics*, Oxford, Oxford University Press.
- Guthrie, Louise, et al. (1996), *The Role of Lexicons in Natural Language Processing*, *Communications of the ACM* 30:1, 63–72.
- Hanks, Patrick (2003), *Lexicography*, in: Ruslan Mitkov (ed.), *The Oxford Handbook of Computational Linguistics*, Oxford, Oxford University Press, 48–69.
- Hobbs, Jerry, et al. (1987), *Commonsense Metaphysics and Lexical Semantics*, *Computational Linguistics* 13, 241–250.
- Jackendoff, Ray (1990), *Semantic Structures*, Cambridge MA, The MIT Press.
- Lyons, John (1977), *Semantics*, Cambridge, Cambridge University Press.
- Marrafa, Palmira (2001), *WordNet do Português – Uma base de dados de conhecimento linguístico*, Lisboa, Instituto Camões.
- Marrafa, Palmira (2002), *The Portuguese WordNet: General Architecture and Semantic Internal Relations*, *DELTA* 18:spe, 131–146.
- Marrafa, Palmira, et al. (2006), *WordNet.PT – Rede Léxico-Conceptual do Português 1.6*, Lisboa, CLG – CLUL, <http://www.clul.ul.pt/clg/wordnetpt/index.html> (20.10.2015).
- Mel'čuk, Igor A. (1988), *Semantic description of lexical units in an Explanatory Combinatorial Dictionary*, *International Journal of Lexicography* 1, 165–188.
- Mel'čuk, Igor A. (1998), *Collocations and lexical functions*, in: Anthony P. Cowie (ed.), *Phraseology. Theory, Analysis, and Applications*, Oxford, Clarendon Press, 23–53.

- Mendes, Sara (2009), *Syntax and Semantics of Adjectives in Portuguese: analysis and modeling*, tese de Doutorado, Universidade de Lisboa.
- Mendes, Sara/Chaves, Rui Pedro (2001), *Enriching wordnet with qualia information*, in: *Proceedings of the NAACL Workshop on WordNet and other Lexical Resources*, Pittsburgh, PA, North American Association for Computational Linguistics, 108–112.
- Miller, George A., et al. (1990), *Introduction to WordNet: An On-line Lexical Database*, International Journal of Lexicography 3:4, 235–244.
- Miller, Katherine J. (1998), *Modifiers in WordNet*, in: Christiane Fellbaum (ed.), *WordNet: an electronic lexical database*, Cambridge MA, The MIT Press, 47–68.
- Móia, Telmo (2000), *Identifying and computing temporal locating adverbials: with a particular focus on Portuguese and English*, tese de Doutorado, Universidade de Lisboa.
- Nirenburg, Sergei/Raskin, Victor/Tucker, Alien (1987), *The Structure of Interlingua in TRANSLATOR*, in: Sergei Nirenburg (ed.), *Machine Translation: Theoretical and Methodological Issues*, Cambridge, Cambridge University Press, 90–113.
- Pustejovsky, James (1995), *The Generative Lexicon*, Cambridge MA, The MIT Press.
- Pustejovsky, James (2001), *Type construction and the logic of concepts*, in: Pierrette Bouillon/Federica Busa (edd.), *The Syntax of Word Meanings*, Cambridge, Cambridge University Press.
- Pustejovsky, James (2007), *The Mechanics of Selection and Coercion in Grammar*, Paris, École Normale Supérieure.
- Talmy, Leonard (1985), *Lexicalization patterns: semantic structure in lexical forms*, in: Timothy Shopen (ed.), *Language typology and syntactic description*, vol. 3, Cambridge, Cambridge University Press, 57–149.
- Trier, Jost (1931), *Der deutsche Wortschatz im Sinnbezirk des Verstandes*, Heidelberg, Winter.
- Trier, Jost (1934), *Das sprachliche Feld: Eine Auseinandersetzung*, Neue Fachbücher für Wissenschaft und Jugendbildung 10, 428–449.
- Vassilyev, Leon M. (1974), *The theory of semantic fields: a survey*, Linguistics 137, 79–93.
- Vendler, Zeno (1967), *Linguistics in Philosophy*, Ithaca, Cornell University Press.
- Vossen, Piek (ed.) (2002), *EuroWordNet General Document*, EuroWordNet Project LE2-4003 & LE4-8328 report, Amsterdam, University of Amsterdam.
- Wierzbicka, Anna (1972), *Semantic Primitives*, Frankfurt, Athenaeum.
- Wierzbicka, Anna (1996), *Semantics. Primes and Universals*, Oxford, Oxford University Press.
- Wittgenstein, Ludwig (1965), *Philosophical Investigations*, New York, Macmillan.

João Paulo Silvestre

8 Lexicografia

Abstract: A lexicografia é uma área das ciências da linguagem, que analisa e propõe métodos de descrição do léxico, repercutindo os resultados de investigação de disciplinas contíguas como a lexicologia, a sintaxe, a linguística de corpus e a linguística histórica. Contempla também uma dimensão histórica, na medida em que trabalha sobre o corpus textual acumulado nos dicionários anteriores, refletindo sobre as descrições lexicais neles registadas e interpretando os dados de acordo com as especificidades da técnica lexicográfica usada em cada um desses dicionários. A dicionarística produzida em Portugal não apresenta a diversidade tipológica nem o apoio teórico e documental que se tem desenvolvido para a lexicografia do castelhano. Por outro lado, tem beneficiado do exemplo da lexicografia brasileira, que vem suprindo as falhas da descrição lexical especializada no que é comum a ambas as variedades. Este capítulo trata de aspetos que caracterizam e distinguem a dicionarística contemporânea em Portugal, nomeadamente a terminologia, as tipologias lexicográficas mais representadas, a técnica lexicográfica dos dicionários de informação linguística e o caso da onomástica como exemplo de um domínio do léxico que coloca dificuldades de descrição.

Keywords: terminologia lexicográfica, técnica lexicográfica, onomástica

1 Terminologia de lexicografia em português europeu (em confronto com o português do Brasil)

Na bibliografia e nos trabalhos académicos sobre lexicografia, publicados em Portugal e no Brasil, recolhe-se uma terminologia nem sempre convergente, com influências ora da tradição inglesa, ora da tradição francesa. As principais diferenças na terminologia em uso em Portugal e no Brasil verificam-se nos termos relacionados com a representação das unidades discursivas (*lexema, lema, forma-lema*) e com a estrutura dicionarística (*nomenclatura, entrada, artigo*).

Na segunda metade do século XX, nos dicionários de língua produzidos em Portugal, explicitou-se muito sumariamente a fundamentação teórica, apenas exposta em prefácios curtos ou breves indicações ao leitor. O *Dicionário da Língua Portuguesa Contemporânea* (DLPC) (2001), organizado pela Academia das Ciências de Lisboa, foi o primeiro dicionário moderno a apresentar uma extensa justificação da terminologia e técnica lexicográficas. No Brasil essa preocupação com a fundamentação teórica está presente em dicionários gerais preparados para o grande público (cf. o prefácio de Houaiss/Villar 2001). A lexicografia brasileira tem registado mais rápidos

progressos no que respeita à informatização da escrita dos dicionários e à constituição de corpora de extração, mas também na investigação em metalexicografia, com trabalhos de síntese sobre a terminologia da disciplina (Borba 2003; Biderman 2005; Isquardo/Finatto 2010; Bugueño Miranda 2014), que por sua vez incorporam a investigação desenvolvida para o castelhano (Alvar Ezquerro 1993; Porto Dapena 2002; Medina Guerra 2003; Haensch/Omeñaca 2004).

No caso de Portugal, foi mais sistemática a investigação sobre a história da lexicografia e a historiografia linguística, desde os dicionários do Renascimento até à intensa produção de dicionários bilingues no século XIX (Verdelho 1995; Verdelho/Silvestre 2007; 2011). Os primeiros trabalhos de descrição do património dicionarístico e de análise metalexicográfica seguem bibliografia francesa, adaptando e incorporando a terminologia de textos fundamentais como Quemada (1968), Imbs (1971) ou Rey (1977).

O corpus terminológico ampliou-se substancialmente com a produção profissional e normalizada de dicionários, com parâmetros afinados para a lexicografia que envolve a língua inglesa. A informatização da redação dos dicionários é um processo estreitamente relacionado com a expansão da linguística computacional e da terminografia: se estas duas disciplinas operam com muitos termos da lexicografia, também é verdade que são responsáveis pela ampliação da disciplina nos últimos 20 anos. A lexicografia de corpora admitiu naturalmente os termos que dão conta das relações entre unidades lexicais e, tal como a linguística computacional, assumiu como ponto fundamental a definição da unidade lexicográfica e os limites da palavra.

O conceito de *lexema* refere a unidade básica no estudo do léxico, que pode ser objeto de descrição num dicionário. O *lexema* é uma combinação de aspetos formais (a forma gráfica, a sua realização fonética, por exemplo) e de aspetos semânticos (a interpretação num determinado contexto gramatical). É também designado por unidade lexical (na terminologia inglesa, *lexical unit* ou *lexical item*).

O *lema* é a forma que se encontra registada no dicionário e que representa uma unidade lexical, correspondendo à redução, a uma forma canónica, de todas as variantes flexionais de uma palavra, mas também de expressões ou de combinações lexicais. Este processo de redução designa-se por *lematização* e é essencial para uma avaliação da frequência das palavras num corpus.

A acumulação de possibilidades terminológicas começa justamente na designação da forma de acesso à definição. Em lexicografia, *lema*, *entrada*, *forma de entrada* são possibilidades equivalentes para uma categoria que em inglês também tem vários termos: *entry*, *entry head*, *headword*, *keyword*, *lemma*, *main entry*, *word entry*.

O conjunto de elementos ordenados num dicionário que permitem o acesso à informação linguística designa-se por *macroestrutura*. Esse conjunto de elementos é a *nomenclatura* (no Brasil, *nominata*; *word list* ou *vocabulary* em inglês) e corresponde à lista de formas-*lema* que iniciam os artigos.

O termo *entrada* suscita equívocos. Na terminologia do português europeu equivale a *artigo* (cf. francês *article*), enquanto na terminologia em uso no Brasil significa o conjunto da informação fornecida acerca de um lema. No Brasil, *entrada* e *verbete*

são termos equivalentes (cf. inglês *entry*), admitindo-se pontualmente o termo *artigo* quando se trata de texto enciclopédico. *Verbete* está em desuso em Portugal e apenas significa o registo de informação reunida sobre um item lexical, para apoiar a redação do dicionário, seja sob a forma de fichas de papel ou em base de dados. É justamente um sentido metonímico de *verbete* que continuamos a encontrar na terminologia brasileira.

O *artigo* é uma unidade de informação, que se organiza numa *microestrutura* composta pela *forma-lemma* e pela *glosa*, sendo esta o conjunto de informações relacionadas com a forma-lemma. Na terminologia brasileira, a informação que preenche o artigo designa-se por *definição*. Em Portugal, reserva-se o termo *definição* para a parte do artigo que explica o significado; no Brasil *glosa* é uma definição muito sucinta, que pode complementar um equivalente num dicionário bilingue, por exemplo.

As terminologias diferem pontualmente em outros aspetos nos domínios da tipologia e da técnica lexicográfica, de novo em resultado de diferentes soluções de tradução ou adaptação dos termos ingleses.

Quadro 1: Diferenças na terminologia metalexigráfica entre Portugal e o Brasil.

Português europeu	Português do Brasil
Dicionário breve	Dicionário conciso, pequeno dicionário
Dicionário bilingualizado	Dicionário semibilingue
Nomenclatura, macroestrutura	Nominata, nomenclatura, macroestrutura
Artigo	Verbete
Entrada	Cabeça de verbete, palavra-entrada
Forma canónica	Forma básica
Glosa	Minidefinição
Remissão	Remissiva
Etiquetas	Marcas de uso

2 Tipologias lexicográficas

A investigação em tipologia lexicográfica trata da categorização dos dicionários, reconhecendo que diferentes línguas e culturas podem desenvolver tradições lexicográficas distintas. A normalização dos dicionários apoia-se nas descrições tipológicas, procurando estabelecer critérios que acomodem as tradições nacionais e a intercomunicação com dicionários similares de outras línguas.

Seguindo as características de uma determinada categoria, os dicionaristas asseguram não só a coerência e a usabilidade, mas também a atualização do próprio dicionário, que pode integrar a todo o momento novos materiais linguísticos disponí-

veis em formatos compatíveis. Por outro lado, na perspectiva do utilizador, o reconhecimento de categorias facilita o processo de consulta, pois é recorrendo a dicionários tipologicamente equivalentes que se podem encontrar informações complementares para a mesma necessidade. Por exemplo, um dicionário histórico e um dicionário etimológico, que são tipologicamente distintos, podem fornecer informações diferentes sobre a mesma palavra.

A tipologia mais tradicional, que remonta ao período das grandes enciclopédias do século XVIII, considera três categorias principais e tem como critério a presença ou ausência da informação linguística: dicionários de língua, dicionários de informação extralinguística e dicionários enciclopédicos, estes últimos combinando informação linguística e extralinguística. As tipologias mais recentes consideram a *função lexicográfica*, isto é, o objetivo com que o utilizador usa o dicionário: a resolução de dificuldades linguísticas, o acesso a conhecimentos gerais, a síntese do conhecimento em domínios específicos, etc. Os dicionários podem classificar-se como monofuncionais ou polifuncionais (Svensén 2009, 22) consoante ofereçam uma ou várias valências.

A lexicografia do inglês ensaiou, pelo menos desde o *Oxford English Dictionary* (OED), um conjunto de procedimentos de normalização técnica reconhecidos pela comunidade internacional de lexicógrafos, pelos quais se afere a qualidade e a tipologia dos dicionários das restantes línguas. Nesse sentido, espera-se que os dicionários de língua reflitam os progressos em linguística de corpus e que os dados sejam ponderados quanto à frequência, registo e periodização do uso.

Na maioria dos dicionários do português europeu não se pode garantir, por enquanto, que os dados refletem uma avaliação da frequência e uso na sincronia, pelo que o conceito de «dicionário de língua contemporânea» terá de ser admitido com algumas reservas sobre a descrição estatística e semântica da língua portuguesa desde o início do século XX.

Para as tipologias de dicionários de língua em português europeu, podemos partir de categorias estabelecidas por Hartmann/James (2001, 147–148), tentando aplicá-las aos dicionários existentes. Estes autores consideram uma primeira distinção entre dicionários gerais (que procuram uma descrição extensiva) e dicionários especializados (focados numa descrição restrita, incidindo sobre um domínio ou subdomínio). A segunda distinção diz respeito ao predomínio ou de informação linguística (sobre diferentes características do signo linguístico) ou de informação factual (descrição de realidades extralinguísticas). Estas categorias combinam-se em diversos graus de prevalência, descrevendo produtos dicionarísticos usuais, como os dicionários de língua com informações complementares de tipo enciclopédico, ou géneros mais focados na descrição linguística, como os dicionários terminológicos ou dicionários bilingues.

2.1 Dicionários gerais

A tradição lexicográfica do português europeu – e conseqüentemente a oferta editorial – apresenta pouca diversidade tipológica, com dicionários que procuram ser instrumentos polifuncionais, na medida em que dão informações de diversos tipos (semântica, etimologia, conhecimentos extralinguísticos), mas sem que nenhuma destas categorias seja sistematicamente abordada. Ou seja, são tipologicamente sincréticos, sem serem especializados em nenhuma categoria de informação. Os dicionários gerais procuram recolher várias categorias de informação linguística, sem todavia oferecer uma descrição especializada em cada uma dessas categorias. A qualidade de um dicionário geral pode ser avaliada pelo número de aspetos descritos na microestrutura do artigo:

- informação ortográfica (estrutura silábica, distribuição dos diacríticos);
- informação fonológica;
- etimologia (identificação do étimo) e dados diacrónicos (datação a partir de primeiras atestações; identificação de palavras em desuso, arcaísmos, neologismos);
- informação gramatical: morfologia, morfossintaxe, classes e categorias gramaticais;
- informação semântica: significado, definição, sinonímia, antonímia, paronímia, hiperonímia;
- informação sintática (valências sintáticas, estrutura e construção da frase);
- informação diatópica, diastrática, diafásica;
- integração em lexias complexas (estruturas combinatórias fixas, expressões idiomáticas, expressões formulares, provérbios).

Outro índice de qualidade é a inclusão de exemplos para a informação semântica e sintática, que podem ser exemplos redigidos pelo lexicógrafo, ou citações documentadas (*abonações*).

Os dicionários gerais classificam-se como monolíngues, bilingues ou multilíngues, consoante o número de línguas que são objeto de descrição. Os dicionários monolíngues são concebidos para utilizadores nativos da língua objeto, ou pelo menos com elevada proficiência. Nos dicionários bilingues, distingue-se entre a *língua de partida* que constitui a nomenclatura e a *língua de chegada*, que está presente em definições ou equivalentes.

Se na descrição linguística se supõe uma distinção entre língua materna (L1) e língua estrangeira (L2), o dicionário é unidirecional. Os dicionários que exploram a direção L1 > L2 são dicionários de codificação, usados como auxílio à produção de enunciados escritos e orais; inversamente, a orientação L2 > L1 resulta em dicionários de descodificação. Classifica-se como bidirecional um dicionário em que se pressupõe que o consulente tem uma competência de nativo quando acede à informação sobre cada uma das línguas em confronto. Dos dicionários bilingues espera-se uma descri-

ção das colocações que assinale as diferenças de seleção lexical, de que resultam os problemas de tradução ou equivalência entre L1 e L2. Na prática, como estes dicionários se baseiam em nomenclaturas monolíngues, não é incomum que prevaleça a perspectiva de uma das línguas em comparação. No seguinte exemplo, um artigo de um dicionário português-espanhol (PORTO EDITORA), exige-se uma competência linguística em L1 que torna o dicionário pouco adequado para falantes de L2.

- (1) *brilhante*
adjetivo de 2 géneros
 1. (*luz, cor*) brillante, reluciente, lustroso
 2. *figurado (notável)* brillante
 3. *figurado* brillante, prometedor

A intuição de falantes nativos e o conhecimento do léxico que suporta as equivalências são essenciais para a descodificação do artigo. Assim, sobre o adjetivo *brilhante*, percebe-se que em 1) *luz* e *cor* não são unidades coocorrentes, mas uma indicação dos domínios semânticos (interpretável como «referindo-se a luz e cor») em que *brilhante* está presente em estruturas combinatórias livres (*carro+brilhante; cabelo+brilhante*). Em 2), *notável* indica um quase-sinónimo que permite construções com a mesma interpretação (*um orador notável, um orador brilhante*). Por fim, em 3) parece referir-se a combinação fixa *futuro brilhante*, que equivale ao espanhol *futuro brillante, futuro prometedor*. Trata-se de uma interpretação intuitiva, pois parece ser o único contexto frequente em que *brilhante* pode ser traduzido por *prometedor* (sobre combinações pluriverbais em dicionários bilingues, cf. Iriarte Sanromán 2000, 155–160).

O mesmo dicionário, na direção espanhol–português, apenas indica a equivalência *brillante–brilhante*, inferindo-se que a palavra estrangeira tem o mesmo significado e os mesmos contextos de uso que a palavra portuguesa *brilhante*.

- (2) *brillante*
adjetivo de 2 géneros
 1. brilhante

Os exemplos confirmam que esta técnica de redação lexicográfica é especialmente útil para a codificação, permitindo ao consulente identificar os contextos de uso das palavras na língua de chegada, através de analogias com a língua de partida.

É no campo das expressões idiomáticas que os dicionários bilingues têm necessariamente uma redação bidirecional, pois os artigos são constituídos a partir da tradução de estruturas semanticamente equivalentes. Cf. os artigos *prometedor* (espanhol–português) e *promissor* (português–espanhol) em PORTO EDITORA:

- (3) *prometedor*
adjetivo
 promissor, prometedor
 un futuro prometedor *um futuro promissor*

promissor
adjetivo
 (futuro) prometedor
um futuro promissor
 un futuro prometedor

2.1.1 Caracterização da técnica lexicográfica

Tradicionalmente, a técnica lexicográfica dos dicionários do português europeu assenta numa descrição lexical condicionada pelas relações etimológicas e por hipóteses de explicação diacrónica da evolução semântica, que acumulam informações de diferentes momentos da diacronia e perturbam a descrição do português contemporâneo.

É de assinalar uma discrepância entre a prática lexicográfica do português europeu e as técnicas aplicadas na lexicografia do inglês e de outras línguas românicas, que seguem padrões descritivos que privilegiam a informação sincrónica. Além da opção explícita pela orientação prescritiva ou descritiva, é nas diversas categorias de informação linguística que os dicionários validam a mudança e a variação na língua, com admissão ou exclusão de formas não reconhecidas pelas gramáticas normativas.

A forma-lemma é representativa de um paradigma e pode variar de acordo com a língua e a tradição dicionarística, mas os utilizadores interpretam a forma-lemma como recomendação normativa. A fixação das formas-lemma segue a norma da língua escolarizada, com pouco espaço para a representação de variantes dialetais, ainda que estas sejam comumente reconhecidas pelos falantes. Em contrapartida, a tradição consagra a dicionarização de variantes lexicais de termos técnicos em que há diferenças na sílaba tónica, consoante se defenda uma etimologia grecizante ou latinizante (*abstemia* ou *asbtémia*, *erisipela* ou *erisípela*), ou admitindo interferências de outras línguas modernas (*absinto* ou *absíntio*, do fr. *absinthe* ou do lat. *absinthium*). Também estão integradas na norma variantes como *açaima/açaimo*, *eriçar/erriçar*, *assobio/assovio*, *escalpo/escalpe*; ou a variação em sufixos derivacionais (*bebedeiro/bebedouro*, *vindouro/vindoiro*; *vendedora/vendedeira*). Todavia, não há comentários sobre a distribuição das formas, do ponto de vista dialetal, diastrático ou diacrónico, pelo que são classificadas como equivalentes.

As formas dialetais, ainda em uso mas não representadas na escolarização da língua, são omitidas. Trata-se de formas reconhecidas pelos falantes e classificadas como erradas na perspetiva normativa, ainda que tenham lugar na história da evolução das palavras em português (*oivir* = *ouvir*; *coive* = *couve*). É certo que a ortografia normalizada pretende representar algumas variantes de pronúncia sob uma mesma configuração gráfica, como sucede nos casos em que a vogal *e* evoluiu para *a* quando seguida de consoante palatal (e.g. *lenha*, *cereja*). Todavia, a transcrição fonética num dicionário como o DLPC aponta apenas para a pronúncia padrão, com a vogal *a*; o mesmo se observa em *dezoito*, transcrito com *o* aberto, ou em *leite*, em que a possibilidade de monotongação não é considerada (cf. DLPC).

O lema é uma forma canônica que representa uma unidade discursiva. Na descrição lexicográfica, a forma-lema representa a palavra que é objeto de análise num artigo, bem como as unidades superiores à palavra que não têm artigo autônomo (expressões, colocações) e assim são convencionalmente indexadas. A inclusão das palavras compostas em posição de forma-lema é condicionada por critérios gráficos, pois tradicionalmente selecionam-se unidades que na escrita são delimitadas por espaço em branco. Esta opção introduz incoerências do ponto de vista da descrição lexical, pois podem encabeçar artigo as palavras compostas que admitem hifenização, por convenção ortográfica. Por exemplo, no DLPC, *pão* e *pão-de-ló* são formas-lema, mas não *pão ázimo* (cf. Quadro 2). Estas palavras compostas são remetidas para a microestrutura e são indexadas sob uma das palavras simples que as formam, de que resultam equívocos quanto à sua localização e, sobretudo, dúvidas quanto ao seu estatuto de palavras da língua. Analisando mais alguns exemplos de unidades pluriverbais lexicalizadas recolhidos no DLPC, constata-se que podem ser registadas em entrada ou subentrada:

Quadro 2: Localização de unidades pluriverbais lexicalizadas na estrutura do dicionário.

Subentrada (microestrutura)	Entrada (macroestrutura)
pão atrasado (= pão duro)	pão-de-ló = bolo fofo
pão ázimo (= pão sem fermento)	pão-duro = avarento
pão de Deus (= bolo com cobertura de creme de ovo e coco)	pão-por-Deus = peditério
pão de primeira (= pão feito com farinha de melhor qualidade)	
pão branco (= pão de trigo)	
pão de mistura (= pão de trigo e centeio)	
pão de segunda (= pão de mistura de cereais)	
pão (= homem bonito)	
ganhar o pão (nosso) de cada dia (= trabalhar para viver)	
pão, pão, queijo, queijo (= com clareza)	

2.1.2 Técnica de redação das definições

O modo como os dicionários gerais descrevem o léxico é muito condicionado pelo formato tradicional do dicionário impresso, que obriga à concentração de dados no mesmo artigo, ou em artigos próximos. Por isso, as formas-lema de um dicionário não são o resultado direto do processo de lematização de um corpus. A seleção das formas-lema implica uma decisão sobre a inclusão de palavras compostas e expressões, mas também a resolução de casos em que uma mesma forma canônica corresponde a diferentes lexemas (homonímia) e o agrupamento de palavras com contiguidades semânticas perceptíveis (polissémicas).

Tratando-se de uma relação de homonímia (na prática, uma coincidência formal), os dicionários consagram artigos diferentes para cada lexema. Quando se reconhece uma relação semântica de tipo polissêmico (extensão de sentido, metonímia ou metáfora), os dicionários costumam agrupar no mesmo artigo as diferentes aceções. Assim, a homonímia é registada na macroestrutura do dicionário, enquanto a polissemia é descrita na microestrutura do artigo.

Esta configuração é própria dos dicionários que têm como utilizador típico os falantes nativos de uma língua. A polissemia não deixa de ser uma relação entre lexemas que têm significados distintos, mas os dicionários procuram explicá-la sob a forma de uma rede de contiguidades, que apela a uma consciência etimológica partilhada pela comunidade de falantes. Este processo é pouco eficaz se o dicionário se destina a não nativos, pois a rede polissêmica da língua de partida não encontrará paralelo na língua de chegada. Assim, em dicionários monolíngues ou dicionários bilíngues para aprendentes, a localização da informação e a compreensão da relação entre palavras polissêmicas implica uma aprendizagem metalinguística (Villalva/Silvestre 2014, 200–202).

Os dicionários de língua descrevem a polissemia com critérios diacrónicos, construindo uma definição inicial que corresponde a um sentido primitivo e subsequentes aceções que correspondem a etapas da mudança semântica. Segundo este critério, agrupam-se as palavras de acordo com o étimo, tratando no mesmo artigo palavras de categorias gramaticais diferentes. A etimologia justifica o estatuto de palavra homónima e por isso haverá diferentes artigos para *cunhado* ('irmão do cônjuge', do lat. COGNATU-) e para *cunhado*, participípio de *cunhar*, do lat. CUNEARE), ou para *cabo* ('corda', do lat. tardio CAPULU-) e *cabo* ('comandante', do lat. CAPUT). No artigo *cunhado* no dicionário PORTO EDITORA reconhece-se a organização diacrónica:

(4) *cunhado*

cunhado (1) nome masculino

1. irmão de um dos cônjuges ou companheiros (em relação ao outro cônjuge ou companheiro)
2. marido ou companheiro da irmã ou do irmão de alguém

Do latim COGNĀTU-, «aparentado».

cunhado (2) adjetivo

1. amoadado
2. *figurado* criado

Participípio passado de *cunhar*.

A descrição torna-se mais opaca quando o mesmo artigo inclui as diferentes etapas de uma longa evolução semântica, de que resultaram diversas palavras com uso contemporâneo. Por exemplo, para a palavra *barra* o dicionário HOUAISS apenas identifica um étimo (do lat. vulgar *barra*, 'travessa ou tranca'), pelo que sob um mesmo lema se agrupam significados cuja relação não é evidente, entre os quais 'entrada do porto',

‘desembocadura do rio’, ‘peça alongada e estreita’, ‘tira de tecido que serve de acabamento ou enfeite’, ‘bainha’.

Além disso, a interpretação dos dados etimológicos está sujeita a alterações se novos ou diferentes étimos forem estabelecidos. Para o substantivo *dado* (‘peça usada em jogos’) identifica-se ou uma origem árabe *DAD* (‘peça usada em jogos’), ou uma relação com o particípio latino *DATU-* (< *DARE*). De acordo com esta última hipótese, os lexemas relevantes seriam entendidos como polissêmicos na diacronia.

Também pode haver divergência etimológica caso se opte por considerar apenas a relação de contacto linguístico que está na origem de um determinado significado. Os dicionários PORTO EDITORA e HOUAISS consideram que *apartamento* (= ‘afastamento’, der. do verbo *apartar*) e *apartamento* (‘parte de moradia coletiva’, do fr. *appartement*) são palavras homónimas. Mas, numa perspetiva interlinguística, a palavra francesa teve origem no italiano *appartamento* (‘parte de moradia coletiva’), que por sua vez provém do substantivo castelhano *apartamiento* (‘ação de se afastar’).

Os lexicógrafos tentam redigir os artigos de modo a tornar a relação polissémica perceptível. Mas esse *continuum* semântico pode assentar em significados que não são comuns na língua contemporânea e que por isso têm de continuar a ser dicionarizados. Assim, em dicionários gerais, é possível encontrar definições diacrónicas que descrevem um estádio antigo da língua, sem informações que restrinjam o uso. A tradição dicionarística e os hábitos de consulta são as principais justificações para a preferência pelo critério diacrónico em dicionários de uso geral. As desvantagens são particularmente evidentes na perspetiva dos utilizadores sem conhecimentos de etimologia, com especiais dificuldades no caso de utilizadores de L2, para quem as relações etimológicas se tornam ainda menos perceptíveis.

Outros critérios podem ser adotados se o objetivo é uma descrição sincrónica da língua ou a compilação de dicionários bilingues ou para aprendentes.

A abordagem morfossemântica distingue-se pelo facto de a forma-lema ser englobante, recobrando os diversos lexemas homónimos. Estes dicionários exploram a informação morfológica ao incluir em cada subentrada as palavras derivadas e compostos que dão continuidade ao significado iniciado pelo lexema. O artigo *cunhado* apresentaria esta configuração:

(5) *cunhado*

1. Subst. masc., irmão de um dos cônjuges ou companheiros.
2. Subst. masc., marido ou companheiro da irmã ou do irmão de alguém
3. Adj. figurado. 3.1. Cunhar. Figurar
4. Adj. amoeado. 4.1. Cunhar. Fazer moeda.

Este método propõe uma descrição por famílias de palavras, que em alguns casos se pode converter numa rede complexa e tornar menos evidente a localização de palavras derivadas, pois à partida tem de se conhecer alguns dos traços semânticos que as definem. Pode também acontecer que em diferentes artigos se repitam as mesmas

formas originadas por derivação, que são também homônimos entre si (cf. *cunhar*, no exemplo *supra*).

A ordenação das aceções pode obedecer a critérios de frequência, principiando a descrição pelas aceções mais comuns, o que altera a sequência diacrônica das derivações semânticas. Este critério é adequado à compilação de dicionários bilíngues (L2>L1), pois as possibilidades de descodificação de um lexema estão concentradas num único artigo.

Todavia, os redatores de dicionários gerais sentem dificuldade em assinalar as unidades lexicais e aceções que estão em desuso no português contemporâneo. Para o português não existe um dicionário (ou base de dados lexicográfica) que descreva o léxico e a mudança lexical numa perspectiva diacrônica, rigorosamente documentado com testemunhos textuais datados, uma avaliação da frequência das unidades e da ocorrência em fontes de natureza diversa.

Sem o auxílio de tesouros lexicais para o período clássico ou para o português medieval, as datações dos dicionários etimológicos portugueses têm sido sucessivamente corrigidas com antedatações de vários séculos, mas permanece a incerteza sobre a frequência relativa das ocorrências e sobre os processos de formação derivacional nos largos hiatos que os corpora não documentam. Para a análise diacrônica da formação de palavras acresce o óbice de muitos corpora eletrônicos se basearem em edições não diplomáticas, o que dificulta o estabelecimento de um nexos temporal. Em Portugal e no Brasil existem corpora coerentemente editados e anotados, que preservam a informação linguística essencial, e com uma seleção criteriosa e representativa das variações diacrônicas e de registo (Corpus UNESP; Corpus de Referência do Português Contemporâneo). Para documentar a dicionarização do léxico, está disponível um corpus diacrônico com o texto integral dos principais dicionários latim-português e português-latim publicados entre os séculos XVI e XIX, incluindo as obras dos lexicógrafos Jerónimo Cardoso e Rafael Bluteau (cf. Corpus Lexicográfico do Português; Verdelho/Silvestre 2014).

2.2 Dicionários especializados

Os dicionários especializados têm por objeto um elemento da descrição linguística. Incidem sobre um recorte lexical específico, sobre um domínio da informação metalinguística, ou exploram relações sintagmáticas ou paradigmáticas. Nas secções seguintes referem-se os principais trabalhos de lexicografia em dicionários combinatórios, paradigmáticos, ortográficos e etimológicos, com uma referência final à dicionarização da onomástica e à antroponímia portuguesa.

2.2.1 Relações sintagmáticas e relações paradigmáticas

A informação sobre relações sintagmáticas está presente em dicionários combinatórios, que indicam, para cada lema, que tipos de complemento são sintaticamente possíveis e as eventuais restrições; por sua vez, os dicionários de colocações especificam, para cada lema, que palavras podem ocorrer com ele para formar unidades de sentido distintas da soma dos significados individuais. O trabalho mais consistente neste campo tem sido desenvolvido para o português brasileiro, primeiro com um dicionário de descrição dos verbos (Borba 1990) e posteriormente ampliado à descrição de outras categorias gramaticais (Borba 2002). Para o português europeu, estão disponíveis dicionários de orientação prescritiva, sem uma sistemática documentação em corpora orais da língua contemporânea (Busse et al. 1994; Ramalho 1999; Casteleiro 2007).

A informação paradigmática explora a possibilidade de comutação das palavras que têm uma relação semântica entre si. Os dicionários de sinónimos são os modelos mais antigos, porque as relações de sinonímia são um recurso para apoio à composição literária, explorando a comutação como elemento de coesão textual. Distinguem-se os dicionários de sinónimos cumulativos, que agregam as palavras com sentido semelhante sem indicar quais são mais adequadas a cada contexto, e os dicionários de sinónimos distintivos, que esclarecem as particularidades semânticas ou restrições à comutação.

Os principais dicionários de sinónimos disponíveis para o português europeu são de tipo cumulativo e com pouca informação metalinguística. A obra mais difundida na segunda metade do século XX, sucessivamente reimpressa, tem origem no dicionário de Melo (1949), que foi organizado por uma associação de charadistas. A informatização dos dicionários motivou uma segunda edição (Porto Editora 1997), que apresenta de forma mais consistente as relações de sinonímia e antonímia, mas também numa perspetiva cumulativa.

Os dicionários analógicos exploram relações lógico-semânticas (a associação de conceitos), pelo que, ao contrário do que acontece com os sinónimos, a relação de comutação não está limitada à correspondência no interior de uma mesma categoria gramatical. As associações podem ser meramente temáticas, estabelecendo-se a partir de factos extralinguísticos e relações ontológico-enciclopédicas. A primeira tentativa de organização de um amplo dicionário analógico para o português é Bivar/Ferreira/Ferreira (1948). Este dicionário procurou integrar inúmeros termos técnicos em campos semânticos, pelo que rapidamente se tornou desatualizado face à evolução das linguagens de especialidade. Para o português brasileiro, foi publicado recentemente um dicionário que se inspira na técnica lexicográfica de Bivar (Pessek 2010). O espaço dos dicionários analógicos é atualmente ocupado por recursos paralexigráficos moduláveis como as bases de dados Wordnet, que descrevem as unidades lexicais de acordo com as relações semânticas que estabelecem entre si (7 Lexicologia e linguística computacional).

2.2.2 Dicionários ortográficos

Os dicionários ortográficos registam as formas básicas do lema e a sua escrita em função das regras ortográficas em vigor, dispensando as palavras derivadas de formação regular. No caso do português europeu, a normalização de casos duvidosos compete à Academia das Ciências de Lisboa.

A última edição (Academia das Ciências 2012) tem cerca de 70 mil entradas, que funcionam como uma nomenclatura básica para a compilação de um dicionário prático. Não inclui os termos técnicos com baixa frequência, mas regista formas derivadas que apresentam desvios semânticos em relação à raiz morfológica (como *altamente*) ou participípios de verbo ausente (como *impensado*).

O *Vocabulário Ortográfico do Português* (VOP), concluído pelo ILTEC (Instituto de Linguística Teórica e Computacional) ainda antes da edição do vocabulário da Academia, é uma base de dados lexical de acesso online, que indica para cada forma a categoria, informação flexional, divisão silábica e marcação de sílaba tónica. Caracteriza-se pela acumulação de formas (200 mil) sem esclarecer quanto à frequência ou uso contemporâneo da palavra. Está em curso a integração do VOP numa base de dados que reúne vocabulários nacionais próprios dos membros da Comunidade de Países de Língua Portuguesa (*Vocabulário Ortográfico Comum da Língua Portuguesa*).

O VOP diverge do vocabulário da Academia no que respeita a opções na adaptação de estrangeirismos (por exemplo, a Academia fixa a grafia *córpus*) e pelo facto de considerar um maior número de variantes (por exemplo, *espetador* e *espectador*, ao passo que a Academia fixa somente a grafia *espetador*).

A coerência do sistema ortográfico exigiria a convergência com as soluções apresentadas pela Academia, que é legalmente o órgão consultivo do governo português em matéria linguística. Todavia, o período entre a ratificação do Acordo Ortográfico em 2008 e a publicação do vocabulário da Academia deixou espaço para que os dicionários entretanto editados propusessem soluções ortográficas que consignaram usos diversificados. A palavra *corpus*, que neste manual segue a grafia comumente usada na bibliografia especializada de linguística portuguesa, é um exemplo de uma decisão normativa vencida pelo uso. Acresce que, apesar das recomendações da Academia, o Acordo Ortográfico prevê o uso facultativo das variantes europeia e brasileira, de modo aberto para todo o espaço lusófono, e alarga a margem de opção livre entre várias formas alternativas.

2.2.3 Dicionários históricos e etimológicos

Os dicionários patrimoniais são recolhas lexicais seletivas, com uma perspetiva diacrónica, que tradicionalmente se documentam na memória da língua fixada em textos literários. Os dicionários patrimoniais continuam e atualizam o trabalho iniciado desde o século XVIII por instituições académicas ou universitárias, com especial

referência para o *Diccionario de la lengua española* (Real Academia Española) e para o *Dictionnaire de l'Académie française* (Académie française). Em Portugal a recolha equivalente foi iniciada pela Academia Real das Ciências de Lisboa, mas deu origem apenas ao volume da letra A (Academia Real das Ciências 1793). Em 1976 tentou-se uma segunda edição, também interrompida na letra A. Previa-se um dicionário seletivo que reunisse a língua portuguesa contemporânea (literária e corrente) e a língua literária dos séculos XIX e XX, com excertos de autores canónicos para exemplificar os valores estilísticos e os contextos de uso. Em 2001 a Academia das Ciências de Lisboa completou um dicionário que se apoia sistematicamente em abonações de texto literário (DLPC), mas que se afasta do plano de um dicionário histórico por incidir sobre o uso contemporâneo. No que respeita à tipologia, aproxima-se do modelo do *dicionário académico* (correspondendo à categoria que em inglês se designa *scholarly dictionary*) pelo facto de se documentar em exemplos autênticos (de texto literário e jornalístico), criteriosamente transcritos e com referência exata das fontes. Todavia, não fornece atestações para todas as entradas ou propõe exemplos de uso redigidos pelo lexicógrafo, pelo que só parcialmente cumpre a principal função esperada de um dicionário académico, que é registar e descrever os usos autorizados.

Na tradição lexicográfica portuguesa do século XX não se distinguiram os limites tipológicos entre dicionários etimológicos, que descrevem a origem da palavra recuando aos mais antigos testemunhos do seu uso, e dicionários históricos, que descrevem a sua evolução diacrónica. As fontes dicionarísticas sobre etimologia são em certa medida complementares, pois não são coincidentes na nomenclatura e na quantidade de documentação apresentada; destacam-se Nascentes (1932); Machado (1956); Messner (1976); Cunha (1986). O dicionário de Machado é o mais extenso, mas é uma fonte datada, revista sem incorporar as análises críticas ou correções sugeridas após a primeira edição (ignorando, por exemplo, as contribuições em Lorenzo 1968).

A dicionarística do português encarregou-se de polir ou corrigir muitas das informações veiculadas por Machado, à medida que os dicionários gerais de língua começaram a robustecer os campos dedicados à origem das palavras. Todavia, a propósito da «origem», invocam-se informações díspares, pois confundem-se conceitos como *radical* (que refere a parte de uma palavra que é partilhada pelos elementos pertencentes a uma determinada família de palavras), *étimo* (que é a forma a partir da qual uma outra palavra provém) e *empréstimo* (que é uma inovação lexical a partir do contacto com outra língua).

Nos artigos do DLPC (2001) a informação sobre famílias de palavra exemplifica esta acumulação de critérios. Sobre a palavra simples dá-se informação sobre o étimo, enquanto para as palavras derivadas se dá informação sobre o radical e afixos. Esta técnica de descrição supõe que a derivação ocorreria em português, quando de facto a introdução de palavras por empréstimo pode facilmente reproduzir a variedade existente na(s) língua(s) de origem. Por exemplo, *fotografia*, *fotográfico* e *fotografar*, ainda que facilmente datáveis da primeira metade do século XIX, não são descritos

como empréstimos que se adaptaram ao português apoiados na analogia com formas pré-existentes em outras línguas modernas (cf. DLPC).

- (6) fotografia (Do fr. *photographie*, do gr. φως, φωτος ‘luz’ + γραφω ‘escrever’ suf. *-ia*)
 fotográfico, a (De *fotografia* + suf. *-ico*).
 fotografar (De *foto-* + gr. γραφω ‘escrever’ + suf. *-ar*).

O dicionário Houaiss representa um progresso na informação etimológica, não só porque recorre aos trabalhos etimológico-históricos e ao arquivo de António Geraldo da Cunha, mas porque inclui sistematicamente a datação das entradas, indicando o ano da primeira forma documentada, quer para o léxico antigo, quer para o moderno. Esta obra, que tem uma edição adaptada à ortografia e à descrição semântica do português europeu (Houaiss 2011), pode ser considerada o dicionário geral com a informação etimológica mais atualizada.

O português antigo é o período da diacronia que foi objeto de uma descrição histórica mais documentada: por um lado, através do extenso vocabulário compilado por António Geraldo da Cunha, que reúne variantes atestadas e contextos (Cunha 2014); por outro, com o *Dicionário de Verbos do Português Medieval* (DVPM), online, que interage com um corpus de textos medievais e oferece informação sintática. Os dados recolhidos pela equipa portuguesa fazem parte de uma base mais alargada, que desde 2012 integra textos galegos e portugueses (*Tesouro Medieval Informatizado da Língua Galega* – TMILG).

São promissoras as investigações de linguística de corpus em curso, com abordagem diacrónica. Existem bases de dados que podem ser parcialmente utilizadas e de que, a longo prazo, resultarão dados essenciais para a descrição semântica e sintática do português antigo e clássico.

2.2.4 Lexicografia onomástica

Os estudos de onomástica portuguesa não são abundantes e a descrição lexicográfica reflete essa insuficiência de fontes críticas. O denso inventário de Leite de Vasconcelos (Vasconcelos 1928) fornece um esquema teórico para explicar a formação do nome em Portugal e inclui dados desde o período medieval até ao início do século XX. Além do dicionário de nomes medievais de Cortesão (1912), deve citar-se uma série de trabalhos de linguística histórica que Joseph Piel dedicou aos nomes no espaço galego e português, bem como a dicionarização dos nomes antigos de origem visigótica (Piel/Kremer 1976).

Antenor Nascentes, no seguimento do dicionário etimológico publicado em 1932, compilou um volume sobre onomástica (Nascentes 1952). Continua a merecer consulta pela qualidade da informação diacrónica, mas a categorização de nomes próprios tornou-se desatualizada face às classificações dos dicionários modernos: além de topónimos e antropónimos, encontra-se uma acumulação enciclopédica de nomes

culturalmente significativos, em domínios heteróclitos como mitónimos, constelações, ventos, palácios, eventos históricos, nomes de tribos, entre muitos outros.

O *Dicionário onomástico etimológico da língua portuguesa* (Machado ¹1984) é a mais extensa recolha de topónimos e antropónimos em uso ou documentados em português. A nomenclatura é abundante e sobredimensionada para um dicionário especializado, saindo do âmbito estrito da onomástica nacional: inclui nomes de organizações e nomes de pessoas recolhidos em listas ou jornais, sem avaliar se os portadores do nome são portugueses (cf., e.g., as entradas *Internacional* ou *Voigt*). A multiplicação de entradas perturba a clareza das relações etimológicas entre nomes comuns, antropónimos e topónimos, mas também a perceção da evolução da palavra.

No caso dos topónimos, podem ser entrada tanto as formas modernas como as suas variantes na diacronia; quando considerados como antropónimo, geralmente opta-se pela forma moderna. A datação, a atestação e a etimologia são categorias de informação frequentemente confundidas, especialmente quando explica a origem de nomes claramente antigos como derivados de formas mais recentes. Nos casos em que um substantivo comum dá origem a um topónimo e, posteriormente, a um apelido ou alcunha, a descrição desdobra-se para admitir as duas possibilidades. Veja-se o exemplo dos nomes *Bacelo* e *Bacelar*, *Bacelinho* e *Bacelete* em Machado (¹1984):

- (7) *Bacelo* 1 top. < s.m. *bacelo*
Bacelo 2 apel. < *bacelo* 1 ou < s.m. *bacelo*
Bacelada top. < s.f. *bacelada*
Bacelar 1 top. < s.m. *bacelar*, lugar onde há bacelos
Bacelar 2 apel. < top. *Bacelar* ou < s.m. *bacelar*
Bacelete top. < dim. de *bacelo*
Bacelinho top. < dim. de *bacelo*

Como a informação onomástica está fora da nomenclatura dos dicionários gerais de língua, muito pouco se avançou na revisão ou documentação cronológica dos dados de Machado. Atualmente encontra-se na base de dados da Porto Editora um conjunto de entradas de toponímia que parece ter recuperado a informação do dicionário de Machado, eliminando as atestações, e que procura simplificar as relações que Machado estabelecia entre antropónimos, alcunhas e topónimos, fazendo-os derivar de uma única forma, sem distinções diacrónicas, ou geográficas.

- (8) *Bacelar*
 De *bacelar*, ‘terreno onde há bacelos’. Existe também na Galiza, e tem os derivados *Bacelada*, *Bacelares*, *Bacelete*, *Bacelinho*, *Bacelinhos*, *Bacelo* e *Bacelos*. (PORTO EDITORA)

A partir da década de 80, a investigação em onomástica portuguesa encontra um suporte crítico e metodológico com a participação do Centro de Linguística da Universidade de Lisboa no projeto internacional *Patronymica Romanica* (PatRom), que desenvolve as bases metodológicas e documentais de um dicionário etimológico antropónimoico. Tem por objetivo recolher étimos latinos que se revelam produtivos

em dois ou mais domínios linguísticos românicos, recolhendo nomes em toda a documentação histórica e nos repertórios demográficos medievais. Os artigos integram a informação histórica do português, com atestações do étimo em nomes simples, derivados e compostos. Trata-se de um longo projeto, de redação demorada, mas que vem publicando parcialmente os resultados, as discussões metodológicas, com dados relevantes para o português (Castro 1990; Kremer 1997; Nunes/Kremer 1999; Becker 2009).

2.2.5 Informação sobre onomástica ausente nos dicionários

A dicionarização da onomástica suscita problemas de descrição que podem ser enquadrados numa tipologia que distinga as propriedades referenciais e semânticas de antropónimos, nomes próprios e epónimos.

Os dicionários gerais incluem os nomes não tanto pelo facto de serem unidades linguísticas, mas por o seu uso depender de fatores extralinguísticos. As palavras são registadas nos dicionários porque os seus referentes podem ser identificados através de uma categoria representada por essa mesma palavra, que tem por isso conteúdo semântico. Em contrapartida, os nomes próprios não têm conteúdo semântico, servindo para referir diretamente um ser ou um local concretos. Ou seja, os nomes próprios nada dizem sobre as características do referente que invocam, sendo essa relação estabelecida por convenção. A arbitrariedade não é invalidada pelo facto de existirem, numa língua, convenções onomásticas: *João* é tipicamente reconhecido como um nome de homem, mas pode ser parte do nome de uma mulher (*Maria João*) ou de um topónimo (*São João*), e até ser atribuído como nome de animal de estimação. Da mesma forma, as considerações etimológicas sobre a origem de um nome não interferem com o seu uso em contexto comunicativo nem descrevem propriedades do referente: um enunciado como *Ele é o João* apenas identifica um indivíduo.

A dicionarização de nomes próprios que adquirem conteúdo semântico sai do âmbito habitual dos dicionários de língua, na medida em que é necessário descrever factos extralinguísticos sobre entidades, reais ou fictícias, cujos nomes (e palavras deles derivadas) adquirem um significado no interior de uma cultura ou comunidade. Encontram-se geralmente em obras de tipo para-lexicográfico, dedicadas à descrição de alusões, frases feitas ou à história cultural dos nomes próprios (e.g. Santos 1990; Neves 2000). Ainda assim, são unidades lexicais relevantes, pois o seu uso e descodificação implicam uma competência que pode ser comparada ao conhecimento das expressões idiomáticas (sobre tipologia dos nomes próprios, cf. Langendonck 2007, 96–106).

Podemos identificar três casos de transformação de nomes próprios que não figuram em dicionários de língua. Em primeiro lugar, as unidades lexicais multideno-tativas que, por processos de metonímia, podem ser nomes próprios para referentes de diferentes espécies. Por exemplo, a descodificação do enunciado *Um Almada é um*

bom investimento supõe o conhecimento da informação extralinguística *Almada é um pintor*.

Em segundo lugar, as unidades que resultam da transformação de um nome próprio em nome comum com uma interpretação conotativa (processo de apelativização). Esta atribuição de significado explora as características estereotipadas que uma comunidade atribui a um nome, por uma relação metafórica com as características do referente. O nome comum assim criado perde capacidade referencial, sendo possível usá-lo com artigo indefinido ou modificadores (*Eça de Queirós é um escritor português. O João não é um Eça de Queirós. Aquele novo escritor não é um verdadeiro Eça de Queirós.*)

Por fim, também não se regista num dicionário de língua o uso autonímico dos nomes próprios (*Ele chama-se João. João é um nome português. Há muitos Joões nesta lista de alunos*). Nestes contextos, os nomes próprios têm uma função metalinguística, sem interpretação semântica, e como tal apenas se encontram registados em dicionários especializados onomásticos (cf. Langendonck 2007, 246–249).

Nos dois primeiros casos, as unidades lexicais têm o estatuto de epónimos, pois em rigor são novas palavras (com valor semântico) que têm origem em nomes próprios. Paradoxalmente, os epónimos derivados de nomes próprios por processos de sufixação são uma parte estatisticamente relevante dos neologismos introduzidos nos dicionários de língua portuguesa: no dicionário Houaiss existem cerca de 2.400 entradas que correspondem a derivações em torno de antropónimos tipicamente registados em dicionários enciclopédicos. Esta abundância parece acompanhar a apetência dos utilizadores por terminologias, uma vez que a maioria das unidades são adjetivos relacionais que invocam um nome de uma entidade culturalmente significativa, associados às terminologias e à apetência dos utilizadores por referentes de informação de tipo enciclopédico (cf. Bragina/Lubensky 2003).

Todavia, a dicionarização de derivados de antropónimos suscita problemas de consistência de informação linguística em aspetos como a etimologia e os critérios de regularidade morfológica na adaptação de empréstimos.

Um dicionário de língua não inclui uma entrada referente a *Wagner* (apelido que refere o compositor alemão Wilhelm Richard Wagner, 1813–1883), mas o DLPC acolhe o adjetivo e o substantivo *wagneriano*, bem como o substantivo *wagnerismo*; a estes Houaiss acrescenta o verbo *wagnerizar* e o substantivo *wagnerista*. Este entumescimento encontra paralelo na tradição de dicionários patrimoniais, que dão conta do uso historicamente documentado de palavras. O OED, por exemplo, documenta *Wagneresque*, *Wagnerianism*, *Wagnerianly*, *Wagnerism*, *Wagnerist*, *Wagnerite*, *Wagneritis*, *Wagnerize*.

O registo destas unidades nos dicionários portugueses, inclusive nos dirigidos ao público escolar, parece ser um automatismo de amplificação de nomenclatura, adaptando o estrangeirismo com recurso ao sufixo morfológicamente equivalente. Tratando-se de antropónimos de origem estrangeira, a diversidade nas soluções de adaptação morfológica tem como resultado a aceitação simultânea de formas como

Kafkaesco, *Kafkesco*, *Kafkiano*, que denunciam a introdução de formas ora derivadas do inglês, ora do francês, por exemplo.

Todavia, como a maioria destes dicionários não oferece atestação (o próprio DLPC não apresenta exemplos das entradas acima citadas), subsiste a dúvida sobre os contextos de uso.

2.2.6 Antroponímia portuguesa e formação do nome

A antroponímia não é objeto de nenhum dicionário especializado que descreva a diversidade de nomes próprios, apelidos e alcunhas em uso no português contemporâneo, ou mesmo as peculiaridades da onomástica nacional. A composição dos nomes de pessoas em Portugal segue regras legais que estabelecem limites ao número de unidades lexicais e ao espectro de nomes que podem ser selecionados (Decreto-Lei 36/97, artigo 103). Não obstante existirem milhares de nomes reconhecidos e autorizados, a lista dos nomes mais frequentes tem-se apresentado relativamente estável ao longo das últimas décadas, constituindo um repertório de nomes que se pode caracterizar como conservador no que respeita à variedade (cf. Castro 2003).

Historicamente, durante a Idade Média os indivíduos eram identificados apenas com o nome próprio, a que se podia acrescentar um patronímico sob uma forma de genitivo (*Fernandes*, filho de Fernando), a posposição do nome do pai, ou a evocação de um nome de santo. A partir do século XVI, o nome de família individualiza-se, recuperando como apelido os patronímicos, nomes próprios, nomes religiosos, alcunhas, nomes de profissões, topónimos, referências geográficas. Os movimentos populacionais estão testemunhados nas adaptações portuguesas de apelidos holandeses, flamengos e franceses (Castro 1990).

A manutenção de apelidos maternos e paternos, e a sua transmissão integral à geração seguinte, foi percebida como um traço socialmente valorizado, inclusive fora das famílias com tradição aristocrática. Esta acumulação caracterizou a composição do nome durante o século XIX e durante boa parte do século XX. Atualmente, a lei define que os nomes completos se compõem, no máximo, de seis vocábulos gramaticais, simples ou compostos, dos quais só dois podem corresponder ao nome próprio e quatro a apelidos. Os apelidos, ou nomes de família, destinam-se a identificar um indivíduo como pertencente a um grupo; os nomes próprios asseguram a individualidade dentro desse mesmo grupo.

O primeiro nome indica o sexo do registando, pelo que o segundo nome próprio pode ser tipicamente feminino ou masculino (*José Maria*, *António Rita*, *Maria José*). A posição de segundo nome tem admitido também nomes ambíguos quanto ao género (*Cris*) e adaptações de nomes que não constam na tradição onomástica portuguesa (*Gerardo*, *Darci*). Também se admitem evocações de inspiração religiosa (*Maria de São José*, *José de Santa Maria*), em que o hierónimo equivale a uma unidade lexical, para efeitos de registo. A escolha de alcunhas familiares e topónimos como segundo nome

próprio não é incomum, e pode também ser enquadrada na função evocativa desta posição na estrutura do nome completo. Os patronímicos e os nomes de profissão são hoje classes fechadas, admitindo-se a inclusão de alcunhas familiares no campo dos apelidos.

Aquando do assento de nascimento, os quatro nomes de família, assim como a sua ordem, são escolhidos livremente entre os apelidos de um ou de ambos os progenitores, ou a cujo uso tenham direito. Nas alterações posteriores, em virtude de casamento, não se pode exceder o número de apelidos, pelo que pode ser necessário renunciar a nomes de família para a formação do novo nome.

Os nomes próprios devem ser portugueses, o que inclui todos os que surgem documentados como tal em algum momento da história. Não há impedimento de recuperar nomes antigos, raros ou sem qualquer uso atestado no momento do registo; todavia, a ortografia deve seguir normas vigentes, pelo que as grafias sofrem a correspondente atualização (*Sophia*, *Theóphilo*, *Luiz* são hoje registados *Sofia*, *Teófilo* e *Luís*).

Os nomes estrangeiros, para serem admitidos, devem ser adaptados grafica e foneticamente à língua portuguesa. Podem manter a forma originária se o registando (ou um dos progenitores) for estrangeiro ou tiver outra nacionalidade além da portuguesa. A consulta de listas de nomes, publicadas pelo Instituto dos Registos e do Notariado, demonstra que a onomástica portuguesa é permeável às alterações sociais motivadas pelos movimentos migratórios e pelo contacto interlinguístico. Os serviços públicos de registo tendem a admitir variações ortográficas que não perturbam a identificação do nome (*Eveline*, *Evelina*, *Evelin*, *Evelyn*) ou a facilitar a adaptação de nomes estrangeiros que os falantes se habituaram a reconhecer como nomes próprios (*Lucy*, *Sheila*). Garante-se também o direito à onomástica religiosa da fé professada pelos pais, ao abrigo da lei de liberdade religiosa (Lei n.º 16/2001), o que possibilita a escolha de nomes hindus e islâmicos para cidadãos portugueses.

3 Conclusão

A lexicografia da língua portuguesa tem beneficiado, nas últimas décadas, com a decisão dos editores de incluir nos dicionários cada vez mais informação sobre as variedades nacionais. Esta confluência de vontades responde, em primeiro lugar, à passagem dos dicionários para as plataformas digitais: o dicionário não se limita ao espaço físico em que pode circular o livro impresso e alcança utilizadores de todas as variedades. Em segundo lugar, a atualização da norma ortográfica foi aproveitada, em Portugal e no Brasil, para rever a nomenclatura e avaliar quais as variantes preferenciais em cada espaço linguístico.

A descrição semântica e sintática não foi revista ao ponto de se poder incluir, sob um mesmo artigo, informação sobre o que às duas variedades é comum e distinto. Sem esse trabalho prévio, é difícil conceber um dicionário comum a ambos os espaços

linguísticos e reconhecido pelos respectivos sistemas de escolarização. As diferenças de terminologia metalexigráfica entre Portugal e o Brasil são pontuais, mas a incorporação e fixação de novos termos poderá mais facilmente surgir a partir do português brasileiro, como natural consequência da maior quantidade de textos teóricos sobre lexicografia e de investigação académica.

No que respeita às tipologias lexicográficas, os dicionários gerais do português europeu têm desenvolvido a qualidade da descrição linguística e melhorado a coerência da técnica lexicográfica. Todos ampliaram o número de entradas, por via da incorporação de terminologias e pela manutenção de um fundo de léxico antigo. Hoje, os dicionários distinguem-se sobretudo pelo maior ou menor espaço que concedem a exemplos de uso (seleccionados em corpora literários) e pela quantidade da informação linguística que disponibilizam (informações etimológicas e informação gramatical recuperada de dicionários anteriores). Se, por um lado, a admissão de novas entradas e aceções parece seguir os métodos aceites pela lexicografia internacional, por outro lado os lexicógrafos não têm meios expeditos para identificar de aceções que podem estar fora de uso ou restritas à memória literária. Continua a faltar uma verdadeira interação entre a memória latente do léxico e a descrição do português antigo, clássico e contemporâneo, com evidentes repercussões na qualidade da lexicografia histórica e patrimonial do português. Nessa lacuna reside o perigo de os dicionários gerais prosseguirem numa acumulação de unidades, que descaracteriza a representação do léxico do português contemporâneo. Assim, os dicionários da língua contemporânea tendem a construir uma descrição desproporcional do léxico, que sobrevaloriza o testemunho ancilar dos dicionários anteriores e que os torna desatualizados como base para a dicionarística bilingue.

4 Referências

- Academia das Ciências de Lisboa (2012), *Vocabulário ortográfico atualizado da língua portuguesa*, Lisboa, Imprensa Nacional–Casa da Moeda.
- Academia Real das Ciências (1793), *Diccionario da lingua portugueza*, Lisboa, Academia Real das Ciências.
- Académie française (1694, 1992), *Dictionnaire de l'Académie française*, Paris, Académie française, <http://www.academie-francaise.fr/le-dictionnaire/la-9e-edition> (06.02.2016).
- Alvar Ezquerro, Manuel (1993), *Lexicografia descriptiva*, Barcelona, Bibliograf.
- Becker, Lidia (2009), *Hispano-romanisches Namenbuch*, Tübingen, Niemeyer.
- Biderman, Maria Tereza (2005), *Unidades complexas do léxico*, in: Graça Rio-Torto/Olívía Figueiredo/Fátima Silva (orgs.), *Estudos em Homenagem ao Professor Doutor Mário Vilela*, vol. 2, Porto, Faculdade de Letras da Universidade do Porto, 747–757.
- Bivar, Artur/Ferreira, Manuel/Ferreira, Maria (1948–1958), *Dicionário geral e analógico da língua portuguesa*, Porto, Ouro.
- Borba, Francisco da Silva (1990), *Dicionário gramatical de verbos do português contemporâneo*, São Paulo, UNESP.
- Borba, Francisco da Silva (2002), *Dicionário de Usos do Português do Brasil*, São Paulo, Ática.

- Borba, Francisco da Silva (2003), *Organização de Dicionários: uma introdução à lexicografia*, São Paulo, Editora da UNESP.
- Bragina, Natalia/Lubensky, Sophia (2003), *Eponyms as cultural key words and their lexicographic description in English and Russian*, in: Anna Braasch/Claus Povlsen (edd.), *Proceedings of the Tenth EURALEX International Congress 2002*, vol. 1, Copenhagen, Center for Sprogteknologi, 419–427.
- Bugueño Miranda, Félix (2014), *Da classificação de obras lexicográficas e seus problemas: proposta de uma taxonomia*, Alfa: Revista de Linguística 58, 215–231.
- Bugueño Miranda, Félix/Farias, Virginia (2011), *Princípios para o desenvolvimento de uma teoria da definição lexicográfica*, Alfa: Revista de Linguística 55, 31–61.
- Busse, Winfried, et al. (1994), *Dicionário sintático de verbos portugueses*, Coimbra, Almedina.
- Casteleiro, João Malaca (dir.) (2007), *Dicionário Gramatical de Verbos Portugueses*, Lisboa, Texto Editores.
- Castro, Ivo (1990), *Actes du 1^{er} Colloque du Dictionnaire Historique des Noms de Famille Romans (Trèves, 1987)*, Tübingen, Niemeyer, 10–13.
- Castro, Ivo (2003), *O linguista e a fixação da norma*, in: *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (Porto, 2002)*, Lisboa, APL, 11–24.
- Corpus de Referência do Português Contemporâneo (CRPC)*, Lisboa, CLUL, <http://alfclul.clul.ul.pt/CQPweb> (06.02.2016).
- Corpus Lexicográfico do Português*, Lisboa/Aveiro, CLUL/Universidade de Aveiro, <http://clp.dlc.ua.pt/> (06.02.2016).
- Corpus UNESP, Usos do Português* [200 milhões de palavras, português brasileiro escrito], Araraquara, UNESP.
- Cortesão, António Augusto de (1912), *Onomástico medieval português*, Lisboa, Imprensa Nacional.
- Cunha, António Geraldo da (1982, ²1986), *Dicionário Etimológico Nova Fronteira da Língua Portuguesa*, Rio de Janeiro, Nova Fronteira.
- Cunha, António Geraldo da (2014), *Vocabulário histórico-cronológico do português-medieval*, Rio de Janeiro, Fundação Casa de Rui Barbosa.
- Decreto-Lei n.º 36/97, *Diário da República*, I-A, n.º 26 (31.01.1997), <https://dre.pt>.
- DLPC = Academia das Ciências de Lisboa (2001), *Dicionário da Língua Portuguesa Contemporânea da Academia das Ciências de Lisboa*, Lisboa, Verbo.
- DVPM = Xavier, Maria Francisca/Vicente, Graça/Crispim, Maria de Lurdes (orgs.) (2011), *Dicionário de Verbos Portugueses Medievais*, Lisboa, Centro de Linguística da Universidade Nova de Lisboa, <http://cipm.fcsh.unl.pt/gencontent.jsp?id=5> (03.03.2016).
- Ferreira, Aurélio (2004), *Novo Dicionário Aurélio da Língua Portuguesa*, Rio de Janeiro, Positivo.
- Haensch, Günther/Omeñaca, Carlos (2004), *Los Diccionarios del Español en el Siglo XX*, Salamanca, Universidad de Salamanca.
- Haensch, Günther, et al. (edd.) (1982), *La lexicografía: de la lingüística teórica a la lexicografía práctica*, Madrid, Editorial Gredos.
- Hartmann, Reinhard/James, Gregory (2001), *Dictionary of Lexicography*, London, Routledge.
- Houaiss, António/Villar, Mauro de Salles (2001), *Dicionário Houaiss da Língua Portuguesa*, Rio de Janeiro, Objetiva.
- HOUAISS = Villar, Mauro de Salles (coord.) (2011), *Dicionário do Português Atual Houaiss*, Lisboa, Círculo de Leitores.
- Imbs, Paul (1971), *Préface*, in: *Trésor de la langue française. Dictionnaire de la langue du XIXe et du XXe siècle*, t. 1, Paris, CNRS – Gallimard, IX–XLVII.
- Iriarte Sanromán, Álvaro (2000), *A Unidade Lexicográfica: Palavras, Colocações, Frasemas, Pragmáticas*, tese de Doutoramento, Braga, Universidade do Minho.

- Isquierdo, Aparecida/Finatto, Maria José (orgs.) (2010), *As Ciências do Léxico, Lexicologia, Lexicografia, Terminologia*, Porto Alegre, Editora da UFRGS.
- Kremer, Dieter (1997), *Dictionnaire historique de l'anthroponymie romane (PatRom)*, Tübingen, Niemeyer.
- Langendonck, Willy van (2007), *Theory and Typology of Proper Names*, Berlin/New York, Mouton de Gruyter.
- Lei n.º 16/2001, *Diário da República*, I-A, n.º 143 (22.06.2001), <https://dre.pt>.
- Lorenzo, Ramón (1968), *Sobre cronología do vocabulário galego-português. Anotações ao «Dicionário Etimológico» de José Pedro Machado*, Vigo, Galaxia.
- Machado, José Pedro (1977, 1956), *Dicionário etimológico da língua portuguesa: com a mais antiga documentação escrita e conhecida de muitos dos vocábulos estudados*, Lisboa, Livros Horizonte.
- Machado, José Pedro (2003, 1984), *Dicionário onomástico etimológico da língua portuguesa*, Lisboa, Livros Horizonte.
- Medina Guerra, Antonia María (coord.) (2003), *Lexicografía española*, Barcelona, Ariel.
- Melo, Alfredo Leite Pereira de (1949), *Dicionário de Sinónimos da Língua Portuguesa*, Lisboa, Tertúlia Edípica.
- Messner, Dieter (1976), *Dictionnaire chronologique portugais*, Heidelberg, Winter.
- Nascentes, Antenor (1932), *Dicionário Etimológico da Língua Portuguesa*, Rio de Janeiro, Alves.
- Nascentes, Antenor (1952), *Dicionário Etimológico da Língua Portuguesa. Nomes Próprios*, vol. 2, Rio de Janeiro, Livraria Acadêmica.
- Neves, Orlando (2000), *Dicionário de nomes próprios*, Lisboa, Editorial Notícias.
- Nunes, Naidea/Kremer, Dieter (1999), *Antroponímia primitiva da Madeira e Repertório onomástico histórico da Madeira*, Tübingen, Niemeyer.
- OED = (1884–1928), *Oxford English Dictionary*, Oxford, Oxford University Press, www.oed.com.
- Pessek, Kurt (2010), *Dicionário de palavras interligadas. Analógico e ideias afins*, Brasília, Thesaurus.
- Piel, Joseph Maria/Kremer, Dieter (1976), *Hispano-gotisches Namenbuch*, Heidelberg, Winter.
- Porto Dapena, José Álvaro (2002), *Manual de técnica lexicográfica*, Madrid, Arco/Libros.
- Porto Editora (1997), *Dicionário de sinónimos*, Porto, Porto Editora.
- PORTO EDITORA = *Dicionário da Língua Portuguesa com Acordo Ortográfico*, Porto, Porto Editora, 2003–2016, <http://www.infopedia.pt/dicionarios/lingua-portuguesa/assecla> (19.01.2016).
- Quemada, Bernard (1968), *Les dictionnaires du français moderne, 1539–1863. Étude sur leur histoire, leurs types et leurs méthodes*, Paris, Didier.
- Ramalho, Énio (1999), *Novo dicionário Lello estrutural, estilístico e sintático da Língua Portuguesa*, Porto, Lello.
- Real Academia Española (1739, 23/2014), *Diccionario de la lengua española*, Madrid, Real Academia Española, <http://dle.rae.es> (06.02.2016).
- Rey, Alain (1977), *Le lexique: images et modèles. Du dictionnaire à la lexicologie*, Paris, Colin.
- Santos, António (1990), *Novos dicionários de expressões idiomáticas*, Lisboa, João Sá da Costa.
- Svensén, Bo (2009), *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*, Cambridge, Cambridge University Press.
- TMILG = Varela Barreiro, Xavier (dir.) (2004-), *Tesouro Medieval Informatizado da Língua Galega (TMILG)*, Santiago de Compostela, Instituto da Língua Galega, <http://ilg.usc.es/tmilg> (06.02.2016).
- Vasconcelos, José Leite de (1928), *Antroponímia portuguesa*, Lisboa, Imprensa Nacional.
- Verdelho, Telmo (1995) *As origens da gramaticografia e da lexicografia Latino-Portuguesas*, Aveiro, INIC.

- Verdelho, Telmo/Silvestre, João Paulo (2007), *Dicionarística Portuguesa. Inventariação e estudo do património lexicográfico*, Aveiro, Universidade de Aveiro.
- Verdelho, Telmo/Silvestre, João Paulo (2011), *Lexicografia bilingue. A tradição dicionarística português-línguas modernas*, Lisboa/Aveiro, Centro de Linguística da Universidade de Lisboa/Universidade de Aveiro.
- Verdelho, Telmo/Silvestre, João Paulo (2014), *El Corpus Lexicográfico do Português: la memoria de los diccionarios en la historia de la lengua y de la cultura*, in: Félix Córdoba Rodríguez/Ernesto González Seoane/María Dolores Sánchez Palomino (edd.), *Lexicografía de las lenguas románicas*, vol. 1: *Perspectiva histórica*, Berlin/Boston, De Gruyter, 283–298.
- Villalva, Alina/Silvestre, João Paulo (2014), *Introdução ao Estudo do Léxico. Descrição e Análise do Português*, Petrópolis, Vozes.
- VOP = Correia, Margarita (coord.) (2010), *Vocabulário Ortográfico do Português*, Lisboa, Instituto de Linguística Teórica e Computacional, <http://www.portaldalinguaportuguesa.org/vop.html> (04.03.2016).

Amália Mendes

9 Linguística de corpus e outros usos dos corpora em linguística

Abstract: Este capítulo discute a noção de corpus e as diferentes formas como é utilizado nos trabalhos em Linguística: quer, na abordagem da linguística de corpus e de perspetivas teóricas afins, como uma fonte de dados que fornece informações sobre fenómenos que poderiam não ser detetados com base apenas na intuição e cuja quantificação permite apontar padrões e tendências, quer como metodologia de trabalho utilizada em trabalhos que se situam em teorias linguísticas várias. Discutem-se algumas das etapas de compilação de corpora, os princípios de representatividade e equilíbrio, os problemas de direitos de autor envolvidos e algumas ferramentas de pesquisa. Apresentam-se, ainda, os diferentes tipos de corpora existentes para o português europeu (podendo alguns incluir outras variedades do português) e os níveis de anotação linguística desenvolvidos. Finalmente, são referidos alguns estudos em diferentes áreas da linguística que tiveram como base corpora compilados para o português.

Keywords: compilação de corpora, anotação linguística, pesquisa de informação, análise quantitativa e estatística, estudos baseados em corpora

1 Introdução

A linguística de corpus baseia o estudo da língua em ocorrências extraídas de um *corpus*, isto é, de um conjunto de textos escritos (ou excertos de textos) ou de transcrições de registos orais, tipicamente em formato eletrónico. A utilização de corpora permite fundamentar análises linguísticas com base em contextos variados e num conjunto alargado de dados. Além disso, a disponibilidade de ocorrências produzidas em contextos naturais de comunicação vem com frequência chamar a atenção para fenómenos que não seriam perceptíveis com outro tipo de dados, pelo que «corpus linguistics has the potential to reorient our entire approach to the study of language» (McEnery/Hardie 2012, 1). Os dados de corpora não só fornecem informação sobre os fenómenos que ocorrem na linguagem escrita e oral, como indicam ainda com que frequência ocorrem os fenómenos analisados: «Corpus linguistics are concerned typically not only with what words, structures or uses are possible in a language but also with what is probable – what is likely to occur in language use» (Kennedy 1998, 8). Com corpora diversificados, permite ainda uma análise quantitativa comparativa entre géneros textuais e registos.

Embora os corpora mais recentes tendam a ser da ordem dos vários milhões de palavras, podem ainda assim não fornecer ocorrências de alguns elementos lexicais,

ou de alguns significados ou construções sintáticas. Por esse motivo, o corpus não fornece verdadeiramente dados negativos, isto é, informação sobre fenómenos agramaticais na língua, uma vez que uma palavra ou construção podem não ocorrer no corpus sem por isso ser agramaticais. Os dados de corpora funcionam assim a par de outras fontes de dados disponíveis para a análise linguística, que pode ainda recorrer a dados de introspeção e à eliciação.

Em Portugal, os trabalhos com corpora iniciaram-se com o projeto *Português Fundamental* nos anos 1970, no Centro de Linguística da Universidade de Lisboa, sob a direção de Luís Filipe Lindley Cintra, muito antes de a designação «linguística de corpus» passar a ser usada. O objetivo último era fornecer informação sobre o vocabulário português mais frequentemente utilizado em situações da vida corrente. Para tal, constituiu-se um corpus de gravações em situação de comunicação oral espontânea, realizadas entre 1970 e 1974, em todo o continente e ilhas de Portugal, o chamado *Corpus de Frequência* (Bacelar do Nascimento/Garcia Marques/Segura da Cruz 1987; Bacelar do Nascimento/Rivenc/Segura da Cruz 1987). Na sequência deste trabalho, teve início a compilação do *Corpus de Referência do Português Contemporâneo*, em 1988, um corpus escrito e oral que abrange sobretudo a variedade europeia do português, mas também inclui dados das variedades do português no mundo. Estes primeiros trabalhos foram seguidos de muitas outras iniciativas de compilação de corpora para a língua portuguesa que serão revistos na secção 3 deste capítulo.

Embora a utilização de dados de corpora, como metodologia de trabalho, seja compatível com diferentes teorias linguísticas, a designação «linguística de corpus» está profundamente associada a perspetivas linguísticas centradas no uso da língua e nas situações de comunicação (com a observação de registos desse uso), isto é, a tradições de cunho funcionalista, aqui tomadas no sentido lato, e abrangendo a linguística funcionalista, cognitiva e abordagens tipológicas (cf. McEnery/Hardie 2012). Os corpora são usados como fontes de dados também em teorias formalistas, que pretendem verificar hipóteses de trabalho com base nos dados, mas focam tipicamente áreas para as quais o investigador não pode dispor da sua intuição, como seja o estudo das variedades dialetais ou nacionais de uma língua ou dos processos de aquisição de uma língua por crianças ou por falantes estrangeiros. Os trabalhos em linguística de corpus partem, pelo contrário, da perspetiva segundo a qual os dados do corpus são úteis mesmo quando o investigador trabalha sobre a sua variedade nativa e permite revelar novos fenómenos que não seriam visíveis se este apenas se baseasse na sua intuição. Como indica o título deste capítulo, pretende-se aqui abrançar estes diferentes tipos de usos dos corpora nos estudos linguísticos.¹

1 Para além de uma abordagem que pretende verificar hipóteses com base em dados de corpus (designada como *corpus-based*), existe ainda quem assuma uma perspetiva do tipo *corpus-driven*, que se distingue por não assumir nenhuma hipótese prévia a verificar, partindo sempre primeiro dos dados para levantar fenómenos e propostas de análise.

Para além do trabalho de compilação de dados, a linguística de corpus preocupa-se ainda com a anotação de informação linguística sobre os textos que compõem o corpus. A anotação permite localizar de forma rápida e automática os fenómenos que se pretende observar, sem exigir a inspeção manual de todos os textos. Os fenómenos a anotar dependem dos objetivos das equipas de investigação, mas centram-se habitualmente num primeiro nível de categorização morfossintática (ao qual se pode juntar informação de flexão nominal e verbal) e lematização dos itens lexicais, sobre o qual se implementa, num segundo nível, a anotação sintática (cf. secção 4).

O trabalho de preparação dos textos, a sua anotação linguística e a sua pesquisa requerem uma colaboração estreita entre a linguística de corpus e a linguística computacional, que visa o desenho e implementação de programas informáticos que permitam automatizar alguns processos, treinar etiquetadores e criar ferramentas de pesquisa. Os corpora, para além de constituírem fontes de dados para a análise linguística, são ainda fonte de informação para a criação de aplicações várias, como, por exemplo, redes conceptuais, sistemas de sumarização automática, de extração de informação, de tradução automática, de reconhecimento da fala e síntese de voz (cf. Branco et al. 2012). A área das humanidades em geral pode beneficiar das metodologias e aplicações desenvolvidas para os corpora de língua pela linguística de corpus e pela linguística computacional (↗10 Avanços nas humanidades digitais).

Neste capítulo, a secção 2 fornece informação sobre o trabalho de compilação de corpora, a secção 3 apresenta alguns corpora para o português compilados com diversos objetivos; finalmente, a secção 4 apresenta exemplos de níveis de anotação linguística nos corpora do português, a par de estudos linguísticos que utilizaram essa informação.

2 Compilação e pesquisa de corpora

Nesta secção, discutem-se alguns aspetos relacionados com o trabalho de compilação e pesquisa de um corpus, que permitirão situar os casos concretos de corpora para o português, que serão apresentados na secção 3.

2.1 Tipo de corpus, representatividade e equilíbrio

O objetivo do corpus irá determinar o tipo de material selecionado, que varia consoante: (i) a modalidade de produção: corpus escrito ou oral; (ii) o recorte sobre a língua portuguesa: standard ou dialetal, contemporânea ou histórica; (iii) os géneros textuais; (iv) o perfil dos autores/informantes: produções de falantes nativos, de crianças a adquirirem a sua primeira língua, de aprendentes de português língua estrangeira; (v) o registo formal ou informal dos textos e das gravações.

Os corpora que visam representar uma língua na sua variedade standard contemporânea, por vezes chamados *corpora de referência*, têm como objetivo dar conta da diversidade da língua e incluem geralmente uma subparte escrita e uma subparte oral, em repartições variáveis, e visam incluir o maior número possível de géneros textuais e registos. Esse objetivo faz com que estes sejam geralmente *corpora fechados* (ou *estáticos*), no sentido em que foram planeados e compilados para obedecer a alguns princípios de equilíbrio entre géneros textuais (veja-se o exemplo do *British National Corpus – BNC*). Por oposição aos corpora fechados, os *corpora abertos* (ou *dinâmicos*) podem continuar a integrar textos, quer mantendo as proporções inicialmente fixadas para cada tipo textual, quer integrando todos os textos disponíveis, sem considerações de equilíbrio, o que se designa então como *corpus monitor*, de que é paradigma o *Bank of English*. Existem ainda corpora que apenas incluem um género textual ou textos de especialidade, que apenas incluem dados de fala, que focam variedades dialetais ou ainda variedades de uma língua no mundo (o caso do *International Corpus of English – ICE*), que apresentam dados diacrónicos ou ainda dados de aquisição de uma língua como língua primeira (L1), língua segunda (L2) ou língua estrangeira (LE). Finalmente, os corpora podem visar especificamente estudos contrastivos, sendo compilados vários corpora que seguem a mesma estrutura interna, isto é, *corpora comparáveis*, de diferentes dialetos ou variedades de uma língua, ou ainda de diferentes línguas de especialidade. Para estudos contrastivos entre diferentes línguas, existem ainda alguns *corpora paralelos*, constituídos por versões dos mesmos textos em diferentes línguas.

A constituição de um corpus tem sempre subjacente o princípio de que o corpus irá fornecer dados que permitem algum grau de generalização sobre uma língua ou sobre algum aspeto da variação de uma língua, isto é, pressupõe algum princípio de *representatividade* (Leech 1991). Desse ponto de vista, o corpus seria uma amostragem da totalidade da diversidade encontrada em cada sistema linguístico visado. No entanto, não será fácil determinar de que forma atingir essa representatividade e a dificuldade em confiar nas generalizações produzidas com base numa seleção de textos é uma crítica por vezes apresentada contra o uso de corpora na análise linguística. O desenho cuidado do corpus é, por conseguinte, um aspeto fulcral. A representatividade é atingida através da diversidade de géneros textuais/gravações incluídos e também através do *equilíbrio* na proporção destes géneros. Essas proporções variam muito de acordo com cada projeto e dependem dos objetivos iniciais, bem como do tempo e financiamento disponíveis. Por exemplo, o BNC contém 10% de língua falada, enquanto o ICE apresenta uma percentagem de 60% de oralidade, mas a dimensão dos dois corpora é bem diversa: 100 milhões de palavras no caso do BNC e 1 milhão no caso do ICE.

A dimensão dos corpora tem aumentado à medida que aumenta o número de textos disponíveis em formato digital, chegando nalguns casos a ultrapassar os 1.000 milhões de palavras (*Bank of English*). Será realmente necessário assegurar a compilação de corpora tão extensos? Na verdade, uma análise dos dados de frequência de

corpora aponta para o facto de 40% das ocorrências de um corpus de 5 milhões de palavras serem casos de *hapax legomena*, isto é, casos de formas que apenas ocorrem uma vez; e ainda que num corpus de 1 milhão de palavras uma nova palavra aparece aproximadamente de 30 em 30 palavras (Kennedy 1998, 100). Esta dificuldade afeta especialmente estudos lexicais e semânticos, uma vez que alguns sentidos de palavras polissémicas podem ser raros ou nunca ocorrerem, mesmo em corpora de grandes dimensões. Enquanto o número de tokens (cada forma que ocorre no corpus) é sempre crescente no corpus, o aparecimento de formas novas (*type*) diminui gradualmente e essa diminuição é ainda mais acentuada quando se contabiliza a ocorrência de novos lemas (Sánchez Pérez/Cantos Gómez 1997).

Nesta perspetiva, tem vindo a ganhar importância a tendência para a utilização do conjunto de documentos disponíveis na internet como corpus linguístico, conhecida como *web as corpus* (Kilgarriff/Grefenstette 2003). Neste caso, a ausência de controlo sobre os textos incluídos no corpus e a ausência de equilíbrio é compensada pelo elevado número de contextos disponíveis para análise, numa abordagem que se aproxima do corpus monitor.

2.2 Compilação

A compilação dos textos é precedida de uma fase de desenho da estrutura do corpus, que estabelece quantos géneros textuais devem estar representados, quantos textos de cada género e o tamanho de cada amostragem incluída (embora esta fase possa não existir ou ser menos significativa na compilação de um corpus monitor). Um corpus de fala será organizado em termos de gravações de situações públicas (entrevistas, média, aulas, conferências, etc.) e de situações privadas (diálogos, conversas) e visa a diversidade sociolinguística do falante: diferentes faixas etárias, níveis de escolaridade, profissões. Um corpus escrito irá integrar textos de ficção e não ficção, de diferentes fontes (livro, revista, jornal, etc.), registos (formal ou informal) e cobrindo diversos tópicos e variedades.

Os textos podem ser já obtidos em formato eletrónico ou podem ser descarregados da internet. Neste último caso, o ficheiro inclui etiquetas em formato html que dificultam as fases posteriores de pesquisa e de anotação dos dados, pelo que estão disponíveis ferramentas de limpeza dos textos. Num corpus que se quer diversificado e representativo, nem sempre a totalidade dos textos a incluir está disponível em formato digital, o que requer um trabalho de localização de textos e digitalização da versão impressa com reconhecimento de caracteres (OCR – Optical Character Recognition) e sua posterior revisão. No caso de instâncias de língua falada, a gravação deve assegurar boas condições de captação de som, embora haja sempre um compromisso necessário entre naturalidade e qualidade sonora. A filmagem dessas situações poderá complementar o registo sonoro com imagem e permitir a análise da relação entre prosódia e gesto.

À gravação da situação de comunicação segue-se a transcrição ortográfica, por vezes acompanhada de transcrição fonética. A transcrição do oral requer o estabelecimento de normas que reproduzam graficamente pausas, fronteiras entoacionais, hesitações, sobreposições de turno de fala, palavras fragmentadas, elementos extralinguísticos (tosse, riso, choro), elementos paralinguísticos (*ah, hum, nanana*). Para além da transcrição, é ainda possível alinhar o texto transcrito com o som, usando um dos programas existentes para o efeito – Transcriber (Barras et al. 2000), EXMARaLDA (Schmidt 2012), Praat (Boersma/Weenink 2013), entre outros. Apresenta-se de seguida um exemplo de transcrição de uma conversa informal do corpus C-ORAL-ROM (Baccalar do Nascimento et al. 2005), em que as chavetas indicam sobreposição de fala, & marca uma palavra incompleta, *hhh* a produção de um elemento paralinguístico, / uma pausa breve, // o final de enunciado e o ponto de interrogação uma curva prosódica característica de um enunciado interrogativo. Note-se que a marcação de informação prosódica tenta evitar o uso de pontuação ortográfica para evitar que a transcrição seja interpretada como um texto escrito.

- (1) *JOS: olhe / a / a / <a sua viagem>\$
 *RIT: [<] <hhh>\$
 *JOS: / ao Brasil / como é que correu ?\$
 *RIT: hhh / a viagem ao Brasil / &corr / correu bem // \$

Num corpus construído a partir de fontes escritas, é necessário assegurar que todas as informações que constam do original (em formato manuscrito, impresso ou digital) são mantidas, como as marcas de negrito, itálico, sublinhado, palavras rasuradas ou inseridas acima da linha, etc., e codificadas numa linguagem que possa ser interpretada por diferentes programas informáticos. O capítulo 10 Avanços nas humanidades digitais apresenta informação detalhada sobre este aspeto.

As propriedades de cada texto e de cada transcrição são descritas ao nível dos *metadados*, que podem constituir uma base de dados independente do corpus, ou podem estar integrados em cada ficheiro num cabeçalho. Os metadados dos ficheiros escritos incluem a informação bibliográfica disponível e informação sobre a fonte, o género e o(s) tópico(s), a variedade do português, nacionalidade do autor, bem como informação de gestão do ficheiro (código de identificação, estado do tratamento do ficheiro, localização, existência de autorização para distribuição do texto, etc.). Por sua vez, os metadados das transcrições devem documentar o tipo de situação de comunicação gravada, incluir dados sobre os intervenientes e informação de gestão do corpus (códigos dos ficheiros de som, de texto e de alinhamento e, eventualmente, de vídeo, autor da transcrição e alinhamento). Existem várias iniciativas internacionais para estabelecer metadados estandardizados para cada tipo de texto, que permitam a interoperabilidade entre diferentes projetos. As normas da *Text Encoding Initiative* (TEI) incluem uma secção sobre codificação de metadados e outras propostas foram compiladas no âmbito das redes internacionais META-SHARE (*Multilingual Europe Technology Alliance*) e FLARENET (*Fostering Language Resources Network*).

2.3 Direitos de autor e distribuição do corpus

Um aspeto essencial do trabalho de compilação e distribuição de um corpus consiste em assegurar, sempre que possível, autorizações por parte dos detentores dos direitos de autor dos textos e, ainda, por parte dos informantes que disponibilizam a gravação da sua voz/imagem. Um corpus totalmente coberto por autorizações de uso e divulgação pode ser distribuído integralmente, com total acesso aos ficheiros e possibilidade de descarregamento e uso num computador local, com ferramentas próprias. Um corpus com restrições de acesso apenas poderá estar acessível para pesquisas *online* de concordâncias de tamanho restrito, ao abrigo da lei da citação (400 palavras). Este acesso pode ser suficiente e até amigável para trabalhos em linguística, mas não permite o uso do corpus para treino e desenvolvimento de novas ferramentas para o processamento da linguagem. Existem duas agências de distribuição de recursos: ELRA (European Language Resources Association) e LDC (Linguistic Data Consortium), que asseguram a distribuição de recursos para investigação e uso comercial mediante licenças. Estas duas instituições contribuíram para lançar o *ISLRN – International Standard Language Resource Number* (www.islrn.org), um número de identificação único para cada recurso linguístico. A apresentação dos corpora para o português na secção 3 terá em consideração a existência de autorizações e o tipo de distribuição.

2.4 Pesquisa de corpora

Subjacente ao trabalho de compilação de um corpus está sempre a possibilidade de pesquisar os textos para localizar ocorrências de fenómenos específicos que se pretende analisar. Para esse efeito, foram criados programas chamados *concordanciadores* que visam facilitar a pesquisa sobre corpora e que permitem obter uma *concordância*, isto é, o conjunto das ocorrências de determinado fenómeno no corpus. Os resultados apresentam tipicamente o item ou expressão pesquisado (por vezes chamado o *nó*) centrado numa linha de contexto, num formato KWIC (*Key Word In Context*), como se ilustra na Figura 1. A possibilidade de ordenação dos resultados pelo nó, pelo contexto à esquerda ou à direita ou ainda por uma combinação destas opções permite identificar de forma rápida regularidades de coocorrência lexical e padrões sintáticos. A utilização de expressões regulares permite pesquisas dirigidas e complexas, que incidem sobre uma palavra ou expressão, ou ainda sobre os vários níveis de anotação linguística que foram adicionados ao texto original (cf. secção 4). Além disso, os concordanciadores permitem obter informação sobre a frequência dos itens lexicais no corpus e, nalguns casos, sobre as colocações de uma palavra (os seus coocorrentes privilegiados).

ir brincar com as meninas lá da rua e atirou -se da	janela	abaixo . O resultado foi partir os vidros todos e uma das
e é já ! Mal disse isto , saltámos todas da	janela	abaixo . Ou , por outra , como todas nós estávamos enfiadas
há executivos a pensarem que talvez seja melhor atirarem -se de uma	janela	abaixo . Porquê ? Porque deitaram contas à vida e resultado deu
bastava também que lhe batessem à porta para ir logo tudo da	janela	abaixo : calhaus de que se munira , os ladrilhos do fogão
de Salazar apaixonada -se por uma presa política que se atira da	janela	abaixo e morre . Porém , este romance pretende mostrar mulheres snobes
vestir , ao fundo , onde ela costura : os batentes da	janela	abanam como sacudidos por mão oculta , rapidamente ; os dois guarda-fatos
voltou para o escritório . Bruto ! , então não deixara a	janela	aberta ! Fechou -a e atirou -se com um suspiro para a
com um sorriso de superioridade . - E só durmo com a	janela	aberta ! Pois sim , mas não se lava senão uma vez
" . Somos , por isso , por uma Académica de "	janela	aberta " , cuja vivência em pormenor não seja tabu só desvendado

Figura 1: Concordâncias de *janela* no CQPweb em formato KWIC, ordenadas pelo contexto à direita do termo pesquisado.

Alguns corpora de português podem ser pesquisados *online* através do sistema IMS CQP – Corpus Query Processor (Evert/Hardie 2011). Os corpora totalmente disponíveis podem ser usados localmente com concordanciadores livres, como o AntConc (Anthony 2013), ou com programas mais completos e pagos, como o MonoConc Pro (Barlow 2000) ou o WordSmith (Scott 2011). Alguns corpora orais alinhados permitem uma pesquisa *online* que fornece uma linha de contexto acompanhada da audição desse fragmento. Outros corpora podem ser explorados localmente com os mesmos programas que permitem a transcrição e alinhamento. No caso do EXMARALDA, para além do módulo de transcrição (PartiturEditor), existe ainda um módulo de gestão de corpus (Coma) e um módulo de pesquisa (Exakt), que permite a extração das concordâncias, a audição dos contextos e a filtragem dos resultados com base nos metadados associados.

3 Corpora de português e estudos linguísticos baseados em corpora

Nesta secção apresenta-se uma visão geral dos corpora existentes para o português europeu (PE) que não se pretende seja exaustiva, uma vez que muitos projetos envolvem como objetivo principal ou secundário a compilação de corpora do português. Os corpora aqui referidos e o seu endereço web são listados no final do capítulo.

3.1 Corpora de grandes dimensões

Existem três corpora de grandes dimensões para o PE (sendo que nalguns casos estes corpora abrangem igualmente outras variedades do português no mundo).

O primeiro corpus desta natureza a ser compilado para o português foi o *Corpus de Referência do Português Contemporâneo – CRPC*, iniciado em 1988 no Centro de Linguística da Universidade de Lisboa (CLUL) (Bacelar do Nascimento 2000; Génèreux/Hendrickx/Mendes 2012). O CRPC é um corpus de português contemporâneo

com cerca de 311 milhões de palavras, de textos da segunda metade do séc. XIX até 2006, embora a grande maioria seja posterior a 1970. Caracteriza-se pela diversidade de géneros textuais incluídos: textos literários, didáticos, técnicos e académicos, folhetos, cartas, Acórdãos do Supremo Tribunal de Justiça, Diários da Assembleia da República, e uma subparte oral com diferentes registos e cobrindo situações de comunicação diversas (os subcorpora orais *Português Fundamental*, *Português Falado* e *C-ORAL-ROM*, serão apresentados na secção sobre corpora orais). Embora planeado como um corpus de referência, evoluiu para um corpus de tipo monitor: isto é, apresenta preocupações com a diversidade dos textos incluídos, embora não almeje o equilíbrio entre os diferentes géneros textuais. Trata-se de um corpus aberto, que foi e continua a ser objeto de novas inclusões de materiais. Foi automaticamente anotado com informação de classe de palavras (e de flexão verbal e nominal) e lematizado.

Por restrições de direitos de autor, a parte escrita do CRPC apenas está disponível para pesquisa *online* na plataforma CQPweb (Hardie 2012) e foi processado com o IMS CQP, pelo que possibilita um conjunto alargado de opções de pesquisa de concordâncias: por palavra, parte de palavra, sequência de palavras, expressão regular. Todas elas podem ser combinadas com a pesquisa de lema, classe de palavra e flexão, existindo a possibilidade de filtrar os resultados por algumas categorias textuais latas. Os resultados podem ser ordenados e descarregados. Fornece ainda informação sobre a frequência de palavras, a sua distribuição no corpus, as suas colocações mais significativas e ainda uma lista total de palavras do corpus. Foram isolados, no CRPC, dois subcorpora de textos livres de direitos de autor, que puderam por isso ser distribuídos sem restrições de acesso no catálogo ELRA. É o caso do Corpus Literário (Corpus LT), com 1,7 milhões de palavras, que inclui 70 obras literárias de PE e de português do Brasil (PB) que caíram entretanto no domínio público; e também do Corpus de Política (PTParl), com 1 milhão de palavras dos Diários da Assembleia da República.

O *Corpus do Português*, com 45 milhões de palavras, inclui textos escritos e transcrições do oral para PE e PB desde o séc. XIV ao séc. XX, com origem em vários outros corpora compilados para o português (Davies 2014). O corpus foi anotado com informação de classe de palavras e lema. Não há acesso integral ao corpus por questões de direitos de autor, mas a plataforma permite pesquisas *online* por século, com informação de variedade e género. Fornece concordâncias, informação de frequência total por séc./género/variedade, e ainda informação sobre colocações (ver mais informação sobre o subcorpus de português não contemporâneo em 3.3).

O corpus *CETEMPúblico* (Linguatca) contém 190 milhões de palavras extraídas de edições do jornal diário *Público* entre 1991 e 1998, tendo os artigos sido subdivididos em excertos de algumas frases por questões de direitos de autor (Rocha/Santos 2000). O corpus foi anotado com vários níveis de informação: classe de palavras, flexão verbal e nominal, lematização e constituintes sintáticos com o programa PALAVRAS (Bick 1999). Foi desenvolvida uma plataforma de pesquisa *online* para vários corpora da Linguatca no âmbito do projeto AC/DC, que usa o sistema IMS CQP e permite pesquisas complexas que combinam os itens lexicais e sua anotação, e

ainda a distribuição no corpus. Para além destas funcionalidades de pesquisa que permitem obter dados para análise linguística, o corpus CETEMPúblico está ainda disponível de forma integral, a pedido, sendo por isso uma fonte de dados muito importante para o desenvolvimento de ferramentas para o processamento automático do português. O CETEMPúblico foi utilizado em inúmeros trabalhos, tanto como fonte de dados para análise linguística, como como fonte de materiais para a construção de aplicações para o processamento da língua portuguesa.

A Linguateca produziu ainda alguns corpora de géneros textuais específicos, como o corpus *CoNE (Correio Não Endereçado)*, um corpus de 675.000 palavras de PE e PB, com mensagens de correio eletrónico com publicidade ou informação recebidas pela equipa da Linguateca e não endereçadas a um destinatário específico. Ou ainda o corpus *Avante*, um corpus de 6,8 milhões de palavras do jornal *Avante!*, órgão oficial do Partido Comunista Português, e o corpus *Natura/Minho*, com 1,7 milhões de palavras, com edições do jornal regional *Diário do Minho*. Estes corpora estão anotados com os mesmos níveis de anotação do CETEMPúblico e através da mesma plataforma AC/DC (veja-se o site da Linguateca para uma descrição destes corpora e sua pesquisa).

3.2 Corpora de fala

As dimensões dos corpora de fala para o português são muito diferentes das que foram referidas para o escrito na secção anterior. De facto, a compilação e preparação destes dados é morosa.

Referimos acima que o CRPC tem uma subparte oral, constituída no âmbito de dois projetos que deram lugar a corpora específicos e independentes para o PE: o *Português Fundamental*, iniciado ainda antes do CRPC, nos anos 1970, e o corpus *C-ORAL-ROM*. O *Português Fundamental* inclui 1800 gravações de conversas espontâneas de falantes com idades, níveis de escolaridade e profissões diversificadas (Bacelar do Nascimento/Garcia Marques/Segura da Cruz 1987; Bacelar do Nascimento/Rivenc/Segura da Cruz 1987), tendo 1400 destas gravações sido transcritas (700.000 palavras). As transcrições foram recentemente atualizadas para o formato XML, alinhadas com o programa EXMARaLDA e anotadas com informação de classe de palavras. Essa nova versão está disponível gratuitamente para investigação no catálogo da ELRA. O corpus C-ORAL-ROM foi constituído mais recentemente e constitui-se como um conjunto de corpora orais comparáveis para 4 línguas românicas (português, espanhol, francês e italiano) (Bacelar do Nascimento et al. 2005). A transcrição da parte portuguesa contém 300.000 palavras, alinhadas com o som no programa WinPITCH e recentemente transpostas para o formato EXMARaLDA. O corpus das 4 línguas está também disponível no catálogo ELRA.

O corpus *REDIP – Rede de Difusão Internacional do Português: rádio, televisão e imprensa* (Instituto de Linguística Teórica e Computacional – ILTEC, CLUL, Universi-

dade Aberta) inclui 330.000 palavras de textos escritos e de programas de media portugueses. A subparte oral está pesquisável *online* no site do ILTEC, para obtenção de concordâncias. O corpus *Corp-Oral* (ILTEC) contém 56 gravações (50 horas) de conversas entre falantes da área metropolitana de Lisboa com perfis muito diversificados. O corpus pode ser pesquisado *online* para obtenção de concordâncias através do programa Spock, que permite a audição dos segmentos pesquisados, e pode ainda ser obtido, a pedido, através da *Isle MetaData Initiative* (Max Planck Institute for Psycholinguistics).

Existem vários corpora especificamente compilados e utilizados para o processamento automático da fala, embora nem sempre estejam disponibilizados. O Corpus *HESITA* (Instituto de Telecomunicações – IT) contém 27 horas de gravações de fala de telejornais diários da televisão portuguesa, transcritas e especificamente anotadas ao nível das hesitações e outras disfluências (Candeias et al. 2013). As transcrições do corpus podem ser obtidas *online* e os ficheiros de som a pedido. Vários corpora orais foram ainda compilados e explorados pelo Instituto de Engenharia de Sistemas e Computadores (INESC-ID) para sistemas de síntese e reconhecimento automáticos da fala, como o corpus de aulas universitárias *LECTRA*, o corpus de notícias *ALERT* e ainda o corpus de diálogos *CORAL*.

A Linguateca disponibiliza pesquisas sobre o *Corpus Museu da Pessoa*, que contém 1,4 milhões de palavras de entrevistas de falantes de PE e PB posteriores a 2000. O corpus está acessível para pesquisas através da plataforma AC/DC, com os mesmos níveis de anotação e funcionalidades de pesquisa acima referidas. Note-se que apenas é pesquisável a transcrição escrita, não havendo possibilidade de audição do registo sonoro associado.

3.3 Corpora de português não contemporâneo

Foi já aqui referido o *Corpus do Português*, que abarca estádios antigos da língua portuguesa e permite pesquisas dirigidas por século. Contém os seguintes corpora diacrónicos, compilados por outras instituições e, nalguns casos, pesquisáveis noutros acessos: (i) o *Corpus Informatizado do Português Medieval* (Centro de Linguística da Universidade Nova de Lisboa – CLUNL); (ii) o *Corpus Tycho-Brahe* (Universidade Estadual de Campinas – Unicamp); (iii) o *Corpus Lexicográfico do Português* (CLUL/Universidade de Aveiro); (iv) o *Corpus Eletrónico de Textos Históricos* (Universidade do Minho); (v) o *Corpus Eletrónico do CELGA – Português do Período Clássico* (cujos textos do séc. XVI e XVII também estão disponíveis em pdf no site do Centro de Estudos de Linguística Geral e Aplicada – CELGA); (vi) e o *Corpus Eletrónico de forais de Vila Real e Bragança* (Universidade de Trás-os-Montes e Alto Douro).

O *Corpus Informatizado do Português Medieval* (CIPM) inclui textos latino-romances do século IX ao século XII e textos portugueses do século XII ao século XVI (Xavier/Crispim 2002). Permite o acesso aos textos integrais com alguma normaliza-

ção gráfica e possibilita a pesquisa de palavras sobre a totalidade do corpus, devolvendo uma lista de textos nos quais ocorre o termo pesquisado. O *Corpus Tycho-Brahe* (Unicamp) é um corpus eletrônico anotado, composto de textos escritos por autores portugueses nascidos entre 1380 e 1881 (Galves/Faria 2010). Estão disponíveis 65 textos (2.792.217 palavras) para pesquisa livre, com um sistema de anotação linguística em duas etapas: anotação morfossintática (aplicada em 34 textos, num total de 1.508.757 palavras); e anotação sintática (aplicada em 16 textos, num total de 671.694 palavras).

No âmbito do projeto *WOChWEL*, que teve recentemente início no CLUL, está em fase de compilação um novo corpus de *Textos Antigos do Português*, cronologicamente anteriores aos que integram o *Tycho Brahe*, anotados com informação morfossintática e sintática, de acordo com o modelo do corpus *Tycho-Brahe*, e disponibilizados em versões integrais. O *Corpus Lexicográfico do Português* (CLUL/Universidade de Aveiro) contém dicionários do séc. XVI ao séc. XIX, na sua maior parte bilingues (os mais importantes são de português-latim e de latim-português). O corpus pode ser pesquisado, obtendo-se indicação do número de ocorrências do termo pesquisado nas diferentes obras incluídas, podendo ainda obter-se concordâncias desses contextos. O corpus *Post Scriptum* (Arquivo Digital de Escrita Quotidiana em Portugal e Espanha na Época Moderna) constitui um caso particular no âmbito dos corpora diacrônicos pelo tipo de textos incluídos. Sendo uma continuação dos projetos CARDS/FLY, é constituído por cartas particulares portuguesas e espanholas, do séc. XVI ao séc. XIX, escritas por pessoas de diferentes estratos sociais (Vaamonde et al. 2014). Os materiais manuscritos foram digitalizados e codificados em formato XML, podendo ser pesquisados pelo extenso conjunto de metadados disponíveis (data, local, palavras-chave, tipo de carta: amor, amizade, informação, etc.) em duas plataformas distintas.

3.4 Corpora de variedades regionais de PE

Paralelamente ao desenvolvimento do *Atlas Linguístico e Etnográfico de Portugal e da Galiza* (ALEPG), do *Atlas Linguístico do Litoral Português*, do *Atlas Linguístico e Etnográfico dos Açores*, e do projeto *Fronteira Dialectal do Barlavento Algarvio*, foi compilado um conjunto de gravações áudio que podem ser consultadas a pedido no CLUL. Uma parte destas gravações foi reunida no corpus *CORDIAL-SIN – Corpus Dialectal para o Estudo da Sintaxe*, que contém 600.000 palavras (Carrilho 2010). Para cada gravação, existe informação sobre os metadados do informante e da gravação, uma transcrição conservadora com informação de pausas, sobreposições de turno de fala, hesitações, repetições, e uma transcrição normalizada sem estas marcas da oralidade, sobre a qual foi acrescentada anotação morfossintática e anotação sintática ao nível da oração, pesquisável através da ferramenta *CorpusSearch*. Além disso, a totalidade dos materiais pode ser descarregada no *site* do projeto.

Outro corpus, o *CPE-Var – Corpus de Português Europeu – Variação* (CLUL), incide especificamente sobre dados de fala recolhidos entre 1996 e 1997 nas cidades de Lisboa e Braga. O corpus, que visa estudos fonéticos e fonológicos, contém, para cada informante, uma parte de conversa formal, seguida de leitura de palavras isoladas, de frases isoladas e de textos e, para concluir, uma parte de conversa informal (Rodrigues/d'Andrade 1998). Os informantes analfabetos são gravados em contexto de conversa espontânea. O estudo da fala de Braga é também o tema do corpus *Perfil Sociolinguístico da Fala Bracarense* (Centro de Estudos Humanísticos da Universidade do Minho – CEHUM), que contém 60 entrevistas compiladas de acordo com critérios sociolinguísticos, transcritas e alinhadas com o programa EXMARaLDA, e disponíveis de forma integral.

3.5 Corpora de variedades do português no mundo

Alguns corpora acima referidos, como o *CRPC* e o *Corpus do Português*, incluem subpartes de variedades do português no mundo, embora não pretendam constituir-se como corpora comparáveis, com a mesma composição e dimensão. Há ainda alguns corpora acessíveis através do AC/DC da Linguateca que incluem as variedades portuguesa e brasileira. É o caso, por exemplo, do *corpus CONDIVport* (Silva 2008), com 5,6 milhões de palavras extraídas de jornais e revistas portuguesas e brasileiros (3,3 milhões de PE), que foi anotado com os mesmos níveis de anotação do *CETEM-Público*.

Tal como o *CONDIVport*, outros corpora foram compilados tendo já em vista fornecer materiais para estudos contrastivos entre as variedades do português. No CLUL, foi compilado o corpus *Português Falado: documentos autênticos*, com 92.000 palavras de conversas informais e formais registadas entre 1970 e 1990, que ilustram todas as variedades do português no mundo (cf. Bacelar do Nascimento 2001). Uma nova versão do corpus tem alinhamento som-texto no formato EXMARaLDA e está disponível no catálogo ELRA. O corpus *VARPORT – Análise Contrastiva de Variedades do Português* (Universidade Federal do Rio de Janeiro – UFRJ/CLUL) inclui textos escritos e orais de PE e PB. A estrutura do corpus é comparável em termos de dimensão e de constituição e inclui, para o escrito, anúncios, editoriais e notícias de jornais do séc. XIX e XX. Embora acessíveis *online*, os materiais não estão associados a um motor de busca.

Existe ainda um conjunto de corpora que visam especificamente as variedades africanas do português. Um deles é o *Corpus África*, um corpus oral e escrito compilado no CLUL, constituído por um conjunto de 5 corpora comparáveis das variedades de Angola, Cabo Verde, Guiné-Bissau, Moçambique e São Tomé e Príncipe. Cada subcorpus tem dimensões comparáveis (24.000 palavras de oral e c. 580.000 de escrito), e uma constituição interna semelhante (Bacelar do Nascimento et al. 2008). Para além do *Corpus África*, há ainda a referir no CLUL o projeto em curso *VAPOR – Variedades*

Africanas do Português, no âmbito do qual continuam as recolhas de dados escritos e orais das variedades africanas do português, estando em preparação um corpus de português de S. Tomé.

3.6 Corpora paralelos

Alguns corpora referidos na secção anterior são comparáveis, no sentido em que foram compilados segundo os mesmos princípios de dimensão e constituição interna para poderem dar lugar a estudos contrastivos, como o *Corpus África* ou o *Corpus VARPORT*. No caso dos corpora paralelos, pretende-se incluir os mesmos textos em línguas diferentes. Para além das iniciativas da Comunidade Europeia (corpus *Euro-parl*, com as Atas das sessões do Parlamento Europeu nas diferentes línguas de trabalho; *JRC-Acquis*, com legislação europeia multilingue), foi compilado pela Linguateca o corpus *COMPARA*, um corpus paralelo bidirecional de português e inglês. O corpus *COMPARA* tem cerca de 1 milhão de palavras de textos literários dos séc. XIX–XX, traduzidos de e para inglês, anotados com informação morfossintática e lema, e alinhados ao nível da frase. Está disponível *online* um motor de pesquisa sobre o corpus que permite delimitar os textos a pesquisar e obter concordâncias alinhadas dos textos.

3.7 Corpora de aquisição e corpora de aprendizagem

Outra área de estudo para a qual é fundamental dispor de dados para análise é a aquisição do português, como língua primeira e língua segunda. Refira-se a *Base de Dados de Aquisição do Português (AcEP)*, do CLUL, com dados longitudinais espontâneos de crianças portuguesas entre 0 e 4 anos de idade, recolhidos entre 1990 e 2000. O *Corpus Freitas* e o *Corpus Santos* (Freitas et al. 2013; Santos 2006), incluídos na AcEP, estão disponíveis na plataforma CHILDES – Child Language Data Exchange System, estando o corpus Freitas também pesquisável na plataforma CQPweb. Também o *Corpus Batoréo* se encontra acessível no CHILDES: contém narrações elicítadas com base numa série de imagens, contadas por 30 adultos e 30 crianças, nos anos 1992–1993 (cf. Batoréo 2000). Existe ainda a base de dados *LumaLIDA* com gravações diárias e transcrições fonética e ortográfica (Frota/Vigário/Jordão 2012).

Os esforços de compilação de dados de aprendizagem do português por falantes estrangeiros levaram à criação de vários corpora. A *Recolha de dados de PLE* na Faculdade de Letras da Universidade de Lisboa (FLUL) produziu um corpus de 470 produções escritas por aprendentes de português língua estrangeira, num total de 70.500 palavras transcritas. A natureza dos materiais compilados, os seus metadados e as normas de transcrição são apresentados na página do corpus, que permite ainda descarregar a totalidade dos materiais em formato *plain text*. A mesma metodologia

é seguida pelo *Corpus de Produções Escritas de Aprendentes de PL2 – PEAPL2* (CELGA) e resultou num corpus de 516 produções escritas, num total de 119.381 palavras, que pode ser descarregado. O *Corpus de Aquisição de L2 – CAL2* (CLUNL) contém 281.301 palavras e difere dos anteriores por incluir produções escritas de falantes adultos e de crianças, embora o número de textos do primeiro tipo seja muito superior (1380 vs. 103) e por incluir produções orais: 192 entrevistas de adultos e 95 entrevistas de crianças (o corpus está acessível a pedido). Mais recentemente, teve início na FLUL a compilação do *Corpus de Português Língua Estrangeira/Língua Segunda* (COPLE2), constituído por um conjunto de materiais escritos e orais que integram um acervo de aproximadamente 1000 textos produzidos por cerca de 500 alunos de PLE/L2 que frequentaram a FLUL (Instituto da Cultura e Língua Portuguesa – ICLP e Centro de Avaliação de Português Língua Estrangeira – CAPLE) em 1999–2011 e continuará, no futuro, a ser alimentado por novas produções de alunos. Os originais manuscritos foram digitalizados e transcritos em formato XML e são visualizáveis e pesquisáveis através da plataforma TEITOK (Mendes et al. 2014 e Janssen 2016).

Com o objetivo de analisar a produção académica de estudantes de licenciatura e fornecer a base de uma intervenção didática linguisticamente orientada, foi iniciada a compilação do *CUTE – Corpus of Portuguese Undergraduate Texts* (CLUL/Escola Superior de Educação de Lisboa – ESELx), constituído por artigos de divulgação e artigos de opinião e com anotação de erros que envolvem os níveis micro e macroestruturais (Cardoso et al. 2014).

4 Anotação linguística

A anotação de fenómenos linguísticos sobre o corpus permite uma rápida identificação e extração de contextos de uma categoria ou estrutura específicas. Nos casos de corpora de grandes dimensões, a anotação é o resultado de um processo automático, que pode basear-se num sistema de regras contextuais ou numa aprendizagem automática de base estatística sobre um corpus de treino anotado manualmente. A qualidade e dificuldade do esquema de anotação (e consequente consistência esperada na anotação manual) podem ser verificadas através de uma avaliação do grau de concordância entre anotadores (*inter-annotator agreement*). A anotação automática posterior implica inevitavelmente uma percentagem de erros de anotação, que incide sobretudo nas palavras com várias categorias.

4.1 Anotação morfossintática e lematização

A anotação linguística é frequentemente acrescentada em diferentes níveis sucessivos sobre um corpus, sendo um nível elementar a lematização (atribuição de lema a cada

forma), a informação de classe de palavra ou anotação morfossintática (geralmente referida em inglês como *part-of-speech* – POS) e a flexão nominal e verbal. Apresenta-se, em (2), um excerto de livro literário do corpus CRPC anotado com etiquetas de classe de palavra (_CJ (conjunção), _QNT (quantificador), etc.) seguidas de etiquetas de flexão (#fs – feminino singular, #ms1 – masculino singular, 1.^a pessoa, #pi-3s – presente do indicativo, 3.^a pessoa singular, etc.) – cf. manual do CRPC *online*.

- (2) Mas_CJ toda_QNT#fs a_DA#fs obra_CN#fs de_PREP Auster_PNM ,_PNT admirável_ADJ#fs ,
_PNT me_CL#ms1 perturba_V#pi-3s um_UM pouco_IND pelo_PREP+DA#ms modo_CN#ms
tão_ADV explícito_ADJ#ms como_CJ se_CL#ms3 insere_V#pi-3s neste_PREP+DEM#ms pa-
radigma_CN#ms (CRPC-L0774)

A anotação morfossintática do corpus CRPC (Généreux/Hendricks/Mendes 2012) foi realizada automaticamente por um etiquetador treinado num corpus manualmente revisto, o corpus *CINTIL-Corpus Internacional do Português* (NLX/CLUL), por sua vez uma adaptação do *Corpus PAROLE*, um corpus de 250.000 palavras disponível no catálogo ELRA (Bacelar do Nascimento et al. 1998). O corpus *CINTIL* tem um milhão de palavras (de textos escritos e de transcrições de oralidade) lematizadas e anotadas com etiquetas morfossintáticas e flexionais que foram manualmente revistas. O corpus está disponível *online* para pesquisas e ainda no catálogo ELRA.

Uma fase prévia à anotação consiste no processo de *tokenização* do corpus, isto é, da sua segmentação em *tokens* ou unidades linguisticamente significativas. Este processo permite identificar duas formas lexicais numa sequência gráfica sem espaços como verbo-clítico (e.g. *cantei-o*, *cantá-lo-ia*). As opções de tokenização poderão variar consoante os corpora. Por exemplo, no CRPC, em (2), a contração da preposição *por* com o artigo definido *o* (*pelo*) foi tratada como uma unidade gráfica, tal como no corpus CETEMPúblico, enquanto no corpus *CINTIL* é transformada em duas unidades (*por_ o*). Estas opções condicionam por sua vez a anotação morfossintática: a contração *pelo* recebe uma etiqueta conjunta PREP+DA em (2) e no sistema de anotação do CETEMPúblico (PRP+DET_ARTd), enquanto no corpus *CINTIL* cada unidade recebe uma etiqueta morfossintática única. O próprio conjunto de etiquetas morfossintáticas varia consoante os projetos e não existe um conjunto estável e padronizado para este nível de anotação. Por exemplo, a classe dos verbos pode ser etiquetada como uma única categoria ou pode ser subdividida em verbos plenos e verbos auxiliares.

4.2 Anotação sintática e semântica

O tipo de anotação sintática varia consoante o tipo de teoria sintática subjacente: gramática baseada em *estrutura de constituintes* ou gramática *valencial* (ou gramática de *dependências*). A diferença entre os dois tipos de anotação é ilustrada na Figura 2 com uma mesma frase (*Washington acompanhou os movimentos de Saddam desde a*

primeira hora.), anotada no CINTIL-Treebank, com anotação em estrutura de constituintes, e no CINTIL-DependencyBank, com anotação de relações de dependência gramatical (ambos do grupo NLX-Natural Language and Speech, da Faculdade de Ciências da Universidade de Lisboa – FCUL).

A primeira representação estrutura a frase na divisão entre sujeito e predicado e identifica constituintes nos níveis intermédios (NP e N, por exemplo). A segunda representação assinala dependências entre palavras da frase, isto é, relações binárias assimétricas, sendo o verbo o elemento central da frase, que estabelece relações de dependência com o sujeito, com o núcleo nominal do objeto direto e com a preposição que introduz o adjunto temporal. O núcleo nominal *movimentos* estabelece, por sua vez, relação de dependência com o artigo definido e com a preposição, e esta com o nome.

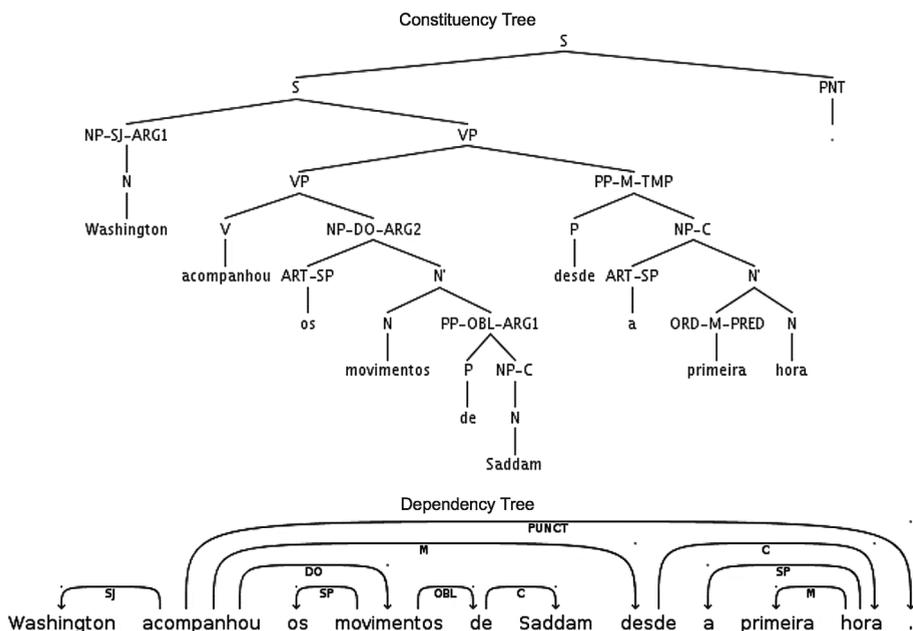


Figura 2: Anotação no CINTIL-Treebank e no CINTIL-DependencyBank.

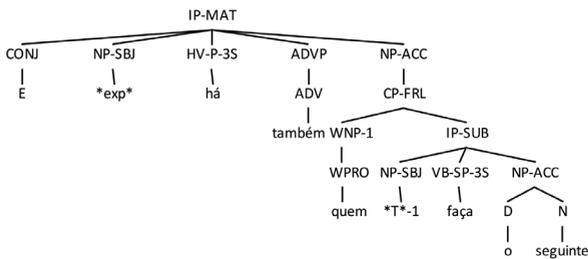
A anotação sintática em estrutura de constituintes varia ainda de acordo com os objetivos dos projetos. Veja-se a anotação da sequência *E há também quem faça o seguinte*, no corpus oral CORDIAL-SIN (anotação que segue os princípios do Penn-Helsinki Parsed Corpus of Middle English 2) e a anotação automática da mesma sequência usando a ferramenta LX-parser (NLX), desenvolvida com base no CINTIL-Treebank (cf. Figura 3). Contrariamente ao resultado do LX-parser, a anotação sintática do CORDIAL-SIN permite múltiplas ramificações num mesmo nível da estrutura (sujeito, verbo e objeto direto estão, por exemplo, ao mesmo nível); não representa os

níveis intermédios (como N', ADJ'); considera uma posição de expletivo nulo com verbos impessoais; e representa categorias vazias, como a posição NP-SBJ da forma verbal *faça*. Muitas das opções tomadas no CORDIAL-SIN estão associadas ao facto de se tratar de um corpus de oralidade e beneficiam de vários estudos sintáticos de membros da equipa.

Um excerto de um milhão de palavras do CETEMPúblico e do CETENFOLHA (o equivalente para o PB) foram anotados sintaticamente e constituem a Floresta Sintá(c)tica (Afonso et al. 2002), anotada com o etiquetador PALAVRAS. É possível obter concordâncias ou visualizar a estrutura em árvore, a partir de editores como o Milhafre.

Os níveis de anotação semântica e discursiva (por exemplo: papéis semânticos, dependências referenciais, localização temporal de eventos, modalização de eventos, relações discursivas) foram ainda pouco desenvolvidos para o PE (Branco et al. 2012). Alguns exemplos a este nível são: o CINTILDependencyBank que contém informação de papéis semânticos; o CETEMPúblico, que permite uma pesquisa da distribuição de determinada forma por uma lista limitada de campos semânticos; a anotação de valores modais num corpus escrito de 2000 frases (160.000 palavras) que será brevemente disponibilizado (Hendrickx/Mendes/Mencarelli 2012).

CORDIAL-SIN



LX-parser

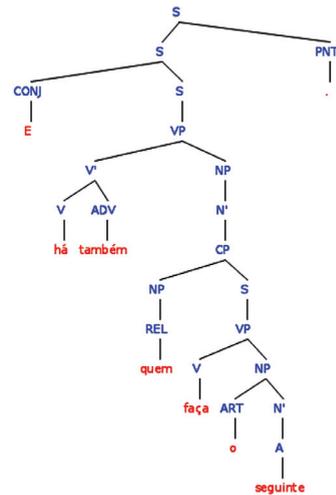


Figura 3: Exemplo de anotação sintática no corpus CORDIAL-SIN (CORDIAL-1, INF-AAL02.35) e no CINTIL-Trebank.

4.3 Anotação do erro

Os corpora de aprendizagem de português L2/LE são, nalguns casos, anotados com etiquetas que identificam o tipo de erro cometido pelo aprendente. Trata-se de uma anotação abrangente, que contempla diferentes níveis de informação linguística. Por exemplo, o corpus *CUTE – Corpus of Portuguese Undergraduates’ Texts*, é anotado nos seguintes níveis de análise: ortografia, pontuação, morfologia, sintaxe, morfossintaxe, sintaxe/semântica, semântica e gralhas.

5 Alguns estudos baseados em corpora

A disponibilidade dos corpora de fala referidos em 3.2 veio trazer novas perspetivas para o estudo da oralidade, até aí encarada como uma produção sem estruturação interna, ao contrário da escrita, e permitiu apontar propriedades sintáticas e discursivas próprias da oralidade (Bacelar do Nascimento 1996), ou ainda verificar a existência de variação interna ao sistema linguístico, como o caso das diferentes estratégias de relativização para além da consagrada pela norma, identificadas no corpus CRPC (Alexandre 2000), no corpus REDIP (Arim/Freitas 2004) e no corpus CORDIAL-SIN (Cardoso/Alexandre 2013), tendo este corpus sido fonte de dados para diversas análises, como a da expressão pronominal *a gente* (Pereira 2003), do gerúndio flexionado (Lobo 2008), do expletivo *ele* (Carrilho 2005), da interpolação (Magro 2007), da construção de duplo sujeito com *se* (Martins 2009), da concordância sujeito-verbo (Cardoso/Carrilho/Pereira 2011) e das estruturas clivadas (Vercauteren 2010).

O corpus *Português Fundamental* foi ainda fonte de muitos estudos contrastivos entre PE e PB, como a análise de aspetos morfossintáticos apresentada em Duarte/Vasco/Freire (2003). Os fenómenos de concordância sujeito-verbo e concordância interna ao sintagma nominal foram objeto de estudos contrastivos com base no corpus VARPORT, no corpus África e no corpus de S. Tomé do projeto VAPOR (Mota/Vieira 2013), tendo este corpus e o de Cabo Verde sido ainda tema de vários estudos de sintaxe, como o caso das estruturas relativas (cf. Alexandre/Gonçalves/Hagemeijer 2011). O *Corpus África* serviu de base ao estabelecimento de um léxico contrastivo das 5 variedades do português em África e a trabalhos contrastivos sobre as construções com *SE* nas variedades europeia, africanas e brasileira (Mendes et al. 2015). Os corpora das variedades africanas do português foram ainda fonte de dados para o reconhecimento automático das variedades no INESC-ID (projeto POSPORT). O corpus de variedades nacionais CONDIVport possibilitou uma análise dos fenómenos de convergência e divergência lexical entre as variedades portuguesa e brasileira do português, explorados em Silva (2010) e outras publicações do mesmo autor. Para além dos estudos contrastivos das variedades do português, os corpora permitem estudos de variação diacrónica da língua: o corpus *REDIP*, analisado em comparação com outros dados, permitiu a identificação de neologismos nos media portugueses

(Freitas/Arim 2004); o CIPM foi fonte de dados para muitos estudos, especialmente ao nível da sintaxe (e.g. Xavier/Fiéis/Silva 2002), tal como o corpus *Tycho Brahe* (e.g. Galves 2012).

Em semântica lexical, os contextos e os dados de frequência do CRPC apontaram para padrões de polissemia regular na classe dos verbos psicológicos (Mendes 2004) e os dados dos corpora provaram ser fundamentais na análise das diferentes funções e valores dos marcadores discursivos, como a análise da reinterpretação de *ai* e *daí* de advérbios anafóricos a conectores com valor consecutivo (Mendes/Bacelar do Nascimento 2007) e o levantamento dos valores semânticos e pragmáticos do marcador *assim* (Macário Lopes/Carapinha 2004), entre outros exemplos. Os corpora diacrónicos e a conjugação de fontes de diversos corpora possibilitam uma análise de processos de gramaticalização e discursivização na história da língua portuguesa, veja-se, entre outros, a análise de *pois* no CIPM (Lima 2002) e no corpus *Post Scriptum* (Costa 2013).

Um dos tópicos privilegiados na linguística de corpus é a análise das unidades multilexicais (UML) da língua portuguesa, isto é, sequências de palavras bem formadas sintaticamente, que apresentam algum grau de fixidez lexical e/ou sintática e, nalguns casos, também de idiomaticidade (Bacelar do Nascimento 2013). A disponibilidade de um corpus de dimensões significativas evidencia a ocorrência frequente de determinadas sequências de palavras, que podem ser totalmente fixas, como as locuções e os provérbios, terem significado não composicional, como as expressões idiomáticas, ou não serem totalmente fixas nem idiomáticas, embora tendam a ocorrer com elevada frequência, sendo nesse caso associações lexicais preferenciais designadas como *colocações* (sequências lexicais como *janela de oportunidade*, *puramente circunstancial*, *impávido e sereno*, *pessoal e intransmissível*). As colocações são especialmente interessantes por desafiar a divisão entre léxico e sintaxe, pois embora apresentem algum grau de lexicalização continuam a ser processadas composicionalmente: «the principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments» (Sinclair 1987, 320). Os concordanciadores integram geralmente uma funcionalidade que permite obter informação sobre os coocorrentes privilegiados de uma palavra pesquisada no corpus (cf. CRPC no CQPweb, e ainda o AntConc e Wordsmith).

Vários estudos apontam para o facto de as unidades multilexicais nos seus vários tipos serem muito frequentes no discurso: cerca de 70% das palavras de um texto corrido no corpus do inglês London-Lund de meio milhão seria parte de uma UML recorrente (Kennedy 1998). A quantificação das UML no discurso é dificultada pelo facto de o conceito de UML abranger fenómenos muito diversos, como provérbios, locuções gramaticais, nomes compostos, sequências de verbo suporte (ou verbo leve) seguido de elemento predicativo, colocações. É, por exemplo, especialmente difícil distinguir entre as colocações e as sequências totalmente livres. O processo de seleção das UML e a sua classificação é discutida em Mendes et al. (2006), Bacelar do Nascimento/Mendes/Antunes (2006) e Baptista/Correia/Fernandes (2004). A elabora-

ção de uma proposta de anotação de UML em textos com base num léxico estruturado é proposta em Hendrickx/Mendes/Antunes (2010).

Os corpora podem ainda criar pontes para a compreensão de fenómenos relacionados com aspetos históricos, políticos e sociológicos em Portugal, como o caso da análise do período pré e pós 25 de Abril no subcorpus da Assembleia da República integrado no CRPC (Généreux et al. 2010) ou o caso do corpus *Post Scriptum*, que fornece informação de caráter histórico e sociológico sobre a vida em Portugal e Espanha nos últimos séculos (Marquilhas 2014).

6 Principais siglas utilizadas

CHILDES	Child Language Data Exchange System
CQP	Corpus Query Processor
ELRA	European Language Resources Association
HTML	Hypertext Markup Language
ISLRN	International Standard Language Resource Number
KWIC	Key Word in Context
L1	língua primeira
L2	língua segunda
LDC	Linguistic Data Consortium
LE	língua estrangeira
POS	Part-of-Speech
TEI	Text Encoding Initiative
UML	Unidades Multilexicais
XML	Extensible Markup Language

6.1 Corpora

CRPC – Corpus de Referência do Português Contemporâneo

<http://alfclul.clul.ul.pt/CQPweb>

Subcorpora de acesso livre de direitos de autor: Corpus Literário (Corpus LT) e Corpus de Política (PTParl) <http://catalog.elra.info>

CINTIL – Corpus Internacional do Português

<http://cintil.ul.pt>; <http://catalog.elra.info>

CINTIL-Treebank, CINTIL-DependencyBank: <http://lxcenter.di.fc.ul.pt>

Corpus do Português

<http://www.corpusdoportugues.org>

Corpus CETEMPúblico

Corpus CoNE (Correio Não Endereçado)

Corpus Avante

Corpus Natura/Minho

<http://www.linguateca.pt/ACDC/>

Floresta Sintá(c)tica

<http://www.linguateca.pt/Floresta/>

6.2 Oralidade

C-ORAL-ROM – Integrated Reference Corpora for Spoken Romance Languages

<http://catalog.elra.info>

Corpus Português Fundamental

<http://www.clul.ul.pt/en/resources/84-spoken-corpus-qportugues-fundamental-pfq-r>

<http://catalog.elra.info>

REDIP – Rede de Difusão Internacional do Português: rádio, televisão e imprensa

<http://www.iltec.pt/?action=concord>

Corp-Oral

<http://www.iltec.pt/spock/>

http://corpus1.mpi.nl/ds/imdi_browser/

Corpus HESITA

<http://lsi.co.it.pt/spl/hesitation/downloads.html>

LECTRA, ALERT, CORAL

<https://www.l2f.inesc-id.pt/wiki/index.php/Resources#Corpora>

Corpus Museu da Pessoa

<http://www.linguateca.pt/acesso/corpus.php?corpus=MUSEUDAPESSOA>

6.3 Diacrónicos

CIPM – Corpus Informatizado do Português Medieval

<http://cipm.fcsh.unl.pt>

Tycho-Brahe – Parsed Corpus of Historical Portuguese

<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html>

Corpus Lexicográfico do Português

<http://clp.dlc.ua.pt/Inicio.aspx>

Corpus Eletrónico de Textos Históricos

Corpus Eletrónico do CELGA – Português do Período Clássico

Corpus Eletrónico de forais de Vila Real e Bragança

Cf. Corpus do Português

Corpus Post Scriptum – Arquivo Digital de Escrita Quotidiana em Portugal e Espanha na Época

Moderna

<http://alfclul.clul.ul.pt/cards-fly/index.php?page=mainen>

<http://ps.clul.ul.pt/index.php?page=downloads>; plataforma TEITOK <http://cards-fly.clul.ul.pt/teitok/postscriptum/index.php?action=home>

<http://ps.clul.ul.pt/index.php?page=downloads>; plataforma TEITOK <http://cards-fly.clul.ul.pt/teitok/postscriptum/index.php?action=home>

Corpus Wochwel

<http://alfclul.clul.ul.pt/wochwel/oldtexts.html>

6.4 Dialectais

CORDIAL-SIN – Corpus Dialectal para o Estudo da Sintaxe

<http://www.clul.ul.pt/pt/recursos/226-corpus-syntax-oriented-corpus-of-portuguese-dialects-cordial-sin>

Gravações no âmbito das recolhas dos projetos *Atlas Linguístico e Etnográfico de Portugal e da Galiza (ALEPG)*, *Atlas Linguístico do Litoral Português*, *Atlas Linguístico e Etnográfico dos Açores*, *Fronteira Dialetal do Barlavento Algarvio*

<http://www.clul.ul.pt/en/research-teams/516-related-projects>

Corpus Perfil Sociolinguístico da Fala Bracarense

<https://sites.google.com/site/projectofalabracarense/corpus>

6.5 Variedades do português

Corpus Português Falado: documentos autênticos

<http://www.clul.ul.pt/pt/recursos/83-spoken-portuguese-geographical-and-social-varieties-r>

<http://catalog.elra.info>

VARPORT – Análise Contrastiva de Variedades do Português

<http://www.letras.ufrj.br/varport/>

Corpus África

<http://www.clul.ul.pt/pt/investigacao/186-linguistic-resources-for-the-study-of-the-african-varieties-of-portuguese>; plataforma TEITOK <http://alfclul.clul.ul.pt/teitok/corpusafrica/>

Corpus de português de S. Tomé (projeto VAPOR)

<http://www.clul.ul.pt/pt/investigacao/184-vapor-african-varieties-of-portuguese>

Corpus CONDIVport

<http://www.linguateca.pt/ACDC/>

6.6 Paralelos

Corpus Europarl, Corpus JRC-Acquis

<http://ipsc.jrc.ec.europa.eu/index.php?id=198>

Corpus COMPARA

<http://193.136.2.104/COMPARA/Bem-vindos.html>

6.7 Aquisição / Aprendizagem

Base de Dados de Aquisição do Português (AcEP)

<http://www.clul.ul.pt/en/research-teams/476-acquisition-of-european-portuguese-databank>

Corpus Freitas

<http://childes.psy.cmu.edu/data/Romance/Portuguese/>

<http://alfclul.clul.ul.pt/CQPweb/child/>

Corpus Santos

<http://childes.psy.cmu.edu/data/Romance/Portuguese/>

LumaLiDa – Linguistic Diary of Luma

<http://labfon.letras.ulisboa.pt/LumaLiDa.htm>

Recolha de dados de PLE

<http://www.clul.ul.pt/pt/recursos/314-corpora-of-ple>

PEAPL2 – Corpus de Produções Escritas de Aprendentes de PL2

<http://www.uc.pt/fluc/rcpl2/>

CAL2 – Corpus de Aquisição de L2

<http://cal2.clunl.edu.pt/>

COPLE2 – Corpus de Português Língua Estrangeira/Língua Segunda

<http://www.clul.ul.pt/en/research-teams/547>; plataforma TEITOK <http://alfclul.clul.ul.pt/teitok/learnercorpus/>

CUTE – Corpus of Portuguese Undergraduates' Texts

<http://www.cute.org.pt>

7 Referências

- Afonso, Susana, et al. (2002), *Floresta sintá(c)tica: um treebank para o português*, in: Anabela Gonçalves/Clara Nunes Correia (edd.), *Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística (APL 2001)*, Lisboa, APL, 533–545.
- Alexandre, Nélia (2000), *A Estratégia Resumptiva em Relativas Restritivas do Português Europeu*, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Alexandre, Nélia/Gonçalves, Rita/Hagemeijer, Tjerk (2011), *A formação de relativas de PP no português oral de Cabo Verde e de S. Tomé*, in: Armanda Costa/Pilar Barbosa/Isabel Falé (edd.), *Textos Seleccionados do XXVI ENAPL 2010*, Lisboa, APL, 17–34.
- Anthony, Laurence (2013), *Developing AntConc for a new generation of corpus linguists*, in: *Proceedings of the Corpus Linguistics Conference (CL 2013)*, Lancaster, UCREL Lancaster University, 14–16.
- Arim, Eva/Freitas, Tiago (2004), *Estratégias de relativização nos meios de comunicação social portugueses*, in: *Actas do XIX Encontro Nacional da Associação Portuguesa de Linguística (APL)*, Lisboa, APL, 279–288.
- Bacelar do Nascimento, Maria Fernanda (1996), *Aspectos da sintaxe do português falado (repetições lexicais e de estruturas sintáticas em produções orais: fenômenos de deslocação)*, in: Inês Duarte/Isabel Leiria (edd.), *Actas do Congresso Internacional sobre o Português*, vol. 1, Lisboa, APL, 203–223.
- Bacelar do Nascimento, Maria Fernanda (2000), *Corpus de Référence du Portugais Contemporain*, in: Mireille Bilger (ed.), *Corpus, Méthodologie et Applications Linguistiques*, Paris, Champion / Presses Universitaires de Perpignan, 25–30.
- Bacelar do Nascimento, Maria Fernanda (coord.) (2001), *Português Falado, Documentos Autênticos, Gravações audio com transcrições alinhadas*, Lisboa, Centro de Linguística da Universidade de Lisboa e Instituto Camões [CD-ROM].
- Bacelar do Nascimento, Maria Fernanda (2013), *Processos de Lexicalização*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 9, Lisboa, Fundação Calouste Gulbenkian, 215–246.
- Bacelar do Nascimento, Maria Fernanda/Garcia Marques, Maria Lúcia/Segura da Cruz, Maria Luísa (1987), *Português Fundamental, Métodos e Documentos*, tomo 1: *Inquérito de Frequência*, Lisboa, INIC, CLUL.
- Bacelar do Nascimento, Maria Fernanda/Mendes, Amália/Antunes, Sandra (2006), *Typologies of Multi-Word Expressions Revisited: A Corpus-driven Approach*, in: Yuji Kawaguchi et al. (edd.), *Spoken Language Corpus and Linguistic Informatics*, Amsterdam/Philadelphia, Benjamins, 227–244.
- Bacelar do Nascimento, Maria Fernanda/Rivenc, Paul/Segura da Cruz, Maria Luísa (1987), *Português Fundamental, Métodos e Documentos*, tomo 2: *Inquérito de Disponibilidade*, Lisboa, INIC, CLUL.
- Bacelar do Nascimento, Maria Fernanda, et al. (1998), *LE-PAROLE – Do corpus à modelização da informação lexical num sistema-multifunção*, in: *Actas do XIII Encontro da Associação Portuguesa de Linguística*, Lisboa, APL, 115–134.

- Bacelar do Nascimento, Maria Fernanda, et al. (2005), *The Portuguese Corpus*, in: Emanuela Cresti/Massimo Moneglia (edd.), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*, Amsterdam/Philadelphia, Benjamins, 163–207.
- Bacelar do Nascimento, Maria Fernanda, et al. (2008), *Aspectos de unidade e diversidade do português: as variedades africanas face à variedade europeia*, *Revista Veredas* 9, 35–60.
- Baptista, Jorge/Correia, Anabela/Fernandes, Graça (2004), *Frozen Sentences of Portuguese: Formal Descriptions for NLP*, in: *Second ACL Workshop on Multiword Expressions: Integrating Processing*, Barcelona, Association for Computational Linguistics, 72–79.
- Barlow, Michael (2000), *MonoConc Pro*, Houston, Athelstan.
- Barras, Claude, et al. (2000), *Transcriber: development and use of a tool for assisting speech corpora production*, *Speech Communication* 33:1–2, 5–22.
- Batoréu, Hanna Jakubowicz (2000), *Expressão do Espaço no Português Europeu. Contributo psicolinguístico para o Estudo da Linguagem e Cognição*, Lisboa, Fundação Calouste Gulbenkian/Fundação para a Ciência e a Tecnologia.
- Bick, Eckhard (1999), *The parsing system PALAVRAS*, Aarhus, Aarhus University Press.
- Boersma, Paul/Weenink, David (2013), *Praat, a system for doing phonetics by computer*, *Glott International* 5:9/10, 341–345.
- Branco, António, et al. (2012), *The Portuguese Language in the Digital Age – A Língua Portuguesa na Era Digital*, Berlin, Springer.
- Candeias, Sara, et al. (2013), *HESITA(tions) in Portuguese: a database*, in: *DiSS 2013, ISCA endorsed Interspeech 2013 satellite workshop*, Stockholm, KTH Royal Institute of Technology, 13–16.
- Cardoso, Adriana/Alexandre, Nélia (2013), *Relativas clivadas em variedades não padrão do português*, in: Fátima Silva/Isabel Falé/Isabel Pereira (edd.), *XXVIII Encontro Nacional da Associação Portuguesa de Linguística – Textos selecionados (APL 2012)*, Lisboa, APL, 205–227.
- Cardoso, Adriana/Carrilho, Ernestina/Pereira, Sandra (2011), *On verbal agreement variation in European Portuguese: syntactic conditions for the 3SG/3PL alternation*, *Diacrítica* 25:1, 137–159.
- Cardoso, Adriana, et al. (2014), *CUTE: Corpus of Portuguese Undergraduates' Texts – Um recurso para a investigação em escrita académica em português*, in: A. Moreno et al. (edd.), *XXIX Encontro Nacional da Associação Portuguesa de Linguística – Textos selecionados (APL 2013)*, Porto, APL, 169–184.
- Carrilho, Ernestina (2005), *Expletive «ele» in European Portuguese dialects*, tese de Doutoramento, Universidade de Lisboa.
- Carrilho, Ernestina (2010), *Tools for dialect syntax: the case of CORDIAL-SIN (an annotated corpus of Portuguese dialects)*, in: Gotzon Aurrekoetxea/José Luis Ormaetxea (edd.), *Tools for Linguistic Variation*, Bilbao, Universidad del País Vasco, 57–70.
- Costa, Ana Luísa (2013), *Um «pois» comentador*, in: João Veloso et al. (edd.), *XXIX Encontro Nacional da Associação Portuguesa de Linguística – Textos selecionados (APL 2013)*, Porto, APL, 199–211.
- Davies, Mark (2014), *Creating and using the Corpus do Português and the Frequency Dictionary of Portuguese*, in: Tony Berber Sardinha/Telma Ferreira (edd.), *Working with Portuguese Corpora*, London, Continuum, 89–110.
- Duarte, Maria Eugénia/Vasco, Sérgio Leitão/Freire, Gilson C. (2003), *Português europeu e português brasileiro: alguns aspectos morfossintáticos*, in: Cláudio César Henriques (ed.), *Linguagem, Conhecimento e Aplicação*, Rio de Janeiro, Europa, 253–266.
- Evert, Stefan/Hardie, Andrew (2011), *Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium*, comunicação apresentada em *Corpus Linguistics 2011*, University of Birmingham, UK (disponível em: http://cwb.sourceforge.net/files/EvertHardie_CL2011_paper.pdf, 30.09.2015).

- Freitas, Maria João, et al. (2013), *Child-Adult Interaction: A Database on European Portuguese*, versão 2.0, Lisboa, CLUL, Anagrama.
- Freitas, Tiago/Arim, Eva (2004), *Os neologismos nos meios de comunicação social portugueses*, in: *Actas do XIX Encontro Nacional da Associação Portuguesa de Linguística (APL)*, Lisboa, APL, 443–454.
- Frota, Sónia/Vigário, Marina/Jordão, Raquel (2012), *LumaLiDaOn (with prosodic labelling)*, versão 2, Lisboa, Lab.Fonética, CLUL/FLUL.
- Galves, Charlotte (2012), *Periodização e competição de gramáticas: o caso do português médio*, in: Tânia Lobo/Zenaide Carneiro/Silvana Ribeiro (edd.), *ROSAE: linguística histórica, história das línguas e outras histórias*, Salvador, Edufba, 74–88.
- Galves, Charlotte/Faria, Pablo (2010), *Tycho Brahe Parsed Corpus of Historical Portuguese*, <http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html> (20.10.2015).
- Généreux, Michel/Hendrickx, Iris/Mendes, Amália (2012), *Introducing the Reference Corpus of Contemporary Portuguese On-Line*, in: Nicoletta Calzolari et al. (edd.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation – LREC 2012*, Istanbul, ELRA, 2237–2244, <http://www.lrec-conf.org/proceedings/lrec2012/index.html> (09.01.2016).
- Généreux, Michel, et al. (2010), *Lexical analysis of pre and post revolution discourse in Portugal*, in: *Proceedings of the Third Workshop on Building Comparable Corpora*, LREC 2010, 65–71, <http://www.fb06.uni-mainz.de/lk/bucc2010/documents/Proceedings-BUCC-2010.pdf> (09.01.2016).
- Hardie, Andrew (2012), *CQPweb – combining power, flexibility and usability in a corpus analysis tool*, *International Journal of Corpus Linguistics* 17:3, 380–409.
- Hendrickx, Iris/Mendes, Amália/Antunes, Sandra (2010), *Proposal for Multi-word Expression annotation in running text*, in: *Proceedings of the Fourth Linguistic Annotation Workshop (LAW IV)*, Uppsala, Association for Computational Linguistics, 152–156.
- Hendrickx, Iris/Mendes, Amália/Mencarelli, Silvia (2012), *Modality in Text: a proposal for corpus annotation*, in: Nicoletta Calzolari et al. (edd.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation – LREC 2012*, Istanbul, ELRA, 1805–1812, <http://www.lrec-conf.org/proceedings/lrec2012/index.html> (09.01.2016).
- Janssen, Maarten (2016), *TEITOK: Text-Faithful Annotated Corpora*, in: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, ELRA, 4037–4043, <http://www.lrec-conf.org/proceedings/lrec2016/index.html> (08.06.2016).
- Kennedy, Graeme (1998), *An Introduction to Corpus Linguistics*, London/New York, Longman.
- Kilgarriff, Adam/Grefenstette, Gregory (2003), *Introduction to the Special Issue on Web as Corpus*, *Computational Linguistics* 29:3, 333–347.
- Leech, Geoffrey (1991), *The state of the art in corpus linguistics*, in: K. Aijmer/B. Altenberg (edd.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, London, Longman, 8–29.
- Lima, José Pinto de (2002), *Grammaticalization, subjectification and the origin of phatic markers*, in: Ilse Wischer/Gabriele Diewald (edd.), *New Reflections on Grammaticalization*, Amsterdam/Philadelphia, Benjamins, 363–378.
- Lobo, Maria (2008), *Variação morfo-sintática em dialetos do português europeu: o gerúndio flexionado*, *Diacrítica* 22:1, 25–55.
- Macário Lopes, Ana Cristina/Carapinha, Conceição (2004), *Contributos para uma análise semântico-pragmática das construções com «assim»*, *Cadernos de Linguística* 5, 57–80.
- Magro, Catarina (2007), *Clíticos: variações sobre o tema*, tese de Doutorado, Universidade de Lisboa.
- Marquilhas, Rita (2014), *Under-privileged social actors: their texts, their language, their history*, comunicação apresentada em *International symposium The Lower Classes, Scriptuality, and the*

- History of Language. An Interdisciplinary Balance*, Christian-Albrechts-Universität, Kiel, 6–7 November 2014.
- Martins, Ana Maria (2009), *Subject doubling in European Portuguese dialects: the role of impersonal «se»*, in: Enoch O. Aboh et al. (edd.), *Romance Languages and Linguistic Theory*, vol. 1, Amsterdam/Philadelphia, Benjamins, 179–200.
- McEnery, Anthony/Hardie, Andrew (2012), *Corpus Linguistics*, Cambridge, Cambridge University Press.
- Mendes, Amália (2004, ¹2001), *Predicados Verbais Psicológicos do Português. Contributo para o estudo da polissemia verbal*, Lisboa, Fundação Calouste Gulbenkian/Fundação para a Ciência e a Tecnologia.
- Mendes, Amália/Bacelar do Nascimento, Maria Fernanda (2007), *Grammaticalization Processes in a Spoken Portuguese Corpus: space, time and discourse*, in: María Carmen Campoy-Cubillo/María Jose Luzon (edd.), *Spoken Corpora in Applied Linguistics*, Berna, Lang, 147–160.
- Mendes, Amália, et al. (2006), *COMBINA-PT: a Large Corpus-extracted and Hand-checked Lexical Database of Portuguese Multiword Expressions*, in: *Proceedings of the V International Conference on Language Resources and Evaluation – LREC2006*, Genoa, LREC, 1900–1905, <http://www.lrec-conf.org/proceedings/lrec2006/> (09.01.2016).
- Mendes, Amália, et al. (2014), *Corpus de Português Língua Estrangeira / Língua Segunda – COPLE2*, póster apresentado no XXX Encontro Nacional da Associação Portuguesa de Linguística, Porto.
- Mendes, Amália, et al. (2015), *Pronominal constructions and subject indetermination in varieties of Portuguese*, in: *Pluricentric Languages Worldwide and Pluricentric Theory*, Frankfurt/Wien, Lang.
- Mota, Maria Antónia/Vieira, Sílvia Rodrigues (edd.) (2013), *Patterns of verbal and nominal agreement in Portuguese varieties*, *Journal of Portuguese Linguistics special issue*.
- Pereira, Sandra Maria de Brito (2003), *Gramática comparada de «a gente»: variação no Português europeu*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Rocha, Paulo/Santos, Diana (2000), *CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa*, in: Maria das Graças Volpe Nunes (ed.), *Actas do V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)*, São Paulo, ICMC/Universidade de São Paulo, 131–140.
- Rodrigues, Celeste/Andrade, Ernesto d' (1998), *CPE VAR (Corpus de Português Europeu – Variação)*, in: *Actas do XIV Encontro Nacional da Associação Portuguesa de Linguística*, vol. 2, Braga, APL, 627–629.
- Sánchez Pérez, Aquilino/Cantos Gómez, Pascual (1997), *Predictability of Word Forms (Types) and Lemmas in Linguistic Corpora. A Case Study Based on the Analysis of the CUMBRE Corpus: An 8-Million-Word Corpus of Contemporary Spanish*, *International Journal of Corpus Linguistics* 2:2, 259–280.
- Santos, Ana Lúcia (2006), *Minimal Answers. Ellipsis, Syntax and Discourse in the Acquisition of European Portuguese*, tese de Doutoramento, Universidade de Lisboa [publicado Amsterdam/Philadelphia, Benjamins, 2009].
- Schmidt, Thomas (2012), *EXMARaLDA and the FOLK tools – two toolsets for transcribing and annotating spoken language*, 236–240, <http://www.lrec-conf.org/proceedings/lrec2012/index.html> (09.01.2016).
- Scott, Mike (2011), *WordSmith Tools version 6*, Liverpool, Lexical Analysis Software.
- Silva, Augusto Soares da (2008), *O corpus CONDIV e o estudo da convergência e divergência entre variedades do português*, in: Luís Costa/Diana Santos/Nuno Cardoso (edd.), *Perspectivas sobre a Linguateca/Actas do encontro Linguateca: 10 anos*, Linguateca, 25–28, <http://www.linguateca.pt/LivroL10/Livro-Costaetal2008.pdf> (14.01.2016).
- Silva, Augusto Soares da (2010), *Measuring and parameterizing lexical convergence and divergence between European and Brazilian Portuguese*, in: Dirk Geeraerts/Gitte Kristiansen/Yves Peirsman (edd.), *Advances in Cognitive Sociolinguistics*, *Cognitive Linguistics Research* 45, 41–83.

- Sinclair, John (1987), *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- Vaamonde, Gael, et al. (2014), *Post Scriptum: Archivo Digital de Escritura Cotidiana*, in: Sagrario López Poza/Nieves Pena Sueiro (edd.), *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, Janus, Anexo 1, 473–482.
- Vercauteren, Aleksandra (2010), *Como é que é com o «é que»? Análise de estruturas com «é que» em variedades não standard do português europeu*, dissertação de Mestrado, Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa.
- Xavier, Maria Francisca/Crispim, Maria de Lourdes (2002), *Corpus Informatizado do Português Medieval – CIPM Constituição e Processamento Automático*, in: *Actas do 6.º Congresso da Associação Internacional de Lusitanistas, Agosto de 1999*, Rio de Janeiro, ALL.
- Xavier, Maria Francisca/Fiéis, Alexandra/Silva, Maria Cristina (2002), *Word Order Alternation in Medieval Portuguese*, in: *Galician/Portuguese Poetry, Prose & Linguistics*, Santa Barbara, UCSB, 307–325.

Rita Marquilhas e Iris Hendrickx

10 Avanços nas humanidades digitais

Abstract: Neste capítulo acompanham-se os avanços da filologia do português desde que o ambiente digital se começou a anunciar como o contexto mais apropriado para a circulação do conhecimento. Remonta-se às primeiras experiências de processamento mecânico de textos portugueses, quando se entreviam já duas grandes vantagens no auxílio informático para efeitos de estudo histórico da língua: prevenção de erro humano em transcrições e edições, prevenção de abandono de tarefas demasiado gigantescas para a capacidade humana. Acompanha-se uma fase ulterior, em que os académicos, a nível internacional, deixaram de instrumentalizar apenas o digital para passarem a harmonizar-se com ele, tentando compreender quantos conceitos e métodos é preciso revolucionar para que a filologia possa continuar a cumprir a responsabilidade de disciplina que se ocupa da peritagem dos textos e do seu diálogo com a história da cultura e a história da língua. Analisam-se aqueles modelos de edição académica que correspondem, por terem codificação explícita e consistente, ao imperativo da legibilidade por máquina, ao mesmo tempo que permitem, fruto da linguagem de marcação e da anotação rica que adotam, uma crescente manipulação das suas representações computacionais. E demonstra-se como a filologia do português ganhou um ritmo acelerado de experimentação a este nível.

Keywords: filologia, texto, linguagem de marcação, edição académica digital, corpus histórico

1 Introdução

A investigação em linguística, tal como acontece em muitas áreas de investigação científica, procede obrigatoriamente à validação das suas hipóteses junto de dados empíricos. Quando a investigação em causa envolve o estudo histórico de uma língua, falar em dados empíricos é o mesmo que falar em textos de épocas passadas, textos que se procura conhecer de uma forma que não pode envolver a mais pequena cedência a processos de mistificação. Os artefactos (*testemunhos*) que conservam as formas mais genuínas desses textos, não necessariamente os mais antigos, são geralmente manuscritos com o formato de documento ou de códice. Guardam-se em espaços físicos, como os arquivos ou as áreas reservadas das bibliotecas, e só são acessíveis a leitores

Agradecimentos: Este trabalho foi produzido no âmbito de uma investigação financiada pelo European Research Council (ERC Advanced Grant 2011 Grant Agreement 295562).

Queremos agradecer a Maarten van Gompel, Louis Onrust e Stephen Parkinson os úteis comentários que nos ofereceram ao longo da composição do capítulo.

acreditados. É preciso, portanto, publicá-los, o que corresponde a preparar uma edição rigorosa, comparada e completa, cujo conteúdo serve na composição de glossários, dicionários, gramáticas, antologias e estudos de linguística diacrônica.

A forma de publicação em causa é um tipo de edição tão enriquecida que contém, baseada no que vem escrito nas páginas iniciais das introduções, nos rodapés das notas, ou ainda nas páginas em espelho, uma proposta de leitura empiricamente defensável. Corresponde, em algumas tendências de edição inspiradas no método do alemão Karl Lachmann, àquilo que se julga simbolizar melhor, por aproximação, o conteúdo textual da fonte original (Timpanaro 2005, ¹1963). Em outras tendências, mais influenciadas pelo método do francês Joseph Bédier, corresponde ao conteúdo do testemunho menos modificado por uma história de transmissão textual (Bédier 1970). São edições que recebem o rótulo de *textos, que são*. Ao tipo de peritagem que a elas conduz chama-se *crítica textual*, ou *filologia*, termo que preferiremos aqui.

Desde o século XIX, tem-se experimentado e teorizado sobre a forma mais segura de se fazerem boas edições críticas, se as referidas Lachmanniana ou Bédieriana (Dionísio 2006). A discussão girou sempre em torno da oposição entre atitudes objetivas e subjetivas perante testemunhos históricos que foram herdados, lidos e transmitidos com a intervenção de múltiplos acasos, difíceis de reconstituir e causadores de incessante variação. Discutiu-se, inclusivamente, se a objetividade em filologia seria alguma vez possível. Uma coisa, no entanto, não foi preciso nunca discutir, até finais da década de 1980: o formato e o suporte de publicação para uma edição crítica. Seriam sempre, indiscutivelmente, o formato do livro, o suporte do papel. Esse tempo chegou ao fim, até porque se percebeu logo, ainda antes da explosão da internet, que a informática ia ser instrumental num novo tipo de edição crítica, que passava agora pelo «elogio da variante», título de um trabalho de Bernard Cerquiglini em que se reconhecia o determinismo da tecnologia impressa nos extremos filológicos assumidas no passado (Cerquiglini 1989).

Neste capítulo, apresentaremos uma síntese das transformações resultantes da entrada da filologia portuguesa num mundo em que o ambiente mais natural para a circulação do conhecimento é precisamente o ambiente digital. Foi um processo que, como é inevitável em investigação científica, esteve sempre integrado em experiências internacionais, pelo que em certos aspetos seria artificial separá-lo de tal contexto. Por outro lado, foi um processo experimentado, com crescente solidariedade, por todas as áreas do conhecimento que interrogam textos e memórias coletivas, as quais são, principalmente, estas três: i) os estudos literários, porque veem nos textos a manifestação da capacidade humana de criação artística, ii) a história, porque se documenta nos textos, que são testemunho das sociedades e das culturas no tempo, e iii) a linguística histórica, porque toma os textos escritos como realizações antigas das línguas naturais, impossíveis de documentar por outra forma. Na sua faceta de *peritagem dos textos*, a filologia funciona hoje como disciplina auxiliar de qualquer destes estudos. Com efeito, cada filólogo é normalmente, também, ora um historiador da língua, ora um estudioso da literatura e da cultura. O próprio termo *filologia*, antes

da autonomização disciplinar oitocentista que levou à distinção entre linguística, história da literatura e crítica textual, serviu para cobrir todas elas.¹

Contudo, a partilha do método filológico, no passado, serviu sobretudo de pretexto para se definirem as diferenças entre as modalidades de edição que as três áreas exigiam. Ora a adoção de métodos digitais e a criação de recursos levaram a que as «paredes» entre aquelas áreas, numa dinâmica exemplar de interdisciplinaridade, começassem a cair (McCarty 2005, 118). Buscou-se, inclusivamente, um termo suficientemente genérico que a todas englobasse, na sua faceta de adesão às práticas computacionais e à exploração da publicação em linha. Foi busca que parece ter terminado quando, em 2001, se cunhou a expressão *humanidades digitais* (Kirschenbaum 2010, 2–3).

Em termos de problemas, o mais desafiante que as humanidades digitais têm encontrado é o de perceberem que certas metodologias e conceitos, que tomavam como adquiridos, se revelam, afinal, incomodamente dependentes de simplificações criadas pelas tradicionais culturas do manuscrito e do impresso. É portanto necessário embarcar numa aventura crítica que se espera venha ajudar a entender melhor o lugar do texto, e da comunicação escrita em geral, dentro das lógicas da sociedade tradicional, da sociedade moderna e da nova sociedade da informação.

2 Os começos da filologia digital

2.1 Nota sobre filologia portuguesa tradicional

A tarefa de dedicar cuidados filológicos a textos portugueses foi inicialmente cumprida pelos fundadores Gonçalves Viana, Epifânio da Silva Dias, Adolfo Coelho e Jules Cornu e prosseguida por Leite de Vasconcelos, Carolina Michaëlis de Vasconcelos, Júlio Moreira, José Maria Rodrigues, José Joaquim Nunes, Manuel Said Ali, David Lopes, Álvaro da Silveira, Cláudio Basto, João da Silva Correia e Manuel Rodrigues Lapa, todos nascidos ainda no século XIX. Consolidaram-na Joseph Piel, Paiva Boléo, Serafim da Silva Neto, Paul Teyssier, Celso Cunha e Lindley Cintra, para registar apenas os que nasceram no primeiro quartel do século XX (para a maioria das suas biografias e publicações, cf. Prista/Albino 1996).

O propósito da ladainha de nomes, praticamente exaustiva, é o de demonstrar que, ao longo de mais de cem anos, esteve toda uma pequena elite «livresca» encarregada da preparação de materiais eminentemente didáticos (antologias, gramáticas, dicionários, glossários, histórias da língua portuguesa), baseados em edições críticas que os

¹ Foi August Schleicher o primeiro a sugerir, no século XIX, que a *linguística* era uma ciência natural e a *filologia* (ainda no sentido de estudo dos textos herdados, sobretudo os literários) um tipo de história (Timpanaro 2005, 1963).

mesmos, ou pares muito próximos, elaboravam. A sua matéria empírica era constituída por manuscritos e impressos antigos, que tinham cuidadosamente escrutinado dentro de exemplares de acesso muito restrito. O sistema pressupunha uma escala hierárquica: no topo, as instituições (e alguns colecionadores privados), guardiões físicos dos artefactos que continham a memória textual da língua nacional na forma mais genuína, e na quantidade mais numerosa, que se conseguiu conservar. As credenciais para aceder a esse nível de topo estavam reservadas a académicos com credibilidade suficiente para lhes ser confiado o manuseio das relíquias. O seu papel era o de converter a credibilidade que as instituições lhes reconheciam, bem como o privilégio de acesso a objetos valiosos, em publicações com estatuto de autoridade. A autoridade não era, contudo, automática: discutia-se em correspondência com os pares, transbordava para artigos de revistas científicas, e, uma vez aprovada, acabava por atingir o escalão de base, o do ensino, sobretudo o secundário e o universitário. De permeio, atuavam as casas editoras (as imprensas nacionais, portuguesa e brasileira, as imprensas de Universidade, de biblioteca ou de fundação), nada alheias ao sistema porque se tratava normalmente de publicações subsidiadas. Eram, precisamente, edições dispendiosas, sem grande correspondência entre o investimento que envolviam e o público que as consumia. Mas eram tidas como necessárias por se entender, desde que a ideologia nacionalista de Oitocentos tinha triunfado, que cabia ao Estado velar pela repetida celebração da herança cultural da Nação, cristalizada nos seus diferentes monumentos, incluindo os textos históricos e os textos literários. Este panorama ainda se mantinha quando a informática e a cultura se começaram a cruzar, mas tornou-se claramente desajustado num mundo de onde já parece ter desaparecido o problema da conservação da memória textual, para dar lugar ao da sua sobre-representação.

2.2 O começo da filologia portuguesa digital

O início da filologia digital remonta à apropriação da tecnologia informática no sentido de se conseguir o processamento de textos de grandes proporções, considerados representativos de uma cultura através do léxico da sua língua. Chamou-se aos produtos desse trabalho, sintomaticamente, *Tesouros*. Veja-se como o caso se passou em Portugal, entre as décadas de 1960 e 1980.

As primeiras experiências resultaram da colaboração entre a linguista Maria Helena Mira Mateus e o Centro de Cálculo Científico da Fundação Calouste Gulbenkian (Mateus 1968, 227). A preocupação era lexicográfica e filológica. Pretendia-se transpor para o caso português o que se praticava em França com a construção do *Trésor de la Langue Française*,² i.e., «elaborar um dicionário histórico que desse, em

² Em 1960, três anos depois de o projeto ter sido pensado, o CNRS decidiu criar em Nancy, e com direito a edifício novo, o *Centre pour un Trésor de la Langue Française*. O *Trésor* foi publicado primeiro

períodos sucessivos, um quadro tão completo quanto possível do vocabulário da língua desde a origem aos nossos dias» (Mateus 1974, 3).

O *Tesouro* português começou a ser construído com o *Glossário da Vida e Feitos de Júlio César, tradução portuguesa quatrocentista de Li Fet des Romains*, e o resultado foi a publicação da obra em papel, primeiro em sucessivos números do *Boletim de Filologia*, entre 1974 e 1992, cobrindo o léxico das letras A a S, depois em edição integral, numa publicação encadernada (Mateus 2010).

Na verdade, a intervenção de matemáticos, informáticos, programadores (e máquinas) do Centro de Cálculo Científico consistiu, estritamente, num auxílio externo, destinado a tornar mais eficiente um processo que podia ter também realização manual (Mateus 1974, 5): «É evidente que a elaboração de um Tesouro da Língua se não restringe à etapa (informática) indicada. Mas é evidente, também, que esta etapa, vital em relação à essência da própria obra, se não poderá realizar com eficiência sem esse importante contributo». À época, porque ainda não havia microcomputadores, o termo «computador» era usado em itálico, havia necessidade de o descrever recorrendo à perífrase «complexo mecanográfico que inclui o computador como unidade principal», e não se hesitava em encará-lo enquanto «instrumento», se bem que a noção da sua progressiva indispensabilidade para a investigação já fosse óbvia: «[tem-se] tornado indispensável na coadjuvação do esforço do homem para o progresso científico». Podia-lhe ser entregue, no caso do estudo do português, «a tarefa de evidenciar, alfabetizar e organizar todos os vocábulos, não de um único texto, mas de milhares de obras consideradas representativas [da] língua» (Mateus 1974, 5).

O *Tesouro* não se fez na altura: só o *Glossário da Vida e Feitos de Júlio César* ficou pronto. Em termos de descendência, o recurso em linha que atualmente mais se aproxima dos objetivos do *Tesouro* pertence à área da lexicografia. Trata-se do *Corpus Lexicográfico do Português*, da Universidade de Aveiro e do Centro de Linguística da Universidade de Lisboa, da responsabilidade de Telmo Verdelho e João Paulo Silvestre. Foi lançado em 2003 e disponibiliza, em modelo crescentemente pesquisável dentro do sistema DICIweb, a dicionarística portuguesa dos séculos XVI a XIX (<http://clp.dlc.ua.pt/DICIweb/>).

O trabalho que se seguiu à experiência de Maria Helena Mateus decorreu no início da década de 1980, empreendido pelo lusitanista britânico Stephen Parkinson, que lançou o AOPT, *Archive of Old Portuguese Texts* (Parkinson 1983). Parkinson contava com a colaboração do *Computing Centre* da Universidade de Aberdeen e com a máquina Honeywell H66, que já usava a fita magnética em substituição dos cartões perfurados. Já estava também disponível um programa de concordâncias, o *Oxford*

em papel, em 16 volumes saídos entre 1971 e 1994. Em 2002 saiu o *Trésor de la Langue Française Informatisé*, nas versões CD-ROM e em linha (<http://atilf.atilf.fr/>) (Mateus 1974; Del Mancino/Pierrel 2009).

Concordance Program (Hockey/Marriott 1980) e um editor para conversão automática de texto.

Aparecia claramente a Parkinson a grande vantagem que já acima referimos, ao mencionar Cerquiglini, e que se tornou depois banal para todos os filólogos que praticam codificação de texto para publicação eletrónica: deixou de haver razões para a aplicação de normas de edição irreversíveis, sempre de difícil coerência interna. Eram normas que se tinham usado no passado porque era preciso assegurar que as edições mantivessem alguma legibilidade do ponto de vista do leitor comum. No caso português, um exemplo que se tornou clássico foi decidido por um grupo de trabalho formado em 1970 no Centro de Linguística da Universidade de Lisboa e dirigido por Luís Filipe Lindley Cintra. Inclui regras de conversão deste tipo, relativo à transcrição das vogais nos manuscritos medievais (Castro et al. 1973, 418):

«[vogais] não etimológicas: são sempre transcritas como vogais simples, mesmo que, para alguns textos, se possa suspeitar que, na intenção do escriba, a geminação representava abertura ou tonicidade da vogal. (Em casos destes, o facto será devidamente registado na Introdução.)

Ex: *taes* por *taaes*, *ó* ou *oh* por *oo*, *doe* por *dooe*, *ceo* por *ceeo*, *som* por *soom* (<SUM)».

A tecnologia veio permitir que se passasse a proceder de outro modo, como foi experimentado e relatado por Parkinson. Perante um manuscrito antigo, e linguisticamente relevante, passou a ser possível progredir por etapas. Primeiro, procedia-se à recolha de dados crus, para usar um anacronismo, já que a expressão da altura era «transcrição diplomática pré-editada». Para o problema das abreviaturas, por exemplo, havia a solução dos símbolos compostos, «sequências de símbolos que o computador interpretará como símbolo único». A sua resolução era indicada à parte, numa lista de regras de expansão que podia mudar consoante o resultado editado que o filólogo pretendesse.

Na notícia que Parkinson deu deste trabalho, em artigo do *Boletim de Filologia*, o autor reconhecia a inspiração retirada de trabalhos pioneiros, ensaiados para o alemão medieval (Murdoch 1971), e expunha a virtude da fidelidade e da reversibilidade nas edições assistidas por computador (Parkinson 1983, 242–243):

«[É] possível dar instruções para a conversão de quaisquer símbolos ou sequências de símbolos em quaisquer outros símbolos ou sequências, inclusive os «control characters». Neste poder transformacional reside uma solução para o problema de transcrição de textos medievais. O editor de textos medievais teve sempre que escolher uma forma de transcrição, bem cónscio de que esta decisão afastava um ou outro sector do público do documento. Numa edição modernizadora, aliás, os efeitos da decisão eram permanentes e irreversíveis porque não seria possível deduzir as formas manuscritas sem o apoio de aparatos críticos. O poder transformacional do computador abre caminho para uma prática editorial reversível. (...) O computador fará uma tal tradução de transcrição diplomática para edição com abreviaturas desenvolvidas com muito mais fidelidade que um amanuense».

Os exemplos aqui evocados, dos trabalhos desenvolvidos por Mateus e por Parkinson antes da generalização dos computadores pessoais, demonstram claramente as duas

grandes oportunidades que a filologia entreviu no recurso ao auxílio informático. Eram as oportunidades de prevenir a hipótese de erro humano em tarefas de escrita (ao longo de transcrições e edições) e de prevenir a hipótese de abandono de trabalhos demasiado pesados para a capacidade humana, ainda que teoricamente necessários do ponto de vista da investigação filológica. Um *robot*, na sua aceção etimológica de «escravo»,³ poderia fazê-los.

O destino da experiência de Stephen Parkinson já foi, ao contrário do que acontecera com o trabalho de Maria Helena Mateus, o de um armazenamento digital. Os documentos tabeliônicos medievais que editou passaram a integrar o primeiro corpus histórico de textos portugueses, o CIPM ou *Corpus Informatizado do Português Medieval*, constituído a partir de 1994 no Centro de Linguística da Universidade Nova de Lisboa, precisamente em torno do resultado do trabalho de Parkinson e com a sua colaboração (Xavier/Brocardo/Vicente 1995). Hoje são pesquisáveis em linha quer no site do próprio CIPM (<http://cipm.fcsh.unl.pt>), quer no *Corpus do Português*, já com anotação morfossintática, (<http://www.corpusdoportugues.org/>).

Entretanto, na transição entre as décadas de 1980 e 1990, as condições mudaram. Surgiram dois desenvolvimentos que levariam ao abandono deste tipo de experiências individuais de codificação dirigidas para uma só língua. Aconteceu que a codificação de texto e a codificação de caracteres se tornou numa experiência de crescente dimensão coletiva, da qual muitas diferentes filologias puderam começar a beneficiar: criaram-se o Unicode, de iniciativa empresarial, e o TEI, de iniciativa académica. Antes de abordarmos tais codificações, convém contudo analisar as consequências desta primeira fase da informatização nas humanidades.

2.3 A tecnofobia e a crítica digital

Referimos na secção anterior que a conversão de materiais filológicos num formato mecanicamente legível, bem como a correspondente constituição de bases de dados, resolveu dois problemas clássicos inerentes à atividade da edição crítica de textos: o dos erros humanos e o das limitações humanas. Teoricamente, uma máquina bem programada não se engana, uma máquina bem mantida não se cansa.

Acontece que estas mesmas duas virtudes da mecanização (o automatismo das operações e a viabilização dos empreendimentos gigantescos) é também o nó de um problema cultural e social, inevitavelmente criado quando há abandono de atividades manuais mediante a adoção da tecnologia. É o problema da «proletarização», muito utilizado pelo ceticismo antitecnológico, o que não quer dizer que não seja um problema real.

³ O termo *robotics* pertence ao campo semântico do trabalho porque começou no empréstimo ao próprio inglês do termo checo *robotnik* ‘escravo’.

A posição mais cética em relação ao trabalho em humanidades digitais aponta-lhe o seguinte: é trabalho que parece levar apenas a poupar horas e recursos, mas não leva a nada de novo ao nível da construção do conhecimento. O próprio Roberto Busa, reconhecido por muitos como o pai das humanidades digitais, discutiu tal perversão (Busa 2004).

Sociologicamente analisado, o perigo parece ser o de se estar a proletarizar uma atividade cultural (Stiegler 1998, 1994; Robertson 2015). Proletarização tem aqui vários dos sentidos que Marx atribuiu ao termo na sua crítica à ligação entre capitalismo e industrialização nas sociedades ocidentais. O paralelismo é o seguinte: tal como o produtor tinha sido afastado da terra, seu antigo meio de produção, e obrigado a entrar como operário não especializado na engrenagem industrial, também o académico se afastou do conhecimento com a chegada da digitalização às humanidades, sendo arrastado para um trabalho mecânico que ele, académico, só pode desenvolver a níveis pouco especializados já que não tem uma preparação de engenheiro. Ao mesmo tempo, o engenheiro informático que se ocupa da infraestrutura da edição académica (ou do corpus histórico, ou do museu virtual) vai lidar com temas de cultura ao nível da modelização dos programas, só que não tem preparação humanística para tanto. O resultado combinado será, argumenta-se em setores mais conservadores da esfera pública, uma perda para a língua, a história, a literatura, a cultura, enfim, para o conhecimento.

Esta é também a linha da tradicional crítica à técnica, que tem origem num tipo de ressentimento com a mesma antiguidade da própria filosofia. Nas palavras de Bernard Stiegler, «desde a sua origem e até agora, a filosofia tem reprimido a técnica como objeto de pensamento. A técnica é o não pensamento» (Stiegler 1998, 1994, ix).⁴ Contudo, como aponta o mesmo autor, precisamos muito pragmaticamente de uma rápida reação a tal repulsa, por maior tradição que ela tenha: «A mudança de perspetiva e de atitude torna-se necessária, obrigando a uma capacidade de reação tão urgente quanto inevitável» (Stiegler 1998, 1994, x).

A reação mais urgente, que é precisamente a que estão a assumir as humanidades digitais, é a de compreender a máquina, investindo na busca de uma solução para a falta de harmonia entre cultura e técnica. Tal falta de conciliação nasce sempre que se olha para as máquinas enquanto meros substitutos do indivíduo-artesão, «portador de ferramentas». Supera-se vendo nelas «indivíduos técnicos», que é preciso conhecer mediante a construção de uma «mecnologia» que detete até que ponto a máquina vai ganhando «capacidade de regulação» (Stiegler 1998, 1994, 69, retomando ideias de Gilbert Simondon).

Na primeira fase de informatização da filologia, o computador era simplesmente instrumentalizado. Pedia-se-lhe para, repetindo o que escreveram Maria Helena Mateus e Stephen Parkinson, «evidenciar, alfabetizar e organizar vocábulos», ou então

4 Ao longo do capítulo, as sequências citadas foram traduzidas para português por Rita Marquilhas.

para fazer a conversão de símbolos «com muito mais fidelidade que um amanuense». Na segunda fase, estão a ser os próprios académicos a harmonizar-se com o digital, a envolver-se nele de forma a compreenderem quantos conceitos é preciso revolucionar para que a filologia possa continuar a cumprir a responsabilidade de disciplina que se ocupa da peritagem dos textos e do seu diálogo com a cultura e a língua.

Ao mesmo tempo, este envolvimento dos académicos com a tecnologia dentro do movimento das humanidades digitais é um processo que equivale a resistência. Resistência a duas forças: a da vetusta tradição académica e a da ameaçadora agressividade empresarial. Por um lado, as humanidades digitais caracterizam-se por uma cultura que, reconhecidamente, envolve «colaboração, abertura, relações não hierárquicas e agilidade» (Kirschenbaum 2010, 5), por contágio das práticas entretanto naturalizadas por programadores informáticos e por frequentadores da blogosfera. Por outro lado, um maior protagonismo daquele movimento significará que o controlo da edição digital não fica tacitamente entregue às empresas, enfrentando a crítica textual uma nova missão, de se converter também em *crítica digital*. Se os filólogos tivessem olhado mais cedo para a constituição de bibliotecas e livrarias em linha, argumenta Jerome McGann, a *Google Books* prestaria hoje um serviço muito diferente, documentalmente mais responsável (McGann 2014, cap. 7).

3 O conceito de texto na esfera digital

Como foi referido anteriormente, uma das grandes vantagens que a revolução digital trouxe a qualquer área do conhecimento que lide de perto com textos, cultura e memória foi a de ter oferecido instrumentos valiosos no auxílio à reflexão sobre «o que é um texto?». Na formulação de Maria Clara Paixão de Sousa, que é, na esfera da filologia portuguesa, e com João Dionísio (Dionísio 2006), quem mais se tem dedicado a refletir sobre o tema, «a difusão digital exige, [pelo] menos, transformações profundas nas nossas perspetivas conceituais sobre o texto» (Sousa 2013a, 20; cf. também Sousa 2013b).

O termo *texto* está carregado de conotações antigas, associadas às tecnologias da escrita e à noção de autoridade intelectual, bem como de conotações modernas, ligadas à história recente da defesa do direito de autor. Sintomático da sua polissemia é o facto de as mesmas palavras, por exemplo, *obra*, *discurso*, *mensagem*, *enunciado*, serem explicadas muitas vezes como equivalentes de *texto*, e igual número de vezes como distinguindo-se dele.

O uso comumente dicionarizado é o de «palavras fixadas pela escrita», que acaba por ser também recolhido por alguma bibliografia didática de filologia (Roncaglia 1975, 23; Blecua 1983, 17). Mas tanto na mesma área da filologia, como em teoria literária e em linguística do texto, o termo é visivelmente incómodo, levando certos autores a conduzir ensaios inteiros só para o definirem e outros a optarem por o retirar da terminologia mais precisa.

Roland Barthes, num conhecido e influente ensaio intitulado *De l'œuvre au texte*, alongou-se na busca das melhores metáforas para explicar a complexa distinção entre texto literário – um «campo metodológico», um «espaço social», um «plural irreduzível» – e obra literária – «um fragmento em substância [que] ocupa uma porção do espaço dos livros (numa biblioteca, por exemplo)» (Barthes 1994, 1971, 1212).

Paul Ricœur, aquele cujo esforço de definição é porventura dos mais citados, entregou-se ao paradoxo: explicou texto como aquilo que é um discurso não oral – «discurso fixado pela escrita» –, precisando, no entanto, que se trataria de um discurso que vivia «no ar», «fora do mundo» ou «sem mundo» (Ricœur 1986, 154 e 158).

Esta contradição entre a materialidade inevitável de tudo quanto seja «fixado» e a imaterialidade daquilo que vive «fora do mundo» diz bem da dificuldade em explicar um termo que remonta a Quintiliano, autor que escolheu dar um sentido figurado à designação latina de «tecido» (*textus*), referindo-se às palavras que os humanos compõem enlaçando-as num tecido («verba qua compositione in textu iungatur») (*apud* Roncaglia 1975, 23).

Em linguística do texto, uma disciplina que, tal como a crítica textual, recusa *a priori* a equivalência entre texto e texto literário, a questão da materialidade é posta em termos de *canal* ou *modo*. Tal facto decorre de se conceber, aqui, que os textos tanto podem pertencer à oralidade como à escrita. São tomados como instância de comunicação: situacionalmente, podem adotar um modo que pode ser transiente (tipicamente, o falado), ou permanente (tipicamente, o escrito). Numa das abordagens mais difundidas, fala-se menos de texto e mais de *textualidade*, uma propriedade das expressões linguísticas que se agregam de forma a apresentarem coesão, coerência, intencionalidade, aceitabilidade, informatividade, situacionalidade e intertextualidade (Beaugrande/Dressler 2005, 1972, 46). Outra abordagem formula a existência de um campo abstrato que os textos ocuparão, campo esse que é cruzado por uma oposição de meio, ou modo, (oralidade vs. escrita) e por uma escala espacial (da extrema proximidade à extrema distância comunicativa). Pretende-se assim dar conta do fenómeno da variação textual, com a conversa face a face num dos polos, e a obra literária canónica, no outro. Em qualquer ponto do campo pode surgir um género textual diferente, com outras tantas idiosincrasias (as *tradições discursivas*) que se manifestam em todos os níveis de descrição da língua (Koch/Oesterreicher 2007, 1999).

O papel da materialidade na constituição dos textos é, por conseguinte, um fator que reúne pouca unanimidade, suscitando diferentes reações dentro da reflexão teórica sobre o texto. Mas quando a informação passou a circular digitalmente, uma coisa se tornou bem clara. Percebeu-se que nem se pode dizer que os textos são desprovidos de matéria específica, nem se pode considerar que eles são constituídos por matéria específica. A componente da materialidade está instalada nos textos, sim, mas ocupa um nível abstrato, «subespecificado», no sentido de latente e imprevisível,

para usar termos da Fonologia teórica (cf. Steriade 1995).⁵ Tornou-se mais claro como a atividade textual envolve a nível abstrato, para quem fala e escreve, uma conceção da materialidade dos seus textos que só deixa de ser polivalente quando há convergência de mais planos, i. e., quando o plano físico e o interpretativo se vêm juntar ao conceptual. Referindo-se ao que se aprende sobre texto na análise dos textos literários já nascidos-digitais, N. Katherine Hayles resume (Hayles 2004, 72):

«Nesta perspetiva de materialidade, ela não é uma coleção inerte de propriedades físicas; é antes uma qualidade dinâmica que emerge da articulação entre o texto enquanto artefacto físico, o seu conteúdo conceptual e as atividades de interpretação de leitores e escritores [...]. A materialidade não pode ser especificada à partida».

O carácter subespecificado, emergente e dinâmico da materialidade dos textos estava intuído nos trabalhos de Donald F. McKenzie, publicados nos anos 1970 e 1980, quando pugnou por uma filologia do livro impresso enquanto *bibliografia histórica*, ou *sociologia do texto*. Referindo-se à edição setecentista das peças reunidas do dramaturgo inglês William Congreve (*Works*, 1710), McKenzie proclamava (McKenzie 2002, ¹1977, 200):

«O livro em si [Works 1710] é um meio de expressão. Aos olhos, as suas páginas oferecem uma agregação de significados, tanto verbais como tipográficos, a serem traduzidos para o ouvido [na encenação da peça]; mas temos de aprender que a forma que ele toma na nossa mão também nos fala do passado. A explicação plena desses significados, em toda a sua riqueza contextual, é a principal função textual da bibliografia histórica».

Concluía McKenzie que «qualquer ênfase na estrutura integrada de um texto é portanto bem vinda»; para o conjunto integrado da edição textual participavam, além de todos os «detalhes» e «defeitos», as «versões diferentes», as «formas não verbais», os «comportamentos de leitura» e as «decisões históricas feitas por autores, desenhadores e artesãos» (McKenzie 2002, ¹1977, 223).

As conceções de McKenzie, que pertenciam sobretudo à defesa de uma modalidade mais culturalmente empenhada da prática da filologia, revelam-se penetrantes porque no caso dos textos que o filólogo analisava, saídos da baixa tecnologia da imprensa manual, o carácter subespecificado da materialidade não estava tão obviamente presente. Os artefactos manuscritos e impressos são, com efeito, muito sólidos na sua aparência física, e facilmente se cria a ilusão de que suscitam representações idênticas em qualquer pessoa, momento ou lugar da história. No mundo da comunicação digital, a subespecificação da materialidade, pelo contrário, só não se torna óbvia para os infoexcluídos profundos. Qualquer variação de software ou de hardware, qualquer caminho diferente que se tenha seguido na navegação, qualquer históri-

5 A autora explica também os diferentes sentidos que o termo *subespecificação* (de traços fonológicos) pode ter.

co que tenha ficado na memória dos seus computadores pessoais vai afetar a forma como um texto se materializa para leitores e escritores, que de qualquer forma também variam entre si na agilidade com que se adaptam a novidades e *upgrades*.

A lição para os filólogos da edição digital tem sido dupla: não podem sonhar com uma edição crítica estável, até porque ela é impossível; mas podem lidar de frente com tal impossibilidade. Primeiro, harmonizando convenções entre si, para que o grau de subespecificação material das suas edições (que são, elas próprias, textos) não seja tão elevado quanto a tecnologia o permite. Depois, convertendo as edições em abordagens integradas, onde se codifique o que no texto é físico, o que é conceptual e o que é social. Um seguidor da *sociologia dos textos* de McKenzie, Jerome McGann, foi precisamente o pioneiro deste tipo de experiência, quando lançou a edição académica digital das obras plásticas e literárias de Dante Gabriel Rossetti (<http://www.rossettiarchive.org>).

Nas secções que se seguem procuraremos demonstrar a viabilidade técnica destas edições. O argumento é o de, como expôs Willard McCarty, ser preciso construir modelos de edição que correspondam ao imperativo da legibilidade computacional, pelo que têm de ser completamente explícitos e consistentes. Por outro lado, tem de se aceitar que serão edições sempre manipuláveis, por ser isso mesmo que acontece nas representações computacionais (McCarty 2004).

4 Questões de codificação e de marcação

4.1 Codificação e linguagens de programação

Em termos de linguagem, os computadores apenas conseguem lidar com código binário: trata-se de unidades básicas de computação de informação, ou bits (*binary digits*), que são ligadas e desligadas. Representam-se numa série de 0s e 1s, que correspondem frequentemente em termos físicos aos sinais elétricos de alta e de baixa voltagem que dão entrada no computador. Todos os dados que aí se contenham, quer se trate de programas quer de conteúdo, são representados desta maneira, em sequências de cadeias binárias.

Em termos de estrutura interna, o núcleo do computador é constituído pela sua unidade de processamento central, ou CPU (acrónimo de «Central Processing Unit»), aquela parte que executa o código-máquina. Código-máquina, por sua vez, é a linguagem de programação mais primitiva de todas, uma linguagem em que se codificam as instruções para o computador; não é legível por humanos, portanto a maior parte dos programas de computador (o chamado *software*) está escrita em linguagens de programação de nível superior. Exigem uma ulterior compilação, ou interpretação, em código-máquina para poderem ser executados. Pode dizer-se, portanto, que uma linguagem de programação é nada mais nada menos do que uma linguagem formal na qual se podem escrever instruções para um computador executar.

São linguagens que existem aos milhares, variando entre as que se destinam a tarefas específicas, as que servem todos os propósitos e as que são independentes do componente físico, ou «ferro», o *hardware*. Compreensivelmente, interessam aos programadores de humanidades digitais só aquelas linguagens que permitem manipular conteúdos textuais, as quais variam em termos de sistemas operativos. Em computadores do tipo Unix, i.e., computadores que têm como sistema operativo o Linux, o BSD, ou mesmo o Mac OS X, há uma parte das suas ferramentas básicas que é muito apropriada à manipulação de texto. Com uns poucos e simples comandos Unix, podem-se extrair, por exemplo, listas de frequências de todas as diferentes ocorrências (tokens) de uma palavra, ou de todos os diferentes tipos de palavras numa coleção de textos.

Já para computadores com outros sistemas operativos, ou para tarefas mais complexas, usa-se muito uma linguagem de programação para manipulação de texto chamada Perl (Wall/Christiansen/Orwant 2004), que foi desenvolvida em finais da década de 1980. Em Perl podem-se empregar expressões regulares muito sofisticadas, sendo que uma expressão regular é uma sequência padronizada de caracteres que permite encontrar e substituir (*find and replace*) cadeias de sinais e de palavras. A máquina de expressões regulares do Perl está hoje reimplementada em linguagens de programação mais recentes, como é o caso das linguagens Python e Java. Na Tabela 2 da secção 4.2, abaixo, mostraremos um exemplo simples de um programa escrito em Perl.

Como já referimos, há uma discrepância grande entre as cadeias binárias que um computador consegue entender e os textos com a escrita própria de uma língua natural. Para que estes últimos possam ser trabalhados por uma máquina, os caracteres legíveis por humanos têm de ser transformados em cadeias binárias e vice-versa. Tal passo de conversão recebe o nome de codificação de caracteres e implica a tradução de bits para caracteres. Um dos mais antigos sistemas de codificação de caracteres foi o que se usou na década de 1960 com os primeiros computadores: chama-se ASCII, *American Standard Code for Information Interchange*. O código ASCII tem espaço para 27 bits, o que lhe permite codificar até 128 caracteres diferentes, incluindo um conjunto de caracteres de controlo, tais como os que codificam «tabulação» e «nova linha». É o suficiente para codificar o alfabeto que se emprega na escrita do inglês, mas é recurso demasiado escasso quando se pretende abranger todos os sistemas de escrita do mundo. O facto gerou problemas à medida que, na era da computação, se foi assistindo à criação de múltiplos sistemas de codificação alternativos, por parte de produtores em países de língua não inglesa. Viu-se um exemplo de tal iniciativa quando, acima, se referiu a criação, por parte do lusitanista Stephen Parkinson, de um sistema de combinações de símbolos destinadas à codificação de abreviaturas em textos medievais portugueses. Com experiências desse tipo, a troca de programas e de ficheiros entre diferentes computadores foi-se tornando crescentemente difícil, o que conduziu, na década de 1980, à definição de um novo padrão unificador, o Unicode. É um sistema descritivo para codificação de caracteres que

cobre, atualmente, mais de 100.000 caracteres diferentes, desde o dos hieróglifos do Antigo Egito até aos mais variados ícones, incluindo os *emojicons* nossos contemporâneos. Note-se que o Unicode é um padrão descritivo, pelo que o valor final de cada carácter, i.e., o seu corpo, fonte e estilo, ainda é atribuído pelo programa que o interpreta, seja ele o sistema de programas do navegador da internet (*web browser*) ou o de um editor de texto.

A implementação de Unicode que hoje mais se usa é a codificação em UTF-8, que é compatível com ASCII. Para o inglês, codifica os caracteres em códigos ASCII de 7 bits, mas recorre-se a mais bits para o caso de outros caracteres. Na perspetiva das humanidades digitais, o padrão Unicode tem uma série de inconvenientes, como foi já salientado, por exemplo, por Fiorimonte (2012). Há que ser-se cuidadoso e crítico em virtude de se tratar de um padrão industrial orientado para as necessidades do Ocidente, sobretudo do Ocidente moderno, o que o torna potencialmente desajustado para muitos sistemas de escrita. São limitações de que estão bem cientes muitos investigadores da área das humanidades digitais, que prosseguem, aliás, com a criação original de tipos para computador quando as suas edições o exigem. É o caso, em Portugal, do *Notator Mono*, um tipo medieval para computador criado para permitir edições diplomáticas com «uma representação tipográfica rica, complexa e fidedigna dos sistemas de escrita usados em Portugal entre os séculos IX e XIV na produção de documentação notarial» (Emiliano 2005, 139). A título de exemplificação, reproduz-se abaixo na Tabela 1 o caso da representação em *Notator Mono* das letras traçadas por traços verticais e oblíquos previstas para todas as combinações possíveis entre letras minúsculas e sinais de *-UM*, uma terminação muito frequente em palavras latinas.

Tabela 1: Representação no tipo *Notator Mono* das abreviaturas de terminações latinas em *-mum*, *-num*, *-rum* e *-tum* (reproduzido de Emiliano 2005, 152).

CAR#	DESIGNAÇÃO	VALOR
#156	M minúsculo traçado	-m (um)
#159	N minúsculo traçado	-n (um)
#168	R minúsculo redondo traçado	-r (um)
#171	R minúsculo traçado	-r (um) [carácter da letra visigótica]
#178	T minúsculo visigótico traçado	-t (um) [carácter da letra visigótica]
#179	T minúsculo traçado	-t (um)

A codificação de caracteres mantém-se um obstáculo problemático em muitos projetos de humanidades digitais. Surge em dois contextos. Por um lado, quando se quer retomar documentos digitais já antigos, codificados em formato pré-Unicode, como é o caso de muitas edições preparadas em computador nas décadas de 1960, 1970, 1980 e mesmo ainda em 1990, cuja codificação não é corretamente reconhecida pelos

programas atuais, além de que não se converte facilmente para outros formatos. Por outro lado, quando se coligem documentos de proveniências diferentes, é sempre possível que eles tragam codificações variadas. É essencial, por conseguinte, criar uma versão estandardizada em que tal variação desapareça, mas nem sempre o responsável pela coleção consegue controlar o formato exato que os seus documentos digitais receberam na origem. No momento de recorrer a dados textuais fornecidos por voluntários, por exemplo, pode acontecer que alguns dados tenham sido obtidos por cópia a partir de uma qualquer aplicação, seguida de colagem num editor de texto, sendo que o documento pode ser enviado já no formato desse programa de edição, sem consciência, muitas vezes, da codificação que se pode ter perdido ou corrompido no momento da cópia.

A propriedade da subespecificação, dinamismo e fluidez material dos documentos e programas feitos em computador é portanto motivo para que na prática das humanidades digitais haja um alerta constante em relação a problemas de codificação e um investimento crescente em técnicas para a prevenção dos mesmos.

4.2 Marcação

Os documentos de texto digitais têm, para além do seu conteúdo codificado concreto, uma camada adicional de informação sobre o texto em si e respetivo contexto, i.e., sobre a fonte arquivística, o género, o autor, a data, as revisões, etc., bem como sobre o formato geral que a visualização deve ostentar quando publicada, em pormenores como os das fronteiras de parágrafos, das cores, do corpo da letra, etc. O procedimento mais usual que se adota para indicar esta informação adicional é o da marcação de texto (*textual markup*), o que implica usar uma linguagem de marcação. É, tal como as de programação, uma linguagem formal, mas tem propósitos específicos. Enquanto uma linguagem de programação se usa para formular instruções para computador, uma linguagem de marcação usa-se para anotar um documento com informação extra, dando continuidade à tradição da inclusão de símbolos descritivos e de notas de aparato na publicação em papel de edições diplomáticas, críticas e genéticas (Roncaglia 1975, 75–78; Blecua 1983, 147–152; Castro 2001). Com efeito, é na camada da marcação de uma edição crítica eletrónica que se inclui a indicação de lacunas, rasuras ou acrescentos no original do texto que se edita, conjeturas de editor ou existência de variantes dentro de uma tradição textual.

Marcação significa emprego de etiquetas para distinção efetiva, a níveis diferentes, entre a camada de texto e a camada de informação de natureza editorial. As etiquetas podem revestir diferentes formas em função do programa que se estiver a usar, podendo ser simplesmente **negrito**, *_sublinhado_*, */cursivo/* ou *<etiqueta>*. Uma linguagem de marcação fornece uma descrição, e depois é sempre necessário dispor de programas específicos e de ferramentas capazes de a interpretar e de lhe dar um destino útil.

Uma das linguagens de marcação mais conhecidas é o HTML (*HyperText Markup Language*), usada na descrição das páginas da internet, padrão que é mantido pela organização *World Wide Web Consortium* (W3C) (<http://www.w3.org/>). Nas etiquetas de marcação dessas páginas, incluem-se indicações para o navegador disponibilizar a visualização do conteúdo por meio de chamadas de atenção para o início e o fim da informação adicional. Veja-se um exemplo em (1) e (2), onde se demonstra o uso de `` e `` para assinalar as partes de um texto que devem ser visualizadas a negrito.

- (1) Dentro desta frase em particular, `esta secção deve ser a negrito` para efeitos de disponibilização na internet.
- (2) Dentro desta frase em particular, **esta secção deve ser a negrito** para efeitos de disponibilização na internet.

Na visualização oferecida em (2), a qualidade do negrito é percebida analogicamente, como sempre aconteceu na tradição impressa, i.e., os caracteres são eles próprios diferentes, neste caso mais carregados em termos de preto. Pelo contrário, em (1), a qualidade do negrito indica-se e visualiza-se digitalmente, constituindo uma informação com existência discreta em relação aos caracteres em que o texto vem escrito.

Uma outra linguagem de marcação, esta extremamente relevante para o universo das humanidades digitais, é o XML (*eXtensible Markup Language*), uma linguagem descritiva de aplicação generalizada, descendente de um padrão anterior, o do SGML (Goldfarb 1999). Ao contrário do HTML, o XML não fornece um conjunto fixo de etiquetas, limitando-se a definir um formato. São depois os utilizadores quem escolhe que nomes dar e que significado associar a cada etiqueta de marcação. Por exemplo, uma indicação de negrito em XML tanto pode ser ``, como `<n>`, como `<negrito>`, como `<meunegrito>`, desde que se estabeleça no lugar apropriado o que significa a etiqueta e como ela precisa de ser interpretada pelos programas que reconhecem tal formato. É por isso que cada documento XML tem de ser acompanhado por definições externas que especificam o significado e a estrutura das etiquetas ali particularmente usadas, definições essas que recebem o nome de *XML schema*, ou então o nome do modelo seu antecessor, DTD (*Document Type Definition*).

As etiquetas XML organizam-se dentro de uma estrutura em árvore. Por exemplo, querendo transpor o conteúdo de um livro para linguagem XML, as etiquetas de parágrafo ficam em ramos inferiores, filhos dos nós de secção, por sua vez filhos dos de capítulo; todos juntos, organizam-se num todo ramificado que representa o conteúdo do livro.

Dado que a linguagem XML é, como o nome indica, muito maleável ou «extensível», torna-se potencialmente infinita a variedade dos seus elementos, indicados com parênteses angulares `<xxx>`, e seus atributos, indicados entre aspas dentro da categoria dos elementos `<xxx y="zzz">`. Assim, não tardou a surgir, manifestada por académicos, bibliotecas e arquivos, a necessidade de se dispor de um formato estandardizado para o uso desta linguagem de marcação na descrição e edição eletrónicas de

conteúdos textuais. Foi por isto que se impôs na comunidade das humanidades digitais o padrão TEI (*Text Encoding Initiative*), desenvolvido desde finais da década de 1980. O objetivo foi o de criar e manter um padrão independente de marcação para a codificação de dados em humanidades digitais. A versão atual do TEI segue definições XML e criou um Manual muito detalhado que permite a adoção de um mesmo padrão em resultados digitais tão diferentes como são a edição crítica de uma obra literária, a transcrição de diálogo num arquivo de registo oral, a codificação detalhada de metainformação sobre a relação entre objetos digitais e seus originais físicos, ou a codificação sobre a proveniência e a anotação de corpora linguísticos. Em relação a estes, foram publicadas em 1996 as recomendações para se passar a dispor de corpora linguísticos estandardizados em TEI (Ide/Priest-Dorman/Véronis 1996), as quais têm tardado, contudo, a ser aplicadas a corpora portugueses.

Cada documento XML-TEI divide-se em duas partes, a do conteúdo, etiquetada como <text>, e a do cabeçalho, etiquetada como <TeiHeader>. Esta última descreve a metainformação relativa ao documento de que se trate, a qual pode ser usada por motores de busca desenhados para pesquisar este tipo de XML. Tudo quanto se registre em termos de metadados – relativos à história, aos suportes e às edições do texto em causa e à edição eletrónica que se estiver a elaborar – fica disponível para pesquisas avançadas, articuláveis com as que incidam sobre os conteúdos textuais em si, anotados com marcação que pode ser filológica, linguística, geográfica, cronológica ou outra.

Na Tabela 2 incluímos a demonstração de como algumas etiquetas TEI podem servir a codificação, dentro do elemento <text>, das abreviaturas e leituras difíceis de um manuscrito (cf. também as Tabelas 4 e 5):

Tabela 2: Codificação de abreviaturas, conjeturas e *loci desperati* segundo o protocolo TEI P5 (TEI Consortium 2015).

1	<pre><expan> <abbr>Ill</abbr> <ex>ustrissi</ex> <abbr>mo</abbr> </expan></pre>
2	<pre><choice> <expan>Illustrissimo</expan> <abbr>Illmo</abbr> </choice></pre>
3	<pre><supplied resp="CA" reason="damage">pois</supplied></pre>
4	<pre>Tinha <unclear>vinurazer</unclear> humbocado de contrabando</pre>

Linha 1: codificação da abreviatura «Illmo» e indicação do seu desenvolvimento em «Illustrissimo».

Linha 2: codificação alternativa do desenvolvimento da mesma abreviatura.

Linha 3: codificação da conjuntura «pois» feita pelo editor «CA», devidamente identificado no cabeçalho do documento XML-TEI; esta é uma conjuntura motivada por danificação do manuscrito

Linha 4: codificação de um *locus desperatus* em «vinurazer», uma palavra cujos caracteres se leem bem no manuscrito, mas que desafiam o entendimento do editor.

A reprodução facsimilada do manuscrito de onde provêm as formas acima codificadas, uma carta particular escrita provavelmente em 1827 por um criado que assim pedia ajuda ao seu antigo patrão, pode observar-se no sítio em linha do projeto P.S. *Post Scriptum*.⁶ Na mesma localização escolhe-se a visualização de uma edição diplomática, crítica ou modernizada, com ou sem indicação, conforme desejado, de lemas e de anotação morfossintática. O significado variado das etiquetas, mesmo sem se sair do padrão TEI, ajuda não só a transformar o documento em causa na base para vários formatos simultâneos de saída, que vão da edição mais ou menos diplomática à mais ou menos modernizada, mas podem ser também, entre muitos outros, um glossário, uma lista de abreviaturas desenvolvidas, uma amostra de treino para operações automáticas ou um corpus linguisticamente anotado.

Ao mesmo tempo, torna-se possível, com este tipo de recurso, correr testes automáticos à qualidade do trabalho já executado, desencadeando-se campanhas de correção, quer manual quer automática, dos procedimentos adotados ao longo da edição digital. Para tanto, são de grande ajuda os programas escritos em Perl, acima referidos. Ilustra-se um caso na Tabela 3, que contém um programa em Perl destinado a extrair e listar em separado o conteúdo textual dos parágrafos do corpo de um texto codificado em XML-TEI.

Tabela 3: Programa Perl exemplificado à esquerda e respetiva explicação à direita.

1	<code>#!/usr/bin/perl</code>	1	Declaração de que se trata de um programa Perl.
2	<code>use strict;</code>	2	Instrução dada à interpretação do Perl para que seja estrita...
3	<code>use warnings;</code>	3	...e para que emita avisos sempre que surjam erros no código.

⁶ Localização: <http://ps.clul.ul.pt/pt/index.php?action=edit&cid=CARDS0002>.

<pre>4 use XML: :LibXML;</pre>	<p>4 Instrução de utilização de uma biblioteca Perl, uma biblioteca preexistente, formada especificamente para lidar com o formato XML.</p>
<pre>5 my \$filename = "CARDS0001.xml";</pre>	<p>5 Especificação do ficheiro de entrada, aquele onde se deseja que os comandos venham a ser executados, no caso, o ficheiro CARDS0001.xml.</p>
<pre>6 my \$xmlparser = XML: :LibXML->new();</pre>	<p>6 Instrução de criação de um analisador de sintaxe (um parser) capaz de lidar com XML.</p>
<pre>7 my \$doctree = \$xmlparser-> parse_file(\$filename);</pre>	<p>7 Instrução para o parser analisar a estrutura do ficheiro de entrada: o resultado é a distribuição por uma estrutura em árvore de toda a informação desse ficheiro.</p>
<pre>8 foreach my \$par (\$doctree-> findnodes('/TEI.2/text/body/p')) { print \$par->to_literal, "\n"; }</pre>	<p>8 Instrução para que no documento com a estrutura em árvore (\$doctree) sejam encontrados os nós com a etiqueta <i>p</i>, hierarquicamente inferiores aos nós etiquetados com <i>body</i>, <i>text</i> e <i>TEI.2</i>, sucessivamente, sendo que para cada <i>p</i> encontrado, o respetivo conteúdo textual deve ser impresso e fechado, no final, com um código significando 'nova linha' (\n)</p>

No exemplo, usa-se a função *findnodes* para chegar à informação contida dentro de certos elementos, os elementos *p* (no caso, com o valor de 'parágrafo'). Para tanto, indica-se concretamente o caminho a seguir até se encontrar a localização de *p*, que, dentro da estrutura do documento, é filho de *body*, por sua vez filho de *text*, e este filho de *TEI.2*. Para cada *p* encontrado, extrai-se para impressão em ficheiro separado o respetivo conteúdo textual, fechando-se o resultado com uma instrução para mudança de linha (barra n).

Neste exemplo, o conjunto de instruções em Perl é relativamente curto e simples, mas a mesma linguagem pode servir para inúmeras manipulações, impostas pelas necessidades de investigação ou tão-só pela imaginação. Pode-se conciliar todo o género de variáveis extratextuais previamente registadas ao nível de metadados (e.g. género, cronologia, classificações sociológicas, coordenadas geográficas) com a extração de partes ou da totalidade do texto. Podem-se procurar padrões indetetáveis a olho nu, determinados pelo contexto em que as palavras ou os caracteres ocorrem. Podem-se esconder para efeitos de busca todas as formas conjeturadas de uma edição crítica, ou todas as leituras provenientes de um testemunho menos fiável.

5 A edição acadêmica digital

Esta última secção destina-se a referir recursos para o estudo histórico do português que estejam a ser construídos no âmbito das humanidades digitais. O seu formato emblemático é o da edição acadêmica digital (*scholarly digital edition*), um «recurso de informação que oferece uma representação crítica de documentos ou textos (normalmente) históricos», na definição de Patrick Sahle (Sahle 2014). Seguindo o mesmo autor, responsável por uma pormenorizada elaboração de critérios para avaliação e descrição de edições acadêmicas digitais e pelo respetivo catálogo (<http://www.digitale-edition.de/>), não cabem nesta classificação as meras publicações em formato digital por não serem «sistemas de informação que seguem uma metodologia determinada pelo paradigma digital», como é o caso da metodologia que envolve cuidado na codificação de caracteres e nas linguagens de programação e de marcação que acabámos de ver. As simples publicações em formato digital seguem, por seu lado, metodologias do paradigma impresso. Não se podem considerar edições acadêmicas digitais, por conseguinte, nem as edições impressas digitalizadas, ainda que críticas, nem a maioria dos projetos de digitalização empreendidos por bibliotecas e arquivos, mesmo que acompanhados de descrição, transcrição e indexação. Igualmente excluídos estão os casos das edições digitais que não envolvam representação crítica, i.e., que não respeitam critérios filológicos mínimos, despreocupadas que estão com a responsabilidade de oferecer leituras não mistificadas dos textos históricos publicados.

Em sentido estrito, os corpora anotados hoje disponíveis para o estudo histórico do português (CIPM,⁷ *Corpus do Português*,⁸ *Tycho Brahe*,⁹ *P.S. Post Scriptum*,¹⁰ *WOChWEL*¹¹ e *Colonia*¹²) também não foram montados para funcionarem como edições acadêmicas digitais, desprovidos que estavam todos eles, no início, de linguagem de marcação textual de natureza filológica e da respetiva ancoragem à anotação gramatical, fosse ela POS, morfossintática ou sintática. Cumpriam a prática consagrada da linguística de corpus de separar, enquanto objetos fisicamente autónomos, os documentos que contêm o texto-fonte dos documentos cujos tokens são acompanhados de anotação linguística (*stand-off annotation*, cf. McEnery/Wilson 2001, 1996, 38).

A anotação em *stand-off* tem inegáveis vantagens, reconhecidas mesmo no âmbito das edições acadêmicas digitais (cf., por exemplo, Schmidt 2010). Permite i) o

7 Localização: <http://cipm.fcsh.unl.pt>.

8 Localização: <http://www.corpusdoportugues.org/>.

9 Localização: <http://www.tycho.iel.unicamp.br/~tycho/corpus/index.html>.

10 Localização: <http://clul.ul.pt>.

11 Localização: <http://alfclul.clul.ul.pt/wochwel/index.html>.

12 Localização: <http://corporavm.uni-koeln.de/colonia/>.

cruzamento de ramos ao longo de níveis diferentes de anotação, o que não pode acontecer em XML, por desencadear malformações estruturais, ii) a ausência de interferência entre os vários níveis de anotação adotados, iii) diferentes versões de uma mesma anotação, iv) o acrescento de níveis ulteriores de marcação, sem alteração do texto-fonte, v) o trabalho simultâneo de várias pessoas sobre os mesmos dados, primários e inalterados e vi) a salvaguarda em relação a problemas gerados pela legislação que protege o direito de autor, uma vez que os documentos primários estão arquivados separadamente.

No entanto, algumas destas vantagens convertem-se em desvantagens quando se passa da teoria à prática. Ao anotar informação a níveis diferentes, a ocorrência de erros é muito comum, como observam Grover/Matthews/Tobin (2006), no caso, para a anotação de entidades biomédicas: surgem constantemente incompatibilidades entre as dependências das fronteiras de palavra e as dependências das etiquetas para as diferentes entidades identificadas. Nas edições académicas digitais, por seu lado, acontece constantemente surgirem novas interpretações de natureza filológica a propósito das fontes primárias, o que conduz a necessidades de modificação da anotação de uma mesma ou de várias palavras, já entretanto registadas em níveis diferentes. Pode tornar-se, por conseguinte, muito pouco prático o procedimento do *stand-off*, já que os níveis de anotação se distinguem aí, precisamente, pela independência recíproca e pela não-interferência. A mesma mudança tem de ser repetida, nesta modalidade, tantas vezes quantos os níveis que houver. Há a possibilidade técnica, que está inclusivamente a ser experimentada (cf. por exemplo Druskat 2014), de criar dependências entre os vários níveis de uma anotação em *stand-off*. No entanto, quanto mais complexas e adaptadas, no sentido de não-estandardizadas, forem as camadas de uma anotação, mais difícil e demorada se torna a escrita de programas capazes de lidar simultaneamente com todas elas. Noutra sentida, e uma vez que é muito comum haver investigadores diferentes a trabalhar em simultâneo sobre a mesma coleção de textos, torna-se essencial que as anotações e revisões de cada um sejam sistematicamente visíveis para os outros, o que não acontece quando se trabalha em planos independentes. Também por causa disto, a anotação alinhada (*embedded* ou *inline*), alternativa à anotação em *stand-off* por fazer coincidir num mesmo documento todos os níveis de anotação, pode tornar-se mais desejável (cf. 79 Linguística de corpus e outros usos dos corpora em linguística).

No sentido de mudar do sistema em *stand-off* para um sistema alinhado, o projeto *Tycho Brahe* da Universidade de Campinas, coordenado por Charlotte Galves, construiu a ferramenta *eDictor* (<http://edictor.net/>), a combinação de um editor de XML e de um etiquetador morfossintático, que gera automaticamente edições navegáveis, i.e., em HTML, que podem ser ora diplomáticas, ora semidiplomáticas, ora modernizadas, bem como versões com anotação morfossintática, tanto em texto simples como em XML. Na retaguarda de todos estes diferentes formatos de saída estão documentos em XML que integram, alinhadas em nós irmãos, tanto a informação textual, como a

filológica e a morfossintática (cf. Tabela 4). Qualquer mudança ou revisão dos processos de registo dessas informações só tem de ser introduzida uma vez, o que permite não só o trabalho em equipa em torno de uma mesma edição, mas também a prevenção de desalinhamentos e incompatibilidades entre os formatos de saída, que podem ser gerados de novo na sequência dos processos de correção.

Tabela 4: Estrutura de um XML editado em eDictor, correspondente à anotação alinhada do mesmo exemplo da Tabela 2.

```

1 <w id="20">
  <o>Illmo</o>
  <e t="exp">Illustrissimo</e>
  <e t="norm">Ilustríssimo</e>
  <m v="ADJ-S"/>
  </w>

```

```

2 <w id="48">
  <o>pois</o>
  <comment author="CA" date="09/24/15" title="supplied">damage</comment>
  <m v="C"/>
  </w>

```

```

3 <w id="93">
  <o>vinurazer</o>
  <comment author="CA" date="09/24/15" title="unclear">unclear</comment>
  <m v="VB"/>
  </w>

```

A ferramenta *eDictor* foi anunciada em 2010 (Faria/Kepler/Sousa 2010; Sousa 2013b) e tem estado a ser usada no sentido de criar corpora históricos anotados a partir de edições académicas digitais e vice-versa (o *Corpus Anotado do Português Tycho Brahe* da Universidade de Campinas, os projetos do Grupo de Pesquisas Humanidades Digitais da Universidade de São Paulo, os do Laboratório de História do Português Brasileiro da Universidade Federal do Rio de Janeiro, o *Corpus Eletrónico de Documentos Históricos do Sertão*, CE e o corpus WOChWEL do Centro de Linguística da Universidade de Lisboa (CLUL), entre os que têm já materiais disponíveis).

No projeto P.S. *Post Scriptum* do CLUL, está a ser usada desde 2014 uma tática alternativa à do recurso ao *eDictor*. Dado que esta ferramenta não aceita ainda ficheiros originalmente elaborados em XML-TEI (só aceita as suas próprias definições de XML), dado que também ainda não é utilizável em linha, que não tem lematizador e que não oferece maleabilidade para a inclusão de anotadores automáticos diferen-

tes dos do *Tycho Brahe*,¹³ o *P.S. Post Scriptum* passou a utilizar o sistema TEITOK,¹⁴ desenvolvido no CLUL por Maarten Janssen. Tal como o *eDictor*, o TEITOK cria um mesmo suporte em XML para o corpus linguisticamente anotado e para a edição académica digital, mas em TEITOK esse suporte edita-se num ambiente em linha. Trata-se de um sistema baseado na web para visualizar, criar e editar textos com marcação filológica rica acompanhada de anotação linguística. O sistema contém uma interface gráfica em que o documento anotado pode ser visualizado em formatos diferentes, dependendo dos interesses do visitante. Já para os administradores do mesmo sistema, o TEITOK permite que na mesma interface se edite, transforme e anote o XML-TEI subjacente. Pode-se assim modernizar automaticamente a ortografia dos textos transcritos sem perder a marcação de origem, pode-se lematizar e anotar morfossintaticamente os mesmos textos com recurso a anotadores automáticos definidos pelo utilizador e consultar todo o resultado em função das variáveis extratextuais registadas ao nível de metadados. Mais uma vez, na retaguarda de toda esta maleabilidade está uma anotação alinhada e não uma anotação em *stand-off*. Estão ficheiros em linguagem XML que alinham em torno da leitura crua das palavras da fonte primária, sucessivos atributos com as informações paleográficas, filológicas, lexicais e gramaticais que o investigador queira registar (cf. Tabela 5).

Tabela 5: Estrutura de um XML editado em TEITOK, correspondente à anotação alinhada do mesmo exemplo das Tabelas 2 e 4.

```
1 <tok id="w-20" form="Illmo" pos="ADJ-S" fform="Illustrissimo" #x201D;
  nform="Ilustrissimo" lemma="ilustre">Illmo</tok>
```

```
2 <supplied resp="CA" reason="damage">
  <tok id="w-48" form="pois" pos="C" lemma="pois">pois</tok>
  </supplied>
```

```
3 <unclear>
  <tok id="w-93" pos="VB" lemma="vinurazer">vinurazer</tok>
  </unclear>
```

O sistema tem também a vantagem de poder ser usado na edição e anotação de textos de qualquer língua (já está a anotar espanhol, por exemplo, ainda dentro do *P.S. Post Scriptum*), mas apresenta a desvantagem de exigir a presença de programadores experientes nas equipas que o usam, técnicos que consigam modificá-lo no sentido da inclusão de corpora de treino e de etiquetadores automáticos adicionais.

¹³ Localização: <http://www.tycho.iel.unicamp.br/~tycho/corpus/manual/tags.html>.

¹⁴ Localização: <http://alfclul.clul.ul.pt/teitok/site/index.php>.

Tanto o *eDictor* como o TEITOK continuam a ser constantemente apurados e expandidos em termos de sofisticação. No horizonte mais próximo, por exemplo, está o desafio do alinhamento da anotação sintática. Mas ambos demonstram para já, aliados a exemplos referidos nas secções anteriores, como a filologia do português sempre se manteve ativa ao nível da experimentação em humanidades digitais.

6 Observações finais

Vimos neste capítulo como as humanidades digitais se caracterizam por uma principal virtude: são abertas. Abertas no sentido de serem acessíveis, discutíveis, colaborativas, democráticas. Caracterizam-se também por uma imposição: são velozes. Os procedimentos mecânicos têm uma velocidade intrínseca que é, à partida, incompatível com o ritmo pausado da interpretação, ou exegese, método incontornável na construção de conhecimento em humanidades. Entrou também, por conseguinte, na agenda das humanidades digitais a reflexão sobre novos métodos, adaptados à tecnologia, ou transformadores da mesma. Também se revisitam velhos temas, convertidos em tópicos atuais pela sociedade de informação: oposição entre técnica e conhecimento, diferença entre língua natural e língua artificial, relação entre texto e gramática.

7 Referências

- Barthes, Roland (1994, ¹⁹⁷¹), *De l'œuvre au texte*, in: Éric Marty (org.), *Œuvres complètes*, vol. 2, Paris, Seuil, 1211–1217.
- Beaugrande, Robert-Alain/Dressler, Wolfgang Ulrich (2005, ¹⁹⁷²), *Introducción a la lingüística del texto*, Barcelona, Ariel.
- Bédier, Joseph (1970), *La tradition manuscrite du Lai de l'ombre: réflexions sur l'art d'éditer les anciens textes*, Paris, Champion.
- Blecu, Alberto (1983), *Manual de crítica textual*, Madrid, Castalia.
- Busa, Roberto A. (2004), *Foreword: Perspectives on the Digital Humanities*, in: Susan Schreibman / Raymond George Siemens/John Unsworth (edd.), *A Companion to Digital Humanities*, Malden MA, Blackwell, xvi–xxi.
- Castro, Ivo (2001), *Metodologia do Aparato Genético*, in: Manuel Simões/Ivo Castro/João David Pinto Correia (edd.), *Memória dos Afectos: Homenagem a Giuseppe Tavani*, Lisboa, Colibri, 69–81.
- Castro, Ivo, et al. (1973), *Normas de Transcrição para Textos Medievais Portugueses*, Boletim de Filologia 23, 417–425.
- Cerquiglioni, Bernard (1989), *Éloge de la variante. Histoire critique de la philologie*, Paris, Seuil.
- Del Mancino, William/Pierrel, Jean-Marie (2009), *Du trésor de la langue française à l'ATILF et au CNRTL*, La revue pour l'histoire du CNRS [en ligne] 24, mis en ligne de 05 octobre 2009, URL: <http://histoire-cnrs.revues.org/9133> (30.09.2015).
- Dionísio, João (2006), «*Ab da dolchor del temps novel?*», in: Olga Pombo/António Guerreiro/António Franco Alexandre (edd.), *Enciclopédia e Hipertexto*, Charneca da Caparica, Edições Duarte Reis, 376–387.

- Druskat, Stephan (2014), *An Open-Source Software Platform for Multi-Level Corpus Annotation*, in: Stephan Druskat et al. (edd.), *Proceedings of the 12th Konferenz zur Verarbeitung natürlicher Sprache (KONVENS 2014)*, Hildesheim, Universitätsverlag Hildesheim, 228–234.
- Emiliano, António (2005), *Tipo Medieval para Computador: uma Ferramenta Informática para Filólogos, Historiadores e Paleógrafos*, Signo. Revista de Historia de la Cultura Escrita 15, 139–176.
- Ertuna, Irmak (2009), *Stiegler and Marx for a Question Concerning Technology*, Transformations 17, URL: http://www.transformationsjournal.org/journal/issue_17/article_07.shtml (30.09.2015).
- Faria, Pablo P. F./Kepler, Fábio N./Sousa, Maria Clara Paixão de (2010), *An Integrated Tool for Annotating Historical Corpora*, in: *Proceedings of the Fourth Linguistic Annotation Workshop*, Stroudsburg PA, The Association for Computational Linguistics, 217–221.
- Fiormonte, Domenico (2012), *Towards a Cultural Critique of the Digital Humanities*, in: Manfred Thaller (ed.), *Controversies Around the Digital Humanities*, Historical Social Research/Historische Sozialforschung, special issue, 59–76.
- Goldfarb, Charles F. (1999), *Future Directions in SGML/XML*, in: Wiebke Möhr/Ingrid Schmidt (edd.), *SGML und XML*, Berlin/Heidelberg, Springer, 3–25, http://dx.doi.org/10.1007/978-3-642-46881-0_1 (21.10.2015).
- Grover, Claire/Matthews, Michael/Tobin, Richard (2006), *Tools to Address the Interdependence between Tokenisation and Standoff Annotation*, in: *Proceedings of the 5th Workshop on NLP and XML: Multi-Dimensional Markup in Natural Language Processing*, Stroudsburg PA, Association for Computational Linguistics, 19–26, URL: <http://dl.acm.org/citation.cfm?id=1621034.1621038> (30.09.2015).
- Hayles, N. Katherine (2004), *Print is Flat, Code is Deep: The Importance of Media-specific Analysis*, Poetics Today 25:1, 67–90.
- Hockey, Susan/Marriott, Ian (1980), *The Oxford Concordance Program. Version 1.0. Users' Manual*, Oxford, Oxford University Computing Centre.
- Ide, Nancy/Priest-Dorman, Greg/Véronis, Jean (1996), *Corpus Encoding Standard*, <http://www.cs.vassar.edu/CES/> (30.09.2015).
- Kirschenbaum, Matthew G. (2010), *What is Digital Humanities and what's it doing in English Departments?*, Association of Departments of English Bulletin 150, 1–7.
- Koch, Peter/Oesterreicher, Wulf (2007, ¹1999), *Lengua hablada en la Romania: español, francés, italiano*, Madrid, Gredos.
- Mateus, Maria Helena Mira (1968), *Informática e Linguística: a Mecanografia nos Estudos da Linguagem*, Revista de Portugal, série A 33, 217–232.
- Mateus, Maria Helena Mira (1974), *Glossário da Vida e Feitos de Júlio César, tradução portuguesa quatrocentista de Li Fet des Romains*, Boletim de Filologia 23, 1–80.
- Mateus, Maria Helena Mira (2010), *Vida e Feitos de Júlio César*, vol. 3, Lisboa, Fundação Calouste Gulbenkian.
- McCarty, Willard (2004), *Modeling: A Study in Words and Meanings*, in: Susan Schreibman/Raymond George Siemens/John Unsworth (edd.), *A Companion to Digital Humanities*, Malden, MA, Blackwell, 254–270, <http://www.digitalhumanities.org/companion/> (30.09.2015).
- McCarty, Willard (2005), *Humanities Computing*, New York, Palgrave.
- McEnery, Tony/Wilson, Andrew (2001, ¹1996), *Corpus Linguistics*, Edinburgh, Edinburgh University Press.
- McGann, Jerome J. (2014), *A New Republic of Letters: Memory and Scholarship in the Age of Digital Reproduction*, Cambridge, MA/London, Harvard University Press, <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=706819> (21.10.2015).
- McKenzie, Donald F. (2002), *Making Meaning: «Printers of the Mind» and Other Essays*, ed. por Peter D. McDonald/Michael Felix Suarez, Amherst MA, University of Massachusetts.

- Murdoch, Brian O. (1971), *The Production of Concordances from Diplomatic Transcriptions of Early Medieval German Manuscripts: Some Comments*, in: Roy A. Wisbey (ed.), *The Computer in Literary and Linguistic Research*, Cambridge, Cambridge University Press, 35–44.
- Parkinson, Stephen (1983), *Um Arquivo Computorizado de Textos Medievais Portugueses*, *Boletim de Filologia* 28, 241–252.
- Poisson, Esther (2011), *Le Trésor de la langue française informatisé: une ressource d'une valeur insoupçonnée*, *Correspondance* 17, n. 1, URL: <http://correspo.ccdmd.qc.ca/Corr17-1/Capsule.html> (30.09.2015).
- Prista, Luís/Albino, Cristina (1996), *Filólogos Portugueses entre 1868 e 1943: Catálogo da Exposição organizada para o XI Encontro Nacional da Associação Portuguesa de Linguística, Faculdade de Letras de Lisboa, 1995*, Lisboa, Colibri.
- Ricœur, Paul (1986), *Du texte à l'action. II, Essais d'herméneutique*, Paris, Seuil.
- Robertson, Benjamin J. (2015), *The Grammatization of Scholarship*, *Amodern* 1, URL: <http://amodern.net/article/the-grammatization-of-scholarship> (30.09.2015).
- Roncaglia, Aurelio (1975), *Principi e applicazioni di critica testuale*, Roma, Bulzoni.
- Sahle, Patrick (2014), *Criteria for Reviewing Scholarly Digital Editions, Version 1.1.*, IDE, Institut für Dokumentologie und Editorik, URL: <http://www.i-d-e.de/reviews/criteria-version-1-1> (30.09.2015).
- Schmidt, Desmond (2010), *The Inadequacy of Embedded Markup for Cultural Heritage Texts*, *Literary and Linguist Computing* 25:3, 337–356.
- Sousa, Maria Clara Paixão de (2013a), *Texto Digital: uma Perspectiva Material*, *Revista da ANPOLL* 35, 17–60.
- Sousa, Maria Clara Paixão de (2013b), *A Filologia Digital em Língua Portuguesa: alguns Caminhos*, in: Maria Filomena Gonçalves/Ana Paula Banza (edd.), *Património Textual e Humanidades Digitais: da Antiga à Nova Filologia*, Évora, CIDEHUS, 113–138.
- Steriade, Donca (1995), *Underspecification and Markedness*, in: John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, Oxford, Blackwell, 114–174.
- Stiegler, Bernard (1998, ¹1994), *Technics and Time, 1, The Fault of Epitheus*, Stanford CA, Stanford University Press.
- TEI Consortium (2015), *Guidelines for Electronic Text Encoding and Interchange* [last modified 2015], URL: <http://www.tei-c.org/P5> (30.09.2015).
- Timpanaro, Sebastiano (2005, ¹1963), *The Genesis of Lachmann's Method*, Chicago, Chicago University Press.
- Wall, Larry/Christiansen, Tom/Orwant, Jon (2004), *Programming Perl*, s.l., O'Reilly, 2004.
- Xavier, Maria Francisca/Brocardo, Maria Teresa/Vicente, Maria da Graça (1995), *CIPM: Um Corpus Informatizado do Português Medieval*, in: *Actas do X Encontro Nacional da Associação Portuguesa de Linguística*, [Lisboa], APL, 599–612.

João Costa, Maria João Freitas e Anabela Gonçalves

11 Linguística clínica: alguns dados sobre o português

Abstract: A linguística clínica é geralmente referida como a aplicação da linguística ao estudo da incapacidade linguística em qualquer uma das suas formas. Estabelece-se, assim, entre teoria linguística e linguística clínica, uma via de dois sentidos, que se explora no presente capítulo: os contextos de perturbação linguística são janelas para um melhor entendimento do funcionamento das línguas naturais e, por sua vez, a teoria linguística fornece-nos instrumentos claros para diagnósticos e intervenções mais eficazes. Assim, procede-se a uma reflexão sobre casos de contributos da teoria linguística para uma compreensão fina de perturbações sintáticas e fonológicas, bem como de contributos dos estudos de perturbações para uma compreensão de fenómenos linguísticos. Apresentam-se, ainda, dois instrumentos de avaliação linguística direcionados para componentes gramaticais específicas (sintaxe e fonologia), auxiliares do trabalho do terapeuta e utilizáveis após aplicação de instrumentos gerais de avaliação da linguagem.

Keywords: teoria linguística, linguística clínica, desenvolvimento linguístico atípico, instrumentos de avaliação linguística

1 Linguística clínica e teoria linguística: uma estrada com dois sentidos

Desde os anos 50, com o nascimento das ciências cognitivas, sabemos que a linguagem é entendida como um objeto de estudo da cognição, isto é, como um produto das atividades do cérebro e da mente humana. Esta afirmação implica perceber que entender o funcionamento das línguas naturais tem como consequência descobrir o seu lugar no organismo, como a linguagem interage com outras componentes internas e externas à cognição.

É fácil entender que há aspetos externos à cognição e ao próprio ser humano que condicionam comportamentos linguísticos. Por exemplo, condicionantes culturais afetam o conjunto de palavras de cada língua. Condicionantes afetivas condicionarão o uso que cada falante faz da língua em situações diferenciadas. Níveis de educação condicionarão a capacidade de domínio e adequação de diferentes registos a situações contextuais diferenciadas.

Há, contudo, aspetos internos ao organismo, mas externos à cognição, que também podem explicar comportamentos linguísticos diferenciados. Por exemplo, uma deficiência no aparelho articulatório, ainda que simples, como a falta de alguns

dentos, afeta a forma como alguns sons são pronunciados. Problemas periféricos de natureza articulatória podem afetar a deglutição de alimentos e condicionar a forma como se articulam alguns sons, o que justifica que muitos problemas desta natureza sejam tratados por terapeutas da fala.

Quando estudamos a cognição, podemos perceber que há fatores estritamente linguísticos e fatores cognitivos gerais que estão envolvidos no funcionamento da linguagem, na distinção entre o que Chomsky/Hauser/Fitch (2002) designam de Faculdade da Linguagem em Sentido Amplo (*Broad Faculty of Language*) e Faculdade da Linguagem em Sentido Estrito (*Narrow Faculty of Language*). A primeira compreende não apenas as propriedades intrínsecas ao funcionamento das línguas, mas também todos os subsistemas da cognição que com eles interagem. Por exemplo, a memória, a atenção, os sistemas de organização neuronal são componentes da cognição fundamentais para bons desempenhos linguísticos, mas não estão especificamente alocados ao funcionamento da linguagem. Sem eles não há linguagem, mas eles não explicam o funcionamento das línguas. Já quando estudamos a Faculdade da Linguagem em Sentido Estrito, descobrimos aquelas propriedades da língua que são estritamente linguísticas. Por exemplo, a possibilidade que temos de pronunciar um constituinte de uma frase longe da posição em que é interpretado decorre apenas de mecanismos linguísticos. Consideremos o seguinte exemplo. Em (1), interpretamos *a Maria* como complemento da forma verbal *viu*.

- (1) Todos disseram que o Pedro viu a Maria.

Em (2), apesar de estar numa posição distinta, continuamos a interpretar *a Maria* como complemento da mesma forma verbal:

- (2) A Maria, todos disseram que o Pedro viu.

Somos, portanto, capazes de estabelecer uma relação de dependência entre formas que não estão adjacentes. Como demonstrado nos últimos 50 anos de investigação em linguística, a partir do trabalho seminal de Ross (1969) e Chomsky (1975), são restrições unicamente linguísticas e que não derivam de nenhum outro aspeto da cognição que explicam a nossa capacidade de produzir e interpretar enunciados como o exemplificado em (2). São, portanto, propriedades linguísticas em sentido estrito.

Durante várias décadas, apesar do desiderato de entender a língua como um objeto da cognição, a teoria linguística precisou de se estabelecer e desenvolver e nem sempre foi clara a interação entre os níveis de abordagem dos fenómenos linguísticos. Há três níveis a distinguir, conforme enunciado em Sprouse/Lau (2013), entre outros: computacional, algorítmico e implementacional. O nível computacional é o nível da teoria gramatical, que nos permite descrever como é o sistema de geração de enunciados linguísticos e por que motivo tem o formato que tem. O nível algorítmico, estudado na psicolinguística, estuda a forma como o sistema é usado e quais os

mecanismos e processos ativados para a produção e compreensão dos enunciados que o sistema computacional tem o potencial de gerar. Finalmente, o nível implementacional relaciona-se com a representação física da língua no cérebro, isto é, com a atividade eletroquímica e com a representação neuronal concreta da linguagem. Este é o nível tradicionalmente estudado pela neurolinguística.

Como dizíamos, a comunicação entre estes três níveis tem sido, por vezes, tímida, mas tem vindo a ganhar crescente atenção na última década, não só por razões teóricas (Chomsky 2004 formula o desejo de ir além da adequação explicativa e entender a linguagem como um todo), mas também por se perceber as vantagens de uma aproximação entre estes três níveis no quadro da *linguística clínica*.

A linguística clínica é geralmente referida como a aplicação da linguística ao estudo da incapacidade linguística em qualquer uma das suas formas (Crystal 1981). Tradicionalmente, incluem-se, nos tipos de incapacidade linguística, três tipos maiores: as dificuldades que advêm do sistema auditivo, do sistema articulatório e do sistema central.

Entende-se que as dificuldades do sistema auditivo resultam em incapacidades na perceção. Os casos mais claros de perturbações desta natureza incluem os diversos tipos de surdez (de condução ou neuro-sensorial), que podem ter consequências diferenciadas, na medida em que dos diferentes tipos podem advir dificuldades de descodificação fonética, mas também problemas mais profundos de natureza sintática.

As dificuldades no sistema articulatório induzem problemas na fluência, na voz ou na articulação. Um exemplo de problema de fluência é a gaguez; a perturbação na voz pode ser exemplificada por problemas de ressonância e as dificuldades articulatórias podem gerar substituições de uns sons por outros.

As perturbações no sistema central, isto é, no cérebro enquanto mediador entre a fala e a perceção, estão na base das afasias. Importa salientar que um dos principais argumentos para perceber a linguagem enquanto atividade da mente advém da observação de que lesões em regiões diferentes do cérebro produzem dificuldades específicas da linguagem, sem que haja necessariamente outras perturbações associadas. Os exemplos clássicos são as afasias de Broca (induzida por lesão na parte anterior do cérebro) e Wernicke (induzida por lesão na parte posterior do cérebro), designadas segundo o nome dos neurologistas que as descobriram, que induzem dificuldades de produção de estruturas gramaticais ou de compreensão e parafasias fonémicas, respetivamente.

Porque a teoria linguística permitiu avanços enormes na caracterização, descrição e explicação dos comportamentos linguísticos, é hoje possível dar um contributo para os casos clínicos relacionados com linguagem de uma forma muito mais clara e detalhada. Em vez de se fazerem caracterizações globais, é possível perceber com grande pormenor o tipo de distúrbio associado a cada tipo de perturbação.

A linguística clínica tem cinco objetivos claros:

- (a) *Descrição* – através dos instrumentos da linguística, é possível descrever os comportamentos linguísticos associados a diferentes patologias, tendo em vista *diagnósticos diferenciais*, isto é, uma caracterização das propriedades específicas associadas a cada subtipo de perturbação.
- (b) *Clarificação* – decorre de uma descrição mais fina e detalhada uma clarificação dos aspetos envolvidos em cada tipo de perturbação.
- (c) *Diagnóstico* – como foi referido, pretende-se que da interação entre a teoria linguística e o conhecimento clínico se obtenham diagnósticos para definir caminhos para a intervenção.
- (d) *Avaliação* – para que o diagnóstico seja possível, é necessária a criação de instrumentos que o viabilizem. Esta é uma das áreas de maior carência para os profissionais, que pedem instrumentos de avaliação adequados à língua com que trabalham. Só com um bom conhecimento da língua é possível a construção de instrumentos de avaliação precisos. Por outro lado, só com um conhecimento preciso do que está afetado em cada contexto de perturbação é possível construir instrumentos de avaliação finos e apropriados. Há uma óbvia interdependência entre o diagnóstico e a avaliação, que requer muitos estudos exploratórios até ao apuramento dos instrumentos mais fiáveis.
- (e) *Intervenção* – como será evidente, o objetivo último desta disciplina será contribuir para a reabilitação das perturbações linguísticas dos pacientes. Pretende-se, portanto, entender quais os contributos que a linguística pode dar para a construção dos melhores instrumentos de intervenção terapêutica e para a monitorização dos resultados da intervenção.

Se estes objetivos são claros e focados numa dimensão muito prática e aplicada – a orientação para a intervenção e para o bem-estar dos pacientes –, pode ser menos claro que a linguística clínica abre uma estrada de dois sentidos. O sentido do contributo da linguística para a prática clínica está já demonstrado. Mas há outro sentido não menos importante: através da linguística clínica, a própria teoria linguística recebe um *input* crucial para o seu desenvolvimento. O estudo das perturbações linguísticas é uma janela para o funcionamento da linguagem e um instrumento para testar as propostas de modelos de análise da língua. Consideremos dois casos clássicos, provenientes da sintaxe e da fonologia.

Chomsky (1957) propôs a autonomia da sintaxe face a outras componentes da gramática. Isto implica que a teoria gramatical deve ser capaz de gerar enunciados bem formados gramaticalmente, independentemente do seu significado. Por exemplo, uma frase como «A minha televisão está grávida.» é bem formada sintaticamente, embora semanticamente anómala. Esta autonomia entre as componentes da gramática é fortemente criticada noutros modelos teóricos. As afasias de Broca, em que os pacientes exibem perturbações de natureza sintática, sem que haja problemas de natureza semântica, com dificuldades específicas na produção dos elementos sintáticos dos seus enunciados (conjunções, flexão, etc.), constituem um teste excelente para validar a hipótese de autonomia relativa entre componentes da gramática.

Um outro exemplo pode vir da fonologia. A sílaba, enquanto unidade prosódica de organização dos sons, é comumente assumida na teoria linguística, mas há modelos que contestam a existência deste nível de análise (como a *Government Phonology*, de Kaye/Lowenstamm/Vergnaud 1990). A observação de que, em contextos de Perturbação Específica do Desenvolvimento da Linguagem (PEDL), há crianças

que têm dificuldades específicas com sílabas (Marshall/van der Lely 2007) constitui uma evidência robusta para a realidade psicológica da sílaba e um teste claro para confrontar os modelos que advogam a sua existência com os que defendem a sua dispensabilidade.

Vemos, assim, que o conhecimento das perturbações linguísticas é um instrumento valioso para a própria teoria linguística e que esta pode beneficiar bastante do estudo de casos clínicos para um aperfeiçoamento dos seus próprios modelos e para validação de hipóteses em confronto.

Neste capítulo, na impossibilidade óbvia de fazer um estado da arte exaustivo dos trabalhos desenvolvidos em linguística clínica para o português,¹ pretendemos oferecer uma panorâmica do que pode ser a relação entre perturbações linguísticas e teoria linguística neste mútuo contributo e no cumprimento dos diferentes objetivos da linguística clínica.

O capítulo encontra-se organizado da seguinte forma:

Na secção 2, é feita uma reflexão sobre a produtividade da relação entre teoria fonológica e conhecimento sobre o desenvolvimento linguístico para a avaliação das estruturas fonológicas nos enunciados verbais em contextos típico e atípico.

Na secção 3, apresentamos alguns casos de contributos da teoria linguística para uma compreensão fina de perturbações sintáticas, assim como contributos dos estudos de perturbações linguísticas para uma compreensão de fenómenos linguísticos. Por fim, enunciamos algumas linhas de trabalho que importa desenvolver.

Na secção 4, centrar-nos-emos na apresentação de dois instrumentos de avaliação linguística (um de base sintática e outro de base fonológica) que ilustram a possibilidade de desenvolver instrumentos direcionados para componentes gramaticais específicas, auxiliares do trabalho do terapeuta e utilizáveis após aplicação de instrumentos gerais de avaliação da linguagem.

Na secção 5, apresentam-se as conclusões do capítulo.

2 Linguística clínica e fonologia

2.1 O papel dos segmentos

A formação de base na área da terapia da fala privilegia, tradicionalmente, aspetos relacionados com as perturbações que afetam os sons da fala. Este facto decorre do papel central das unidades fonológicas na comunicação humana, com contributo determinante para a construção do significado. A comutação de segmentos num par de palavras de uma língua permite estabelecer contrastes fonológicos, cruciais para o processamento lexical: a substituição de um segmento por outro ([g]ola / [m]ola;

¹ Neste capítulo, *português* é usado como sinónimo de *português europeu*.

s[a]la / s[ɔ]la) tem impacto na natureza do item lexical produzido, com consequências óbvias para a construção do significado em contexto comunicativo. O facto de uma criança não produzir um ou mais segmentos-alvo numa dada palavra ([kâte] para *fralda* – Inês, 1;5) pode, assim, comprometer a comunicação.

As unidades mínimas que constituem os enunciados de fala são mensuráveis do ponto de vista físico e, como tal, descritas no domínio da fonética. Os *sons da fala* (ou *fonemes* ou *segmentos fonéticos*) são articulados para a construção de enunciados de fala (dimensão articulatória), propagam-se sob a forma de ondas sonoras (dimensão acústica) e são captados pelo aparelho auditivo (dimensão auditiva). Estas entidades fisicamente observáveis constituem o *output* fonético de unidades representacionais, os *fonemas* (ou *segmentos fonológicos*), parte integrante das representações das palavras no léxico mental. Assim, e tendo em conta as dimensões fonética e fonológica acima referidas, estas unidades mínimas (vogais, semivogais e consoantes) surgem nos enunciados de fala como produto tanto do processamento linguístico central como da ativação de mecanismos motores de articulação no aparelho fonador. Deste modo, tanto uma alteração na organização do conhecimento linguístico como um problema anatómico nas estruturas responsáveis pela articulação poderão causar alterações na qualidade dos enunciados de fala, que, em situações de perturbação severa, comprometem a inteligibilidade e, portanto, colocam em perigo a comunicação.

Nem sempre as dimensões fonética e fonológica dos segmentos se encontram claramente diferenciadas, o que pode ter impacto no formato da intervenção e na velocidade da reabilitação. Como referimos, face a uma alteração na produção de um dado segmento, duas hipóteses se colocam:

- (i) trata-se de um problema articulatório, funcional, decorrente de alterações anatómicas na morfologia do aparelho fonador;
- (ii) trata-se de uma alteração que decorre não de deficiência anatómica, periférica, mas de problemas de processamento central do conhecimento fonológico.

As alterações à estrutura segmental foram tradicionalmente classificadas, até à década de 70 do século passado, como perturbações fonéticas, sendo tratadas com base em programas de intervenção de base articulatória, com treino exclusivo da função motora (Lamprecht et al. 2004). No entanto, em muitos casos de alteração da qualidade dos sons da fala, não há diagnóstico de lesões físicas, motoras, que justifiquem os desempenhos atípicos. Esta observação esteve na base de uma nova perspectiva de trabalho, nascida em meados dos anos 70 do século passado (entre outros, Ingram 1976; Grunwell 1982), com distinção clara entre alterações na produção devidas a deficiência articulatória e alterações decorrentes de organizações deficitárias do conhecimento fonológico. Nestes casos, apenas a reabilitação do sistema fonológico permitirá o sucesso da intervenção.

Apesar das mudanças introduzidas à época, alguma indefinição envolve ainda hoje a classificação das perturbações com impacto no formato sonoro dos enuncia-

dos. A própria designação usada para referir cumulativamente todas as manifestações desta natureza foi mudando ao longo das décadas, não havendo consenso quanto a uma tipologia única. Note-se que o uso atual da expressão *perturbações dos sons da fala* (American Psychiatric Association 2013) inclui perturbações de natureza fonética, de base motora e auditiva, e perturbações de natureza fonológica, decorrentes de problemas na organização do sistema linguístico, portanto, de base cognitiva.

As *perturbações dos sons da fala* de base fonética são diagnosticadas, por exemplo, em casos como (Lousada/Alves/Freitas em prep.):

- (i) desenvolvimento inadequado das estruturas motoras, decorrente ou não de hábitos quotidianos como o uso prolongado da chucha;
- (ii) estado inadequado das estruturas motoras na sequência de uma lesão adquirida, por traumatismo;
- (iii) estado inadequado do tônus muscular em contexto de lesão neurológica;
- (iv) incompetência velofaríngea em situações de fenda palatina.

Em todos estes casos, há prejuízo da função articulatória.

As *perturbações dos sons da fala* de base fonológica são observáveis em estratégias de reconstrução correspondentes a padrões de erro sistemáticos no sistema da criança, que envolvem produções não conformes ao alvo.² Estas estratégias de reconstrução são muitas vezes identificadas também nas crianças com desenvolvimento típico, sendo que o que está em causa na perturbação é, nestes casos, um desenvolvimento tardio das estruturas (Bernhardt/Stemberger 1998). A título ilustrativo, refira-se a produção de consoantes não vozeadas por vozeadas, de oclusivas por fricativas, de um ponto de articulação anterior por um posterior (Lamprecht et al. 2004; Lazzarotto-Volcão 2009; Lousada 2012).

2.2 Interação entre teoria fonológica e desenvolvimento fonológico típico e atípico

Até aos anos 70 do século passado, vigorou na teoria fonológica uma visão linear, de base segmental, da organização fonológica das línguas. O trabalho sobre línguas tonais nos anos 70 mostrou que outras unidades para além dos segmentos permitem a caracterização das propriedades sonoras dos sistemas linguísticos. Esta mudança de paradigma corresponde ao surgimento dos modelos não-lineares (ou multilíneares), que se mantêm até hoje. Ao longo dos anos 80 (para revisão da literatura, cf. de

² Estas estratégias de reconstrução são frequentemente tratadas como *processos fonológicos* na literatura sobre desenvolvimento atípico e na prática clínica (Miccio/Scarpino 2008). Usamos o termo *estratégias de reconstrução* uma vez que, em linguística, o termo *processo fonológico* designa uma generalização sobre o funcionamento de determinado segmento ou classe de segmentos na gramática de uma língua e não um padrão de erro na produção de enunciados verbais.

Lacy 2007; Goldsmith/Riggle/Yu 2011), vários modelos de representação do conhecimento fonológico foram propostos, com integração de diferentes tipos de unidades, prosódicas e segmentais, hierarquicamente organizadas: *enunciado*, *sintagma entoacional*, *sintagma fonológico*, *palavra prosódica*, *pé*,³ *silaba*, *segmento*, *traços distintivos* (com base em Nespor/Vogel 1986).

A perspectiva não-linear de tratamento da fonologia das línguas tem sido testada desde os anos 90 na avaliação do desenvolvimento fonológico típico e atípico (entre outros, Fikkert 1994; Bernhardt/Stemberger 1998; 2000; Lamprecht et al. 2004; Marshall/van der Lely 2009; Ferré/dos Santos 2015). Das várias unidades referidas, as mais testadas têm sido a *palavra prosódica*, o *pé*, a *silaba*, o *segmento* e os *traços distintivos*. Nas secções que se seguem, é dada informação sumária sobre a avaliação de cada um destes constituintes, com especial referência aos aspetos testados no português.

Uma vez que os terapeutas da fala lidam quer com o conhecimento implícito quer com a consciência linguística, tanto na avaliação como na reabilitação, os estudos que abaixo se citam relatam resultados para as unidades fonológicas listadas provindos de estudos elaborados numa ou na outra perspectiva, em contexto de desenvolvimento típico e atípico.

2.2.1 A palavra prosódica

O aspeto mais testado no âmbito da observação do impacto da palavra prosódica na avaliação do desenvolvimento fonológico tem sido a *extensão de palavra* (Bernhardt/Stemberger 2000; Marshall/van der Lely 2009; Ferré/dos Santos 2015). Sabemos que as primeiras palavras são de extensão reduzida, sendo privilegiados formatos monossilábicos e dissilábicos, embora os trissílabos surjam precocemente no português, ao contrário do registado para o inglês ou o catalão (Vigário/Freitas/Frota 2006). O predomínio de dissílabos nos estádios iniciais de desenvolvimento (neste último estudo, 46,5%) é consistente com o predomínio deste formato de palavra nos enunciados dos adultos (44%); no mesmo estudo, os restantes formatos de palavra nos estádios iniciais de desenvolvimento situam-se abaixo dos 20%. Tais resultados permitem prever uma preferência por padrões dissilábicos, facilitadores de tarefas de avaliação de conhecimento linguístico.

Num estudo preliminar, Ramalho/Almeida/Freitas (2013) verificaram o impacto da variável *extensão de palavra* numa amostra de crianças com desenvolvimento fonológico típico. Foram observadas produções de itens lexicais em que [r] é a segunda consoante do ataque ramificado em sílaba átona, em três formatos de palavra: dissílabos (*livros*), trissílabos (*palavras*) e polissílabos (*fotografia*). Os resulta-

3 O pé é a unidade responsável pela representação do acento de palavra na teoria fonológica.

dos mostraram uma taxa de sucesso elevada nos dissílabos (84,8%), que desceu nos trissílabos (71,2%) e nos polissílabos (45,8%).

A variável *extensão de palavra* tem sido avaliada também em estudos sobre a consciência fonológica, tanto em tarefas de consciência silábica como segmental. Diferentes resultados têm sido relatados: Alves/Castro/Correia (2010) resumiam trabalhos desenvolvidos no âmbito do projeto que coordenam, reportando dados que mostram o impacto significativo da extensão de palavra em tarefas que avaliam estes dois tipos de consciência fonológica. O efeito da mesma variável fora já observado em Afonso (2008) e em Vicente (2009), com os dissílabos a favorecerem o desempenho das crianças, por oposição aos trissílabos.

Uma outra variável relevante para o processamento fonológico é a *posição na palavra*. A maior facilidade em processar segmentos em posição inicial de palavra é comumente descrita na literatura. A título ilustrativo, Marshall/van der Lely (2009) descrevem o facto de os grupos consonânticos serem mais difíceis de processar em posição medial do que em posição inicial, tanto por crianças disléxicas como por crianças com PEDL.

Costa (2010) mostrou, para uma amostra de crianças portuguesas com desenvolvimento típico, que os pontos de articulação Labial e Dorsal estabilizam primeiro na periferia esquerda da palavra, por oposição à posição medial de palavra; no entanto, no mesmo estudo, as líquidas estabilizam primeiro em posição medial. Está descrita para o português a aquisição precoce da coda fricativa em final de palavra, por oposição à mesma estrutura em posição medial (Freitas 1997). Este facto foi também observado em Nogueira (2007), numa amostra de crianças prematuras nascidas com muito baixo peso, e em Baptista (2015), numa amostra de crianças com historial de otites médias com derrame. Porém, tendo em conta a totalidade das estruturas observadas, a variável não teve um impacto significativo nas amostras clínicas nestes dois estudos, embora o efeito da variável *posição na palavra* no grupo de controlo tenha sido observado em Nogueira (2007).

Os resultados disponíveis permitem questionar o grau de produtividade da variável *posição na palavra* na avaliação de amostras com desenvolvimento atípico, sendo necessária a sua testagem junto de amostras mais alargadas e com outros perfis clínicos.

2.2.2 O acento de palavra

O acento de palavra é tido como promotor da aquisição das estruturas fonológicas (Fikkert 1994), atribuindo proeminência perceptiva às estruturas a que se associa. O formato das primeiras palavras resulta muitas vezes de trunicações com preservação de material segmental da sílaba tónica (*gosta* [ˈgɔ] – Inês, 1;8; *sapato* [ˈpa] – João, 1;7), o que argumenta a favor do efeito do acento no desenvolvimento fonológico. No entanto, está ainda por definir o impacto significativo desta variável na avaliação do conhecimento fonológico infantil em português.

Em Nogueira (2007), o efeito promotor do acento de palavra revelou-se no grupo de controlo (com desenvolvimento típico), mas não no grupo experimental (crianças prematuras nascidas com muito baixo peso), embora, neste último, as líquidas em final de sílaba tônica tenham sido adquiridas antes das suas congêneres em sílaba átona. Contrariamente ao esperado, em Baptista (2015), as crianças com historial de otites médias com derrame não foram sensíveis ao efeito desta variável. Esta ausência de efeito do acento de palavra foi observada por Marshall/van der Lely (2009) para crianças com PEDL, mas não para crianças com dislexia, que processaram mais facilmente grupos consonânticos tónicos do que átonos.

Sabemos que a marcação do plural nos nomes emerge cedo nas crianças portuguesas, tendencialmente antes dos 2 anos; esta estrutura ocorre preferencialmente em posição átona, dado o padrão paroxítono não marcado em português (['patuʃ]). A coda fricativa que marca o plural é o primeiro tipo de coda a estabilizar; dado que esta estrutura ocorre preferencialmente em posição átona, observa-se, neste caso, um efeito morfofonológico imune ao efeito do acento de palavra (Inês, 1;9 – *bolos* ['boloʃ] mas *feŕta* ['tete]).

Na avaliação da consciência linguística, Afonso (2008) relatou maior dificuldade na segmentação de trissílabos proparoxítonos (*médico*), por oposição aos paroxítonos (*banana*), tendência também registada em Vicente (2009). Por outro lado, a natureza prosódica das palavras (tónica ou átona) tem-se mostrado relevante na avaliação da consciência de palavra (Cardoso 2011; Afonso/Gonçalves/Freitas 2014), sobretudo em fases iniciais de desenvolvimento deste tipo de consciência.

Os resultados díspares disponíveis na literatura obrigam, assim, à prossecução da tarefa de observação do grau de produtividade da variável *acento de palavra* na avaliação do conhecimento fonológico de crianças com diferentes diagnósticos clínicos.

2.2.3 A sílaba

A sílaba é a unidade prosódica no domínio da qual se organizam os segmentos (↗25 A sílaba na gramática do adulto e na aquisição de língua materna). Das várias unidades fonológicas, esta é a mais testada nos dados do desenvolvimento fonológico típico e atípico. A realidade psicológica dos seus constituintes internos (*ataque*, *rima*, *núcleo*, *coda*) está amplamente documentada para várias línguas, havendo uma forte interação entre a aquisição de um segmento e o seu estatuto silábico (Fikkert 1994; Bernhardt/Stemberger 1998; 2000; Lamprecht et al. 2004). A título ilustrativo, sabemos que as líquidas em português são adquiridas gradualmente em função da sua posição na sílaba:

- (i) a lateral é adquirida primeiramente em ataque simples (*lua*), entre os 3;6 e os 4;0; segue-se-lhe a lateral em ataque ramificado (*bicicleta*), entre os 4;0 e os 4;6, e, finalmente, a lateral

em coda (*calças*), entre os 5;0 e os 5;6 (Mendes et al. 2009; Amorim 2014); a ordem relativa dos dois últimos constituintes pode, no entanto, ser a inversa (Freitas 1997);

- (ii) a vibrante simples é adquirida em ataque simples (*farinha*) entre os 4;0 e os 4;6; segue-se-lhe a vibrante em coda (*porta*), entre os 4;6 e os 5;0, e a vibrante em ataque ramificado (*prato*), entre os 4;6 e os 5;6 (Mendes et al. 2009).

A consideração do estatuto silábico do segmento tem impacto na avaliação do desenvolvimento fonológico, como se ilustra na Tabela 1:

Tabela 1: Avaliação linear e não-linear do desenvolvimento fonológico.

Alvo	Produção	Análise linear	Análise não-linear
<i>luzinha</i>	[lu'sie] Luís, 1;11		/l/ em ataque simples = 100%
<i>bolachas</i>	['lajɛ] Luís, 1;11		
<i>Natal</i>	[e:'taj] Luís, 1;11	/l/ = 50%	/l/ em coda = 0%
<i>bicicleta</i>	[pisi'ɛtɛ] Luís, 1;11		/l/ em ataque ramificado = 0%

A avaliação linear, baseada apenas no segmento, leva à identificação de 50% de sucesso no desempenho da criança. A avaliação não-linear, que considera, neste caso, os constituintes silábicos, permite promover a autoestima da criança e a confiança dos pais através da referência à taxa de sucesso de 100% numa estrutura (neste caso, o /l/ em ataque simples), conduzindo a um diagnóstico mais preciso e a uma programação mais adequada da intervenção: a criança não tem problemas no /l/, em geral, mas apenas no /l/ em ataque ramificado e em coda, pelo que estas devem ser as estruturas a trabalhar na intervenção clínica.

Em contextos de PEDL, crianças de outras línguas têm mostrado sensibilidade à complexidade silábica (Marshall/van der Lely 2007; Ferré/dos Santos 2015), o que constitui, uma vez mais, evidência para a realidade psicológica dos constituintes silábicos. Os dados do português em Lousada (2012) permitem observar um efeito da complexidade silábica em crianças com perturbações fonológicas, sendo os segmentos em coda e em ataque ramificado mais difíceis do que os que surgem no domínio do ataque simples.

O impacto da estrutura silábica na aquisição das unidades segmentais no desenvolvimento típico e atípico fornece, assim, evidência empírica a favor de modelos que assumem a interface *prosódia – segmentos* na representação do conhecimento fonológico e contra os que questionam a adequação desta unidade prosódica à descrição da estrutura sonora das línguas (Kaye/Lowenstamm/Vergnaud 1990).

2.2.4 O segmento e os traços distintivos

Na secção 2.1, referimos e justificámos o destaque dado à unidade *segmento* na formação e na prática clínica dos terapeutas da fala. Neste contexto, o segmento é frequentemente tratado como a unidade mínima de análise, está na base do diagnóstico e funciona como alvo na intervenção, não se fazendo muitas vezes referência às suas propriedades internas.

Na teoria fonológica, os segmentos são entendidos como conjuntos de traços distintivos, que representam as suas propriedades internas. Os traços distintivos permitem captar contrastes entre segmentos:

- (i) /p/ e /b/ contrastam no valor de [vozeado], traço que representa informação relativa à atividade laríngea: /p/ é [-vozeado]; /b/ é [+vozeado];
- (ii) /s/ e /ʃ/ contrastam no valor de [anterior], traço que representa informação relativa ao ponto de articulação: /s/ é coronal [+anterior]; /ʃ/ é coronal [-anterior];
- (iii) /t/ e /s/ contrastam no valor de [contínuo], traço que representa informação relativa ao modo de articulação: /t/ é [-contínuo]; /s/ é [+contínuo].

A título ilustrativo, a produção do alvo /b/ por [m] no sistema da criança (como em *bonecas* [mi'nekeʃ] Inês: 1;9) pode ser interpretada não como um problema com o /p/, especificamente, mas como uma dificuldade decorrente da não estabilização dos contrastes representados pelos traços [±soante] e/ou [±nasal] no conhecimento fonológico da criança. Estes contrastes ocorrem noutros pares de segmentos (/b/ vs. /m/; /d/ vs. /n/; /ʒ/ vs. /ɲ/), potencialmente utilizáveis em contexto de reabilitação terapêutica.

O recurso aos traços distintivos é também útil na avaliação da consciência fonológica: numa prova de identificação do intruso por crianças portuguesas,⁴ Alves (2012) observou que classes naturais representadas por [+contínuo] (fricativas e líquidas) são mais fáceis de identificar do que classes naturais representadas por [-contínuo] (oclusivas nasais e orais).

Vários estudos têm demonstrado que o recurso aos traços distintivos permite descrever a aquisição dos inventários fonológicos das línguas, avaliar com detalhe os desempenhos segmentais de crianças com perturbação fonológica e obter sucesso mais rápido na reabilitação. Tal ilustra a relevância do trabalho de interface entre teoria fonológica e desenvolvimento típico e atípico na testagem e reformulação tanto de modelos fonológicos como de modelos de avaliação e intervenção (Mota 1996; Bernhardt/Stemberger 1998; 2000; Duarte 2006; Lazzarotto-Volcão 2009; Costa 2010; Yamaguchi 2012; Amorim 2014).

⁴ Nas provas de identificação do intruso, são normalmente apresentados três estímulos linguísticos e a criança deve identificar aquele que não contém a estrutura ou a unidade presente nos outros dois estímulos.

2.3 Perspetivas futuras

Os estudos citados nesta secção ilustraram, por um lado, a relevância da evidência empírica proveniente do desenvolvimento típico e atípico para a avaliação da teoria fonológica e, por outro, o papel da teoria fonológica na construção de instrumentos de avaliação e intervenção cada vez mais precisos. Os trabalhos para o português no domínio da linguística clínica são ainda escassos, sendo necessário dar continuidade ao diálogo já existente entre terapeutas da fala e linguistas. Esta interdisciplinaridade, essencial na área de especialidade em foco, permitirá testar hipóteses junto de amostras mais alargadas, identificar semelhanças e dissemelhanças entre sujeitos com comportamentos verbais típicos e atípicos, identificar marcadores clínicos associados a diferentes patologias e produzir instrumentos de avaliação e intervenção específicos para os módulos gramaticais ou os tipos de conhecimento linguístico a trabalhar, auxiliares da atividade clínica dos profissionais na área da terapia da fala.

3 Linguística clínica e sintaxe

3.1 O diagnóstico das perturbações sintáticas

Tradicionalmente, e como referido na secção anterior, a terapia da fala e a própria formação de terapeutas da fala tem um olhar mais atento sobre perturbações de natureza fonológica ou sobre dificuldades de comunicação do que sobre perturbações de natureza sintática. Isto não decorre de um qualquer desprezo pela sintaxe, mas sim da própria natureza desta componente da gramática e da dificuldade específica de identificação de perturbações deste tipo.

A sintaxe é a componente da gramática que assegura a interface entre som e significado. Tem como função a concatenação de elementos e a realização de operações para garantir a legibilidade das estruturas. Não afetando a interação verbal de uma forma tão visível como as outras componentes, torna-se mais difícil identificar uma disfunção na componente sintática. Acresce que um défice exclusivamente sintático, se for muito profundo, se traduz numa dificuldade global na geração de enunciados bem formados, que se pode confundir com uma perturbação global, que deixa de ser identificada como uma perturbação sintática. Se o défice não for profundo, as dificuldades podem não ser evidentes já que se vão manifestar na incapacidade de produção e compreensão de apenas algumas estruturas específicas. Por exemplo, uma criança com uma perturbação sintática pode não entender uma interrogativa como (3a), mas não ter problemas na compreensão de uma interrogativa como (3b):

- (3) a. Que menina é que a mãe viu?
 b. Que menina é que viu a mãe?

Por este motivo, uma perturbação sintática facilmente se confunde com défices de atenção, como enfatizado em vários trabalhos de Naama Friedmann (Friedmann/Grodzinsky 1997; Friedmann/Novogrodsky 2004, entre outros). Na verdade, é facilmente compreensível que os que interagem com pacientes com perturbações sintáticas simplesmente não notem que há uma perturbação linguística: não é saliente, como uma perturbação fonológica, não gera equívocos evidentes, como uma perturbação pragmática. No entanto, as perturbações sintáticas interferem na compreensão de enunciados e na produção de determinadas estruturas.

A baixa saliência das perturbações sintáticas explica que os próprios instrumentos de avaliação geralmente utilizados para detetar perturbações linguísticas tenham uma insuficiente representatividade dos tipos de estruturas sintáticas relevantes para uma caracterização fina das perturbações desta natureza. Conforme estudado em Afonso (2011), os testes utilizados para o português, como será descrito na secção 4, nem sempre selecionam as construções mais adequadas para uma identificação das propriedades específicas associadas a diferentes perturbações.

Nas próximas subsecções, exemplificaremos, a partir do resultado de alguns estudos, como a teoria linguística contribui para uma caracterização mais fina das perturbações e como os comportamentos de populações específicas podem ajudar a um melhor entendimento do funcionamento da língua. Nos exemplos que daremos, tentaremos também mostrar de que forma se cumprem os objetivos da linguística clínica: caracterização e descrição, diagnóstico, avaliação e intervenção.

3.2 Contributos da teoria sintática para uma caracterização das perturbações

Apresentamos, de seguida, quatro exemplos de perturbações linguísticas que podem ser mais bem entendidas com os contributos da teoria sintática: a compreensão de dependências de objeto por crianças com PEDL; a produção de clíticos por crianças com o mesmo tipo de perturbações; o diagnóstico diferencial em contexto de agramatismo; a seletividade da perturbação em crianças surdas com implantes cocleares.

3.2.1 Compreensão de dependências de objeto por crianças com PEDL

É sabido desde os anos 70 que a compreensão e produção de alguns tipos de frases relativas e interrogativas não se encontram estabilizadas até por volta dos seis anos de idade. A investigação das últimas décadas revela que há uma assimetria entre pares como o que ilustrámos em (3), aqui repetido como (4):

- (4) a. Que menina é que a mãe viu?
 b. Que menina é que viu a mãe?

A frase (4a), em que o constituinte movido para o início da frase é o objeto, é adquirida mais tardiamente do que a frase (4b), que não é problemática para as crianças. O mesmo acontece em orações relativas, criando dificuldades as relativas que ocorrem em frases como «Mostra-me a menina que a mãe viu.» (ver Friedmann/Belletti/Rizzi 2009 e, para o português, Cerejeira 2009; Costa/Lobo/Silva 2011).

A teoria linguística introduziu o conceito de Minimalidade Relativizada (Rizzi 1990), que postula que, na criação de uma dependência sintática, um elemento não pode cruzar outro com o qual partilhe determinadas características:

- (5) a. *X_A ... Y_A... t
 b. X_A ... Y_B... t

Esta hipótese teórica revelou-se central na compreensão da maior dificuldade de aquisição de certas estruturas que envolvem movimento de constituintes sintáticos. Em Ferreira (2008) e Fonseca (2011), sobre relativas do português, e em José (2011), sobre interrogativas, mostra-se que as dificuldades evidenciadas pelas crianças com desenvolvimento típico são replicadas por crianças com PEDL, mas em idades mais avançadas. Também em Martins/Santos/Duarte (no prelo) se identificam dificuldades na produção e na compreensão de relativas por parte de dois grupos clínicos, um constituído por crianças com PEDL e outro por crianças com perturbações do espectro do autismo, por comparação com um grupo de sujeitos com desenvolvimento típico. Conforme se tem vindo a concluir para diversas línguas, em contexto de PEDL, as dificuldades na produção e compreensão de relativas e interrogativas de objeto prolongam-se pelo menos até à adolescência, com impacto no rendimento escolar.

De relevância para a nossa exposição é a forma como a análise sintática fina contribui para uma caracterização do que está envolvido na PEDL. Quando analisamos a dificuldade específica na compreensão destas dependências, podemos observar que há diferenças no comportamento entre frases como (6a), (6b) e (6c):

- (6) a. Que menina é que viu a mãe?
 b. Quem é que a mãe viu?
 c. Que menina é que a mãe viu?

Como já foi dito, a primeira frase é mais fácil para as crianças, o que evidencia uma maior facilidade com dependências de sujeito do que com objeto. Isto mostra que a função sintática importa. O que é mais interessante é que, internamente às dependências de objeto, há assimetrias entre frases como (b) e (c), sendo que (b) é mais fácil. A diferença reside na presença de um nome no constituinte interrogativo apenas em (c). É a partilha de traços nominais entre *que menina* e *a mãe* que induz dificuldades, confirmando a hipótese de que o princípio de Minimalidade Relativizada, ilustrado em (5), é crucial para um entendimento claro do que está envolvido em contextos de perturbação. De acordo com a hipótese de Friedmann/Belletti/Rizzi (2009), as crianças nas fases iniciais, e em contextos de perturbação, assumem a minimalidade de forma grosseira, sem ter em conta as cadeias específicas em que este princípio é operativo.

Note-se como, sem o contributo da teoria linguística, se correria o risco de não diagnosticar com precisão, dado que se poderia não entender a relevância da estrutura, da função sintática e dos traços envolvidos.

3.2.2 O papel da estrutura sintática no agramatismo

Como já foi referido, os pacientes com afasia de Broca têm dificuldades na produção de estruturas sintaticamente bem formadas. Para que haja uma intervenção adequada, é necessário entender-se exatamente a natureza destas dificuldades.

Friedmann/Grodzinsky (1997) observaram, num estudo de pacientes falantes de hebraico, que as dificuldades na produção de estruturas não eram arbitrárias e que poderiam ser entendidas à luz do conhecimento sobre a hierarquia da estrutura sintática. Descritivamente, registaram que os agramáticos poderiam ter três tipos de comportamento:

- (i) dificuldades com a produção de subordinação e de interrogativas, associada a dificuldades na produção de morfologia de tempo e dificuldades na produção de morfologia de concordância;
- (ii) os mesmos comportamentos sem dificuldades na morfologia de concordância;
- (iii) apenas dificuldades na produção de subordinação e de interrogativas.

De acordo com aqueles autores, os três tipos de comportamento entendem-se quando a estrutura sintática é tida em conta. Assumindo-se uma hierarquia do domínio funcional como em (7) e estabelecendo-se as associações entre fenómenos e categorias funcionais em (8), os dados podem ser explicados:

(7) [CP C [TP T [AgrF Agr

- (8) C – complementadores, domínio para onde são movidos constituintes interrogativos
 T – morfologia flexional de tempo
 Agr – morfologia flexional de concordância (*Agr*, do inglês *agreement*)

De acordo com Friedmann/Grodzinsky (1997), o agramatismo pode ser representado por cortes seletivos na estrutura sintática. Se houver um corte abaixo da categoria Agr, todas as categorias em (8) são afetadas, verificando-se o comportamento descrito em (i). Se o corte se der apenas acima de Agr e abaixo de T, o comportamento em (ii) é gerado, sendo poupada a morfologia de concordância; se o corte se der acima de T, apenas os fenómenos relacionados com C são afetados, derivando-se o comportamento descrito em (iii). O poder preditivo e explicativo desta hipótese é interessante e abre janelas para a intervenção, uma vez que permite supor que uma abordagem *top-down* na intervenção pode ser mais produtiva, uma vez que o estímulo da categoria mais alta (CP) poderá desencadear a ativação das posições mais baixas da estrutura.

Em Cerdeira (2006) e Cerdeira/Costa/Trindade (2007), estudou-se o comportamento de agramáticos portugueses e descobriu-se um comportamento parcialmente distinto. Os pacientes estudados nunca evidenciaram uma dissociação entre morfologia de tempo e concordância. Quando uma das categorias flexionais estava afetada, a outra também estava. Este comportamento foi explicado tendo em conta a natureza específica da morfologia flexional verbal do português, que funciona de forma amalgamada em vários contextos. Por exemplo, numa forma como *cantaste*, o morfema *-ste* marca simultaneamente segunda pessoa do singular e o valor temporal-aspectual de pretérito perfeito. Para a explicação do comportamento dos pacientes portugueses, foi preciso adotar a hipótese de cortes na estrutura funcional da frase em articulação com uma descrição mais fina do comportamento específico da morfologia flexional do português.

Este caso é interessante por mostrar não apenas o poder explicativo da teoria linguística na caracterização da perturbação, mas também o papel da descrição da variação interlinguística para uma avaliação clara da forma como as perturbações se manifestam em diferentes línguas.

3.2.3 Clíticos e objetos nulos na Perturbação Específica do Desenvolvimento da Linguagem (PEDL)

O papel da variação interlinguística pode também ser relevante na construção de instrumentos de avaliação. Não raramente os instrumentos de avaliação são simplesmente adaptados ou traduzidos de umas línguas para outras sem haver o cuidado de se ter em conta as propriedades específicas dos sistemas linguísticos em que vão ser aplicados.

Os pronomes clíticos têm vindo a ser identificados como um marcador robusto de PEDL em várias línguas do mundo. Um dos casos mais estudados é o francês, língua em que se tem observado que as crianças com PEDL omitem clíticos acusativos de terceira pessoa até bastante tarde (Jakubowicz et al. 1998, entre outros). Este comportamento contrasta com o comportamento de crianças com desenvolvimento típico, que exibem omissão de clíticos em algumas línguas, mas que, de forma uniforme, aos cinco anos de idade, têm taxas de produção de clíticos em contexto obrigatório próximas dos 100%, conforme demonstrado em Varlokosta et al. (2015) para um vasto conjunto de línguas. Assim, se uma criança desta idade exhibe problemas na produção de clíticos acusativos, podemos assumir com alguma segurança que estamos perante um caso de perturbação linguística.

É importante notar que o português constitui um caso diferente. Nesta língua, os pronomes clíticos estão em variação com a opção por objetos nulos. No mesmo contexto em que um pronome acusativo é produzido, é legítimo não realizar o objeto direto (Raposo 1986), conforme ilustrado no exemplo seguinte:

- (9) A: Sabes das minhas chaves?
 B: a. Pu-las na gaveta.
 b. Pus \emptyset na gaveta.

Conforme argumentado em Costa/Lobo (2006), este facto da gramática do adulto torna a omissão de pronomes pelas crianças suscetível de duas interpretações: ou as crianças omitem pronomes, tal como acontece noutras línguas, ou produzem objetos nulos como os adultos (conforme (9-Bb)). Na verdade, em estudos posteriores (Costa/Lobo 2009; Costa/Lobo/Silva 2012), demonstra-se que as crianças portuguesas omitem clíticos, sobregeneralizando a construção de objeto nulo. De acordo com estas conclusões, é expectável que a omissão de pronomes não seja comparável à que é encontrada noutras línguas. Com efeito, em Varlokosta et al. (2015), mostra-se que o português é a única língua em que são encontradas taxas altas de omissão de pronomes aos 5 anos, mesmo em contextos em que a omissão não é legítima na gramática do adulto (contextos de ilha sintática).

Estes dados são importantes já que significam que uma mera tradução de um teste de avaliação sintática, que inclua a produção de pronomes, pode ser inconclusiva para o português, pois pode sinalizar como problemática a omissão de pronomes, quando é não só parte da gramática alvo, como uma característica tardia no desenvolvimento típico. Este é, portanto, um caso em que a teoria linguística e os estudos sobre desenvolvimento típico da linguagem contribuem para uma mais eficiente construção de instrumentos de avaliação especificamente pensados para uma língua particular.

3.2.4 Seletividade no movimento na surdez

Consideremos agora um outro contexto de perturbação sintática: o que podemos encontrar em casos de surdez em que os falantes só recebem *input* linguístico tardiamente (por não serem falantes de língua gestual), quando recebem implantes cocleares. Alguns estudos têm mostrado que a exposição tardia ao *input* pode comprometer o desenvolvimento sintático, em particular em construções que envolvem movimento. O caso da surdez é particularmente interessante, já que permite olhar para o movimento de forma bastante seletiva. Sabemos que o movimento sintático pode servir fins diferenciados e relacionar-se com diferentes «áreas» da gramática. Por exemplo, o movimento de constituintes interrogativos para o início da frase permite marcar uma frase como interrogativa (Cheng 1991), enquanto o movimento de um verbo para a categoria funcional T se relaciona com a associação da raiz verbal à morfologia flexional.

A existência de défices sintáticos em crianças com perda auditiva que foram treinadas oralmente, mesmo depois de receberem implantes cocleares, está há muito estabelecida na literatura (Brannon 1968, entre outros). Muitos estudos desenvolvidos

nas décadas de 70 e 80 identificaram dificuldades globais gerais na produção e compreensão de estruturas sintáticas, por comparação com crianças com desenvolvimento típico. Contudo, apenas recentemente se iniciou uma caracterização mais precisa das construções afetadas em contextos de défice auditivo. Várias construções que envolvem movimento sintático têm vindo a ser identificadas como problemáticas para crianças com défice auditivo: passivas (Power/Quigley 1973), interrogativas (de Villiers/de Villiers/Hoban 1994) e relativas (Quigley/Smith/Wilbur 1974).

Friedmann/Szterman (2006) propõem que o movimento sintático não está globalmente comprometido no défice auditivo, isto é, a exposição tardia ao *input* compromete a capacidade de estabelecer cadeias por movimento, mas apenas de alguns tipos. A hipótese de que há problemas com movimento é sustentada pela observação de que, em hebraico, é possível criar dependências sintáticas através de estruturas alternativas que não envolvem movimento, como as estratégias resuntivas (ver também Friedmann/Costa 2011). Friedmann/Szterman (2006) mostram ainda que a data de implantação do implante coclear é crítica para que o acesso ao *input* sintático permita a aquisição de movimento sintático. Se a implantação se der depois do primeiro ano de vida, o défice pode ser irreparável.

Em Costa/Moita (2015), a seletividade do défice no movimento sintático na surdez é explorada para o português. Partindo da hipótese de que se espera que o movimento mais diretamente envolvido na interface da sintaxe com a componente morfofonológica da gramática esteja mais comprometido, estuda-se a forma como as crianças respondem a interrogativas totais em português. Como se sabe, no português a resposta afirmativa a uma interrogativa total pode ser verbal ou *sim*, conforme ilustrado em (10) (Martins 1994; ↗22 O sistema responsivo: padrões de resposta a interrogativas polares e a asserções):

- (10) A: Vais à praia?
B: Vou. / Sim.

Santos (2006) mostra que as crianças com desenvolvimento típico usam, tal como os adultos, predominantemente respostas verbais. Estas respostas instanciam casos de elipse do sintagma verbal, que envolvem movimento do verbo para T. Assim, o facto de as crianças com desenvolvimento típico exibirem este tipo de respostas constitui evidência para se dizer que dominam o movimento do verbo para T desde cedo. No caso das crianças com implante coclear, estudou-se a forma como dão respostas afirmativas a interrogativas totais. Costa/Moita (2015) mostram que estas crianças, ao contrário das crianças com desenvolvimento típico, usam maioritariamente resposta *sim*, não recorrendo a respostas verbais. Uma forma de interpretar esta preferência por *sim* é colocar a hipótese de que estas crianças não têm acesso a movimento do verbo para T.

Este caso é importante porque exemplifica como um instrumento da teoria linguística – o movimento sintático – permite descrever um comportamento específico numa perturbação. Crucialmente, nesta situação, nem todo o movimento está

afetado, apenas o movimento de núcleos, o que permite tornar mais fina a caracterização da perturbação. Finalmente, percebemos que, neste caso específico, há uma perturbação do movimento, enquanto noutras perturbações, como a PEDL, se registaram outros tipos de marcadores (como os problemas de Minimalidade, que afetam a relação entre constituintes movidos e não o movimento *per se*), o que permite entender que os instrumentos teóricos são úteis no estabelecimento de diagnósticos diferenciais.

3.3 Contributos das perturbações para a teoria sintática

Conforme enunciámos na introdução, a linguística clínica tem um impacto óbvio no estudo das perturbações linguísticas, mas a comunicação estabelece-se nos dois sentidos, já que as características das perturbações podem constituir evidência para a testagem de hipóteses teóricas.

No domínio da sintaxe, exemplificamos este contributo possível com um estudo de Lobo e Vaz sobre a aquisição da exaustividade (Lobo/Vaz 2012). Nos trabalhos de Schulz/Roeper (2011), entre outros, mostra-se que há interrogativas que requerem respostas exaustivas. Por exemplo, perante uma pergunta como (11), é possível obter uma resposta satisfatória não exaustiva:

- (11) A: Quem tem uma caneta?
 B: Eu tenho...
 (não se requer que todos os participantes no contexto com uma caneta se identifiquem)

Já as interrogativas múltiplas, com mais do que um constituinte interrogativo, induzem resposta com listas de pares exaustivas (todos os pares possíveis têm de ser dados na resposta):

- (12) A: Quem comeu o quê?
 B: O João comeu uma banana, o Pedro comeu uma maçã e a Ana comeu uma pera.

Schulz/Roeper (2011) mostram que as crianças dão respostas não exaustivas até por volta dos 5 anos em contextos exaustivos, o que é corroborado para o português em Vaz (2012) e Lobo/Vaz (2012). Este comportamento pode derivar de um de dois motivos: ou as crianças não dominam as condições discursivas associadas a este tipo específico de interrogativa ou a exaustividade é um traço codificado sintaticamente e tem de ser adquirido.

Em Costa/Lobo/Vaz (2014), testa-se exploratoriamente o comportamento de crianças com PEDL na resposta a interrogativas que envolvem contextos exaustivos. São testadas crianças com diagnóstico de PEDL sintática e crianças com diagnóstico de PEDL semântico-pragmática. O resultado deste estudo exploratório é muito interessante, na medida em que se verificou que as crianças com PEDL sintática tinham dificuldades com as respostas exaustivas, enquanto as crianças com PEDL semântico-

-pragmática não apresentaram qualquer tipo de dificuldade. Este padrão de comportamentos sugere, pois, que a exaustividade é um traço sintático e não uma condição pragmático-discursiva. Crucialmente, pôde-se chegar a esta conclusão através do estudo do comportamento da população com perturbação. Este é, portanto, um caso em que o estudo de uma perturbação linguística contribui para o esclarecimento de uma questão na própria teoria sintática: o *locus* da codificação da exaustividade.

3.4 Perspetivas futuras

Os exemplos dados visaram atestar os dois sentidos em que se estabelece a comunicação entre teoria sintática e estudos sobre perturbação. Os trabalhos para o português são ainda bastante incipientes nesta área e importa aprofundar os estudos existentes, alargar amostras para tornar mais conclusivos muitos dos estudos exploratórios realizados. Importa, sobretudo, conduzir estudos interpopulacionais, que permitam estabelecer diferenças entre o desenvolvimento típico, o comportamento adulto, a aquisição de língua segunda e estrangeira e o que se passa em patologias diferenciadas. Só estes estudos comparados permitirão identificar o que é específico das perturbações e contribuir para diagnósticos específicos cada vez mais finos.

4 Instrumentos de avaliação da linguagem para o português

Como se enunciou na secção 1, um dos objetivos da linguística clínica é a avaliação do desenvolvimento linguístico de forma a tornar possível o diagnóstico e, assim, definir caminhos para a intervenção. Para que tal avaliação possa ser levada a cabo, é necessário criar instrumentos finos e adequados que permitam identificar o perfil de desenvolvimento linguístico das crianças, sabendo-se, porém, que há que ter em conta alguma variabilidade nos desempenhos das mesmas, que é expectável para a definição de parâmetros normativos. É na construção de tais instrumentos de avaliação que a relação entre teoria linguística e linguística clínica se revela de maior importância. Com efeito, como se afirmou anteriormente, só com um bom conhecimento da língua se constroem instrumentos de avaliação precisos, no sentido de serem linguisticamente controlados, e só com um conhecimento do que está afetado em cada contexto de perturbação é possível construir instrumentos de avaliação apropriados.

Na prática clínica, têm sido usadas diferentes metodologias para a recolha de dados necessária para avaliação do desenvolvimento linguístico, no que diz respeito quer à produção quer à compreensão. Uma das técnicas que têm sido utilizadas, no contexto da terapia da fala, para a avaliação da produção é a análise de discurso espontâneo. Trata-se de um método de observação direta, que pode ser aplicado a

qualquer criança, independentemente da sua idade, das suas capacidades linguísticas e das suas capacidades cognitivas (Martins/Vieira em prep.). Adicionalmente, os dados de discurso espontâneo, pela sua riqueza, podem ser analisados sob diferentes perspetivas, permitindo a avaliação de fenómenos linguísticos diversos (Eisenbeiss 2009), associados a diferentes áreas (fonética, fonologia, sintaxe, morfologia, semântica, pragmática). Como assinalam Martins/Vieira (em prep.), o facto de a interferência do clínico/investigador ser mínima nas recolhas de discurso espontâneo torna improvável que a criança desenvolva estratégias de resposta, o que torna os dados altamente fiáveis.

No entanto, a recolha de discurso espontâneo apresenta algumas desvantagens quando usada em contexto clínico: trata-se de uma técnica morosa e difícil de avaliar com base num sistema de cotações, o que pode conduzir a diagnósticos distintos em função do avaliador; não oferece parâmetros de comparação, cruciais quando se pretende um diagnóstico preciso; não permite prever quais as estruturas que a criança irá produzir, pelo que a ausência de uma estrutura pode decorrer de fatores de natureza diversa, não sendo possível saber se a criança não a produz porque não a conhece ou porque o contexto comunicativo não motivou a sua produção (McDaniel/McKee/Cairns 1998; Valian/Aubry 2005; Vieira 2012).

Devido às dificuldades da avaliação do discurso espontâneo, os terapeutas da fala recorrem frequentemente à análise de discurso semiespontâneo, como a descrição de uma figura ou a narração de uma história baseada em figuras (veja-se, para o português, Ramalho/Almeida/Freitas 2014). Tal técnica, podendo complementar a recolha de discurso espontâneo, não deverá, no entanto, constituir o único instrumento de avaliação. Com efeito, a descrição de uma figura pode conduzir a um discurso demasiado sintético ou até a uma mera listagem de nomes, uma vez que se trata de uma tarefa difícil sobretudo para crianças mais novas (Kerr 1993). Por sua vez, contar uma história, mesmo que baseada em figuras, constitui uma tarefa que recruta capacidades cognitivas e linguísticas complexas, como o conhecimento geral de situações sociais, a memória de situações específicas, o conhecimento do contexto comunicativo e das regras que regulam a interação linguística e, obviamente, o conhecimento das estruturas linguísticas (Vieira 2012). A criança tem, pois, de dominar toda esta informação, tarefa que se afigura difícil para as crianças mais novas (entre outros, Applebee 1978; Pinto 1987).

Tendo em conta os desafios que a análise de discurso espontâneo e semiespontâneo pode colocar na avaliação do desenvolvimento linguístico, sobretudo no que diz respeito à elaboração de diagnósticos precisos nos casos de crianças com perturbações específicas de linguagem, têm sido usadas outras técnicas de recolha de dados mais estruturadas que permitem elicitar estruturas e unidades linguísticas de vários módulos gramaticais a partir de estímulos linguisticamente controlados. Tais estímulos têm como objetivo motivar a criança a produzir/interpretar determinadas estruturas através de tarefas de imitação exata, imitação com variação, completamento, juízos de gramaticalidade, ou, no caso dos testes de avaliação da compreensão,

tarefas de julgamento de valor de verdade. Estas técnicas são usadas na construção de instrumentos de avaliação diversificados, como as escalas de avaliação formais e os testes estandardizados, entre outros.

No contexto nacional, encontram-se disponíveis no mercado três testes *estandardizados* que são vulgarmente utilizados pelos terapeutas da fala para a avaliação do desenvolvimento linguístico: o *ALO-Avaliação da Linguagem Oral* (Sim-Sim 2001, 1997), o *TICL-Teste de Identificação das Competências Linguísticas* (Viana 2004) e o *TALC-Teste de Avaliação da Linguagem nas Crianças* (Sua-Kay/Tavares 2007).

Os testes referidos permitem uma avaliação global da linguagem de crianças em idade pré-escolar (*ALO*, *TICL* e *TALC*) e escolar (*ALO*) e são constituídos por um número variável de itens que incidem sobre o conhecimento associado a vários módulos gramaticais ou sobre a consciência linguística de dadas estruturas. Constituindo uma ferramenta importante para uma primeira avaliação do desenvolvimento linguístico das crianças, devem, no entanto, ser complementados por outros instrumentos, que avaliem, de forma mais detalhada, cada uma das áreas linguísticas que, numa primeira abordagem, podem até não parecer afetadas. Com efeito, é necessário dispor, por um lado, de um maior número de itens que avaliem os desempenhos das crianças em áreas mais específicas, o que não é possível no caso de testes que têm uma maior abrangência, e, por outro lado, de itens que permitam diagnosticar crianças com PEDL. Neste último caso, em particular, deverão ser tidos em conta, tanto quanto possível, os marcadores clínicos de PEDL, de forma a incluir, nos testes de avaliação, as construções que poderão ser indicadoras da perturbação linguística, como defendido em Afonso (2011). Torna-se, assim, crucial a relação entre teoria linguística e linguística clínica para a construção de instrumentos orientados para componentes gramaticais específicas e para populações específicas.

Existem já alguns testes direcionados para aspetos específicos do comportamento verbal em crianças portuguesas: Guimarães et al. (2014) (*TAV – Teste de Articulação Verbal*) tem o seu foco na avaliação articulatória; Mendes et al. (2009) (*TFF-ALPE – Teste Fonético-Fonológico; Avaliação da linguagem Pré-Escolar*) avalia o desenvolvimento fonológico; Sua-Kay/Tavares/Santos (2015) (*TAS – Teste de Avaliação Semântica*) centra-se no domínio da semântica lexical.

Nas secções seguintes, apresentam-se, a título exemplificativo, dois testes elaborados para o português, que avaliam desenvolvimento sintático (Vieira 2012) e desenvolvimento fonológico (Ramalho/Almeida/Freitas 2014).

4.1 O Teste Pré-Escolar de Produção Sintática / *Schlichting Test for Sentence Development-PT*

O *Teste Pré-Escolar de Produção Sintática (STSD-PT)*, elaborado no âmbito de uma dissertação de doutoramento (Vieira 2012), pretende avaliar o desempenho sintático de crianças entre os 3 e os 6 anos de idade com um normal desenvolvimento da

linguagem, no sentido de estabelecer valores normativos para esta população, o que permite proceder a comparações que levem ao diagnóstico precoce de perturbações que afetam a componente sintática.

O teste é baseado num subteste de produção sintática holandês (Schlichting et al. 1995), incluindo, como a versão holandesa, 40 itens, organizados por grau crescente de dificuldade. Relativamente à versão original, foram mantidas, com as devidas adaptações, 24 estruturas e introduzidas 16 novas estruturas. Para a seleção das estruturas sintáticas que figuram no teste, foram tidos em conta vários estudos sobre aquisição da sintaxe, com particular ênfase no português, a fim de garantir que os itens eram adequados à faixa etária das crianças a quem o teste se dirige. O teste foi estandardizado, sendo a amostra constituída por 847 crianças provenientes de quarenta e dois jardins de infância dos dezoito distritos de Portugal Continental.

Os itens de teste permitem avaliar o desenvolvimento linguístico relativamente a diversos aspetos sintáticos: (i) frases complexas por subordinação (relativas, adverbiais e completivas) e por coordenação (copulativa e adversativa); (ii) estruturas de topicalização e de clivagem; (iii) concordância internamente ao sintagma nominal e entre os constituintes maiores da frase; (iv) frases passivas verbais e adjetivais; (v) predicção secundária; (vi) estrutura interna do sintagma nominal (especificadores, modificadores; clíticos e elementos nominais), do sintagma verbal (complementos e modificadores) e do sintagma adjetival; (vii) classes de verbos, estruturas argumentais e categoria dos complementos. O teste utiliza quatro técnicas de eliciação de dados: imitação exata, imitação com variação, completamento de frases e resposta provocada a uma questão.

Os resultados psicométricos apontam para um teste válido e fiável para uso clínico, tratando-se, assim, de um teste normativo, na medida em que constitui uma referência a partir da qual é possível avaliar o desempenho sintático da criança.

4.2 *Crosslinguistic Child Phonology Project – European Portuguese (CLCP-PE)*

O *Crosslinguistic Child Phonology Project (CLCP)* é liderado por May Bernhardt e Joseph Stemberger, da University of British Columbia, Canadá,⁵ sendo seu objetivo central o de recolher dados de produção fonológica (típica e atípica) em várias línguas do mundo, de forma sistematizada e linguisticamente controlada, com vista ao desenvolvimento de estudos comparados no domínio da aquisição da fonologia. O instrumento de avaliação usado no projeto para aceder ao conhecimento fonológico foi adaptado ao português (CLCP-PE).

⁵ Projeto financiado pelo *Conseil de Recherches en Sciences Humaines du Canada* (#410-2009-0348), <http://phonodevelopment.sites.olt.ubc.ca/>.

O CLCP-PE (Ramalho/Almeida/Freitas 2014) foi concebido assumindo os princípios da fonologia não-linear adaptados à prática clínica (Bernhardt/Stemberger 2000). Inclui o controlo de todas as unidades do inventário segmental do PE e sua interação com os constituintes prosódicos. Cada segmento é testado mais do que uma vez em várias posições prosódicas na língua-alvo, tendo em conta as variáveis: (i) constituição silábica (ataque, rima, núcleo, coda); (ii) acento de palavra (contraste tónico / átono); (iii) posição na palavra (inicial, medial, final); (iv) extensão de palavra (monossílabos, dissílabos, trissílabos e polissílabos).

O instrumento pode ser aplicado como teste de nomeação de imagens e/ou como teste de recolha de fala semiespontânea, em contexto de reconto. A nomeação ou o reconto são realizados a partir do relato da *História do Coelho Cenourinha*, que vive com uma família de humanos. O formato foi já testado nas várias línguas integradas no projeto e revelou-se apelativo para as crianças das várias nacionalidades. A versão atual do CLCP-PE inclui 157 palavras, agrupadas em 44 cenários, podendo ser apresentada em formato digital ou em papel (Ramalho/Almeida/Freitas 2015). A validação deste instrumento está a ser feita no âmbito de uma tese de doutoramento (Ramalho em prep.). Após a validação, o instrumento será disponibilizado em diferentes versões: (i) uma versão breve, com 100 palavras, e uma versão de *screening*, com 50 palavras; (ii) versões por faixa etária, com as palavras com melhores frequências de nomeação (léxico mais adequado) em cada grupo etário testado (3;0–4;0, 4;1–5;0 e 5;1–6;0).

5 Conclusão

Conforme enunciámos no início deste capítulo, a linguística clínica oferece-nos uma estrada de dois sentidos: os contextos de perturbação linguística são janelas para um melhor entendimento do funcionamento das línguas naturais e, por outro lado, a teoria linguística dá-nos instrumentos claros para uma melhor caracterização e intervenção nestes domínios de perturbação.

Há, contudo, ainda um longo caminho a percorrer e só o trabalho interdisciplinar, que fomente o diálogo entre linguistas, educadores e terapeutas da fala, poderá preencher as muitas lacunas ainda existentes para que esta área de investigação, ainda incipiente, continue a desenvolver-se em contexto nacional.

6 Referências

- Afonso, Alexandra (2011), *Análise de Itens Sintáticos em Provas de Avaliação da Linguagem. Relevância para Identificação de Perturbação Específica do Desenvolvimento da Linguagem*, dissertação de Mestrado, FCSH/Universidade Nova de Lisboa e Instituto Politécnico de Setúbal.

- Afonso, Catarina (2008), *Complexidade Prosódica e Segmentação de Palavras em Crianças entre os 4 e os 6 Anos de Idade*, dissertação de Mestrado, Lisboa, Universidade Católica Portuguesa.
- Afonso, Catarina/Gonçalves, Anabela/Freitas, Maria João (2014), «A princesa ficou *adormir ou a dormir?» *Dados sobre a consciência da unidade palavra em português europeu*, *Linguística – Revista de Estudos Linguísticos da Universidade do Porto* 9, 35–58.
- Alves, Dina (2012), *Efeito das Propriedades Segmentais em Tarefas de Consciência Segmental, de Leitura e de Escrita*, tese de Doutoramento, Universidade de Lisboa.
- Alves, Dina/Castro, Ana/Correia, Susana (2010), *Consciência fonológica – dados sobre consciência fonémica, intrassilábica e silábica*, in: Ana Maria Brito et al. (orgs.), *Textos Seleccionados do XXV Encontro Nacional da Associação Portuguesa de Linguística*, Porto, APL, 169–184.
- American Psychiatric Association (2013), *Diagnostic and Statistical Manual of Mental Disorders – DSM-5 (Fifth)*, Arlington, VA, American Psychiatric Association.
- Amorim, Clara (2014), *Padrão de Aquisição de Contrastes do PE: a interação entre traços, segmentos e sílabas*, tese de Doutoramento, Universidade do Porto.
- Applebee, Arthur N. (1978), *The Child's Concept of Story: Ages Two to Seventeen*, Chicago, University of Chicago Press.
- Baptista, Ana Catarina (2015), *O Desenvolvimento Fonológico de Crianças com Otites Médias Serosas*, tese de Doutoramento, Universidade de Lisboa.
- Bernhardt, May/Stemberger, Joseph (1998), *Handbook of Phonological Development from the Perspective of Constraint-Based Nonlinear Phonology*, San Diego, Academic Press.
- Bernhardt, May/Stemberger, Joseph (2000), *Workbook in Nonlinear Phonology for Clinical Application*, Austin, Texas, Pro-Ed.
- Brannon, John B., Jr. (1968), *Linguistic word classes in the spoken language of normal, hard-of-hearing, and deaf children*, *Journal of Speech and Hearing Research* 11, 279–287.
- Cardoso, Sónia (2011), *Consciência de Palavra em Crianças de Idade Pré-Escolar e Escolar*, dissertação de Mestrado, Instituto Politécnico de Setúbal.
- Cerdeira, Ana Margarida (2006), *Flexão Verbal e Categorias Funcionais no Agramatismo*, dissertação de Mestrado, FCSH/Universidade Nova de Lisboa.
- Cerdeira, Ana/Costa, João/Trindade, Vanessa (2007), *Movimento do verbo em português europeu: consequências para diferenças entre línguas em contextos de aquisição e afasia*, in: Maria Lobo/Maria Antónia Coutinho (edd.), *Textos Seleccionados do XXII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 239–250.
- Cerejeira, Joana (2009), *Aquisição de Interrogativas de Sujeito e de Objecto em Português Europeu*, dissertação de Mestrado, FCSH/Universidade Nova de Lisboa.
- Cheng, Lisa (1991), *On the Typology of Wh-Questions*, PhD dissertation, Cambridge MA, MIT.
- Chomsky, Noam (1957), *Syntactic Structures*, The Hague, Mouton.
- Chomsky, Noam (1975), *Reflections on Language*, New York, Pantheon Books.
- Chomsky, Noam (2004), *Beyond explanatory adequacy*, in: Adriana Belletti (ed.), *Structures and Beyond*, Oxford, Oxford University Press, 104–131.
- Chomsky, Noam/Hauser, Marc D./Fitch, W. Tecumseh (2002), *The faculty of language: What is it, who has it, and how did it evolve?*, *Science* 298, 1569–1579.
- Costa, João/Lobo, Maria (2006), *A aquisição de clíticos em PE: Omissão de Clíticos ou Objectos Nulos?*, in: Fátima Oliveira/Joaquim Barbosa (edd.), *Textos Seleccionados do XXI Encontro Nacional da APL*, Lisboa, APL, 285–293.
- Costa, João/Lobo, Maria (2009), *Acquisition of clitics in European Portuguese: Data from comprehension*, in: Acrisio Pires/Jason Rothman (edd.), *Minimalist Inquiries into Child and Adult Language Acquisition: Case Studies across Portuguese*, Berlin/Boston, De Gruyter, 63–84.
- Costa, João/Lobo, Maria/Silva, Carolina (2011), *Subject-object asymmetries in the acquisition of Portuguese relative clauses: adults vs. children*, *Lingua* 121:6, 987–1158.

- Costa, João/Lobo, Maria/Silva, Carolina (2012), *Which category replaces an omitted clitic? The case of European Portuguese*, in: Pedro Guijarro-Fuentes/María Pilar Larrañaga (edd.), *Pronouns and Clitics in Early Acquisition*, Berlin/Boston, De Gruyter, 105–130.
- Costa, João/Lobo, Maria/Vaz, Stéphanie Dias (2014), *The development of dependencies across populations*, comunicação apresentada no *IV LLDI Workshop*, Faculdade de Letras da Universidade de Lisboa, 31 de janeiro–1 de fevereiro 2014.
- Costa, João/Moita, Mara (2015), *Movimento X⁰ em crianças com implante coclear: evidência para défice sintático em contexto de input tardio*, comunicação apresentada no *XXXI Encontro Nacional da Associação Portuguesa de Linguística*, Universidade do Minho, outubro de 2015.
- Costa, Teresa (2010), *The Acquisition of the Consonantal System in European Portuguese: Focus on Place and Manner Features*, tese de Doutoramento, Universidade de Lisboa.
- Crystal, David (1981), *Clinical Linguistics*, Vienna, Springer.
- de Lacy, Paul (2007), *Handbook of Phonological Theory*, Cambridge MA, Cambridge University Press.
- de Villiers, Jill G./de Villiers, Peter A./Hoban, Esme (1994), *The central problem of functional categories in the English syntax of deaf children*, in: Helen Tager-Flusberg (ed.), *Constraints on Language Acquisition: Studies of Atypical Children*, Hillsdale, NJ, Erlbaum, 9–47.
- Duarte, Sabrina (2006), *Relações de Distância e de Complexidade entre Traços Distintivos na Generalização em Terapia de Desvios Fonológicos*, dissertação de Mestrado, Universidade Católica de Pelotas.
- Eisenbeiss, Sonja (2009), *Contrast is the Name of the Game: Contrast-Based Semi-Structured Elicitation Techniques for Studies on Children's Language Acquisition*, Essex Research Reports in Linguistics 57:7, 1–27, <http://repository.essex.ac.uk/id/eprint/95> (11.02.2016).
- Ferré, Sandrine/dos Santos, Christophe (2015), *Comment évaluer la phonologie des enfants bilingues?*, LIDIL 51– Acquisition du Langage et Phonologie: Méthodes, Objets, Enjeux théoriques, 11–34.
- Ferreira, Elisa (2008), *Compreensão e Produção de Frases Relativas por Crianças com Perturbação Específica do Desenvolvimento da Linguagem e por Adultos com Agramatismo*, dissertação de Mestrado, FCSH/Universidade Nova de Lisboa.
- Fikkert, Paula (1994), *On the Acquisition of Prosodic Structure*, Ph.D. dissertation, HIL dissertations 6, Leiden University – The Hague, Holland Academic Graphics.
- Fonseca, Ana Rita (2011), *Produção e Compreensão de Frases Relativas por Crianças com Perturbação Específica do Desenvolvimento da Linguagem*, dissertação de Mestrado, FCSH/Universidade Nova de Lisboa e Instituto Politécnico de Setúbal.
- Freitas, Maria João (1997), *Aquisição da Estrutura Silábica do Português Europeu*, tese de Doutoramento, Universidade de Lisboa.
- Friedmann, Naama/Belletti, Adriana/Rizzi, Luigi (2009), *Relativized relatives: Types of intervention in the acquisition of A-bar dependencies*, *Lingua* 119, 67–88, doi: 10.1016/j.lingua.2008.09.002.
- Friedmann, Naama/Costa, João (2011), *Last resort or no resort: Resumptive pronouns in Hebrew and Palestinian Arabic hearing impairment*, in: Alain Rouveret (ed.), *Resumptive Pronouns at the Interfaces*, Amsterdam/Philadelphia, Benjamins, 223–240.
- Friedmann, Naama/Grodzinsky, Yosef (1997), *Tense and Agreement in agrammatic production: Pruning the syntactic tree*, *Brain Language* 56, 397–425.
- Friedmann, Naama/Novogrodsky, Rama (2004), *The acquisition of relative clause comprehension in Hebrew: a study of SLI and normal development*, *Journal of Child Language* 31, 661–681.
- Friedmann, Naama/Szterman, Ronit (2006), *Syntactic movement in orally-trained children with hearing impairment*, *Journal of Deaf Studies and Deaf Education* 11, 56–75.
- Goldsmith, John/Riggle, Jason/Yu, Alan (2011), *The Handbook of Phonological Theory*, New Jersey, Wiley-Blackwell.
- Grunwell, Pamela (1982), *Clinical Phonology*, London, Croom Helm.

- Guimarães, Isabel, et al. (2014), *TAV – Teste de Articulação Verbal*, Lisboa, Oficina Didáctica.
- Ingram, David (1976), *Phonological Disability in Children*, London, Arnold.
- Jakubowicz, Celia, et al. (1998), *Determiners and Clitic Pronouns in French-Speaking Children with SLI*, *Language Acquisition* 7, 113–160.
- José, Carla (2011), *A Produção e Compreensão de Interrogativas por Crianças com Perturbação Específica do Desenvolvimento da Linguagem*, dissertação de Mestrado, FCSH/Universidade Nova de Lisboa e Instituto Politécnico de Setúbal.
- Kaye, Jonathan/Lowenstamm, Jean/Vergnaud, Jean-Roger (1990), *Constituent structure and government in phonology*, *Phonology* 7, 193–231.
- Kerr, Jean (1993), *Assessment of acquired language problems*, in: John R. Beech/Leonora Harding/ /Diana Hilton-Jones (edd.), *Assessment in speech and language therapy*, New York, Routledge, 99–127.
- Lamprecht, Regina, et al. (2004), *Aquisição Fonológica do Português. Perfil de Desenvolvimento e Subsídio para Terapia*, Porto Alegre, Artmed.
- Lazzarotto-Volcão, Cristiane (2009), *Modelo Padrão de Aquisição de Contrastes: uma Proposta de Avaliação e Classificação dos Desvios Fonológicos*, tese de Doutorado, Universidade Católica de Pelotas.
- Lobo, Maria/Vaz, Stéphanie Dias (2012), *Aquisição de exaustividade em estruturas interrogativas e clivadas do português europeu: desenvolvimento pragmático ou gramatical?*, in: Armanda Costa/Inês Duarte (edd.), *Nada na Linguagem lhe é Estranho. Estudos em Homenagem a Isabel Hub Faria*, Porto, Edições Afrontamento, 605–618.
- Lousada, Marisa (2012), *Alterações Fonológicas em Crianças com Perturbação de Linguagem*, tese de Doutorado, Universidade de Aveiro.
- Lousada, Marisa/Alves, Dina/Freitas, Maria João (em prep.), *Desenvolvimento atípico: aspetos fonéticos e fonológicos*, in: Maria João Freitas/Ana Lúcia Santos (edd.), *A Aquisição de Língua Materna e não Materna. Questões Gerais e Dados do Português*.
- Marshall, Chloe R./van der Lely, Heather (2007), *The impact of phonological complexity on past tense inflection in children with Grammatical-SLI*, *Advances in Speech–Language Pathology* 9:3, 191–203.
- Marshall, Chloe/van der Lely, Heather (2009), *Effects of word position and stress on onset cluster production: Evidence from typical development, specific language impairment, and dyslexia*, *Language* 85, 39–57.
- Martins, Alexandrina/Santos, Ana Lúcia/Duarte, Inês (no prelo), *Syntactic complexity in children with autism spectrum disorders and specific language impairment*, in: Linda Escobar/Vincenç Torrens/Teresa Parodi (edd.), *Language Processing and Disorders*, Newcastle upon Tyne, Cambridge Scholars.
- Martins, Alexandrina/Vieira, Sónia (em prep.), *Avaliação linguística em contextos de desenvolvimento típico e atípico*, in: Maria João Freitas/Ana Lúcia Santos (edd.), *A Aquisição de Língua Materna e não Materna. Questões Gerais e Dados do Português*.
- Martins, Ana Maria (1994), *Clíticos na História do Português*, tese de Doutorado, Universidade de Lisboa.
- McDaniel, Dana/McKee, Cecile/Cairns, Helen Smith (1998), *Methods for Assessing Children’s Syntax*, Cambridge MA, The MIT Press.
- Mendes, Ana, et al. (2009), *Teste Fonético-Fonológico da Avaliação da Linguagem Pré-escolar – ALPE*, Aveiro, Designed, Lda.
- Miccio, Adele/Scarpino, Shelley (2008), *Phonological analysis, phonological processes*, in: Martin Ball et al. (edd.), *The Handbook of Clinical Linguistics*, Cambridge, Blackwell, 412–422.
- Mota, Helena (1996), *Aquisição Segmental do Português: um Modelo Implicacional de Complexidade de Traços*, tese de Doutorado, Pontifícia Universidade Católica do Rio Grande do Sul.

- Nespor, Marina/Vogel, Irene (1986), *Prosodic Phonology*, Dordrecht, Foris.
- Nogueira, Patrícia (2007), *Desenvolvimento fonológico em crianças dos 3 anos e 6 meses aos 4 anos e 6 meses de idade nascidas com baixo peso*, dissertação de Mestrado, Lisboa, Universidade Católica Portuguesa.
- Pinto, Manuela (1987), *A coesão em histórias contadas a partir de imagens por crianças dos 6 aos 8 anos*, Revista da Faculdade de Letras do Porto – Línguas e Literaturas (série II) 4, 93–113.
- Power, Desmond J./Quigley, Steinkamp (1973), *Deaf children's acquisition of the passive voice*, Journal of Speech and Hearing Research 16, 5–11.
- Quigley, Stephen/Smith, Neil/Wilbur, Ronnie B. (1974), *Comprehension of relativized sentences by deaf children*, Journal of Speech and Hearing Research 17, 325–341.
- Ramalho, Ana Margarida (em prep.), *Aquisição Fonológica na Criança: Tradução e Adaptação de um Instrumento de Avaliação Interlinguístico para o PE*, tese de Doutoramento, Universidade de Évora.
- Ramalho, Ana Margarida/Almeida, Letícia/Freitas, Maria João (2013), *Crosslinguistic Child Phonology Project: Preliminary results on European Portuguese*, comunicação apresentada no *Workshop on Knowledge Mobilization for an International Crosslinguistic Study on Children's Speech Development*, University of British Columbia, Canadá, agosto 2013.
- Ramalho, Ana Margarida/Almeida, Letícia/Freitas, Maria João (2014), *CLCP-PE (Avaliação Fonológica da Criança: Crosslinguistic Child Phonology Project – Português Europeu)*, Registo IGAC 67/2014.
- Ramalho, Ana Margarida/Almeida, Letícia/Freitas, Maria João (2015), *Adaptação ao Português Europeu de um instrumento interlinguístico de avaliação fonológica: CLCP-PE*, in: *Textos Selecionados do XXX Encontro Nacional da APL*, Lisboa, APL, 463–471.
- Raposo, Eduardo (1986), *On the null object in European Portuguese*, in: Osvaldo Jaeggli/Carmen Silva-Corvalán (edd.), *Studies in Romance Linguistics*, Dordrecht, Foris, 373–390.
- Rizzi, Luigi (1990), *Relativized Minimality*, Cambridge MA, MIT Press.
- Ross, John (1969), *Constraints on variables in syntax*, PhD dissertation, MIT.
- Santos, Ana Lúcia (2006), *Minimal Answers. Ellipsis, Syntax and Discourse in the Acquisition of European Portuguese*, tese de Doutoramento, Universidade de Lisboa [publicado em 2009, Amsterdam/Philadelphia, Benjamins].
- Schlichting, Liesbeth, et al. (1995), *Schlichting Test for Language Production*, Lisse, Swets Test Services.
- Schulz, Petra/Roeper, Tom (2011), *Acquisition of exhaustivity in wh-questions: A semantic dimension of SLI?*, *Lingua* 121, 383–407.
- Sim-Sim, Inês (2001, 1997), *Avaliação da Linguagem Oral: um Contributo para o Conhecimento do Desenvolvimento Linguístico das Crianças Portuguesas*, 2.ª edição, Lisboa, Fundação Calouste Gulbenkian.
- Sprouse, Jon/Lau, Ellen F. (2013), *Syntax and the brain*, in: Marcel den Dikken (ed.), *The Handbook of Generative Syntax*, Cambridge, Cambridge University Press, 971–1005.
- Sua-Kay, Eileen/Tavares, Dulce (2007), *Teste de Avaliação da Linguagem na Criança*, Lisboa, Oficina Didáctica.
- Sua-Kay, Eileen/Tavares, Dulce/Santos, Maria Emília (2015), *TAS – Teste de Avaliação Semântica*, Lisboa, Oficina Didáctica.
- Valian, Virginia/Aubry, Stephanie (2005), *When opportunity knocks twice: two-year-olds' repetition of sentence subjects*, *Journal of Child Language* 32, 617–641.
- Varlokosta, Spyridoula, et al. (2015), *A crosslinguistic study of the acquisition of clitic and pronoun production*, *Language Acquisition*, online, doi:10.1080/10489223.2015.1028628.
- Vaz, Stéphanie Dias (2012), *Aquisição de exaustividade em crianças falantes de português europeu*, dissertação de Mestrado, FCSH/Universidade Nova de Lisboa.

- Viana, Fernanda Leopoldina (2004), *TICL – Teste de Identificação de Competências Linguísticas*, Vila Nova de Gaia, EDIPSICO.
- Vicente, Francisco (2009), *Consciência Fonológica no Ensino Básico em Moçambique*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Vieira, Sónia (2012), *A Test for Sentence Development in European Portuguese (STSD-PT)*, tese de Doutoramento, Universidade de Lisboa.
- Vigário, Marina/Freitas, Maria João/Frota, Sónia (2006), *Grammar and frequency effects in the acquisition of prosodic words in European Portuguese*, *Language and Speech* 49 – Special Issue: Katherine Demuth (ed.), *Crosslinguistic Perspectives on the Development of Prosodic Words*, 175–203.
- Yamaguchi, Naomi (2012), *Parcours d'acquisition des sons du langage chez deux enfants francophones*, thèse de Doctorat, Université Sorbonne Nouvelle – Paris 3.

Telmo Mória

12 Semântica e pragmática

Abstract: Neste texto, são descritos, de forma panorâmica, e com foco no português, aspetos centrais da interpretação linguística, quer aqueles (de natureza semântica, em sentido estrito) que envolvem diretamente unidades do sistema linguístico, quer aqueles (de natureza pragmática) que atendem a elementos extralinguísticos, como o contexto situacional ou os conhecimentos e intenções dos intervenientes nos atos de fala. Numa primeira secção, são exploradas preocupações centrais da descrição semântica (ou semântico-pragmática) da língua, como a anomalia, a ambiguidade, a vagueza, a paráfrase, as dependências referenciais e a transmissão de informação por meios implícitos. Numa segunda secção, são discutidos aspetos semântico-pragmáticos nucleares dos grandes sistemas gramaticais do português: a predicação (incluindo a combinação predicados-argumentos e a modificação), a determinação, a quantificação, o tempo (com destaque para a *Aktionsart*, o tempo verbal e o tempo adjunto), a modalidade, a negação e ainda – com especial relevância no plano do texto – as dependências anafóricas e as conexões interproposicionais.

Keywords: semântica, pragmática, interpretação, contexto, subsistemas gramaticais

1 As questões centrais da análise semântica e pragmática

A *semântica* é a área da gramática que se ocupa do significado das expressões linguísticas. Tradicionalmente, as análises semânticas centram-se nos aspetos da interpretação exclusivamente dependentes do sistema linguístico, não considerando elementos extralinguísticos, que constituem o objeto de estudo da *pragmática*. Estes últimos respeitam ao contexto situacional em que ocorre a enunciação, aos intervenientes nesse ato (com as suas relações interpessoais, conhecimentos e intenções) ou aos tipos de ações realizados através da linguagem, por exemplo. Atualmente, verifica-se que os estudos semânticos – pelo menos, em certas correntes – tendem cada vez mais a integrar aspetos do significado relacionados com o uso, que tradicionalmente eram considerados do domínio estritamente pragmático. Ou seja, a fronteira entre a semântica e a pragmática tende a esbater-se, para muitos autores.

Como disciplina que se dedica a estudar a forma como a linguagem representa o mundo, a semântica tem algumas preocupações centrais, de que destacarei seis nesta secção – anomalia, ambiguidade, vagueza, paráfrase, dependências referenciais e informação implícita. Essas preocupações são, como veremos, transversais aos sistemas gramaticais descritos na secção 2 (ainda que, por razões de espaço,

algumas delas sejam ilustradas e exploradas de forma mais superficial que outras nessa secção).

A anomalia semântica diz respeito à impossibilidade (ou dificuldade) de atribuir significado a uma expressão linguística. Pode ser ilustrada através das restrições gramaticais que afetam, por exemplo:

- (i) a presença do quantificador *ambos* em sintagmas de interpretação grupal, como os que ocorrem associados a construções recíprocas (cf. Peres 1987, 285–286) – e.g. frase anómala *ambos os políticos conversaram um com o outro* vs. frases plenamente gramaticais *ambos os políticos conversaram com um jornalista* ou *os dois políticos conversaram um com o outro*;
- (ii) a combinação de verbos de valor negativo intrínseco, como *impedir*, com operadores de negação expletivos nos seus complementos (cf. Peres 2013a, 480–481) – e.g. sequência anómala «O ministro classificou o terreno como zona histórica para *impedir* que a Igreja *não* o destine a outros usos [...]» (CETEMPúblico, par=ext15524-clt-94b-2);
- (iii) a combinação de adjuntos temporais com *desde* com descrições de situações télicas sem quantificadores que operem direta ou indiretamente sobre situações (cf. Mória 2000, 295ss.) – e.g. frases anómalas *o Pedro foi operado desde 1990* e *o padre batizou esta criança desde 1990* vs. frases plenamente gramaticais *o Pedro foi operado três vezes desde 1990* ou *o padre batizou mais de mil crianças desde 1990*.

A ambiguidade consiste na possibilidade de associar mais de um significado a uma mesma expressão linguística. É um fenómeno sempre semântico, por definição, embora possa resultar, isolada ou cumulativamente, de aspetos gramaticais muito diversos como por exemplo:

- (i) fonéticos – e.g. *quem não gostava de ter {avós / a voz} assim?!*;
- (ii) morfológicos – e.g. *partimos no domingo* (que pode representar um situação passada ou futura, expressa pelo pretérito perfeito ou pelo presente, respetivamente);
- (iii) lexicais (remetendo para o conceito clássico de homonímia) – e.g. *no final do romance, a heroína ficou na esquadra* (com *heroína* a poder designar a personagem principal do romance ou um tipo de droga);
- (iv) sintáticos atinentes à estrutura em constituintes – e.g. *o atleta confessou o uso de esteroides em novembro* (com a possibilidade de o adjunto temporal *em novembro* se aplicar no interior do sintagma nominal complemento direto, localizando temporalmente o uso dos esteroides, ou na frase matriz, localizando temporalmente a confissão);
- (v) sintáticos atinentes ao estabelecimento de relações anafóricas – e.g. *a Ana telefonou ao Pedro para lhe falar sobre os seus problemas* (com *seus* a poder remeter quer para *a Ana* quer para *o Pedro*, tornando o sintagma nominal *os seus problemas* de interpretação dúbia).

Há ainda a possibilidade de a ambiguidade não resultar de nenhum dos fatores gramaticais atrás enumerados, sendo por vezes referida como *estritamente semântica*. Destacam-se, entre outros tipos:

- (i) a ambiguidade entre leituras distributivas e grupais (amplamente estudada para o português em e.g. Peres 1998) – e.g. *o Pedro e a Ana venderam um apartamento* (que pode representar a venda individual pelo Pedro e pela Ana de um apartamento cada, ou a venda conjunta de um único apartamento);
- (ii) a ambiguidade entre valores reais e virtuais de sintagmas nominais indefinidos – e.g. *o Pedro quer comprar um castelo medieval com quatro torres* (que pode invocar um edifício específico, existente de facto, e que o Pedro pretende comprar, ou uma entidade que existe apenas na imaginação do Pedro);
- (iii) a ambiguidade de escopo entre sintagmas nominais – e.g. *todos os estudantes tiveram de analisar uma obra de Kafka* (que pode implicar que houve uma única obra em análise, comum a todos os estudantes, ou que cada um deles teve de analisar uma obra, possivelmente distinta).

A ambiguidade é frequentemente eliminada no contexto discursivo ou situacional em que ocorre o constituinte ambíguo. Nalguns casos, a dilucidação é fortemente condicionada por fatores de natureza pragmática, como o conhecimento do mundo. Veja-se o seguinte texto noticioso: «Morte de jovem [...] fez aumentar insegurança entre moradores numa zona onde se produz, por ano, meio milhão de arrobas de alfarrobas. Cada uma vale quatro euros» (*Diário de Notícias*, 22-09-2009, p. 25). Naturalmente, é o conhecimento do mundo que determina que a expressão anafórica *uma* seja interpretada em ligação a *arrobas de alfarrobas* e não (por inverosimilhança) a *alfarrobas*.

Importa não confundir a ambiguidade, que implica – ao nível frásico – condições-de-verdade distintas para uma mesma sequência, com a *vagueza*, que consiste na não especificação de certos aspetos das situações descritas, implicando a sua compatibilidade com diferentes cenários. A vagueza é um fenómeno praticamente omnipresente nas frases das línguas naturais, na medida em que não especificamos normalmente elementos das situações que consideramos discursivamente irrelevantes. Assim, frases tão simples como *o Paulo beijou a Ana* ou *ontem choveu* são vagas, na medida em que não explicitam, entre muitas outras coisas, em que parte do corpo é que a Ana foi beijada ou qual o período exato em que esteve a chover. Há, porém, construções linguísticas em que a vagueza ocupa um lugar central na interpretação. Refira-se, a título meramente ilustrativo, o uso de quantificadores intrinsecamente vagos, como *muitos* ou *poucos*, ou de construções recíprocas com sujeitos plurais, como *os estudantes ajudaram-se uns aos outros*. A redução do grau de vagueza pode ser relevante para os objetivos específicos de um determinado ato de fala, caso em que os falantes podem recorrer a aposições introduzidas por conectores particulares, como *ou seja* ou *isto é*, para efetuar essa redução, como nos seguintes dois exemplos do *CORPUS CETEMPÚBLICO*: «[...] [o mercado luta] contra uma crise de confiança que [...] dificilmente será ultrapassada a médio prazo, isto é, até ao final do corrente ano»

(par=ext415502-eco-92b-2); «Um jornal alemão publicou ontem com o destaque necessário – ou seja, na primeira página – uma fotografia de Sua Majestade, o príncipe de Gales» (par=ext684150-soc-94b-1).

A ambiguidade tem como fenómeno inverso a paráfrase. Assim, diz-se, por exemplo, que duas frases (distintas) são paráfrase uma da outra quando são equivalentes, isto é, quando possuem as mesmas condições-de-verdade (ou se implicam mutuamente). Os fenómenos da ambiguidade e paráfrase corporizam a ausência de uma correspondência biunívoca entre forma e valores, característica das línguas naturais. A paráfrase, fenómeno linguístico de grande interesse para os estudos semânticos, pode resultar de fatores lexicais (nomeadamente, a existência de sinónimos) – e.g. *o Paulo comprou um carro vermelho / o Paulo comprou um automóvel vermelho* – ou de fatores estruturais, do plano da combinação sintática – e.g. *é difícil mudar a lei / a lei é difícil de mudar; parece que o Pedro gosta da Ana / o Pedro parece gostar da Ana*.

As dependências referenciais são outra das grandes questões da análise semântica e pragmática das línguas naturais. Com efeito, a interpretação de certas expressões linguísticas depende da situação de enunciação – expressões dêiticas – ou do contexto linguístico em que ocorrem – expressões anafóricas. O estudo das expressões dêiticas, cujo referente não é fixo, mas depende sistematicamente de elementos da situação de enunciação, tem lugar central na pragmática. Salienta-se a diversidade de entidades e valores representados por mecanismos dêíticos: indivíduos ou objetos comuns e entidades de carácter espacial e temporal, que são os casos mais comumente referidos – e.g. *nós viemos aqui hoje para comprar este quadro* –, mas também situações, modos ou quantidades, como em *não voltes a fazer isso!* (vs. *não voltes a saltar o muro*), *vou deixar o livro assim* (vs. *vou deixar o livro de cabeça para baixo*), *queres mais maçãs?* (vs. *queres mais de três maçãs?*). O estudo das expressões anafóricas tem sido intensamente prosseguido, no âmbito da análise sintática e semântica, com especial relevo para as perspetivas que consideram centralmente o texto, em vez da frase, como unidade de análise (e.g. a Teoria da Representação do Discurso; cf. Kamp/Reyle 1993). Este tópico será desenvolvido adiante na secção 2.7.

A consideração da informação implícita é um aspeto de enorme relevância na descrição semântica e pragmática das línguas. Com efeito, apenas parte da informação transmitida numa frase ou texto é diretamente asserida. Outra parte – que pode ser crucial do ponto de vista da comunicação – é deduzida. Gramaticalmente, podem distinguir-se diversos tipos de informação implícita. Destacam-se:

- (i) a implicação lógica (*entailment*), sempre que o significado de uma proposição implica necessariamente o significado de outra, como acontece em frases com e sem modificadores restritivos – e.g. *o Paulo é sobrinho de um deputado socialista* implica *o Paulo é sobrinho de um deputado* – ou com adjuntos de localização durativa e adjuntos de duração – e.g. *o Paulo esteve a trabalhar das 9.00h até às 13.00h* implica *o Paulo esteve a trabalhar durante quatro horas*;

- (ii) a pressuposição, que decorre de algumas sequências só fazerem sentido num contexto em que se assume determinado tipo de conhecimento prévio; trata-se de uma informação que se caracteriza pela «sobrevivência à negação», isto é, pode inferir-se a informação pressuposta numa frase independentemente de esta ocorrer numa forma afirmativa ou negativa – e.g. *se o filho do Pedro (não) vier, saímos às 10 horas*, sequência que pressupõe, em qualquer dos casos, a informação contida na frase *o Pedro tem um filho*;
- (iii) a implicatura conversacional, termo cunhado por Grice (1975), que refere a informação obtida não através do significado literal do que se diz, mas do ato de asserção, um reflexo evidente das competências pragmáticas associadas à atividade de comunicação verbal humana – e.g. *A: Sabes se a Ana vai à festa? B: Ela está com gripe!*, resposta que implica, conversacionalmente, *a Ana não vai à festa*.

2 Aspectos semânticos e pragmáticos dos grandes sistemas gramaticais do português

Nesta secção, serão discutidos diversos aspectos nucleares na análise semântica e pragmática de alguns dos principais sistemas gramaticais do português: predicação, determinação, quantificação, tempo, modalidade, negação e ainda – com especial relevância no plano texto, de cuja coesão são pilares fundamentais – dependências anafóricas e conexões interproposicionais.

2.1 Aspectos centrais da predicação

Do ponto de vista semântico, a construção de frases – unidades essenciais para a comunicação linguística – implica a combinação de predicadores com argumentos. Os predicadores são expressões que designam propriedades ou relações – sendo tipicamente verbos, adjetivos ou nomes (sublinhados nos exemplos a seguir) – e os argumentos expressões linguísticas que designam as entidades às quais se atribuem essas propriedades ou que se fazem entrar nessas relações – sendo tipicamente sintagmas nominais, sintagmas preposicionais ou frases (parentetizados nos exemplos a seguir): *[o Paulo] é simpático, [a Ana] visitou [o Louvre], [o Pedro] confia [nos políticos], [o Pedro] pensa [que vai chover]*. Através da combinação predicadores-argumentos, formam-se predicações, que são o «núcleo semântico» das frases. Em casos excepcionais – tipicamente com os chamados verbos meteorológicos e afins –, as predicações contêm apenas o elemento predicador sem argumentos – e.g. *já amanheceu, vai chover*. Nos exemplos dados até aqui, o predicador é combinado com o número mínimo, indispensável, de argumentos, formando uma predicação básica; porém, em muitos outros casos, podem ser acrescentadas expressões (sintática e

semanticamente) opcionais, que dão informações adicionais, permitindo expandir a predicação – e.g. *a Ana visitou o Louvre de fugida no Natal com os pais*. Alguns autores denominam todas estas expressões (aplicadas adverbialmente) como modificadores verbais, embora possivelmente elas não formem uma classe única, do ponto de vista semântico e sintático.

Os argumentos relacionam-se com uma das funções primordiais da linguagem: a nomeação (das entidades do mundo). As línguas exploram uma grande diversidade de processos para realizar essa função. Cinjamo-nos, para efeitos ilustrativos, aos sintagmas nominais. Em certos casos, pode haver nomeação direta, como acontece com os nomes próprios – e.g. *Einstein, Lua, Louvre, Amazônia, Holocausto*. Naturalmente, como é inviável atribuir um nome próprio a cada entidade do universo, a nomeação faz-se, na maior parte dos casos por mecanismos mais económicos: parte-se do agrupamento de entidades do mundo, em classes, identificadas por nomes comuns – e.g. *cientista, satélite, museu, floresta, genocídio, balão, brinquedo*; se os nomes forem relacionais, a classificação pressupõe a aplicação de complementos nominais, isto é, argumentos obrigatórios – e.g. *irmãos da Ana*; seguidamente, aplicam-se expressões de valor restritivo, denominadas modificadores nominais, caso seja necessário, e tantas vezes quantas for preciso para bem identificar as entidades sobre que se quer predicar – e.g. *cientista alemão que formulou a Teoria da Relatividade, satélite natural da terra, museu parisiense onde está a Mona Lisa, floresta tropical na bacia do rio Amazonas, genocídio do povo judeu, balão azul, brinquedo com rodas, irmãos da Ana que moram no Brasil*; uma vez identificadas as entidades relevantes (por combinações de nomes, complementos e modificadores), aplicam-se operações de determinação ou quantificação nominal (cf. secções 2.2 e 2.3). Estão então criados os sintagmas nominais, que se podem utilizar como elementos centrais no mecanismo da predicação. Sobre a arquitetura sintático-semântica aqui sumariada, veja-se, por exemplo, Peres/Móia (1995) ou Peres (2013b). Esta arquitetura ilustra, de maneira particularmente perspicua, a atuação de um princípio basilar da semântica nas línguas naturais: o Princípio da Composicionalidade, que – descrito de forma muito simplificada – estatui que a interpretação das expressões linguísticas complexas resulta do significado dos seus elementos constituintes e do modo como estes são combinados, isto é, que a interpretação atende à estrutura sintática e dela depende diretamente (o que acontece na esmagadora maioria dos casos, ainda que haja algumas exceções notáveis).

2.2 Determinação

O sistema da determinação expressa-se, como acima foi dito, no sintagma nominal. Do ponto de vista semântico, a determinação está predominantemente associada à função referencial, isto é, à introdução (ou à retoma anafórica) de entidades no universo de discurso – *Apareceu uma criança na sala. A criança trazia um balão azul e*

um brinquedo com rodas nas mãos. Uma outra criança roubou-lhe o balão. Como se vê, as entidades em causa são tipicamente identificadas por meios nominais, embora sejam possíveis em português nominalizações (definidas) de frases, por exemplo, com infinitivo flexionado, como em «[...] solicitei-lhe [...] uma opinião rápida sobre o estarmos a introduzir em Portugal uma experiência que, nos restantes países da Europa, tem tido resultados tão desanimadores [...]» (CETEMPúblico, par=ext524491-opi-96a-2). Adicionalmente, a determinação – especialmente a indefinida com *um* – pode estar associada à função predicativa, isto é, à atribuição de propriedades a entidades, como em *Liszt era um pianista virtuoso.*

Uma diferença crucial entre a determinação definida e a indefinida reside na necessidade ou não de informação prévia acerca das entidades referidas (por meio das expressões nominais determinadas). Os sintagmas com determinação indefinida – marcada tipicamente pelo artigo *um*, mas também por outras expressões tradicionalmente classificadas como pronomes, como *algum* ou *certo* – não requerem, por norma, qualquer tipo de conhecimento prévio e são frequentemente utilizados para introduzir no discurso entidades novas. Vejam-se os seguintes exemplos: *acabou de chegar um autocarro; tens aí uns jornais em cima da mesa; a criança está com algum problema, vai lá ver o que se passa; segundo o bastonário, certo juiz teria usado a prisão preventiva indevidamente.*

Inversamente, os sintagmas com determinação definida – marcada tipicamente pelo artigo *o*, mas também presente nos sintagmas com demonstrativos (*este, esse, aquele*) ou com o pronome relativo *cujo*, por exemplo – requerem esse tipo de informação, a qual se pode manifestar de pelo menos três maneiras particularmente comuns:

- (i) definitude situacional, quando as entidades referidas estão presentes no contexto situacional em que a frase é enunciada, como em *dá-me o lápis!* ou *esses lápis são bonitos*, em início de discurso; note-se que estas frases são improcessáveis – ao contrário de, por exemplo, *dá-me um lápis!* –, se não houver objetos do tipo em causa no contexto de enunciação;
- (ii) definitude textual, quando as entidades referidas são mencionadas no contexto discursivo (normalmente prévio) da frase e estão por isso acessíveis para efeitos de interpretação, como em *um rapaz desenhou graffiti em vários edifícios públicos; o rapaz foi detido porque os edifícios são património municipal*; note-se que, neste caso, a interpretação das expressões definidas depende da referência prévia, feita na primeira frase, através das sequências *um rapaz* e *vários edifícios públicos*;
- (iii) definitude epistémica, quando as entidades são identificadas a partir do conhecimento do mundo dos falantes, numa evidente interação entre semântica e pragmática, como em *o primeiro-ministro disse que os impostos não vão aumentar* (com interpretação condicionada pela assunção de que, não havendo indicação em contrário, se está a falar do primeiro-ministro atual de Portugal e dos impostos atualmente em vigor no país) ou *a rainha da Dinamarca foi coroada em 1972* ou *o papa é de nacionalidade argentina.*

Para o português, sobre as funções dos sintagmas nominais e os valores de determinação aqui referidos, veja-se, por exemplo, Peres (2013b, 762–766).

O português (especialmente europeu) tem a particularidade de usar determinantes definidos em contextos em que, noutras línguas, como o inglês, esses elementos não são explicitados, apesar de o valor definido estar presente (no que é frequentemente conhecido pela expressão inglesa *bare nominals*). Destacam-se, entre outros, os sintagmas com nomes próprios de indivíduos ou de (alguns) países – *o João chegou* (vs. *John arrived*), *a Suécia é uma monarquia* (vs. *Sweden is a monarchy*) –, com possessivos – *o meu pai avisou-me* (vs. *my father warned me*) –, com adjetivos temporais dêiticos, como *último* ou *próximo* – *o último fim de semana foi muito agitado* (*last weekend was hectic*), *regressarei na próxima semana* (*I will return next week*) –, ou com valor genérico – *a ira é um pecado capital* (vs. *wrath is a deadly sin*), *os computadores mudaram o mundo* (vs. *computers have changed the world*). Na realidade, a expressão morfossintática de valores de definitude varia fortemente entre línguas, mesmo geneticamente afins, como as românicas (cf. e.g. Posner 1996), ou entre variedades de uma mesma língua, como o português europeu e o português brasileiro (cf. e.g. Müller/Oliveira 2004, sobre a distribuição de *bare nominals*), sendo um intenso tópico de estudo em diferentes perspetivas de análise gramatical.

Os determinantes definidos e indefinidos possuem um valor quantificacional intrínseco, universal ou existencial, respetivamente, mas atualmente reconhece-se – seguindo propostas de Hans Kamp ou Irene Heim, no início dos anos 1980, para o inglês (Kamp 1981; Heim 1982) – que a sua função predominante é referencial, e não quantificacional. Além disso, a determinação (essencialmente a definida) pode estar associada a valores quantificacionais explícitos, em sintagmas complexos que conjugam as funções referencial e quantificacional: *os cinco irmãos da Ana que moram no Brasil são meus amigos*, *os 48% de votantes no referendo optaram maioritariamente pelo sim*, *a Ana preside a uma empresa cujos três principais acionistas são estrangeiros*.

2.3 Quantificação

O sistema da quantificação – entendida no sentido lato de «identificação de quantidades» – tem uma expressão gramatical muito ampla no português. Ocorre, de forma especialmente relevante, no sintagma nominal, no sintagma adjetival, no sintagma adverbial e no sintagma verbal. Vejamos cada um destes casos isoladamente.

A quantificação nominal é um dos sistemas gramaticais mais bem estudados do português – cf. em especial Peres (1987; 1992; 2013b), de que aqui se exporão as conclusões principais. Do ponto de vista semântico, o referido autor considera que importa distinguir dois grandes modos de quantificação – a contagem e a medição –, a que é possível juntar um terceiro (próximo da medição), a graduação ou quantificação escalar.

A contagem nominal opera sobre entidades discretas e é expressa tipicamente através de numerais cardinais e fracionários e de unidades tradicionalmente classificadas como pronomes indefinidos (e.g. *muitos*, *poucos*, *todos*), embora possa ser expressa também por outras formas menos típicas (e.g. adjetivos como *numerosos* ou expressões nominais complexas como *uma data de*). No que respeita a esta forma de quantificação, importa fazer uma distinção entre contagem absoluta e contagem relativa, uma distinção paralela à estabelecida por Kamp/Reyle (1993) entre quantificadores cardinais e quantificadores proporcionais. Existe contagem absoluta quando se afirma que o número total de entidades que possuem uma certa propriedade é igual a um dado valor absoluto (eventualmente definido de forma imprecisa ou vaga): *três balões azuis rebentaram, os sindicatos estiveram presentes em mais de seis reuniões, telefonei a uma data de amigos, há inúmeros problemas para resolver*. Existe contagem relativa quando se afirma que esse número é igual a uma dada proporção (eventualmente definida de forma imprecisa ou vaga, e possivelmente correspondente à totalidade, ou 100%, caso em que se fala no subtipo especial de quantificação universal): *dois terços dos balões azuis rebentaram, os sindicatos estiveram presentes em metade das reuniões / na maioria das reuniões, telefonei a todos os meus amigos*. Como se verifica pelos exemplos dados, a contagem relativa opera tipicamente sobre expressões já determinadas (*os balões azuis*, *as reuniões*, *os meus amigos*). Já a contagem absoluta opera frequentemente sobre expressões não determinadas (*três balões azuis*), embora possa também fazê-lo sobre expressões determinadas, em construções mais complexas, denominadas partitivas (*três dos balões azuis*).

A medição nominal opera sobre entidades não discretas de dois tipos: substâncias e partes de entidades. Também aqui se pode fazer uma distinção entre quantificação absoluta e relativa, paralela à que é estabelecida no domínio da contagem. Mais concretamente, existe medição absoluta quando se afirma que a quantidade de uma substância com uma certa propriedade é igual a um dado valor absoluto, identificado numa escala autónoma (e eventualmente definido de forma imprecisa ou vaga): *a joia contém cem gramas de ouro, uma grande quantidade de cocaína foi apreendida pela polícia, transacionaram-se toneladas de café, gastou-se muitíssima água*. Existe medição relativa quando se afirma que a parte de uma entidade com uma certa propriedade é igual a uma dada proporção da entidade total (eventualmente definida de forma imprecisa ou vaga, e possivelmente correspondente à porção total): *um terço do ouro encontrado foi confiscado pelo Estado, uma parte da parede foi pintada de azul, colocou-se um andaime à volta de todo o edifício*. Como acontece com a contagem relativa, a medição relativa também opera tipicamente sobre expressões já determinadas (*o ouro encontrado*, *a parede*, *o edifício*).

A graduação nominal opera sobre propriedades variáveis em escalas, expressas por meios nominais: *os jogadores têm muita confiança na vitória, o treinador tem alguma responsabilidade na derrota*. Trata-se de construções comparáveis às que ocorrem no domínio adjetival, associadas tradicionalmente à variação em grau dos adjetivos – *os jogadores estão muito confiantes na vitória, o treinador é um pouco*

*responsável pela derrota, a sopa está demasiado quente, ele foi um tudo-nada irônico, a Ana não foi nada simpática – e no domínio adverbial – a Ana respondeu muito acertadamente, a Ana fez tudo bastante depressa. Com efeito, a graduação é o modo de quantificação típico dos sintagmas adjetivais, e bem assim dos adverbiais, aplicando-se naturalmente apenas a predicados que denotam propriedades escalares. A marcação é por norma feita através dos tradicionalmente designados advérbios (e locuções adverbiais) de quantidade – *muito*, *demasiado*, *um tudo-nada*. Em português, existem ainda sufixos adjetivais de valor quantificacional, isto é, há expressão genuinamente morfológica da quantificação, como em *este quadro é lindíssimo*.*

Na estrutura interna de sintagmas verbais, também podem ocorrer quantificadores, aplicados à expressão verbal, indicando valores de grau, intensidade, duração e afins: *gosto muito de cinema, hoje dormi bastante*.

Importa ainda destacar três construções sintáticas, em que os valores de quantificação têm um lugar central, e podem ser coletivamente designadas como construções de grau:

- (i) as construções comparativas, cujo elemento oracional é tradicionalmente considerado um tipo de subordinada adverbial (mas, para muitos autores, constitui, discutivelmente, uma forma especial de oração relativa) – e.g. *os Himalaias são mais altos do que os Andes, a Espanha tem menos habitantes hoje do que a França tinha há cinquenta anos atrás, o Pedro esforçou-se tanto como eu; as construções usualmente designadas como *correlative comparatives* (estruturas de correlação de escalas) têm propriedades afins, mas constituem, possivelmente, uma subclasse à parte – e.g. *quanto mais trabalhares, melhores resultados terás*;*
- (ii) as construções consecutivas, cujo elemento oracional também é tradicionalmente considerado um tipo de subordinada adverbial (mas que Peres/Mascarenhas 2006, por exemplo, tratam, juntamente com as comparativas, as completivas e as relativas restritivas, como exemplos de «subordinação presa») – e.g. *a falésia é tão íngreme que poucos a conseguem escalar, a China tem tantos habitantes que teve de aplicar medidas draconianas de controlo do natalidade, ele trabalhou tanto que ficou cansado*;
- (iii) as construções superlativas (relativas), que, ao contrário das duas anteriores, não formam constituintes oracionais, e estão tipicamente associadas a sintagmas nominais definidos, identificando o(s) elemento(s) que ocupam a posição extrema numa dada ordenação; note-se que o quantificador pode estar em posições muito distintas – até bastante encaixadas – dentro do sintagma nominal relevante – e.g. *a Ana é a pessoa mais distraída que conheço, os dois escritores com mais obras publicadas foram condecorados*.

Sobre a semântica das construções de grau, para o português, veja-se, por exemplo, Marques (2004; 2013b).

2.4 Tempo

A semântica temporal do português tem sido um tópico de intensa investigação nos últimos anos, com destaque para cinco áreas: a *Aktionsart* (ou classes aspetuais de situações), o tempo verbal, o tempo adjunto, o tempo argumental e o tempo discursivo (cf. Mória/Alves 2013). Descreverei aqui brevemente apenas as três primeiras, que são aliás as mais estudadas, referindo, no entanto, que têm sido realizados alguns trabalhos sobre a transmissão da informação temporal mediante a combinação predicados-argumentos (cf. e.g. Mória 2006; 2011b) – como em *a invenção da imprensa data de meados do século XV* ou *a leitura da sentença demorou duas horas* – ou em discurso, atendendo às dependências anafóricas transféricas e – seguindo trabalho pioneiro de Nicholas Asher e Alex Lascarides para o inglês (Asher 1993; Lascarides/Asher 1993) – às conexões interproposicionais (cf. e.g. Alves 2003; Mória/Alves 2013, 577–581) – como em *a escola vai contratar um professor novo no próximo ano; entretanto, a Ana irá dar as aulas* (com *entretanto* a designar um intervalo que se estende do momento da enunciação ao tempo futuro referido na primeira frase) ou *a Ana caiu; magoou o braço vs. a Ana caiu; tropeçou numa pedra* (que, variando apenas no verbo da segunda frase, induzem leituras preferenciais distintas, de sequência temporal, no primeiro caso, e de retrocesso temporal, no segundo, com óbvia interferência de aspetos pragmáticos de conhecimento do mundo nessas preferências).

2.4.1 *Aktionsart*

As situações são entidades do universo do discurso que têm um lugar destacado na semântica temporal. Em português, elas são linguisticamente expressas através de frases plenas (i.e. com verbo) – e.g. *o Paulo foi ao Brasil, a estrada esteve cortada* –, estruturas proposicionais sem verbo – e.g. *(é difícil fazer reformas com) os militares no poder* – ou sintagmas nominais com nomes situacionais como núcleo – e.g. *o casamento do Paulo*. A sua importância na semântica temporal resulta de elas serem frequentemente um dos argumentos das principais relações temporais expressas na língua natural, como a localização – e.g. *a estrada esteve cortada, já não está; o Paulo foi ao Brasil há pouco tempo* –, a duração – e.g. *a estrada esteve cortada durante cinco horas; (é difícil fazer reformas com) os militares no poder há tanto tempo* – ou a frequência – e.g. *o Paulo foi ao Porto duas vezes por semana este ano*.

Um facto de grande importância para a descrição gramatical é que as expressões que identificam situações não são todas do mesmo tipo, antes se dividem em classes, conhecidas na literatura com os termos *Aktionsarten* ou categorias/classes aspetuais (de situações), entre outros. Para o português, tem sido considerada relevante a tipologia quadripartida de Vendler (1967), ou a quinqupartida de Moens (1987), estabelecidas para o inglês. As classificações em causa resultam, como sublinha Peres (2003), quer de critérios de base ontológica, que distinguem as situações por proprie-

dades (de natureza predominantemente temporal) como a pontualidade, a telicidade ou a homogeneidade, quer de critérios de natureza distribucional, que atendem aos contextos sintáticos em que os elementos de cada classe podem ocorrer (e a aspetos semânticos da sua interpretação nesses contextos).

A pontualidade distingue situações às quais não é normalmente associada uma duração, como as representadas por *o Pedro tocou à campainha* ou *o Pedro cortou a meta* (eventos pontuais), de situações que ocupam algum tempo, como *o Pedro esteve em Paris*, *o Pedro nadou* ou *o Pedro resolveu a equação*. Uma subdivisão na classe dos eventos pontuais, não presente em Vendler (1967), mas adotada em Moens (1987), distingue culminações – situações sem duração às quais são associadas consequências linguisticamente relevantes, mais exatamente o que se denomina um estado resultante – de pontos – situações pontuais que não evocam tais consequências. A relevância gramatical da distinção culminações-pontos não é porventura tão evidente em português como em inglês, mas é possível que alguns contrastes gramaticais da nossa língua se correlacionem com essa distinção. Por exemplo: Peres (2003, 214) refere a maior facilidade de usar culminações em construções participais absolutas – *nascido o primeiro filho* vs. **tossida a criança*; Cunha (2013, 603) refere a impossibilidade de descrever o estado posterior a um ponto com construções adjetivais precedidas do verbo *estar* – **a Maria está espirrada*.

A telicidade distingue situações que incorporam um fim intrínseco (ou culminação, ou *telos*), como *o Pedro cortou a meta* ou *o Pedro resolveu a equação* (situações téticas, a primeira pontual, a segunda não pontual), daquelas que, podendo naturalmente ter um início ou um fim, não possuem tal ponto culminativo, como *o Pedro esteve em Paris* ou *o Pedro nadou* (situações atélicas, sempre não pontuais). As situações téticas não pontuais designam-se processos culminados: incluem tipicamente três elementos, agregados, mas relativamente autónomos: um processo preparatório, uma culminação e um estado resultante (que correspondem, *grosso modo*, para o segundo exemplo apresentado, à atividade conducente à resolução da equação – e.g. a realização dos cálculos preliminares –, à culminação que constitui a resolução propriamente dita e ao estado em que a equação em causa se encontra resolvida pelo Pedro, respetivamente). As situações atélicas integram duas formas próximas, os estados e os processos, distinguíveis pelo critério ontológico do grau de homogeneidade, descrito a seguir.

A homogeneidade permite separar situações sem esta propriedade (situações heterogéneas), com esta propriedade apenas num certo grau (situações relativamente homogéneas), ou com esta propriedade sem restrições (situações totalmente homogéneas) – cf. e.g. Eberle (1998, 55). As situações heterogéneas coincidem com os processos culminados (e.g. *o Pedro resolveu a equação*) e são aquelas que, verificando-se minimamente num dado intervalo (o compreendido entre o seu início e a sua culminação), não se verificam em nenhum subintervalo deste; assim, se o Pedro começou a resolver a equação às 14h e terminou às 16h, não se pode dizer que tenha resolvido essa equação entre as 14h30 e as 15h, por exemplo. As situações totalmente homogé-

neas coincidem com os estados (e.g. *o Pedro esteve em Paris*) e são aquelas que, verificando-se minimamente num dado intervalo, se verificam em todos os seus subintervalos, independentemente da sua extensão (isto é, verificam-se inclusivamente nos instantes pontuais). Assim, se o Pedro esteve em Paris das 14h às 16h, esteve em Paris entre as 14h30 e as 15h ou às 14h55m10s exatas, por exemplo. Finalmente, as situações relativamente homogêneas coincidem com os processos (e.g. *o Pedro nadou*) e são aquelas que, verificando-se minimamente num dado intervalo, se verificam em todos os seus subintervalos, mas só até um certo grau de granularidade; por outras palavras, os processos correspondem tipicamente a uma sucessão de ciclos (e.g. braçadas, no caso de nadar), realizando-se homogêneamente no tempo até ao nível do ciclo relevante (abaixo do qual não faz sentido falar na existência da atividade em causa). Assim, se o Pedro nadou das 14h às 16h, nadou entre as 14h30 e as 15h, mas é estranho considerar que ele nadou às 14h55m10s exatas, por exemplo.

As diferenças acima notadas refletem-se diretamente na gramática do português, em propriedades distribucionais das expressões denotadoras de situações e em propriedades inferenciais associadas ao seu uso. Na impossibilidade de referir todos os comportamentos sintático-semânticos relevantes, chamo a atenção apenas para dois deles, habitualmente destacados, e que permitem distinguir facilmente, em português, as quatro classes vendlerianas:

- (i) a combinação com expressões adjuntas de duração, com distinções paralelas às que se estabelecem na língua inglesa – as descrições de eventos pontuais não se combinam com expressões de duração (a não ser que haja uma reinterpretação destes como não pontuais), os processos culminados têm a sua duração expressa por adjuntos encabeçados pela preposição *em* – e.g. *o Pedro resolveu a equação em duas horas* – e os estados e atividades por adjuntos encabeçados pela preposição *durante* (entre outras possibilidades) – e.g. *o Pedro esteve em Paris durante duas semanas*, *o Pedro nadou durante meia hora*;
- (ii) a associação de valores de iteratividade ao pretérito perfeito composto – quando este tempo verbal é associado a descrições de situações não estativas, há sistematicamente uma interpretação de repetição – e.g. *o Pedro tem nadado (nesta piscina ultimamente)*, *o Pedro tem lido a carta da namorada (todos os dias)*, *o Pedro tem cortado a meta (em todas as corridas em que tem participado)*; quando ele é associado a descrições de situações estativas, pode haver uma interpretação de repetição ou não – e.g. *o Pedro tem estado em Paris (muitas vezes nos últimos meses)* vs. *o Pedro tem estado em Paris (toda esta semana; nunca lá tinha ido antes)*.

Quanto às propriedades inferenciais, destaco apenas a variação no que pode ser deduzido de frases com descrições de processos culminados vs. com descrições de situações atélicas (estados e processos), em combinação com o verbo aspetual *estar a*, variação essa que naturalmente decorre de diferenças na homogeneidade das situações envolvidas. Assim, de *o Paulo está a resolver a equação* (processo culminado),

podemos inferir o Paulo ainda não resolveu a equação, enquanto que de o Paulo está a nadar (processo) ou o Paulo está a morar em Lisboa (estado) podemos inferir que o Paulo já nadou (alguma coisa) ou o Paulo já morou em Lisboa (durante algum tempo), respetivamente.

Importa ainda ter presente que a classificação aspetual das descrições de situações pode variar com o contexto (e ser especialmente sensível a fatores de natureza pragmática – cf. Mória 2011a, 157–159). Em particular, como foi salientado por Moens (1987) para o inglês, e se verifica igualmente em português, uma expressão linguística com uma dada classificação, chamemos-lhe básica, pode receber uma classificação distinta em certos contextos específicos. Estes processos de reclassificação (ou reinterpretação) são conhecidos pelo termo técnico de comutação aspetual. Por exemplo, *terminar o trabalho* funciona normalmente como a descrição de uma culminação (evento pontual) – cf. *o Pedro terminou o trabalho às 10h15m em ponto* –, mas, na presença de um adjunto de duração, pode funcionar como a descrição de um processo culminado – cf. *o Pedro terminou o trabalho em duas horas*. Entre os processos de comutação aspetual mais produtivos em português, destacam-se:

- (i) a passagem de um processo culminado a um processo, por eliminação da culminação, desencadeada pela combinação com elementos durativos típicos dos processos (como adjuntos com *durante* ou o verbo *passar*) – e.g. *o Pedro leu Os Maias durante meia hora, o Pedro passou meia hora a ler Os Maias*;
- (ii) a passagem de uma culminação a um processo culminado por adição de processo preparatório, desencadeada pela combinação com elementos durativos típicos dos processos culminados (como adjuntos com *em* ou o verbo *demorar*) – e.g. *o Pedro terminou a tese em dois anos, o Pedro demorou dois anos a terminar a tese*;
- (iii) a passagem de uma culminação ou um processo culminado a um estado, por focalização no estado resultante, desencadeada pela combinação com elementos durativos típicos dos estados (como adjuntos com *durante*) – e.g. *a fábrica foi encerrada durante seis semanas, o hospital anunciou que tinha implantado um coração artificial durante 160 dias no peito de um doente de 38 anos de idade*;
- (iv) a iteração de qualquer tipo de situação, que resulta em situações com as propriedades dos processos – e.g. *o Pedro espirrou durante cinco minutos, o padre leu esta passagem da Bíblia durante anos nas suas missas*.

2.4.2 Tempo verbal

O tempo verbal tem sido objeto de grande atenção na descrição gramatical do português, sendo tratado de forma bastante desenvolvida nas gramáticas antigas e tradicionais do português. O tratamento contemporâneo do tempo verbal – cf. e.g. Peres (1993), Oliveira/Lopes (1995) ou Oliveira (2013) – integra geralmente a perspetiva usada por Reichenbach (1947) para a descrição gramatical dos tempos verbais (*tenses*) do inglês.

Muito resumidamente, a ideia central de Reichenbach é que os tempos verbais servem primacialmente para expressar valores de localização de situações relativamente a pontos especiais do eixo do tempo (designados por ele *reference points* e por Kamp/Reyle 1993 *temporal perspective points*). A localização pode envolver três relações possíveis: anterioridade, sobreposição e posterioridade. O ponto de perspectiva pode estar situado em três domínios temporais: o passado, o presente (i.e. o momento da enunciação) e o futuro. Tendo em conta estas duas dimensões (modo de localização e posição do ponto de perspectiva), obtemos nove valores básicos para os tempos verbais (do indicativo), que tendem a corresponder a formas específicas: o pretérito perfeito simples, o presente e o futuro imperfeito sinalizam, respetivamente, a anterioridade, a sobreposição e a posterioridade ao momento da enunciação, comportando-se, por isso, como tempos dêiticos – e.g. *a Ana já esteve no Brasil, a Ana vive em Coimbra, a sonda aterrará em Marte*; o pretérito perfeito composto também indica um valor de anterioridade ao momento da enunciação, mas deixa em aberto se a situação relevante ainda está em curso no presente ou não – e.g. *o Paulo tem estado em Paris ultimamente; não sei se já regressou ou se ainda lá está neste momento*; o pretérito mais-que-perfeito, o pretérito imperfeito e o (tradicionalmente designado) condicional presente, sinalizam, respetivamente, a anterioridade, a sobreposição e a posterioridade a um ponto de perspectiva passado, identificado por norma no contexto discursivo prévio, comportando-se, por isso, como tempos anafóricos – e.g. *quando a Ana chegou, o Pedro já tinha acabado o trabalho; ontem, às 15 horas, eu estava no aeroporto; o jogador entrou em campo aos 60 minutos, mas seria ele que marcaria o golo da vitória*; finalmente, o futuro perfeito e o futuro imperfeito podem comportar-se como os três tempos anafóricos anteriores (com a primeira forma verbal a marcar anterioridade e a segunda a marcar, ambivalentemente, sobreposição ou posterioridade), mas relativamente a um ponto de perspectiva futuro – e.g. *quando a Ana chegar, o Pedro já terá acabado o trabalho; amanhã, às 15 horas, eu estarei no aeroporto; o jogador só vai entrar em campo aos 60 minutos, mas será ele que marcará o golo da vitória provavelmente*.

As correspondências formas-valores indicadas no parágrafo anterior podem ser consideradas básicas, num certo sentido. Porém, os tempos verbais são extremamente polivalentes e admitem valores (de localização) não básicos, como, entre outros: presente com valor de posterioridade ao presente ou sobreposição a futuro, isto é, como forma equivalente ao futuro imperfeito – e.g. *não te preocupes, eu faço (= farei) o que me pediste; amanhã, às 15 horas, já estou (= estarei) em Portugal de novo*; pretérito perfeito simples com valor de anterioridade a futuro, isto é, como forma equivalente ao futuro perfeito – e.g. *possivelmente, quando esta crónica for publicada já perdeu (= terá perdido) atualidade*.

Acresce ainda que os tempos verbais podem marcar – sozinhos ou em combinação com certos elementos gramaticais – outros valores semânticos além dos de localização temporal. Destacam-se valores, por vezes designados aspetuais, como genericidade, habitualidade ou iteração, que frequentemente acumulam com um

valor de localização temporal. Por exemplo, o presente tem um valor genérico em frases como *as baleias são mamíferos*; o presente, o pretérito perfeito simples e o pretérito imperfeito têm um valor habitual (que acumula com um valor de localização temporal concordante com os valores básicos definidos acima) em e.g. *o Pedro jogava tênis*; *o Pedro já fumou, agora já não fuma*; *há dois anos, o Pedro nadava muito, agora não tem feito exercício físico nenhum*; o pretérito perfeito composto tem um valor iterativo quando combinado com descrições de situações não estativas, como em *tenho dormido bem* ou *não tenho ido ao cinema*.

Sobressaem também valores que não são de natureza temporal, o que revela que a designação tradicional «tempos verbais» é, de certo modo, metonímica, já que apenas cobre uma parte da semântica das formas em causa. Entre estes, ocupam um lugar de destaque os valores modais e afins (como a evidencialidade ou a condicionalidade). Ocorrem ainda valores de natureza predominantemente pragmática, como os que resultam da introdução de marcas de cortesia no discurso. Ilustremos as diferentes situações, apenas para formas do indicativo (e do condicional).

Têm um valor modal epistémico (revelando o grau de compromisso do enunciador com a verdade da proposição enunciada) o futuro imperfeito e o futuro perfeito em frases como *a Ana terá uns vinte anos, suponho* ou *a Ana fez anos ontem; terá feito vinte e poucos, imagino*. Tem valor modal desiderativo (expressando desejos do enunciador) o pretérito imperfeito em frases como *eu sentava-me agora a descansar e já não trabalhava mais!* Têm um valor evidencial (sinalizando o facto – de especial relevância no discurso jornalístico – de a informação ser indireta, transmitida por terceiros) o futuro imperfeito, o futuro perfeito e o presente e pretérito perfeito do condicional, em sequências como as seguintes, extraídas do *CORPUS CETEMPÚBLICO*: «Os jornais soviéticos informam [...] que vários altos dirigentes do PCUS [...] estarão neste momento no estrangeiro com o fim de levantar dinheiro para uso pessoal» (par=ext261983-eco-91b-2); «Os detidos, que praticariam os assaltos no intuito de arranjar dinheiro para droga, terão assaltado uma residência de Faro [...]» (par=ext 1228849-soc-93a-1); «Ainda de acordo com o seu depoimento, os membros da resistência armada teriam assaltado uma coluna militar que transportava gado [...]» (par=ext950277-soc-97b-2). Têm um valor condicional (respeitante à dependência de uma situação da realização de uma outra) o presente e pretérito perfeito do condicional em frases como *se eu tivesse possibilidades, iria viver para o campo; se eu tivesse tido possibilidades, teria ido viver para o campo*. Em Portugal, o pretérito imperfeito e o pretérito mais-que-perfeito, podem ocupar o lugar das duas formas verbais acima, respetivamente, sem alteração de valor: *ia* (= *iria*) / *tinha ido* (= *teria ido*) *viver para o campo*. Têm um valor pragmático (sinalizando cortesia, delicadeza ou valores afins) o pretérito imperfeito em e.g. atos de fala diretivos ou expressivos na primeira pessoa – *queria* (= *quero*) *um café, por favor; tinha* (= *tenho*) *muita vontade de te ajudar, mas não posso* – ou o pretérito imperfeito ou o presente do condicional em e.g. atos de fala diretivos nas segunda e terceira pessoas – *poderia* / *podia* (= *pode*) *dizer-me as horas?*.

2.4.3 Tempo adjunto

Os adjuntos temporais formam um sistema altamente complexo da gramática do português. Apresentam grande diversidade de formas – desde unidades lexicais como *ontem* ou *então* até sintagmas de grande complexidade estrutural como *desde há mais de seis semanas atrás* – e acomodam uma paleta muito diversificada de valores. Consideremos alguns dos principais.

A localização de situações no tempo pode ser realizada através do tempo verbal, que, como vimos acima, posiciona os estados-de-coisas relativamente a pontos de perspectiva, mas também – sempre que se pretende um maior grau de precisão – através de adjuntos temporais, como nas frases *o presidente do partido iniciou o discurso às 17h em ponto* ou *a estrada estará cortada entre os dias 12 e 16 do corrente mês*. As situações podem ser localizadas relativamente a intervalos designados diretamente, mediante expressões estritamente temporais (expressões denotadoras de intervalos), como nos exemplos acima, ou relativamente a situações, ou mesmo a indivíduos (cf. Mória 2000, 111), como nos dois exemplos seguintes, respetivamente: *o presidente do partido apresentou o seu programa eleitoral durante o congresso do partido*; «as imagens [...] fazem dos restos da Jugoslávia de Tito uma tragédia histórica, sem par na Europa desde Hitler» (CETEMPúblico, par=ext954822-pol-92b-2). O modo como as situações são localizadas no tempo depende crucialmente de dois fatores linguísticos: a *Aktionsart* (mais especificamente, a telicidade) das expressões que descrevem essas situações e os conectores temporais utilizados (cf. Kamp/Reyle 1993; Mória 2000, 271ss.). Assim:

- (i) as situações téticas são tipicamente localizadas dentro do intervalo de localização da frase (localização inclusiva) – e.g. *Einstein nasceu em 1879*, *Tchaikovsky compôs «O Lago dos Cisnes» na década de 1870*;
- (ii) as situações atélicas associadas a conectores durativos, como *desde* ou *de...a*, preenchem totalmente o intervalo de localização da frase (localização durativa) – e.g. *Einstein viveu na Suíça de 1895 a 1914*, *Timor-Leste é membro das Nações Unidas desde 27 de setembro de 2002*;
- (iii) as situações atélicas associadas a conectores não durativos, como *em* ou *durante*, sobrepõem-se ao intervalo de localização da frase, sem indicação precisa sobre a forma (preenchimento parcial ou total) que essa sobreposição assume (localização de simples sobreposição) – e.g. *a Ana esteve em Paris em agosto*, *choveu durante o fim de semana*.

A duração tem uma expressão linguística muito eclética no português (cf. Mória 2006), com os adjuntos temporais – além dos verbos durativos (e.g. *durar*, *demorar*, *passar*, *levar*) – num lugar de grande destaque. Dado que a duração consiste na associação de quantidades de tempo a situações ou intervalos, ela envolve tipicamente predicados de quantidade de tempo, expressões nominais como *cinco minutos*, *um ano e meio*, *anos e anos* ou *muito tempo*. No que respeita à duração adjunta de situações, ela

depende centralmente – em português como em muitas outras línguas – de valores de *Aktionsart*. Assim, a duração de processos culminados, ou seja, a quantidade de tempo que decorre até ser atingida a culminação, é marcada tipicamente por adjuntos introduzidos por *em* seguidos de um predicado de quantidades de tempo – e.g. *Usain Bolt correu os cem metros em 9,58 segundos*; já a duração de estados e processos (situações atélicas) é marcada tipicamente por adjuntos introduzidos por *durante* (ou *por*, principalmente em português brasileiro, nas variedades contemporâneas) seguidos de um predicado de quantidades de tempo – e.g. *o Rio de Janeiro foi a capital do Brasil durante quase 197 anos*, «*Por anos*, a dama e o cavaleiro viveram em boa paz e união» (in Alexandre Herculano, *Lendas e Narrativas*, apud Mória 2001, 417); em muitos casos (de expressão de duração de situações atélicas), o conector de duração não está explícito, sendo o adjunto coincidente superficialmente com um predicado de quantidades de tempo – e.g. *a Ana descansou várias horas*, *a Ana viveu um ano e meio no Brasil*.

No que respeita à duração adjunta, importa ainda destacar duas modalidades especiais, com propriedades semânticas distintas da «duração comum», e associadas em português – ao contrário do que acontece no inglês, que explora a polivalência de um único conector, *for* – a conectores especializados:

- (i) a duração ancorada, duração atingida por uma situação atélica num dado ponto do eixo do tempo, combinando, por isso, valores de duração e de localização; é expressa tipicamente por adjuntos encabeçados pelos conectores *há/havia* (formas verbais em processo de mudança linguística, com propriedades de preposição temporal – cf. Mória 2011c) – e.g. *Brasília é a capital do Brasil há mais de 44 anos*; *quando Colombo chegou à América, a civilização olmeca já tinha desaparecido há/havia vários séculos*;
- (ii) a duração planeada, duração do estado resultante de uma situação télica, tipicamente pontual, associada a um valor de intencionalidade (valor não factual); é expressa normalmente por adjuntos encabeçados pela preposição *por* – e.g. *o recluso saiu por duas horas (mas só regressou oito horas depois)*, *emprestei um livro à Ana por uns dias (mas ela nunca mo devolveu)*.

No subsistema do tempo adjunto, têm ainda particular relevância relações, de natureza diversa, que aliam tempo e quantificação sobre situações. Destaco quatro delas (cf. Mória/Alves 2013, 561–563):

- (i) a frequência, que respeita ao número de ocorrências de situações por unidade de tempo ou por situação, como nos dois exemplos que se seguem, respetivamente: *a Ana vai ao cinema três vezes por semana*; *ao serviço do Real Madrid, Cristiano Ronaldo marcou uma média de 1,05 golos por jogo*; a frequência resulta de uma divisão do número total de ocorrências relevantes (idas ao cinema ou marcações de golos) pelo número total de unidades consideradas (semanas ou jogos), num intervalo implícita ou explicitamente determinado (o período pré-enunciação, pragmaticamente relevante, em que se faz a contagem de idas ao cinema da Ana,

- ou o período, fixado discursivamente através do adjunto inicial, em que Cristiano Ronaldo esteve ao serviço do Real Madrid);
- (ii) a periodicidade, que respeita ao intervalo médio (dado em unidades de tempo ou situações) que separa ocorrências de situações que se repetem no tempo – e.g. *a Ana vai ao Porto ver os pais de duas em duas semanas, o motorista leva o camião à revisão de cinco em cinco viagens*;
- (iii) a localização associada a quantificação sobre intervalos ou situações correlacionáveis, onde a quantificação induz uma repetição das situações localizadas – e.g. *a Ana foi ao Porto quase todos os fins-de-semana do mês passado, o Cristiano Ronaldo marcou um golo de cabeça em 12% dos jogos que fez ao serviço do Real Madrid, o presidente foi aplaudido de pé de cada vez que discursou*;
- (iv) a delimitação temporal da quantificação, que identifica a totalidade de situações de um determinado tipo que ocorrem dentro de um dado intervalo de tempo (cf. Mória 2000, 295ss.) – e.g. *os chefes de estado do G-20 já se reuniram dez vezes desde 2008, a ponte já foi restaurada três vezes desde a sua construção, a Ana leu sete livros no espaço de cinco dias*.

2.5 Modalidade

O conceito de modalidade tem uma amplitude muito variável consoante os autores, podendo abranger um conjunto vasto e relativamente heterogêneo de valores que incidem sobre uma proposição, alguns dos quais estreitamente relacionados com os valores marcados pelo modo (nas formas verbais). Simplificando bastante, poderíamos considerar que «do ponto de vista linguístico, [...] a *modalidade* é a gramaticalização de atitudes e opiniões dos falantes» (Oliveira 2003, 245).

Entre os valores modais mais estudados e mais comumente referidos (cf. para o português, entre muitos outros, Oliveira 1988; Marques 1995; 2013a; Oliveira/Mendes 2013), destacam-se (ignoradas algumas variações na terminologia e na extensão dos conceitos):

- (i) valores epistêmicos e afins, relacionados com a verdade das proposições (ou o compromisso do enunciador com essa verdade, decorrente do seu estado de conhecimento ou das suas crenças), e abrangendo noções como certeza, possibilidade, probabilidade ou plausibilidade – e.g. *o Paulo deve chegar atrasado, porque está muito trânsito; amanhã é capaz de chover; talvez chova esta semana; possivelmente, vamos chegar atrasados; a Ana terá uns vinte anos*;
- (ii) valores deônticos, relacionados com as condições de realização das situações, dado um sistema de normas, e abrangendo noções como obrigação, permissão ou proibição – e.g. *todos devemos respeitar a lei; temos de chegar a horas; o professor disse que já podemos sair*;
- (iii) valores de capacidade, necessidade e afins, por vezes referidos como «modalidade interna ao participante» nos casos em que esses valores se associam

diretamente a um argumento da predicação – e.g. *o Paulo já não pode passar sem exercício físico, o Paulo tem de fazer exercício físico para se sentir bem disposto* – e «modalidade externa ao participante» nos casos em que se associam a circunstâncias exteriores – e.g. *tem de fazer mais calor para esta árvore frutificar*;

- (iv) valores desiderativos, relacionados com a expressão de vontades, desejos, esperanças e emoções afins – e.g. *oxalá o Paulo não chegue atrasado; eu já me sentava!; não pode chover, estragava-nos os planos*;
- (v) valores avaliativos, relacionados com o juízo que o enunciador faz do conteúdo das proposições descritas – e.g. *lamento que não tenhas vindo; ainda bem que vieste; felizmente, não choveu*.

Como se pode facilmente verificar pelos exemplos acima, a modalidade é codificada linguisticamente em português de maneiras muito diversas. Destacam-se os chamados verbos auxiliares modais (e.g. *poder, dever, ser capaz de, ter de, haver de*), os advérbios (e locuções adverbiais) modais (e.g. *talvez, possivelmente, oxalá, sem dúvida*), os predicados plenos – verbos, adjetivos, substantivos – com argumentos oracionais (e.g. *acreditar, duvidar, lamentar, lamentável, certeza, hipótese*), certos «tempos verbais» (cf. futuro imperfeito ou pretérito imperfeito em exemplos da secção anterior) e – como tem sido amplamente discutido nos trabalhos de Marques (e.g. 1995; também ↗23 O modo conjuntivo) – o próprio modo verbal (cf. variação indicativo / conjuntivo, em *o Pedro não acredita que a Ana é/seja inocente*).

2.6 Negação

O sistema da negação tem sido muito estudado para o português, quer no plano predominantemente sintático (cf. e.g. Martins 1997; 2010; 2014; Matos 1999; Hagemeyer/Santos 2004) quer no plano semântico (cf. e.g. Peres 1997; 2000; 2013a).

Ignorando as manifestações predominantemente morfológicas, morfossintáticas e lexicais da negação, podemos considerar que a negação – sintática – se realiza em português em dois grandes formatos estruturais: a negação oracional e a negação de constituintes.

A negação oracional verifica-se quando o operador negativo se aplica diretamente a estruturas proposicionais, conferindo-lhes um valor negativo. Realiza-se, por exemplo:

- (i) com os operadores *não* ou *nem* (este último associado a um valor enfático), em orações matriz (entre outras) – e.g. *não foi encontrada vida em Marte; a Ana nem pensou duas vezes antes de desistir*;
- (ii) com o operador *sem*, em orações subordinadas – e.g. *corremos até à meta sem parar; é preciso alterar a lei sem que o interesse privado se sobreponha ao interesse público; sem a Ana chegar, não posso sair*;

(iii) com o operador *nem*, em orações coordenadas, com primeiro elemento negativo – e.g. *não choveu nem fez sol, nem choveu nem fez sol.*

A negação de constituintes verifica-se quando o operador negativo se aplica no interior de um constituinte de uma frase (e.g. um sintagma nominal ou adverbial). Realiza-se, por exemplo:

- (i) com a preposição *sem* (antônima de *com*) em construções nominais – e.g. *comprei uma casa sem a ajuda dos meus pais*;
- (ii) com o operador *não*, em estruturas com *mais de*, *menos de*, *muitos*, *poucos* e afins (na negação da superioridade e da inferioridade) – e.g. *não mais de 10% da população lê habitualmente o jornal*;
- (iii) com o operador *nem*, em estruturas de negação da universalidade, negação existencial particularizada e negação de cardinalidade – e.g. *nem sempre / nem em tudo se acerta; nem o Pedro faltou; nem cinco pessoas levantariam este peso*;
- (iv) com o operador descontínuo *nem... nem*, na negação existencial enumerativa – e.g. *nem o Pedro, nem Ana, nem o Rui faltaram*.

Merecem ainda especial destaque, no português, pela sua diversidade e pela complexidade da sua análise gramatical e classificação:

- (i) as construções com as chamadas «expressões autonegativas» – *ninguém, nada, nunca, jamais, nenhum N' ou N' nenhum/algum* –, em que o valor negativo está morfossintaticamente fundido com um valor de quantificação existencial (em sintagmas de valor intrinsecamente negativo) – e.g. *ninguém entrou, nunca / jamais menti, nenhuma pessoa é dispensável, em momento algum hesitei*;
- (ii) as construções com o verbo negativo *deixar*; note-se, por um lado, que delas se podem normalmente construir paráfrases com operadores de negação frásica – por exemplo, o excerto «Houve alguma confusão, mas não foi por isso que deixámos de chegar à final» (CETEMPúblico, par=ext211953-nd-96a-1) mantém o valor se substituirmos a sequência sublinhada por *não chegámos à final*; porém, elas exibem, por outro lado, importantes diferenças relativamente às construções típicas de negação frásica, sendo nomeadamente compatíveis com expressões negativas pré-verbais: *nenhum doente {deixará de ser / *não será} atendido por não estar inscrito na consulta* (Peres 2013a, 469).

O português possui ainda alguns subtipos especiais de negação sintática, de que podemos destacar os seguintes, estudados nas referências acima:

- (i) a negação metalinguística, em que o operador negativo não contradiz a proposição em que está inserido, antes nega a adequação (de forma ou significado) dos elementos linguísticos utilizados – e.g. *o Pedro não «entreviu» na discussão, «interveio»; não gosto de chocolate, adoro!; «o porco fica desta vez livre e quite, porque D. Diogo Lopes não corre, voa para o penhasco»* (in Alexandre Herculano, *A Dama Pé-de-Cabra*);

- (ii) a negação associada a enunciados interrogativos em que o operador negativo não altera o valor do acto de fala realizado – e.g. *não te importas de abrir a porta?* e *importas-te de abrir a porta?* podem ser formas de realizar exatamente o mesmo pedido, ainda que possam divergir em aspetos de natureza pragmática;
- (iii) a negação reforçada, que enfatiza o valor de um operador negativo já presente na frase – e.g. *não quero, não; não vou nada chegar atrasado*;
- (iv) a negação expletiva, que não possui qualquer valor negativo e pode estar associada a construções negativas ou afirmativas (subclassificando-se como negação expletiva redundante e não redundante, respetivamente) – e.g. *o edifício está muito instável, não tarda muito que não caia!, o que eu não daria para estar agora de férias!*;
- (v) a dupla negação, em que dois operadores negativos em sequência (*não sem, não deixar* ou *sem deixar*) se «cancelam», veiculando um valor positivo; naturalmente, o seu uso está condicionado por fatores estilísticos e/ou pragmáticos – e.g. *o político lá ia respondendo, não sem que aqui ou ali transparecesse alguma irritação; tenho a certeza de que o Pedro não deixará de vir à festa; consegue ser uma sinfonia vanguardista, sem deixar de ser melodiosa*.

Note-se, ainda, que podem existir combinações destas construções especiais, tanto quanto sei não descritas na literatura, como na sequência *não deixa de ser interessante, não* (com dupla negação e negação reforçada) ou no excerto, de aceitabilidade marginal, «Mas [...] foi perdendo o tom agressivo e acabou por entregar os kits pedidos, *não sem deixar* de cobrar a taxa, que sublinhou tem de ser sempre paga [...]» (CETEMPúblico, par=ext1246101-soc-93b-2). Note-se que, neste último caso, há três elementos negativos consecutivos (*não, sem, deixar*) e um valor proposicional afirmativo (a taxa foi cobrada), o que parece resultar de uma associação, particularmente pesada, de dupla negação a um terceiro elemento, expletivo.

Finalmente, um dos aspetos gramaticais mais interessantes associados às construções de negação do português é a chamada concordância negativa, que distingue a nossa língua de, por exemplo, o inglês-padrão – cf. *a Ana não viu ninguém* vs. *Ana didn't see anybody/*nobody*. Sobre este tópico, veja-se, entre muitos outros, Peres (1997) ou Matos (1999), para o português, e Swart/Sag (2002) ou Giannakidou (2006), para outras línguas, românicas e não só. Ignorando diversos aspetos mais complexos da concordância negativa, podemos dizer que ela se caracteriza, no português atual, pela ocorrência de expressões com forma negativa (e.g. *ninguém, nada*), mas valor semântico positivo, nomeadamente indefinido/existencial, em posição pós-verbal e sob o escopo de constituintes negativos adequados, denominados «legitimadores da concordância negativa» (negações oracionais, negações existenciais, ou, ainda, negações nominais com *sem*) – *a Ana não viu ninguém, a Ana resolveu a equação sem pedir ajuda a ninguém; nem eu sei nada do que se passa; corres o risco de ficar sem nada*.

No que se refere ainda à concordância negativa, destacam-se duas construções particularmente complexas:

- (i) a concordância negativa múltipla, em que as expressões de forma negativa sem valor negativo em posição pós-verbal são mais de uma – e.g. *eu NUNCA disse nada a ninguém sobre nenhum desses assuntos; o Pedro foi-se embora SEM dizer nada a ninguém*;
- (ii) a concordância negativa transfrásica, em que o legitimador da concordância negativa está numa frase matriz e a expressão de forma negativa sem valor negativo está numa frase encaixada (tipicamente, completiva) – e.g. *eu NÃO pedi que tu falasses com ninguém, falaste porque quiseste*.

2.7 Dependências anafóricas

Como já foi dito acima, a interpretação de certas expressões linguísticas – ditas genericamente anafóricas – depende do contexto linguístico em que ocorrem. No caso mais típico, que abrange o que podemos classificar como anáforas em sentido estrito, a interpretação é realizada mediante a associação a uma expressão precedente, criando-se uma cadeia referencial retrospectiva – e.g. *o presidente já discursou, mas eu não ouvi o que ele disse; a Ana visitou o Japão e gostou muito de lá ter ido; o Paulo viu-se obrigado a escrever com a mão esquerda, mas não se ajeitava nada a escrever assim*. Em certos casos, é possível que a expressão referencialmente dependente preceda aquela que permite a interpretação, formando-se uma cadeia prospetiva ou catafórica – e.g. *o Paulo sugeriu – e isso pareceu-me acertado – que a reunião fosse adiada uns dias*.

Do ponto de vista da sua realização morfossintática, as anáforas têm uma expressão muito variada, destacando-se em português:

- (i) os pronomes e outras unidades tradicionalmente classificadas como advérbios – e.g. *ele, lá, assim, isso*, nos exemplos acima;
- (ii) os sintagmas nominais definidos, incluindo os que integram demonstrativos – e.g. *uma bomba explodiu no mercado central de Bagdade; {a bomba / essa bomba} foi colocada por fundamentalistas islâmicos; eu bem avisei o Paulo, mas o teimoso não me deu ouvidos*;
- (iii) as categorias vazias, ou anáforas nulas ([]), de presença extremamente forte no português, uma língua de sujeito nulo – e.g. *a Ana disse que [] vai a Paris com a mãe [] amanhã, a Ana chegou no dia 12 e o Pedro chegou na véspera []*;
- (iv) as expressões predicativas verbais de carácter anafórico, construídas com os verbos *fazer, acontecer* e afins – e.g. *o Pedro comprou um barco e eu vou fazer o mesmo, a Ana foi induzida em erro e o mesmo aconteceu comigo*.

Do ponto de vista categorial, as expressões anafóricas e seus antecedentes formam cadeias de tipos muito diversos: nominais, frásicas, verbais ou (de sintagmas) preposicionais – já exemplificadas acima – e ainda adjetivais – e.g. *os passageiros estavam ansiosos por chegar e o comandante não o estava menos; a bactéria revelou-se resisten-*

te aos principais antibióticos; bactérias assim são o pesadelo de qualquer médico – ou adverbiais – pensei que ias falar devagar, mas não falaste nada assim.

Do ponto de vista semântico, as expressões anafóricas e seus antecedentes formam cadeias referenciais envolvendo entidades de tipos ontológicos muito distintos: indivíduos ou objetos comuns, propriedades, intervalos de tempo, situações, etc. A relação semântica prototípica, nestas cadeias, é a correferência, que acontece quando duas (ou mais) expressões de um texto representam exatamente a mesma entidade do mundo. Existem, porém, outros casos. Entre as anáforas de tipos especiais, descritas para outras línguas como o inglês, mas igualmente existentes em português, destacam-se:

- (i) a anáfora seccionante, que envolve referência a uma parte de uma entidade (por exemplo, coletiva) mencionada anteriormente – e.g. *entrou um casal na sala; {ela / a mulher} trazia um vestido vermelho;*
- (ii) a anáfora com antecedentes descontínuos, que envolve referência, mediante uma expressão plural, a entidades identificadas por mais de um constituinte precedente – e.g. *o Pedro foi ao cinema com a Ana e a Rita; eu não pude ir com eles;*
- (iii) a anáfora obtida por abstração, que envolve reconstrução do antecedente a partir de material (normalmente quantificado) presente no contexto, sem que haja correferência com um constituinte específico – e.g. *a Ana comprou dois terços dos livros que o professor recomendou; eles estão em cima da mesa* (cf. Kamp/Reyle 1993, 309ss.);
- (iv) a anáfora reconstrutiva temporal/situacional (extensamente estudada para o português por Alves 2003), em que um intervalo de tempo é identificado indiretamente através de uma situação – *o Pedro esteve em Paris; a Ana esteve em Londres na mesma altura;*
- (v) a anáfora ligada, que envolve um processo de ligação de variáveis quantificadas – e.g. *todos os políticos acham sempre que as suas propostas são fantásticas.*

Como se pode verificar pelos exemplos aqui apresentados, a interpretação das expressões anafóricas é fortemente condicionada por questões de natureza pragmática e assenta em grande medida em processos dedutivos.

2.8 Conexões interproposicionais

No plano da construção textual, tem especial relevância gramatical o estudo das conexões entre proposições, que abrange os conceitos tradicionais de subordinação, coordenação e justaposição. Deixando de lado as orações completivas, relativas, comparativas e consecutivas, as conexões interproposicionais, surgem em três grandes formatos sintáticos, na arrumação de Peres/Mascarenhas (2006), que se afasta parcialmente das propostas tradicionais:

- (i) a subordinação livre, que corresponde ao núcleo central da subordinação adverbial tradicional; é formada por aplicação de orações temporais, causais, finais, condicionais ou concessivas a proposições completas, sintaticamente dominantes (*subordinantes*) – e.g. *a jarra partiu-se porque a Ana a deixou cair* – e acrescentando informação semântica (valores de tempo, causa, finalidade, condição e contradição, respetivamente) que pode igualmente ser veiculada por meios não oracionais; a sua aplicação cria uma frase que expande a predicação contida na subordinante; podemos considerar que o uso de orações gerundivas adverbiais, que exprimem um conjunto extraordinariamente diversificado de valores (além dos cinco atrás referidos, que também podem exprimir), se inscreve igualmente neste formato – e.g. *temendo represálias, os manifestantes dispersaram* (cf. Mória/Viotti 2005);
- (ii) a genuína coordenação, que corresponde ao núcleo central da coordenação tradicional; é formada por aplicação de orações coordenadas copulativas, disjuntivas ou adversativas com *mas* a proposições completas, do mesmo nível hierárquico; a combinação das duas frases coordenadas cria uma terceira frase, complexa, com condições-de-verdade distintas das dos seus elementos constituintes, ainda que deles dependentes – e.g. *choveu imenso, mas não houve acidentes*;
- (iii) as conexões que, ao contrário das duas anteriores, não formam frases, mas sim textos ou fragmentos textuais, objetos com propriedades sintático-semânticas não oracionais; são designadas coletivamente como conexões textuais e abrangem, na opinião dos referidos autores, a justaposição e construções afins, como, por exemplo, as que integram conectores adversativos distintos de *mas* (como *porém, todavia, contudo, no entanto*), conectores conclusivos e conectores explicativos – e.g. *choveu imenso; contudo, não houve acidentes*.

Do ponto de vista semântico, os três formatos implicam o estabelecimento de ligações lógicas entre as proposições conectadas, essenciais para o processamento dos discursos (e codificadas no Princípio de Conexidade Discursiva e relações discursivas a ele associadas; cf. e.g. Asher/Lascarides 2003). É de sublinhar que a lista de valores semânticos que ligam proposições ultrapassa largamente a lista de formatos (e subformatos) sintáticos, o que se deve a que, embora alguns destes sejam monovalentes (e.g. as causais ou as conclusivas), muitos são altamente polivalentes, veiculando – em interação estreita com questões pragmáticas – um vasto espectro de valores. Como exemplo de formatos sintáticos extremamente polivalentes, podem referir-se a justaposição simples (sem conectores) – e.g. *a Ana deixou cair a jarra; a jarra partiu-se* (valor consequencial); *a Ana caiu; tropeçou numa pedra* (valor causal); *a Ana esteve em Paris; visitou o Louvre* (valor de explicitação de pormenor, ou «elaboração») – ou a aplicação de gerundivas adverbiais – e.g. *a Ana trancou a porta, impedindo os ladrões de entrar* (valor consequencial); *a Ana esteve em Paris, tendo visitado o Louvre* (valor de explicitação de pormenor); *a Ana abraçou o Pedro, apertando-o com força* (valor de

modo); *a Ana saiu de casa, dirigindo-se apressadamente para o carro* (valor temporal de posterioridade).

3 Referências

- Alves, Ana Teresa (2003), *Sobre a Localização Temporal Adverbial Anafórica em Português*, tese de Doutorado, Universidade dos Açores, Ponta Delgada.
- Asher, Nicholas (1993), *Reference to Abstract Objects in English: A Philosophical Semantics for Natural Language Metaphysics*, Dordrecht, Kluwer.
- Asher, Nicholas/Lascarides, Alex (2003), *Logics of Conversation*, Cambridge, Cambridge University Press.
- CETEMPúblico – *Corpus de Extratos de Textos Eletrônicos MCT/Público*, <http://www.linguateca.pt/CETEMPUBLICO/> (21.10.2015).
- Cunha, Luís Filipe (2013), *Aspeto*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 17, Lisboa, Fundação Calouste Gulbenkian, 585–619.
- Eberle, Kurt (1998), *The Influence of Plural NPs on Aktionsart in DRT*, in: Fritz Hamm/Erhard Hinrichs (edd.), *Plurality and Quantification*, Dordrecht, Kluwer, 55–111.
- Giannakidou, Anastasia (2006), *N-words and Negative Concord*, in: Martin Everaert et al. (edd.), *The Blackwell Companion to Syntax*, vol. 3, Oxford, Wiley-Blackwell, 327–391.
- Grice, H. Paul (1975), *Logic and conversation*, in: Peter Cole/Jerry L. Morgan (edd.), *Speech Acts*, New York, Academic Press, 41–58.
- Hagemijer, Tjerk/Santos, Ana Lúcia (2004), *Itens de polaridade na periferia direita: negação aparentemente descontínua, afirmação enfática e «tags»*, in: *Actas do XIX Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 465–476.
- Heim, Irene (1982), *The Semantics of Definite and Indefinite Noun Phrases*, tese de Doutorado, University of Massachusetts, Amherst.
- Kamp, Hans (1981), *A Theory of Truth and Semantic Representation*, in: Jeroen Groenendijk et al. (edd.), *Formal Methods in the Study of Language*, Amsterdam, Mathematical Centre; reprinted in Jeroen Groenendijk et al. (edd.) (1984), *Truth, Interpretation and Information. Selected Papers from the Third Amsterdam Colloquium*, Dordrecht, Foris, 1–41.
- Kamp, Hans/Reyle, Uwe (1993), *From Discourse to Logic – Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht, Kluwer.
- Lascarides, Alex/Asher, Nicholas (1993), *Temporal Interpretation, Discourse Relations, and Common-sense Entailment*, *Linguistics and Philosophy* 16, 437–493.
- Marques, Rui (1995), *Sobre o Valor dos Modos Conjuntivo e Indicativo em Português*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Marques, Rui (2004), *Para uma Semântica das Construções Comparativas em Português*, tese de Doutorado, Universidade de Lisboa.
- Marques, Rui (2013a), *Modo*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 19, Lisboa, Fundação Calouste Gulbenkian, 673–693.
- Marques, Rui (2013b), *Construções de Grau*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 2, cap. 40, Lisboa, Fundação Calouste Gulbenkian, 2141–2173.
- Martins, Ana Maria (1997), *Aspectos da Negação na História das Línguas Românicas: da Natureza de Palavras como «nenhum», «nada», «ninguém»*, in: *Actas do XII Encontro Nacional da Associação Portuguesa de Linguística*, vol. 2, Lisboa, APL, 179–210.
- Martins, Ana Maria (2010), *Negação Metalinguística («lá», «cá» e «agora»)*, in: *Actas do XXV Encontro da Associação Portuguesa de Linguística*, Lisboa, APL, 225–237.

- Martins, Ana Maria (2014), *How Much Syntax is there in Metalinguistic Negation?*, *Natural Language and Linguistic Theory* 32:2, 635–672.
- Matos, Gabriela (1999), *Negação Frásica e Concordância Negativa em Português Europeu*, in: *Actas do XIV Encontro Nacional da Associação Portuguesa de Linguística (Aveiro, 28–30 de Setembro de 1998)*, vol. 2, Braga, APL, 197–218.
- Moens, Marc (1987), *Tense, Aspect and Temporal Reference*, tese de Doutorado, Universidade de Edimburgo.
- Mória, Telmo (2000), *Identifying and computing temporal locating adverbials with a particular focus on Portuguese and English*, tese de Doutorado, Universidade de Lisboa, http://www.clul.ul.pt/files/telmo_moia/tmoia_phd2000.pdf (30.09.2015).
- Mória, Telmo (2001), *Sobre a Expressão da Duração em Português Europeu e Português Brasileiro: o Uso de Sintagmas com a Preposição «por»*, in: Maria Elías Soares (org.), *Boletim da Associação Brasileira de Linguística, Número Especial, II Congresso Internacional da ABRALIN, Anais*, vol. 1, Fortaleza, ABRALIN, 415–419.
- Mória, Telmo (2006), *Portuguese Expressions of Duration and their English Counterparts*, *Journal of Portuguese Linguistics* 5:1, 37–73.
- Mória, Telmo (2011a), *Aspectos Pragmáticos da Interpretação da Informação Temporal*, in: Maria Auxiliadora F. Lima/Francisco Alves Filho/Catarina Mendes da Costa (orgs.), *Colóquios Linguísticos e Literários: Enfoques Epistemológicos, Metodológicos e Descritivos*, Teresina, EDUFPI, 153–174.
- Mória, Telmo (2011b), *Sobre a Expressão Lexical da Duração e da Localização Temporal em Português*, in: Mathias Arden/Christian Märzhäuser/Benjamin Meisnitzer (edd.), *Linguística do Português. Rumos e Pontes*, München, Meidenbauer, 251–269.
- Mória, Telmo (2011c), *Expressões Temporais com «Haver»: Gramaticalização e Interpretação Semântica*, in: *XXVI Encontro Nacional da Associação Portuguesa de Linguística. Textos Seleccionados 2010. Porto 2010*, Lisboa, APL, 401–419.
- Mória, Telmo/Alves, Ana Teresa (2013), *Tempo Adjunto e Tempo Discursivo*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 16, Lisboa, Fundação Calouste Gulbenkian, 557–581.
- Mória, Telmo/Viotti, Evani (2005), *Sobre a Semântica das Orações Gerundivas Adverbiais*, in: *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística (Lisboa, 13–15 de Outubro de 2004)*, Lisboa, APL, 715–729.
- Müller, Ana/Oliveira, Fátima (2004), *Bare Nominals and Number in Brazilian and European Portuguese*, *Journal of Portuguese Linguistics* 3:1, 9–36.
- Oliveira, Fátima (1988), *Para uma Semântica e Pragmática de «dever» e «poder»*, tese de Doutorado, Universidade do Porto.
- Oliveira, Fátima (2003), *Modalidade e Modo*, in: Maria Helena Mira Mateus et al., *Gramática da Língua Portuguesa*, 5.ª ed. revista e aumentada, Lisboa, Caminho, 243–272.
- Oliveira, Fátima (2013), *Tempo Verbal*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 15, Lisboa, Fundação Calouste Gulbenkian, 509–553.
- Oliveira, Fátima/Lopes, Ana Cristina M. (1995), *Tense and Aspect in Portuguese*, in: Rolf Thieroff (ed.), *Tense Systems in European Languages II*, Tübingen, Niemeyer, 95–115.
- Oliveira, Fátima/Mendes, Amália (2013), *Modalidade*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 18, Lisboa, Fundação Calouste Gulbenkian, 623–669.
- Peres, João Andrade (1987), *Para uma Semântica Formal da Quantificação Nominal Não-Massiva*, tese de Doutorado, Universidade de Lisboa.
- Peres, João Andrade (1992), *Questões de Semântica Nominal*, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Peres, João Andrade (1993), *Towards an Integrated View of the Expression of Time in Portuguese*, Lisboa, Faculdade de Letras da Universidade de Lisboa.

- Peres, João Andrade (1997), *Extending the Notion of Negative Concord*, in: Danielle Forget et al. (orgs.), *Negation and Polarity, Syntax and Semantics, Selected Papers from the Colloquium Negation: Syntax and Semantics, Ottawa, 11–13 May, 1995*, Amsterdam/Philadelphia, Benjamins, 289–310.
- Peres, João Andrade (1998), *Issues on Distributive and Collective Readings*, in: Fritz Hamm/Erhard Hinrichs (edd.), *Plurality and Quantification*, Dordrecht, Kluwer, 339–365.
- Peres, João Andrade (2000), *On the Nature and Licensing Conditions of N-phrases in Portuguese*, DELTA 16, 165–199.
- Peres, João Andrade (2003), *Estrutura das Situações e Semântica Temporal*, in: Ivo Castro/Inês Duarte (orgs.), *Razões e Emoção. Miscelânea de Estudos em Homenagem a Maria Helena Mateus*, vol. 2, Lisboa, INCM, 199–216.
- Peres, João Andrade (2013a), *Negação*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 14, Lisboa, Fundação Calouste Gulbenkian, 461–498.
- Peres, João Andrade (2013b), *Semântica do Sintagma Nominal*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 21, Lisboa, Fundação Calouste Gulbenkian, 735–815.
- Peres, João Andrade/Mascarenhas, Salvador (2006), *Notes on Sentential Connections (Predominantly) in Portuguese*, Journal of Portuguese Linguistics 5:1, 113–169.
- Peres, João/Móia, Telmo (1995), *Áreas Críticas da Língua Portuguesa*, Lisboa, Caminho.
- Posner, Rebecca (1996), *The Romance Languages*, Cambridge, Cambridge University Press.
- Reichenbach, Hans (1947), *Elements of Symbolic Logic*, New York, Macmillan.
- Swart, Henriëtte de/Sag, Ivan (2002), *Negation and Negative Concord in Romance*, Linguistics and Philosophy 25:4, 373–417.
- Vendler, Zeno (1967), *Linguistics in Philosophy*, Ithaca, NY, Cornell University Press.

Armanda Costa

13 Psicolinguística e ciência cognitiva

Abstract: Neste capítulo apresenta-se o contexto de emergência da psicolinguística no âmbito da revolução cognitiva da década de 50, descrevendo-se a sua relevância para o estudo da mente humana, ao focar-se nas representações e nos processos cognitivos envolvidos no processamento da linguagem, na produção e na compreensão, na modalidade oral e na escrita. Referem-se modelos de processamento da linguagem, com foco no processamento de frases, discutindo-se a relação entre teorias da gramática e de processamento. Introduzem-se metodologias de investigação psicolinguística para captação dos processos cognitivos em tempo real. Apresentam-se resultados de investigação psicolinguística em português europeu, através de uma seleção de estudos sobre processamento de frases: interpretação de frases declarativas transitivas com sujeito pós-verbal, processamento de frases relativas ambíguas e interpretação de cadeias de correferência.

Keywords: cognição, linguagem, modelos de processamento, gramática, metodologias de investigação

1 A ciência cognitiva

Compreender o funcionamento da mente humana é objetivo comum de um conjunto de disciplinas que constituem o território da hoje designada *ciência cognitiva*: a antropologia, a filosofia, a inteligência artificial, a linguística, a neurociência e a psicologia. As várias disciplinas, e outras que emergem por interfaces expectáveis, focam-se num objeto comum e partilham objetivos de investigação que ambicionam a obtenção de explicações sobre a natureza das representações mentais e sobre o funcionamento dos processos cognitivos inerentes à construção da identidade cognoscente do ser humano e à sua vida em sociedade. A ciência cognitiva abrange assim um articulado coerente de disciplinas que, tendo um grande objetivo comum, partem de perguntas de investigação de campos epistemologicamente marcados, trazem para a área de estudos perspectivas de abordagem específicas, quadros teóricos próprios e ferramentas de observação e de interpretação dos dados adequadas aos seus propósitos particulares. Sobre a emergência desta área científica e a sua possível datação, diz-nos George A. Miller:

«I date the moment of conception of cognitive science as 11 September, 1956, the second day of a symposium organized by the «Special Interest Group in Information Theory» at the Massachusetts Institute of Technology. [...] I left the symposium with a conviction, more intuitive than rational, that experimental psychology, theoretical linguistics, and the computer simulation of cognitive

processes were all pieces from a larger whole and that the future would see a progressive elaboration and coordination of their shared concerns.» (Miller 2003, 142–143)

Neste simpósio, em que participaram destacados cientistas das áreas da comunicação e das ciências humanas, destacou-se Noam Chomsky, pondo em causa o poder explicativo da teoria da informação então em força (Shannon 1948) para dar conta da faculdade da linguagem tal como ela se manifesta na comunicação, no seu poder representativo e na sua natureza recursiva. Na nova perspectiva que introduzia, Chomsky punha o foco em «The internal mechanisms that generate linguistic expressions and determine their sound and meaning» (Chomsky 2007, 12). Defendia que a linguagem é uma consequência de uma faculdade cognitiva com sustentação biológica, que habilita o ser humano a produzir enunciados linguísticos de formatos infinitos a partir de um núcleo de regras finito, agora entendido como um sistema computacional mental. Até aí, os estudos estruturalistas sobre a linguagem tomavam-na como um sistema complexo, externo ao próprio indivíduo, e o propósito maior era a descrição comparada das línguas naturais através de processos minuciosos e detalhados de categorização e distribuição de elementos linguísticos (cf. Harris 1951). Estas abordagens revelavam-se insatisfatórias quanto à explicação da relação entre mente e linguagem, enquanto a proposta de Chomsky, inovadora e fraturante, gerava grande entusiasmo (ganhando forma desde a publicação de *Syntactic Structures* em 1957 até aos dias de hoje). Numa perspectiva racionalista, aí se inicia a revolução cognitiva no entendimento do que é a linguagem verbal na sua relação mais intrínseca com a cognição. A linguagem verbal não é entendida como um comportamento linguístico modelado a partir dos dados de *input* e de mecanismos mais ou menos complexos de aprendizagem, como defendido pelo dominante psicologismo experimental que Chomsky interpela e questiona na revisão que então faz do estudo de Skinner (Skinner 1957; Chomsky 1959). Na década de 50 do século XX, são duas perspectivas antagónicas que se defrontam para explicar um dos mais fascinantes e distintivos traços da espécie humana: a linguagem enquanto sistema de comunicação e sistema simbólico, que sustenta e estrutura o pensamento. Mas a posição de Chomsky, angular para a evolução da linguística teórica, não surge isolada, antes num quadro epistemológico em que o programa de estudos de um conjunto de disciplinas é orientado para a mente humana. Esse grupo de disciplinas tem como traço unificador o descentramento do estudo do comportamento observável e da sua explicação pelas interações com o meio, para privilegiar uma focalização nas representações mentais e nas capacidades computacionais para lidar com símbolos matemáticos ou linguísticos. Em consequência, surgem duas linhas de interesse crescente forte: uma centrada na biologia, com consequências para as neurociências, outra na área da inteligência artificial. *Mente, cérebro e computador* constituem uma tríade em torno da qual se desenvolvem áreas de conhecimento que tiram vantagem das interações entre disciplinas. Datam desta década decisiva os avanços na Psicologia Cognitiva, com os estudos sobre a memória de trabalho (Miller 1956), ou sobre os

processos mentais e as estruturas do pensamento envolvidos no processamento da informação e na aprendizagem (Bruner/Goodenough/Austin 1956). Há desenvolvimentos na área da Inteligência Artificial e na criação de máquinas à imagem da mente humana, o computador, simulando processos cognitivos (Shannon/McCarthy 1956; Minsky 1961; Newell/Shaw/Simon 1958).

O incremento notável nos avanços do conhecimento deve-se também ao desenvolvimento de técnicas de observação do funcionamento cerebral, com os contributos da eletrofisiologia e da imagiologia, e ao aprofundamento dos estudos sobre lógica, com impacto na matemática e na computação. Da colaboração entre neurocientistas e especialistas da computação, resultam propostas de redes neuronais que modelizam operações da mente humana no processamento de informação linguística ou matemática. Daí ao computador enquanto análogo da mente humana foi um passo que hoje nos parece natural, mas que só foi possível com a colaboração efetiva de investigadores vindos de áreas de início estanques mas que souberam dialogar entre si.

Se na década de 50 há uma explosão de ideias em torno da ambição de entender o funcionamento da mente humana, só a partir da década de 70 é que começa a ser cunhada a expressão *ciência cognitiva*, muito por via do apoio que a *SLOAN Foundation* decide dar a essa área de investigação. A revista *Cognitive Science* iniciou a sua publicação em 1976 e a fundação da Cognitive Science Society data de 1979. No relatório que é feito sobre o estado da arte no respeitante a esta nova área, é incluído um diagrama que representa as ligações e interações entre as várias ciências – o designado *hexágono cognitivo* (Gardner 1985, 37). Cada vértice representa uma das seis disciplinas centrais, mas as possibilidades de interação entre elas levam a que muitos campos de conhecimento interdisciplinares se constituam. A psicolinguística é um deles.

2 A psicolinguística e a ciência cognitiva

Na monografia *Psycholinguistics: A Survey of Theory and Research Problems* (Osgood/Sebeok 1954, 1965), reclama-se para a psicolinguística um campo científico de colaboração ativa entre psicologia cognitiva experimental e linguística. O seu objeto de investigação é o funcionamento da mente, no respeitante ao uso do conhecimento linguístico, na compreensão e na produção. O seu programa de investigação tem como pilares fontes de informação geradas por disciplinas com interesses comuns. No domínio da investigação psicolinguística, não é possível encontrar verdadeiros problemas e formular perguntas acutilantes para relacionar conhecimento com comportamento, sem teorias da gramática que disponham de propostas explicativas para as estruturas lexicais e sintáticas usadas, que expliquem níveis de complexidade que tais estruturas podem possuir, que enunciem princípios de organização do sistema que se possam relacionar com estratégias de processamento. Contudo, para estudar o

uso da linguagem, não chega possuir boas descrições e explicações das unidades linguísticas, que são usadas composicionalmente na formação de palavras e de frases, ou das propriedades sintáticas ou semânticas das construções. É necessário relacionar os enunciados com estruturas do pensamento e com as operações mentais que viabilizam a sua produção e compreensão. Para isso é fundamental conhecer o funcionamento da mente em termos dos processos psicológicos intervenientes, dos mecanismos de percepção auditiva e visual, do funcionamento da memória, dos recursos cognitivos e metacognitivos que sustentam os processos mentais que operam com as representações linguísticas. Por outro lado, compreender e produzir têm bases biológicas partilhadas mas também específicas. O efeito de modalidade no processamento pode explicar-se por bases neuronais dedicadas ao tratamento do sinal linguístico acústico ou visual. É o caso da especificidade do processamento da palavra ouvida ou lida, que depende de áreas corticais distintas, dedicadas ao sinal auditivo, por um lado, e ao sinal visual, por outro (cf. hipótese da *Visual Word Form Area*, Dehaene/Cohen 2011, ou Marinkovic 2004, para evidências de indicadores espacio-temporais do decurso do processamento da palavra).

É perante este quadro que se torna evidente o caráter multidisciplinar da psicolinguística, a necessidade de teorias da gramática consistentes e com poder explicativo forte para dar conta das representações linguísticas que integram o sistema humano de processamento, a necessidade de modelos teóricos sobre os processos cognitivos que intervêm no processamento da informação, de paradigmas experimentais que permitam recolher evidências comportamentais e neurofisiológicas do funcionamento da mente humana no uso da linguagem.

2.1 A investigação psicolinguística

A psicolinguística cobre os domínios de investigação sobre a aquisição da linguagem e seu desenvolvimento, sobre o processamento da linguagem, compreensão e produção, na modalidade oral e escrita, no processamento monolíngue e bilingue, por sujeitos com desenvolvimento linguístico típico e atípico (Sandra 2009). Os estudos sobre a aquisição da linguagem e seu desenvolvimento foram-se autonomizando, constituindo hoje uma área especializada, com enquadramentos teóricos próprios e procedimentos de recolha e análise de dados que evoluíram significativamente, desde a pioneira recolha dos diários de observação e registo da linguagem produzida pela criança em contexto natural, até à recolha controlada de dados em meio experimental. Esta área é responsável pela dimensão dos estudos psicolinguísticos que têm como foco central a variação entre línguas, seja para a identificação do que é universal, seja para avaliar o impacto da variação na aquisição (Norcliffe/Harris/Jaeger 2015). O estudo das perturbações da linguagem pode estar também sob o escopo da investigação psicolinguística, quando apoiado em metodologias experimentais que tentam, através de estudos normativos e de técnicas comportamentais,

obter dados explicativos de usos linguísticos atípicos. Contudo, é à Neurolinguística que cabe o estudo aprofundado das afasias e das perturbações da linguagem de etiologia diversa, com apoio em técnicas de neuroimagem e de eletrofisiologia (cf. Ingram 2007, para uma perspetiva neurolinguística que dá conta de aspetos essenciais do processamento linguístico em condições de danos cerebrais). Também a designada *linguística clínica* se ocupa do estudo das perturbações da linguagem (cf. Cummings 2008).

Em alguns capítulos deste livro são apresentados dados de aquisição: ↗16 A aquisição dos pronomes clíticos no português L1; ↗17 O infinitivo flexionado na gramática do adulto e na aquisição de L1; ↗21 Sujeitos nulos: gramática do adulto, aquisição de L1 e variação dialetal; ↗25 A sílaba na gramática do adulto e na aquisição de língua materna. No cap. ↗11 Linguística clínica: alguns dados sobre o português são abordados aspetos relativos às perturbações da linguagem. Excluindo os estudos de aquisição e das perturbações da linguagem, nas secções seguintes ocupar-nos-emos do processamento da linguagem, sobretudo pelo adulto, incidindo na investigação sobre a compreensão de frases, mais no escrito do que na oralidade.

2.2 A psicolinguística e as teorias da gramática

«Perhaps the closest partnership in cognitive science is the one between psycholinguistics and formal linguistics. The two disciplines were born together and have grown up essentially side-by-side as sister disciplines. But like many family relationships, the dynamics between these fields have been complicated.» (Ferreira 2005, 365)

Nos domínios da compreensão e da produção da linguagem, os desafios para a investigação sobre *mente, cérebro, linguagem* são estimulantes. Explicar a compreensão implica perceber como em frações de segundo são executadas operações de decodificação e reconhecimento do sinal auditivo ou visual, em condições frequentemente adversas e perante *input* degradado, complexo e ambíguo. Implica descrever os mecanismos de discriminação e reconhecimento de unidades fonológicas e da forma acústica ou visual da palavra, perceber como se acede ao significado lexical, como se desenrolam as operações de estruturação sintática (doravante *parsing*) para a criação de estruturas e sua interpretação, como se integra a nova informação extraída do enunciado ouvido ou lido em representações conceptuais pré-existentes (cf. Cutler/Clifton 1999). Na produção, há que perceber os processos iniciais de conceptualização, a posterior seleção lexical e a sequente codificação gramatical através de princípios morfossintáticos, para que haja formulação de enunciados interpretáveis; depois, a implementação dos sistemas motores articulatorio ou manuo-visual para a produção de uma sequência acústica, gestual ou gráfica com significado (cf. Levelt 1999). Todos os processos envolvidos durante a compreensão e a produção são executados com suporte na perceção, atenção, memória e raciocínio e, criticamente, usam conhecimentos ou representações linguísticas. E as questões avolumam-se: que

representações estão envolvidas? Que unidades? Que princípios de organização? Que restrições? Que relação com o comportamento e com as bases biológicas da linguagem? A procura de respostas conduz inevitavelmente às teorias linguísticas. E, sobretudo, leva às que tenham capacidade descritiva e explicativa para caracterizar adequadamente o conhecimento linguístico, àquelas que sejam capazes de captar o que cada língua tem de particular e de universal.

A relação entre a psicolinguística e a linguística teórica tem passado por aproximações e afastamentos ao longo dos tempos, conforme o estado das teorias e a convergência ou não do produto da reflexão teórica e do legado da investigação empírica (Ferreira 2005).

«Do grammatical theories and language processing models describe separate cognitive systems – independent functions of the human mind? Or are they accounts of different aspects of the same system?» (Lewis/Phillips 2015, 28)

No contexto da *revolução cognitiva* de que fala Miller (2003) ou Chomsky (2005), surge a possibilidade de uma nova associação entre a linguística e a psicologia, que então andavam afastadas por via da forte influência das teorias comportamentalistas e estruturalistas. Sob a égide do lema «reunir de novo a linguística e a psicologia» – *The leading idea was ‘reuniting linguistics and psychology’* (Osgood/Sebeok 1954, 1965) –, a teoria linguística que melhor serve então os desígnios da psicolinguística é a proposta por Chomsky (1957 em diante), ao pressupor um modelo cognitivo-biológico para a linguagem humana. O modelo da gramática generativo-transformacional ofereceu desde o início explicações satisfatórias para a natureza do conhecimento linguístico (com diferentes níveis de representação e regras transformacionais), com os conceitos de *competência* e *performance* a distinguir entre conhecimento e o uso do conhecimento (*o que se sabe, o que se faz com o que se sabe*), e o conceito central de *Gramática Universal*, como correspondendo a um sistema computacional de base neuronal que opera com símbolos e operações, produzindo um conjunto infinito de estruturas linguísticas através de mecanismos recursivos (cf. conceitos de faculdade da linguagem em sentido alargado e em sentido estrito, como discutido em Hauser/Chomsky/Fitch 2002). A Gramática Universal é tomada como estado inicial determinado biologicamente, que dará lugar a estados mais especificados em resultado da interação com dados externos, correspondentes às línguas particulares adquiridas em contexto de interação. O modelo chomskiano era tão mais atraente quanto se estava numa altura em que as teorias de cariz behaviorista ainda disputavam o terreno na explicação da linguagem e se revelavam insatisfatórias para explicar comportamentos comunicativos cujo sucesso implica a manipulação de estruturas linguísticas de alta complexidade e de grande diversidade. O programa da gramática generativo-transformacional veio revolucionar a teoria linguística e contribuir fortemente para os estudos sobre cognição na sua relação com a linguagem. Em *Language and Problems of Knowledge*, colocam-se perguntas centrais que vão guiar a investigação linguística numa perspetiva cognitivista.

«1. What is the system of knowledge? What is in the mind/brain of the speaker of English, or Spanish or Japanese? 2. How does this system of knowledge arise in the mind/brain? 3. How is this knowledge put to use in speech (or secondary systems such as writing)? 4. What are the physical mechanisms that serve as the material basis for this system of knowledge and for the use of this knowledge?» (Chomsky 1988, 3)

Estas perguntas envolvem conceitos que vão sendo clarificados face a apropriações menos rigorosas (cf. Chomsky 2015) e são reformuladas tendo em conta avanços do programa biológico para a linguagem em Berwick et al. (2013):

«Perhaps the core question about language is: what is its basic «design»? As with any biological subsystem, the extent to which this question can be answered is indicative of whether one can tackle other basic questions, including how language is acquired and used, how the capacity for language evolved, how languages vary, and what the neural correlates of language are.» (Berwick et al. 2013, 89)

À psicolinguística interessam os conceitos de língua interna e externa, de gramática e de gramática em uso, e cabe-lhe especificamente investigar como é que o conhecimento linguístico se mapeia em comportamentos linguísticos observáveis; interessa-lhe relacionar conhecimentos e processos cognitivos com indicadores comportamentais, fisiológicos e neuronais, cabendo-lhe contribuir para as respostas a algumas das perguntas acima enunciadas. É assumido que a compreensão e a produção de enunciados são possíveis pela existência de um mecanismo de processamento da linguagem, específico da espécie humana, que opera com princípios universais motivados biologicamente, em conformidade com uma gramática de arquitetura modular e sob limitações criadas pelas funções cognitivas gerais.

Nas décadas de 60 e 70 houve um grande entusiasmo ao associar-se a teoria linguística formal com a psicologia cognitiva e os seus procedimentos empíricos, desencadeando-se uma onda experimentalista. Foi testada a hipótese de haver uma correspondência direta entre gramática e processamento, entre processos linguísticos (as regras transformacionais preditas pelo modelo generativo) e os processos psicológicos (as operações cognitivas necessárias à compreensão das construções). Criou-se uma relação estreita, quase de dependência, entre as duas disciplinas, no sentido de que a experimentação psicolinguística serviria para validar as predições da Gramática Teórica. Surge a Teoria da Complexidade Derivacional, com a pretensão de demonstrar uma ligação linear entre comportamento e conhecimento, no caso, comportamento linguístico e conhecimento sintático. Tendo o modelo da gramática generativa por referência, pretendia-se obter evidências de que a estrutura profunda e as regras transformacionais que derivavam estruturas de superfície tinham realidade psicológica; assumia-se que isso se projetava no comportamento observável em situação experimental, através, por exemplo, dos tempos de reação à leitura e compreensão de frases.

A procura de evidências era feita comparando-se uma frase de estrutura básica, o designado *kernel* – uma frase declarativa, ativa e afirmativa, de ordem SVO, do tipo

A Maria agride o Mário –, com frases dela derivadas em consequência de transformações. Considerando um conjunto de três operações (ativa/passiva; afirmativa/negativa; declarativa/interrogativa), criou-se um paradigma estrutural de oito frases derivadas (representadas no cubo de Miller 1962, *apud* Townsend/Bever 2001). A maior ou menor distância entre a forma básica e as formas derivadas dependia do número de transformações e refletir-se-ia em tempos de processamento (a passiva representava um 1.º passo de distância relativamente ao *kernel*, a passiva negativa um 2.º nível, enquanto a passiva negativa interrogativa oferecia a distância máxima). Assumia-se que à medida que a frase era produto de mais operações transformacionais mais difícil seria a sua percepção, mais e mais complicadas seriam as estratégias perceptivas necessárias à sua compreensão, mais custos haveria para a memória de trabalho. Como corolário, predizia-se que a complexidade perceptiva de uma frase era proporcional ao número de transformações envolvidas na descrição gramatical da frase.

«Fodor and Garrett argue that this view leads to an empirical prediction that the perceptual complexity of a sentence is proportional (or at least monotonically related) to the number of transformations involved in the grammatical description of the sentence.» (Bever 1970, 287)

Os resultados obtidos foram entusiasmantes, mas não se mantiveram quando outras estruturas bem mais complexas foram sendo testadas. Apesar da ideia de uma unificação entre sistemas cognitivos – da gramática transformacional e do mecanismo de processamento – ter um forte poder de atração, não foi possível mantê-la por os resultados se mostrarem inconsistentes.

Uma outra abordagem para explicar o que se passa entre «o que se percebe, o que se sabe e o que daí resulta», isto é, como se mapeiam estruturas externas em estruturas internas e se chega a uma interpretação, foi proposta por Thomas Bever (1970), com o que designava por «estratégias perceptivas para o processamento da linguagem». Usada no quadro da argumentação sobre a possibilidade da relação entre as propriedades das estruturas da língua e o comportamento poder indiciar leis cognitivas gerais, esta proposta tem no cerne a assunção de que perante estruturas ambíguas, sempre presentes na língua, o sujeito procede pela aplicação de estratégias motivadas psicologicamente, que preferem estruturas simples a estruturas complexas, para reduzir custos cognitivos. Estas estratégias configuram-se como heurísticos que, guiados por propriedades disponíveis no *input*, asseguram alguma rapidez na análise perante a incerteza de interpretação. De entre um extenso conjunto de estratégias, destacamos uma das mais testadas: «Strategy D: Any Noun-Verb-Noun (NVN) sequence within a potential internal unit in the surface structure corresponds to ‘action-object’» (Bever 1970, 298).

A sua atuação pode exemplificar-se com a frase (1), uma das mais emblemáticas nas teorias do processamento de frases, explicativas do comportamento do mecanismo humano de análise sintática (doravante *parser*) perante a ambiguidade local (em (2) apresenta-se uma frase em português de estrutura equivalente).

- (1) The horse raced past the barn fell.
 (2) Mulher suspeita de rapto de filho foge.

Pelas suas propriedades de ambiguidade temporária, estas construções têm sido designadas de *garden-path* (ou frases labirínticas): o ouvinte/leitor é levado para uma interpretação que, no confronto do material linguístico subsequente, tem de abandonar sob o risco de ficar com uma estrutura agramatical e não interpretável. Numa primeira leitura, o SN pré-verbal – *the horse/mulher* – é tomado como o sujeito de uma frase simples que precede uma forma verbal finita (*raced/suspeita*), frases (3) e (4); quando se encontra o verbo final, a frase tem de ser reanalisada e reinterpretada, resultando numa frase mais complexa, em que o sujeito frásico é um SN modificado por um constituinte adjetival, frases (5) e (6).

- (3) [The horse raced past the barn] ?fell.
 (4) [Mulher suspeita do rapto de filho] ?foge.
 (5) [The horse [raced past the barn]] fell.
 (6) [Mulher [suspeita do rapto de filho]] foge.

As estratégias heurísticas são um bom suporte para o processamento, aceleram-no e dão lugar a interpretações adequadas, mas não o explicam inteiramente e, sobretudo, lidam com a «superfície» das frases, não dando grande importância à gramática não visível das construções.

A partir dos anos 80 inicia-se uma fase em que a gramática passa a ter um papel maior na explicação do processamento da linguagem. Colocam-se hipóteses interessantes, com particular atenção às propriedades estruturais e semânticas que vão sendo explicadas pelos modelos da gramática. É sobretudo intensa a investigação sobre o processamento de frases, com atenção a aspetos de complexidade atribuíveis a propriedades do léxico (nomeadamente verbos e estruturas argumentais), a propriedades sintáticas, como as dependências a longa distância que desafiam a teoria e motivam trabalho laboratorial muito interessante, a relações anafóricas em cadeias complexas (cf. Osterhout/Swinney 1989; Ferreira 2005).

3 Cognição e modelização do processamento de frases

Quando lemos ou ouvimos, temos acesso a uma cadeia de sinais que são primeiramente percebidos como tendo um valor simbólico linguístico e que, para adquirirem um valor comunicativo, um significado num dado contexto, alimentam uma série de operações de tratamento da informação. O que está em causa na investigação psicolinguística sobre o processamento da linguagem verbal é explicar os processos que permitem mapear uma cadeia de natureza auditiva ou visual, constituída por sinais discretos (sons ou letras), numa representação simbólica com substrato neuronal, que codifica o significado extraído do *input*. Os modelos de processamento existentes devem explicar como é que rapidamente chegamos a uma interpretação das frases a despeito de, com frequência, serem complexas, ambíguas, ou até agramaticais. Para isso, espera-se que tenham em conta o decurso temporal das operações de tratamento do sinal, que especifiquem as fontes de informação linguística concorrentes e a forma como esta é integrada (serialmente ou em paralelo; com autonomia ou cooperativamente), que expliquem o concurso de informação não linguística, que expliquem a resolução da ambiguidade e a superação da complexidade.

Todos os modelos existentes assumem que no *input* verbal há informação disponível de natureza linguística diversa – fonológica, lexical, sintática e semântica – que tem de ser integrada para se chegar a uma interpretação; todos assumem que o processador tem de executar uma série de operações de reconhecimento, categorização e integração das unidades linguísticas. Pode dizer-se que os modelos se distinguem crucialmente pela sua arquitetura interna e pelo seu modo de funcionamento. Uns pressupõem um tratamento da informação guiado por representações simbólicas usadas na computação do *input* linguístico, em que há grande proximidade entre modelos da gramática e modelos de processamento (*symbolic models, grammar-based, modular models*). Outros modelos dão relevo a propriedades do *input* com impacto na sua perceção e que criam restrições contextuais; neste caso a experiência do processador é o fator decisivo (*input-driven, constraint-based approach, connexionist models*). Os modelos distinguem-se ainda por usarem princípios de processamento universais através das línguas, ou estratégias específicas, desencadeadas pelas propriedades particulares das línguas (Pickering 1999; Clifton 2000; Keller 2010).

3.1 Modularidade e sequencialidade no processamento

Para os modelos simbólicos, o mecanismo de processamento de frases possui especificidade relativamente aos mecanismos cognitivos gerais, e fala-se então do *Mecanismo Humano para o Processamento de Frases*, encarado como um subsistema cognitivo distinto de outros subsistemas. O processador assume uma estrutura modular, em que cada módulo possui informação, recurso e princípios próprios (na aceção de Fodor

1983). Coerentemente, prediz-se um funcionamento serial, não interativo, em que o processador sintático tem um estatuto privilegiado (cf. Forster 1979, para uma proposta de modelo estritamente modular e serial). O *parser* constrói estrutura, de forma incremental, a partir do *output* do sistema de processamento lexical; observa princípios de boa formação gramatical e princípios de economia cognitiva, não sobrecarregando a memória de trabalho. Os princípios orientadores das operações de *parsing* levam à preferência por estruturas sintaticamente mais simples em alternativa a estruturas mais complexas. Numa perspetiva mais forte da relação entre gramática (enquanto modelo da competência linguística) e sistema de processamento, a hipótese forte reivindica transparência entre sistemas, em que a competência teria capacidades preditivas para o funcionamento do mecanismo de análise sintática.

Um dos modelos que melhor exemplifica este paradigma é o modelo *Garden-path* proposto inicialmente por Frazier e mais tarde revisto em *Construal* (Frazier/Fodor 1978; Frazier/Rayner 1982; Frazier 1985; 1988; 1990; Frazier/Clifton 1996). O processador sintático proposto é um mecanismo em dois estádios, em que primeiro se compõem os itens lexicais em estruturas sintagmáticas na base de informação categorial (*Preliminary phrase Packager*, mecanismo de janela estreita) e, num segundo passo, se integram esses sintagmas numa única unidade estruturada onde são controladas relações de dependência e hierarquia entre sintagmas, atendendo a relações locais ou à distância (*Sentence Structure Supervisor*, mecanismo de janela alargada) (Frazier/Fodor 1978, 291–292). O *parser* opera com apoio em princípios de processamento que respeitam a sintaxe e evitam as limitações da memória de trabalho. Por exemplo, face a frases com ambiguidade local que induz duas possibilidades de interpretação, o processador compromete-se com uma única estrutura e essa será a que for sintaticamente mais simples. Isto coloca problemas de métricas de complexidade de difícil resposta (para uma métrica da complexidade no processamento de frases, ver Frazier 1985). Ao assumir-se que quanto mais estruturado estiver o material menos espaço ocupa na memória, o processador adota um funcionamento incremental, não adiando decisões e estruturando a informação à medida que vai sendo percebida. O *parser* deverá ter um funcionamento uniforme dada a sua determinação cognitiva; assim, os princípios que operam deverão ser robustos através das línguas, isto é, deverão ser universais.

Adotando a premissa *escolher rapidamente e escolher o mais simples*, vários princípios de processamento, gramatical e psicologicamente motivados, são propostos: *Minimal Attachment*, que leva à escolha de uma estrutura sintaticamente menos complexa – «Do not postulate any potential unnecessary nodes»; *Late Attachment*, que leva à integração do material em análise no último e mais recente marcador sintagmático estruturado – «If grammaticality permissible, attach new items into the clause or phrase currently being processed (i.e. the phrase or clause postulated most recently)» (Frazier/Clifton 1996, 9); o *Minimal Chain Principle*, que propõe que na formação de cadeias anafóricas se opte pela cadeia mais simples, com distâncias mais curtas e menos movimentos – «Avoid postulating unnecessary chain members at

S-structure, but not delay required chain members» (De Vincenzi 1991, 13). Vejam-se os exemplos de (7) a (15) (adaptados dos apresentados em Frazier/Clifton 1996, 11–12), que foram testados com várias técnicas (registro do movimento dos olhos, leitura automonitorada, paradigma do mundo visual). Com exceção das frases de (13) a (15) em que há ambiguidade permanente, em todas as outras há ambiguidade temporária ou local entre uma estrutura mais simples e uma estrutura mais complexa. Por vezes há palavras homônimas responsáveis pela ambiguidade estrutural, como *que*, relativo ou completivo, ou formas verbais ambíguas: *paga*, presente do indicativo ou particípio. Nessas frases, recolheram-se evidências de que numa primeira análise é sempre adotada a estrutura mais simples (evitando ramificações adicionais ou mais estrutura). Os constituintes sublinhados assinalam o ponto em que é atribuída uma primeira estrutura à frase, e que corresponde à estrutura mais simples. Note-se que há ausência deliberada de pontuação ou de notação prosódica que, como fontes de informação quando disponíveis, entrarão numa fase ulterior à estruturação sintática.

- (7) Empresária paga acima da média emigra para os Emiratos Árabes.
(preferência por frase simples em vez de SN modificado por constituinte adjetival)
- (8) O João contou ao amigo que resolveu o problema de matemática as suas dificuldades.
(preferência por frase completiva/OD em vez de frase relativa modificadora do SN)
- (9) Enquanto a avó tricotava as meias caíram do regaço.
(preferência por OD de verbo transitivo em vez de sujeito da frase principal)
- (10) A Maria beijou o João e o seu irmão quando saiu.
(preferência por SN coordenado em vez de oração coordenada)
- (11) O João disse que o Tomás partiu ontem.
(preferência por modificador do verbo da oração encaixada em vez de modificador do verbo principal)
- (12) Arrumei o livro que estavas a ler na biblioteca na minha estante.
(preferência por complemento oblíquo/locativo do verbo da oração mais alta, em vez de modificador do verbo da oração relativa)
- (13) Alguém alvejou a criada da atriz que estava na varanda.
(preferência pela adjunção da relativa ao nome mais encaixado)
- (14) Chamou o João.
(preferência pela cadeia (pro)VO, sobre a ordem com sujeito pós-verbal, VS)
- (15) Quem chamou o João?
(preferência pela extração do sujeito (SVO) sobre a extração do objeto (OVS))

3.2 Interatividade no processamento

Há modelos de processamento que, contrariamente aos que defendem um funcionamento estritamente serial em que há autonomia da sintaxe relativamente à semântica e ao conhecimento do mundo, preveem uma interatividade entre todas as fontes de informação linguística desde o início da computação. Entre si, distinguem-se pelo momento em que essa interatividade acontece, desde o início ou em fases mais adiantadas, e também pelo grau de interação previsto, que pode ser fraco ou forte. A hipótese de fraca interação prevê que a semântica seja consultada pela sintaxe em pontos de ambiguidade em que tem de ser escolhida uma entre mais alternativas, a hipótese forte prevê que a informação semântica e contextual entre desde o início e influencie de facto a estruturação da frase, oferecendo entidades, impondo a aplicação de regras ou o cancelamento de outras (Crain/Steedman 1985, 323–326).

Contra a hipótese de um compromisso com uma só estrutura guiada exclusivamente pela sintaxe, a possibilidade de interação implica a ideia de um processamento em paralelo, em que pode ser gerada mais do que uma estrutura. As estruturas em construção correm em paralelo até ao decaimento de uma e à promoção de uma outra por efeito de informação subsequente, semântica ou contextual, que conduz a escolhas de interpretação por plausibilidade. Neste tipo de modelos há propostas que conferem um papel importante à informação lexical para a estruturação da frase: o reconhecimento do verbo dá acesso, de imediato, às suas propriedades de seleção sintática e semântica, logo à sua grelha argumental, deixando antecipar os tipos de constituintes que seleciona e o modo como os constituintes e os papéis temáticos se associam entre si (Tanenhaus/Garnsey/Boland 1990; Tanenhaus/Carlson/Trueswell 1989; Boland 1993; Tanenhaus et al. 1995). Há outras propostas que valorizam a informação contextual, nomeadamente a informação sobre referência prevenindo interpretações ambíguas (*the referential theory*, Crain/Steedman 1985; Altmann/Steedman 1988). Em frases como (17), a estruturação e interpretação da frase relativa como completiva, que levam a *garden-path*, serão evitadas se houver um contexto precedente em que mais do que uma entidade seja introduzida (frase 16).

(16) Havia duas professoras na escola, uma que gostava de cantar, outra que gostava de dançar.

(17) A menina contou à professora que gostava de cantar os seus desejos de estudar música.

Outro tipo de modelos explicativos do processamento de frases são os que assumem um tratamento indiferenciado de toda a informação linguística, não prevenindo níveis de representação intermédios, ocorrendo o processamento da frase numa só etapa para a qual toda a informação concorre. Daí resulta uma estrutura completamente adequada, não havendo necessidade de reanálise. São importantes os conceitos de restrições e de competição entre fontes de informação, de um funcionamento integrativo e cooperativo. Dada a importância da experiência e da aprendizagem no processamento, prevê-se que este seja guiado por propriedades específicas das línguas,

atendendo a indicadores de frequência lexical ou de padrões de ordem de palavras. A hipótese da sintonização (*Tuning Theory*) tem sido usada para apresentar evidências de processamento de frases ambíguas, em que a preferência da análise não é guiada por princípios universais, mas por critérios relativos à frequência das construções na língua. Veja-se o caso do processamento de frases relativas ambíguas enquanto modificadores de um SN complexo, discutido em Mitchell/Cuetos (1991), Mitchell/Brysbaert (1998) e Cuetos/Mitchell (1988), entre outros.

O *modelo de competição para o processamento de frases* é paradigmático de um sistema com funcionamento interativo, probabilístico, que tira partido de restrições contextuais, podendo por isso ser também considerado uma instância dos modelos de satisfação de restrições (Bates/MacWhinney 1987; 1989; Kail 1999). Alinhado nas propostas conexionistas para o tratamento da informação – *modelos de processamento paralelo distribuído* (Rumelhart/McClelland 1986) – funciona pela detecção de pistas linguísticas disponíveis no *input* que, pela sua força e disponibilidade, criam restrições ao estabelecer relações de convergência ou de competição que constroem a análise e a interpretação. A informação extraída do *input* faz-se a partir do reconhecimento de formas linguísticas, que relevam de qualquer componente da gramática, tendo em vista chegar-se a uma única representação da frase de forma rápida e eficiente. A validade das pistas linguísticas para associação a funções é consequência da sua disponibilidade física, da sua localidade e fiabilidade. O processador é reativo ao contexto: avalia a relevância das pistas linguísticas de forma probabilística, fazendo uma atualização constante da informação e ajustando decisões de análise. À semelhança de redes neuronais, o sistema tem capacidade de aprendizagem a partir da experiência e da integração da informação disponível. O conhecimento linguístico, nomeadamente o sintático, perde centralidade relativamente às propostas modulares.

No quadro deste modelo, tem havido uma forte investigação interlinguística no sentido de perceber a validade das pistas de processamento, consideradas as propriedades particulares de cada língua. Um tópico intensivamente explorado tem sido o da interpretação de frases transitivas com ordem de palavras variável (NVN/NNV/VNN), onde se disponibilizam pistas relativas a ordem de palavras, a informação semântica de animacidade e a informação morfológica. Avalia-se a interpretação da frase através de uma tarefa de atribuição da função de Agente. Em Bates/MacWhinney (1989, 44–45) e Kail (1999), encontram-se resultados comparativos entre línguas.

4 Metodologias de investigação em psicolinguística

À investigação psicolinguística interessa crucialmente captar o decurso do processamento em tempo real, mais do que verificar qual é o seu resultado final. As técnicas a usar podem ser escolhidas em função dos propósitos investigativos: identificação dos processos em tempo real (*on-line*) ou o resultado desses processos (*off-line*); controlo do efeito das variáveis dependentes a obter (comportamentais ou neurofisiológicas);

impacto da modalidade da língua usada (oral ou escrita); adequação aos indivíduos que estão a ser estudados (crianças ou adultos).

Assume-se que tempos de leitura ou de reação, movimentos oculares ou neuroimagens possam ser tomados como correlatos de processos cognitivos e neurais inerentes a operações linguísticas em curso. Cruzando as propriedades do estímulo linguístico que desencadearam o comportamento ou a atividade neural com a tarefa que o participante tem de executar (nomeação ou decisão lexical; leitura de frases ambíguas, observação de imagens e audição de frases, entre outras) obtêm-se dados de tempo (em milissegundos), de precisão na resposta (*accuracy*), de padrões de movimentos oculares ou de padrões neurais que se podem relacionar com representações e processos mentais. Pode dizer-se que há uma relação mais ou menos direta entre os dados recolhidos *on-line* e as operações mentais, conforme a existência ou não de tarefas mediadoras: na leitura automonitorada de uma frase, o participante tem de clicar para fazer avançar o material no ecrã; numa tarefa de nomeação lexical tem de articular uma palavra; na leitura com registo do movimento dos olhos só tem de ler. Já quando se recolhem indicadores da atividade neuronal durante a execução de tarefas linguísticas, está-se supostamente mais próximo da captação dos processos cognitivos.

As *técnicas cronométricas* partem do pressuposto de que variações nos tempos de reação ou de leitura podem ser relacionadas com processos cognitivos que incorporam descodificação, acesso e integração. Cabem nesta categoria: a leitura automonitorada ou ao próprio ritmo (*self-paced reading*); audição automonitorada (*self-paced listening*); apresentação serial visual rápida (*rapid serial visual presentation*); *priming*, efeito positivo no processamento de uma estrutura sintática ou de uma palavra (alvo/target), precedida por outra (desencadeador/prime) com a qual partilha propriedades estruturais ou semânticas; *priming* intermodal (*crossmodal priming*), quando *prime* e alvo são apresentados em diferentes modalidades: oral um, escrito o outro; reconhecimento de *palavra sonda* (*probe recognition*), quando, durante o processamento de uma frase, num dado *locus surge* em tempo subliminar uma palavra que deve ser identificada para verificação de informação processada previamente. Dentro deste grupo, podem ainda ser consideradas técnicas que, não captando indicadores *on-line*, implicam tarefas *off-line* com fortes restrições de tempo, como julgamentos rápidos sobre a boa formação ou aceitabilidade de uma frase (*speeded grammaticality judgments*, *speeded plausibility judgments*). Para uma revisão de métodos de investigação psicolinguística, veja-se Sekerina/Fernández/Clahsen (2008), Carreiras/Clifton (2004) ou Garrod (2006), entre outros. O registo do movimento dos olhos durante a leitura (*eyetracking*) e/ou na observação de imagens ou do mundo real (*visual world paradigm*) parte do pressuposto de que as variáveis analisadas (fixações oculares e movimentos sacádicos) se correlacionam com processos cognitivos envolvidos no reconhecimento visual da palavra e na integração de estruturas sintáticas. Para uma revisão sobre *eyetracking* na investigação psicolinguística, ver Rayner 1998 e, particularmente para a leitura, ver também Rayner/Slattery 2009; Luegi 2012b; Luegi/Costa/Faria 2010).

Em termos de técnicas para obtenção de dados sobre o funcionamento cerebral durante a execução de tarefas linguísticas, pode haver recurso a técnicas hemodinâmicas ou a técnicas eletrofisiológicas. As primeiras mapeiam mudanças no fluxo sanguíneo em áreas cognitivamente ativadas e dão indicadores de grande resolução espacial: a tomografia por emissão de positrões (*positron emission tomography, PET*) e a imagem por ressonância magnética funcional (*functional magnetic resonance imaging, fMRI*). As segundas captam mudanças na atividade eletrofisiológica neural e são as mais adequadas para a obtenção de indicadores temporais: potenciais relacionados a eventos (*event-related potentials, ERPs*) e magnetoencefalografia (*magnetoencephalography, MEG*) (cf. Ulmann 2006).

5 Estudos psicolinguísticos em português

Nesta secção damos conta de investigação psicolinguística sobre processamento em português. Dadas as limitações de espaço, apresentaremos apenas alguns estudos que se focalizam no processamento de frases, simples e complexas, com controlo de propriedades sintáticas que distinguem o português, língua pouco estudada em termos psicolinguísticos.

Destacando alguns aspetos que nos interessam para os estudos que vamos referir, é de sublinhar que o português europeu, como língua românica de sujeito nulo consistente, apresenta um conjunto de propriedades interessantes para o estudo do processamento de frases. Não tendo propriamente ordem de palavras livre, permite variações à ordem básica de palavras SVO, como as ordens marcadas VOS e VSO. Frases declarativas transitivas com sujeito pós-verbal configuram-se como condições interessantes para testar a atribuição de uma estrutura e de uma interpretação. O sistema de flexão em português é rico, havendo flexão verbal abundante, flexão nominal e adjetival em género e número. Na expressão da correferência, tem a particularidade de, nas retomas anafóricas, apresentar um elenco de formas que vão das mais informativas às mais despojadas; acresce que, para expressar a correferência entre sujeitos, a forma pronominal do sujeito anafórico pode tomar uma de duas formas que parecem funcionar em distribuição quase complementar: formas nulas assinalam correferência entre sujeitos, formas plenas assinalam referência disjunta. Este conjunto de propriedades, ao distinguir o português de outras línguas e até ao distinguir as variedades europeia e brasileira entre si, permite desenhar condições para controlo do impacto de propriedades gramaticais no processamento e comparar os resultados obtidos para outras línguas. Nos capítulos 76 Morfologia nas interfaces, 714 Ordem dos constituintes frásicos: sujeitos invertidos; objetos antepostos, 721 Sujeitos nulos: gramática do adulto, aquisição de L1 e variação dialetal são desenvolvidos aspetos da gramática do português relevantes para os estudos de processamento que apresentamos.

5.1 Processamento de frases com sujeito pós-verbal

Na perspetiva do modelo de competição para o processamento de frases, se para a interpretação da frase é necessária uma atribuição de papéis semânticos, é preciso que o processador identifique e integre pistas formais que lhe permitam associar formas a funções. Na frase *A rapariga abraça a amiga*, o verbo *abraçar* pode ser entendido como um verbo simétrico (em que argumento interno e externo podem permutar) e recíproco (*A rapariga e a amiga abraçam-se*). Tendo os dois SNs propriedades léxico-semânticas similares, resta a pista sintática para a atribuição de uma estrutura funcional e semântica: à ordem de palavras NVN, haverá grande probabilidade de corresponder a estrutura $S_{\text{Agente}}VO_{\text{Tema}}$. Contudo, se os SNs dispuserem de propriedades morfológicas ou semânticas que os promovam ou despromovam para a função de Agente, a relação de forças entre pistas varia e o processador encontra condições favoráveis (*A rapariga abraça as amigas* / *A rapariga abraça a árvore*) ou desfavoráveis (*A árvore abraça a rapariga* / *As amigas abraça a rapariga*).

Num estudo realizado por Costa (2005), assumiu-se que a frase declarativa transitiva na sua ordem canónica – SVO – com um SN pré-verbal cumulado com informação morfológica e semântica típica da função de sujeito e de Agente configura a condição ótima para o processador atribuir uma interpretação. Em contrapartida, frases transitivas com sujeitos pós-verbais são ambíguas entre uma interpretação VSO ou VOS, se não houver pistas formais ou semânticas que restrinjam a interpretação. Elaborou-se um desenho experimental em que se manipularam traços morfológicos de número [\pm sg] e de animacidade [\pm anim] nos dois SNs de uma frase declarativa com sujeito pós-verbal; criaram-se condições de ambiguidade (*Abraça a rapariga a amiga*), condições de convergência e contraste máximo (*Abraça a rapariga as árvores*) e de competição máxima (*Abraça as raparigas a árvore*).

Na ordem VNN, criaram-se doze condições, fruto do cruzamento dos fatores animacidade (quatro níveis) e número (três níveis), com o verbo sempre na 3.^a pessoa singular. Numa tarefa de interpretação em tempo limitado, cada frase foi apresentada num ecrã de computador durante 3.800ms, após o que o participante tinha de responder à pergunta *Quem fez X?* (sendo X o evento descrito pelo predicador verbal). Foram registados os tempos de resposta e a escolha do primeiro ou do segundo SN. No estudo participaram 43 universitários, falantes nativos de português europeu.

Os resultados mostraram que quer a informação morfológica quer a semântica criaram restrições que produziram efeitos globais significativos. Verificou-se que a manipulação do contraste de número produziu diferenças significativas, mas que o mesmo não aconteceu com o contraste de animacidade. Os dois fatores não mostram a mesma força de processamento: a morfologia tem um papel significativamente mais robusto mas, em competição com a informação semântica de animacidade, a sua força enfraquece. A ordem VOS mostra-se preferencial relativamente a VSO: SN adequado semântica e morfológicamente em posição final de frase é significativamente mais escolhido e em tempo mais curto do que quando ocupa a posição

imediatamente à direita do verbo. Se o resultado da contribuição das pistas de processamento é conforme à teoria do modelo de competição de frases, em que pistas mais disponíveis e mais locais como as morfológicas são mais válidas para conduzir a interpretação, já a preferência pela ordem VOS sobre VSO é problemática. Num estudo exploratório de *corpus* de português literário, em que se codificaram as ocorrências de ordem VNN em VOS e VSO (Costa 2005), verificou-se que a ordem VSO era mais frequente. Os resultados obtidos neste estudo, em que a ordem preferencial é a menos frequente, foram interpretados como refletindo preferência por uma representação gramatical mais simples. Apelando ao *Minimal Chain Principle* (De Vincenzi 1991), a opção por VOS poderá ser, num primeiro momento, decorrente da preferência por uma cadeia mínima: [*proVO(X)*]. Numa língua de sujeito nulo, em *Abraça a árvore...* ou *Abraça a rapariga...* as duas sequências são compatíveis com [*proVO*]. No decurso da leitura, se o segundo SN for um melhor candidato para objeto direto/Tema, então a primeira interpretação tem de ser abandonada, e VSO é atribuída, demorando mais tempo. Em resumo, as pistas morfológicas são mais robustas que as semânticas, o que vai ao encontro das predições do modelo de competição e de resultados encontrados para outras línguas. Já a preferência por uma estrutura sintaticamente mais simples e menos frequente parece ser melhor explicada por princípios gerais de *parsing* em favor de menos complexidade.

5.2 Processamento de frases por monolíngues portugueses e franceses

Kail/Costa/Faria (2010) conduziram um estudo comparativo sobre o processamento de frases, em que participaram indivíduos monolíngues, falantes de português e de francês, duas línguas românicas com propriedades distintas. O objetivo principal foi o de investigar o papel de algumas restrições contextuais no processamento auditivo de frases declarativas numa tarefa de juízos de agramaticalidade. Numa perspetiva interlinguística do processamento de frases, é importante saber que propriedades, em cada língua, afetam mais a integração da informação e de que forma o desenvolvimento do indivíduo pode afetar capacidades de deteção e integração da informação linguística. Partiu-se dos seguintes pressupostos: a natureza das pistas linguísticas (morfológica ou sintática) tem validade distinta no processamento, sendo que pistas mais locais (género e número) são mais válidas do que pistas menos locais (ordem de palavras); violações de gramaticalidade são menos onerosas para a memória de trabalho quando inseridas em unidades sintáticas simples (sintagmas) do que quando requerem relacionar unidades sintagmáticas; a perceção da agramaticalidade depende da posição em que ocorre: no final da frase, a expectativa criada pela informação contextual favorece condições para uma melhor deteção. Há assim três fatores experimentais: tipo de violação gramatical (*sintática ou morfológica*); posição estrutural (*dentro ou entre sintagmas*); posição na frase (*precoce ou tardia*).

Considerando a frase *A vizinha encheu o frigorífico com legumes frescos*, inseriram-se violações:

- (i) de concordância de género entre determinante e nome no interior do SN, como em *o vizinha*;
- (ii) de concordância de número entre sujeito/verbo, como em *a vizinha enchem*;
- (iii) de ordem de palavras no interior do SN, como em *vizinha a*;
- (iv) de ordem de palavras entre SN e SV, como em: *enche a vizinha o frigorífico*.

Note-se que nesta última condição a ordem de palavras não é agramatical em português europeu, apesar de precisar de um contexto discursivo, sintático ou até prosódico que a legitime.

No estudo participaram dois grupos: um de indivíduos franceses e outro de portugueses, testados em escolas no país respetivo. Cada grupo comportou quatro subgrupos de idade com 16 sujeitos – 6;8, 8;6, 10;10 e adultos – que foram testados individualmente, numa tarefa de decisão de gramaticalidade. As variáveis dependentes analisadas foram o tempo de reação em milissegundos e percentagens de respostas corretas.

Considerados os resultados entre línguas, há mais uniformidade no processamento do que diferenças. Em ambos os grupos há um efeito significativo de desenvolvimento: a capacidade de ajuizar sobre a boa/má formação da frase cresce progressivamente com a idade, embora esse efeito seja mais marcado nos falantes de francês. O tipo de violação gramatical não produziu efeito em francês em taxas de precisão, mas produziu diferenças significativas em tempos de reação, com a morfologia a ter vantagem sobre a sintaxe em todos os quatro grupos. Em português, houve maior precisão na deteção da violação morfológica do que na de ordem de palavras (esta reduzida à condição de violação da ordem determinante-nome, por a de verbo-sujeito não ser agramatical).

O lugar estrutural da agramaticalidade produziu efeitos não esperados. Em francês, as crianças reagem mais à agramaticalidade intrassintagma, efeito que se atenua com a idade. Em português, os resultados são semelhantes aos obtidos em francês em taxas de precisão; contudo, nos grupos das crianças, agramaticalidades intrassintagma (*o vizinha/vizinha a*) são detetadas mais lentamente do que quando ocorrem entre sintagmas (*a vizinha enchem*). Quanto à posição na frase, as taxas de precisão não são afetadas (não importa se o erro ocorre no início ou no final da frase), mas há efeitos de grande magnitude no tempo, com latências significativamente mais curtas na reação a violações tardias.

Em resumo, os resultados confirmaram que capacidades cognitivas de integração e utilização da informação contextual são similares entre grupos de língua: o fator posição explicou a maior taxa de variação, com a posição tardia a favorecer a deteção da agramaticalidade. O fator posição estrutural produziu efeitos diferentes entre línguas. Em francês, resolver problemas no interior do sintagma revela-se mais rápido do que resolvê-los quando isso implica relacionar unidades estruturais entre si. Já em português, as crianças têm mais dificuldades na deteção de problemas de concordância ou de ordem de palavras no interior de sintagmas. Atribui-se tais resultados a

efeitos da modalidade auditiva. Se nas duas línguas a morfologia fornece pistas robustas ao processamento de frases (confirmado neste estudo e em Costa 2005), é um facto que a modalidade auditiva relativiza a sua força. Veja-se que em português, para a marcação de género e número sobre o artigo definido temos *o(s)/a(s)*, enquanto que no francês temos: *le/la/les*, formas fonológicas mas marcadas. Em consequência, as violações de concordância ou de ordem de palavras no interior do sintagma nominal em português são menos perceptíveis auditivamente do que em francês. O estudo demonstra que capacidades cognitivas gerais relativas à integração da informação contextual são similares e progridem com o desenvolvimento, e revela que efeitos de perceptibilidade podem interagir com efeitos estruturais e com a natureza da informação gramatical.

5.3 Processamento de frases relativas ambíguas

O estudo da ambiguidade estrutural tem sido uma linha de investigação fértil em psicolinguística, por proporcionar a obtenção de evidências sobre as fontes linguísticas que alimentam o mecanismo de processamento e sobre a existência de princípios de *parsing* universais. Uma condição que tem sido exaustivamente investigada entre línguas é a exemplificada na frase 18 (adaptada de Cuetos/Mitchell 1988), em que há ambiguidade sobre o ponto de adjunção da oração relativa restritiva enquanto modificador do SN que a precede: pode modificar o primeiro nome (N1), núcleo do SN complexo, ou o segundo nome (N2), o nome mais encaixado.

(18) Alguém alvejou o criado da atriz que estava na varanda.

Late closure (ligação baixa, LB), princípio de *parsing* dito universal, prevê que ligações mais locais sejam mais simples do que ligações menos locais. Portanto, em (18), a interpretação de que a relativa modifica o nome mais encaixado (N2, *atriz*) será a preferencial. Contra evidências da atuação de *late closure* para o inglês e outras línguas, Cuetos/Mitchell (1988) mostram que em espanhol a preferência é por *ligação alta* (LA). Estes resultados, ao porem em questão a reclamada universalidade deste princípio, deram lugar a uma intensiva investigação em várias línguas que tenta identificar a fonte da variação (cf. Hemforth/Konieczny/Scheepers 2000 para uma revisão de estudos e teorias com resultados para várias línguas). Colocam-se várias hipóteses sobre o que pode explicar as diferenças entre línguas: a exposição à frequência de ligação baixa ou de ligação alta na produção enviesaria a preferência na compreensão (the *Tuning Hypothesis*, Mitchell/Cuetos 1991; Mitchell/Brysbart 1998); o peso fonológico da oração relativa, por efeitos da prosódia implícita, levaria a que orações relativas curtas levassem a ligação baixa ao constituir um só sintagma entoacional com o nome mais encaixado, enquanto relativas longas levariam a ligação alta (*Implicit Prosody Hypothesis*, Fodor 2002); a proximidade do predicador

da oração mais alta motivaria ligação alta (*Predicate proximity*, Gibson et al. 1996); o estatuto sintático da oração relativa, enquanto adjunto e não complemento, levaria a que fosse analisada numa fase mais tardia do *parsing*, permeável a informação semântico-discursiva (*construal*, Frazier/Clifton 1996); o estatuto sintático da frase relativa, genuína ou pseudo, determinaria o tipo de ligação, sendo que só relativas genuínas levariam a ligação alta (Grillo/Costa 2014). Há ainda hipóteses que atribuem a traços lexicais dos antecedentes das relativas a determinação da ligação alta ou baixa: nomes mais animados atraem a relativa como seu modificador (*The lexically driven parser*, Desmet et al. 2006). Do ponto de vista dos procedimentos experimentais, há evidências de efeitos de segmentação das frases na ligação alta ou baixa (Gilboy/Sopena 1996; Lourenço-Gomes/Lindemann 2012).

Em português realizaram-se vários estudos sobre o processamento de tais construções, com controlo das fontes de informação linguística disponíveis: sintática, morfológica, semântica e prosódica. Os estudos iniciaram-se com o português brasileiro e estenderam-se posteriormente ao português europeu. Maia et al. (2005) realizaram um primeiro estudo contrastivo, no sentido de verificarem até que ponto há convergência de interpretação destas frases ambíguas nas duas variedades do português, assumindo que estas se diferenciam por propriedades gramaticais que podem ter efeitos no *parsing* e na interpretação. Partindo-se da hipótese do efeito da prosódia implícita na desambiguação da frase, usaram-se duas condições de extensão: oração relativa curta ou longa:

- (19) O Alexandre fotografou a amiga da professora que cantava (nas festas da cidade).

Testou-se a hipótese de que a relativa curta induziria ligação mais local, enquanto a longa modificaria todo o SN anterior, levando a ligação alta. Fez-se um estudo de questionário em que os sujeitos (60 adultos em português europeu e 60 em português brasileiro) interpretavam a frase (*Quem cantava? a amiga / a professora*). Em ambos os grupos foi significativo o efeito do tamanho da relativa no sentido esperado: curta, ligação baixa; longa, ligação alta. Interpretaram-se estes resultados como consequência da prosódia implícita, como proposto por Fodor (2002), não havendo quaisquer diferenças entre as duas variedades do português.

Fez-se um segundo estudo comparativo entre as duas variedades do português, com recurso à técnica de leitura automonitorada para captação dos processos *on-line*. Foi mantido o fator *tamanho da oração relativa*, mas usaram-se frases com desambiguação por meio de informação morfológica de número (flexão verbal e nominal). Criou-se um paradigma com quatro condições, em que se força a ligação da relativa ao primeiro ou ao segundo nome (N1 ou N2) em função da concordância verbal da relativa:

- (20) A vítima reconheceu | os cúmplices do ladrão que fugiram (depois do assalto ao banco).

- (21) A vítima reconheceu | o cúmplice dos ladrões que fugiram (depois do assalto ao banco).

Cada frase foi apresentada em dois segmentos, como assinalado por barra vertical. As variáveis dependentes foram o tempo de leitura do segundo segmento e a precisão das respostas. Houve um significativo efeito de tamanho da oração relativa, mas sempre que a ligação era baixa (curta ou longa) os tempos foram significativamente mais curtos. Estes resultados confirmam a atuação de *late closure* em português. Quanto à precisão nas respostas, obtiveram-se percentagens mais elevadas de respostas corretas na ligação alta (N1). Sendo as respostas um indicador *off-line*, em que todos os tipos de informação linguística já foram integrados, podem tais resultados ser atribuídos a efeitos discursivos: o nome mais próximo do predicador verbal da oração matriz ganha saliência, o que leva a maior precisão nas respostas. Obtiveram-se ainda dois efeitos inesperados do fator Grupo: os falantes de português brasileiro foram significativamente mais lentos nas suas respostas e deram significativamente mais erros. Interpretaram-se estes resultados como consequência de um traço marcante diferenciador entre variáveis: em português brasileiro há enfraquecimento da flexão verbal e, portanto, menos saliência na marcação morfológica da concordância sujeito/verbo. Tal poderá explicar a maior lentidão na leitura para apreensão da informação do *input*, assim como a maior taxa de erros relativamente aos resultados do português europeu.

Fernandes (2015) conduziu um conjunto de experiências para avaliar efeitos de *priming estrutural* (Pickering/Branigan 1998; Branigan 2007). Testou o processamento de frases relativas ambíguas em que o antecedente da relativa era um SN complexo, podendo o pronome relativo tomar como seu antecedente o primeiro nome (N1) ou o segundo nome (N2). Desenhou uma experiência usando o paradigma do mundo visual (Tanenhaus et al. 1995) para determinar a magnitude do efeito de *priming* quando este é reforçado por dois *primes* contra um, assim como o efeito de aprendizagem através da quantidade de exposição aos tipos de *prime* durante a experiência. Como *primes* usou frases desambiguadas por número nas condições de ligação alta (LA) (frase 22) e de ligação baixa (LB) (frase 23), e como *alvo* frases cuja desambiguação é feita pragmaticamente (frases 24 e 25). A relação entre *prime* e *alvo* é de repetição da estrutura – LA/LA (22/24) ou LB/LB (23/25) – ou de estrutura diferente – LA/LB (22/25) ou LB/LA (23/24).

- (22) O irmão dos herdeiros que vai ler o testamento é de França.
- (23) Os irmãos do herdeiro que vai ler o testamento são de França.
- (24) O pai do bebé que vai beber a cerveja é alto.
- (25) O pai do bebé que vai beber o biberão é alto.

Os resultados revelaram efeitos clássicos de *priming* com *alvo* facilitado quando antecedido por um *prime* com a mesma estrutura, em consonância com resultados anteriores (Kamide 2012). Contudo, na condição de dois *primes* obteve-se *reversed-priming*: a

repetição do *prime* durante a experiência promove expectativa pela alternativa não *primada* (ativada) e inibe a procura pela estrutura *primada*. Estes resultados são explicados tentativamente como supressão do efeito de repetição (*repetition suppression*), um efeito que tem sido considerado em tarefas cognitivas de diferentes tipos e que pode estar associada a uma redução da atividade neural perante sobre-exposição a um estímulo ou refletir processos menos automáticos (Fernandes 2015, 101).

Fez-se ainda um estudo em que se controlaram aspetos relativos à memória de trabalho dos participantes, no processamento de relativas ambíguas (Fernandes/Costa/Coco 2015). Com efeito, os resultados apontam para diferenças individuais na utilização da informação obtida durante a leitura dos *primes* para a resolução da ambiguidade no alvo: indivíduos com alta capacidade de memória de trabalho prestam mais atenção aos antecedentes correspondentes à alternativa não *primada*, sugerindo que são mais capazes de manter as duas alternativas de interpretação *online* para uma tomada de decisão final. Estes resultados vão ao encontro da *teoria da capacidade da memória de trabalho na compreensão*, proposta por Just/Carpenter (1992) e, de algum modo, contribuem para a argumentação a favor de um processamento em paralelo.

Em síntese, os trabalhos realizados com relativas ambíguas em português europeu e português brasileiro vão no sentido de sustentar a hipótese de que o *parser* sintático procede com apoio em princípios universais: na condição de ambiguidade de adjunção da oração relativa restritiva a uma posição mais ou menos encaixada, os dados do processamento *on-line* confirmam a atuação de ligação baixa, com ligação em tempo significativamente mais baixo do que a ligação alta. Também se confirma um efeito de prosódia implícita quando se manipula o tamanho do modificador relativo: oração curta integra-se no SN anterior formando um único constituinte entoacional, levando a ligação baixa. Por efeitos da desambiguação por flexão nominal ou verbal, assinalando a concordância sujeito/verbo, colheram-se dados que indicam que portugueses e brasileiros estão a integrar de forma diferente a informação de flexão para a estruturação e interpretação da frase.

5.4 Processamento de cadeias anafóricas e correferenciais

Uma linha de investigação em psicolinguística bastante produtiva tem sido a que se interessa pela resolução de cadeias anafóricas, sobretudo quando se colocam problemas de ambiguidade, de competição entre possíveis antecedentes ou de distância entre membros da cadeia. A gramática tem princípios bem definidos que regulam estas relações, sendo a *Teoria da Ligação* (Chomsky 1981) a que explicita as condições para a interpretação de anáforas em sentido estrito, pronomes e expressões referenciais (↗12 Semântica e pragmática). O que se estuda no processamento é de que forma a codificação das expressões referenciais, lexicais ou pronominais, pode criar restrições à interpretação não explicadas pela gramática. Há vários modelos para o processa-

mento de relações anafóricas, que tentam explicar os processos linguísticos necessários e a informação linguística em jogo. Ariel (1996; 2001) propõe a Teoria da Acessibilidade (*The Accessibility Marking Scale*). A sua proposta sugere que há uma relação inversa entre o grau de informatividade da expressão anafórica e a saliência discursiva da entidade que refere: expressões anafóricas menos informativas e mais atenuadas fonologicamente induzem a procura de entidades salientes no discurso, como tópicos e sujeitos. Grosz/Joshi/Weinstein (1995) desenvolvem a Teoria da Centralidade (*Centering Theory*), no âmbito da linguística computacional, propondo um algoritmo que integra informação relativa a saliência estrutural e discursiva: uma entidade referida no discurso torna-se um centro de atenção e, por questões de coerência discursiva, deve ser retomada por expressões anafóricas menos informativas. Na economia de pronomes/expressões lexicais repetidas ou relacionadas, os primeiros são adequados para referir sujeitos frásicos e tópicos (paradigmáticos de centros de atenção), enquanto expressões lexicais são mais adequadas para retomar entidades menos proeminentes. Há mesmo a previsão de uma penalidade no processamento quando uma entidade saliente é retomada por uma expressão lexical repetida (*repeated name penalty*). Reagindo às estipulações da Teoria da Centralidade, Almor (2000) propõe a Hipótese da Carga Informacional (*Informational Load Hypothesis*), postulando que a retoma ótima de uma entidade pode levar ao uso de expressões correferenciais lexicais, desde que se atenda às relações de maior ou menor especificidade na contiguidade semântica; uma retoma por expressão lexical semanticamente relacionada pode não ser penalizadora e até ser conceptualmente adequada (por exemplo, no caso do hiperônimo que retoma anaforicamente uma categoria semanticamente subordinada, como no par *canário/ave*). Também ao nível do discurso, Gernsbacher/Hargreaves (1988) predizem que a primeira entidade referida estará em vantagem para servir de antecedente a uma expressão referencial subsequente, visto que assume o papel de tópico através do qual se integra a informação posterior (*advantage of first-mention*). Numa perspetiva estrutural, Carminati (2002) propõe a Hipótese da Posição do Antecedente (*The Position of Antecedent Hypothesis*), que estipula que numa língua como o italiano, com alternância pronome nulo/lexicalizado, há uma divisão de trabalho para as duas formas pronominais: a forma nula retoma o constituinte que ocupa a posição mais alta na frase, o lugar de especificador de Flexão, enquanto a forma lexicalizada retoma constituintes em posição estrutural mais baixa (Carminati 2002, 51). Em português europeu, na marcação de cadeias correferenciais, é produtivo o designado *Princípio Evitar Pronome*, princípio associado à propriedade do sujeito nulo, que tem sido objeto de análise e considerado quer um princípio da gramática quer um princípio discursivo (Chomsky 1981; Brito 1991; Costa 2005).

Têm-se desenvolvido estudos que tiram partido das propriedades do português europeu, no que tem de relevante para o processamento de cadeias anafóricas. Costa/Faria/Matos (1998) testaram a interpretação de cadeias correferenciais em frases complexas coordenadas e subordinadas. A variável independente foi a forma do pronome, nulo ou pleno, do sujeito da segunda oração.

(26) A Marta viu a Joana mas [-] / ela não a cumprimentou.

(27) A Marta apresentou a Joana à Inês mas [-] / ela não lhe prestou atenção.

Usou-se uma tarefa de resposta a questionário em tempo ilimitado (uma frase/uma pergunta), em que participaram sujeitos adultos universitários de língua materna portuguesa. Os resultados obtidos confirmaram uma preferência regular pelo pronome nulo para retomar sujeito e pleno para retomar objeto. Contudo, verificou-se assimetria na funcionalidade das formas pronominais: nulo sistematicamente retoma sujeito, mas pleno também retoma sujeito de forma não negligenciável. Por outro lado há alguns efeitos de traços de animacidade controlados no objeto da primeira oração: animado é retomado por pronome pleno, não animado por pronome pleno mas também por nulo.

Costa (2005) fez um estudo em que se avaliou a interação entre fontes de informação linguística distintas na interpretação de cadeias anafóricas cujos membros são sujeitos frásicos. Assumindo a funcionalidade do Princípio Evitar Pronome na resolução de cadeias anafóricas, adiciona-se agora o concurso de informação semântica relativa à grelha argumental do verbo da frase onde se localiza o antecedente. Usaram-se frases complexas adverbiais causais, em que na subordinante ocorriam verbos de causalidade implícita. Trata-se de verbos que atribuem o papel de causador a um dos seus argumentos: ao sujeito, como em *assustar*, ao objeto como em *temer*. Pôs-se a hipótese de que o papel temático atribuído modificaria a saliência semântica do argumento, o que afetaria a forma da sua retoma anafórica, frases (28) a (30).

(28) No congresso | a Maria | assustou | a Sofia | porque [-]/ela fez * | um bom discurso.

(29) No congresso | a Maria | admirou | a Sofia | porque [-]/ela fez * | um bom discurso.

(30) No congresso | a Maria | encontrou | a Sofia | porque [-]/ela estuda * | o mesmo assunto.

Usou-se uma técnica de *palavra sonda* e de *apresentação serial visual rápida* da frase, segmentada como indicado pelas barras verticais; o asterisco indica o ponto de apresentação da *sonda* (*Maria* ou *Sofia* ou outro). Os resultados obtidos apontaram para um efeito claro da forma do pronome: pronome nulo é mais adequado para retoma do sujeito, o que se manifestou em tempos mais curtos que os obtidos com pronome pleno. Há impacto da informação de causalidade do verbo: sujeito/causador é melhor antecedente para o sujeito pronominal da oração causal, mais ainda quando este é um pronome nulo (28). Na condição de controlo ou *base-line* (30), em que está ausente a informação semântica de causalidade que cria restrições à interpretação, o sujeito da matriz inequivocamente correferre com pronome nulo e é disjunto do pronome pleno, confirmando-se a produtividade do Princípio Evitar Pronome em português europeu.

Em Luegi (2012a) e Luegi/Costa/Maia (2014), apresentam-se resultados de estudos sobre correferência para controlo do efeito da forma pronominal nula ou plena, mas

agora incorporando uma condição ainda não testada: a da posição estrutural do possível antecedente para o pronome anafórico. Usaram-se duas ordens de palavras na frase matriz: SVO, em que os antecedentes possíveis para o sujeito pronominal de frase adverbial à direita estão na sua posição básica; OVS, em que o complemento oblíquo do verbo é topicalizado e o sujeito ocupa a posição pós-verbal. Testaram-se frases ambíguas na interpretação da cadeia correferencial e frases desambiguadas morfologicamente por traços de gênero (frases 31 a 34).

- (31) A florista negociou com a peixeira no mercado quando [-]/ela pediu a nova banca junto à entrada.
- (32) Com a peixeira negociou a florista no mercado quando [-]/ela pediu a nova banca junto à entrada.
- (33) O André trabalhou para a Pilar no verão quando [-]/ele(a) foi promovido(a) para a comissão.
- (34) Para a Pilar trabalhou o André no verão quando [-]/ele(a) foi promovido(a) para a comissão.

Realizou-se um conjunto de três experiências para a obtenção de medidas em tempo real com recurso a leitura automonitorada, registo do movimento dos olhos e paradigma do mundo visual (no caso, audição de frases e visualização simultânea de um conjunto de imagens representativas das duas entidades referidas na frase matriz mais o espaço do evento referido). Na experiência de leitura automonitorada, estudou-se contrastivamente o desempenho de participantes falantes de português europeu e português brasileiro no processamento de frases não ambíguas com retoma exclusivamente por forma nula; verificou-se uma preferência por nulos para retomar sujeitos em posição básica em português brasileiro, mas não em português europeu, onde o pronome nulo serve igualmente para retomar qualquer dos antecedentes em qualquer posição. Numa experiência com registo do movimento dos olhos apenas em português europeu, com as mesmas frases ambíguas mas com alternância na forma pronominal da retoma anafórica, os resultados mostraram uma tendência para o pronome nulo retomar sujeitos e a forma plena indiciar inadequação para retomar sujeitos em posição básica ou em posição pós-verbal. Os dados da experiência com paradigma do mundo visual, teoricamente a técnica que captará mais finamente os processos envolvidos, mostraram que a forma do pronome funciona distintamente no decurso temporal da atribuição da referência. O pronome nulo só mostrou efeitos significativos de preferência por um antecedente (por sujeito pré-verbal e por oblíquo topicalizado) em fases tardias, a partir dos 800 milissegundos. Pelo contrário, o pronome pleno produziu efeitos de preferência pelo nome mais encaixado (por sujeito em posição pós-verbal ou por oblíquo em posição básica) apenas na fase inicial, entre os 100 e os 200 milissegundos. Esta reatividade da forma do pronome ao antecedente em função da janela temporal de observação permite pôr a hipótese de que a contribuição do Princípio Evitar Pronome seja distinta em diferentes fases do processamento. Enquanto o nulo serviria como retoma para qualquer antecedente acessível

na frase precedente durante a fase de estruturação da cadeia correferencial, o pronome pleno assinalaria desde o início a desvinculação pelo antecedente mais saliente, ao preferir um antecedente mais encaixado estruturalmente. Considerando estes dados do processamento, poderá discutir-se a natureza do Princípio Evitar Pronome, como enunciado em Chomsky (1981) e retomado por Brito (1991) para o português. Se de facto a forma do pronome parece ter uma função mais decisiva na indexação ao antecedente em fases tardias do processamento, o princípio seria mais relevante na interpretação, onde a informação do discurso é integrada.

Dos resultados contrastivos de português europeu e português brasileiro infere-se que na variedade brasileira o Princípio Evitar Pronome perdeu produtividade, não funcionando como pista robusta para o estabelecimento de ligações anafóricas: há uma maior tendência para interpretar a forma plena como retomando o sujeito, sendo a oposição nulo/pleno menos acentuada. Tais resultados deverão ser discutidos à luz das mudanças na gramática do português brasileiro.

Numa tentativa de se avaliar efeitos sintático-semânticos na interpretação da correferência no português europeu, fez-se um estudo controlando relações interfrásicas, usando cadeias em que antecedente e pronome estão em frases distintas separadas por pontuação, e em que não há qualquer informação que desambigue a interpretação. Adicionalmente, pelo contraste entre forma ativa e passiva, controlou-se o efeito do papel temático do possível antecedente na resolução da retoma anafórica pronominal (Morgado 2012).

(35) O Samuel_{SU/Agente} agrediu o Bruno_{OD/Tema} no pavilhão. Horas mais tarde, Ø/ele discutiu o assunto com preocupação.

(36) O Samuel_{SU/Tema} foi agredido pelo Bruno_{OD/Agente} no pavilhão. Horas mais tarde, Ø/ele discutiu o assunto com preocupação.

Através de uma tarefa de leitura em tempo limitado e escolha de antecedente para o pronome por questionário, recolheram-se valores de tempo de leitura e de taxas de precisão nas respostas. Os resultados mostram que o sujeito da primeira frase é retomado em tempo significativamente mais baixo e com percentagens de escolha significativamente mais altas do que a retoma do objeto. Não se encontraram diferenças entre a utilização da forma nula ou plena do pronome anafórico. Estes resultados mostram que, mesmo quando os dois membros da cadeia estão separados por fronteira discursiva, a posição estrutural mais alta continua a ser a pista de processamento mais saliente, e que essa saliência, sinal de acessibilidade, não determina a forma da expressão anafórica. Contudo, note-se que a retoma de sujeito da frase precedente por forma plena de pronome induz tempos mais baixos do que quando retomado por pronome nulo, de onde se pode depreender que a fronteira por pontuação contribui para a diminuição da acessibilidade do antecedente, mesmo quando este ocupa a posição estrutural mais alta. Quanto ao efeito do papel temático apenas se encontraram resultados em tempo significativamente mais curto para retomar sujeito/Agente.

No mesmo grupo de estudos, Morgado (2012) testa correferência entre sujeitos fráscos em condições intrafrásicas com subordinadas concessivas. Os resultados com frases ativas replicam os encontrados com frases em que a cadeia se estabelece entre um sujeito pronominal anafórico, precedido por uma frase subordinada ou coordenada com verbo transitivo e na ordem canônica (SVO): pronome nulo prefere sujeito, pronome pleno prefere objeto. Já com frases passivas, o papel semântico do antecedente modifica a especialização das formas pronominais. Concluiu-se que o pronome nulo é sensível a fatores exclusivamente sintáticos, retomando o sujeito, Agente ou Tema, enquanto que a forma plena se mostrou mais adequada para retomar o papel semântico menos saliente em termos de uma hierarquia temática: Tema, seja sujeito ou objeto.

Em síntese, os vários estudos que foram referidos sobre a interpretação de relações anafóricas em português controlaram o efeito de informação linguística na atribuição de conteúdo referencial a um pronome sujeito, cujo antecedente está na frase precedente, em condições intra ou interfrásicas. Quando apenas informação sintática está disponível, os resultados dos vários estudos vão no sentido dos resultados encontrados para outras línguas românicas com alternância nulo/pleno na forma pronominal de sujeito fráscico. Confirma-se a hipótese de que um sujeito anafórico pronominal codifica a acessibilidade do seu antecedente: a forma nula correferre com o sujeito da frase precedente, a forma plena codifica a coindexação com um antecedente mais encaixado. Tais resultados confirmam a produtividade em português europeu do Princípio Evitar Pronome (Chomsky 1981), e dão sustentação à hipótese da posição estrutural do antecedente (Carminati 2002). Mas quando, numa hipótese sobre o processamento que tente dar conta do concurso de outras fontes de informação linguística que não as exclusivamente sintáticas, se faz entrar em jogo outros níveis de informação e se controla o tempo de entrada dessa informação (em tarefas *off-* e *on-line*), os resultados permitem descobrir problemas, colocar hipóteses para os equacionar e abrir novas janelas de observação sobre os processos mentais intervenientes na compreensão da linguagem.

Os estudos sobre processamento que foram referidos a título ilustrativo (entre outros igualmente interessantes e inovadores, de entre os quais podemos destacar os trabalhos de investigação sobre leitura, apresentados em Costa 2012, ou o estudo de Falé/Costa/Luegi 2015) mostram que há propriedades na gramática da língua portuguesa que favorecem uma discussão alargada sobre a relação entre gramática e processamento. Considerando as propriedades específicas do português europeu, podemos avaliar o seu impacto na adoção de estratégias de processamento e discutí-las no quadro vasto da universalidade do mecanismo humano para o processamento da linguagem.

6 Referências

- Almor, Almit (2000), *Constraints and mechanisms in theories of anaphor processing*, in: Matthew Crocker/Martin Pickering/Charles Clifton (edd.), *Architectures and Mechanisms for Language Processing*, Cambridge, Cambridge University Press, 341–354.
- Altmann, Gerry/Steedman, Mark (1988), *Interaction with context during human sentence processing*, *Cognition* 30, 191–238.
- Ariel, Mira (1996), *Referring expressions and the +/- coreference distinction*, in: Jeanette Gundel/Thorstein Fretheim (edd.), *Reference and Referent Accessibility*, Amsterdam/Philadelphia, Benjamins, 223–259.
- Ariel, Mira (2001), *Accessibility theory: An overview*, in: Ted Sanders/Joost Schliperoord/Wilbert Spooen (edd.), *Text representation: Linguistic and psycholinguistic aspects*, Amsterdam/Philadelphia, Benjamins, 29–87.
- Bates, Elisabeth/MacWhinney, Brian (1987), *Competition, variation, and language learning*, in: Brian MacWhinney (ed.), *Mechanisms of language acquisition*, Hillsdale, NJ, Erlbaum, 157–193.
- Bates, Elisabeth/MacWhinney, Brian (1989), *Functionalism and the competition model*, in: Brian MacWhinney/Elisabeth Bates (edd.), *The cross-linguistic study of sentence processing*, New York, Cambridge University Press, 3–76.
- Berwick, Robert, et al. (2013), *Evolution, brain, and the nature of language*, *Trends in Cognitive Sciences* 17:2, 89–98.
- Bever, Thomas (1970), *The cognitive bases for language structures*, in: John R. Hayes (ed.), *Cognition and Language Development*, New York, Wiley, 279–362.
- Boland, Julie (1993), *The role of verb argument structure in sentence processing: Distinguishing between syntactic and semantic effects*, *Journal of Psycholinguistic Research* 22:2, 133–152.
- Branigan, Holly (2007), *Syntactic Priming*, *Language and Linguistics Compass* 1:1–2, 1–16.
- Brito, Ana Maria (1991), *Ligação, correferência e o princípio evitar o pronome*, in: Ana Brito (ed.), *Encontro de Homenagem a Óscar Lopes*, Maia, Associação Portuguesa de Linguística, 101–121.
- Bruner, Jerome/Goodenough, Jacqueline/Austin, George (1956), *A Study of Thinking*, New York, Wiley.
- Carminati, Maria Nella (2002), *The processing of Italian subject pronouns*, Doctoral Dissertations available from Proquest, Paper AAI3039345, <http://scholarworks.umass.edu/dissertations/AAI3039345> (30.01.2016).
- Carreiras, Manuel/Clifton Jr., Charles (edd.) (2004), *The on-line study of sentence comprehension: Eyetracking, ERPs and Beyond*, New York, Hove, Psychology Press.
- Chomsky, Noam (1957), *Syntactic Structures*, 's-Gravenhage, Mouton and Co.
- Chomsky, Noam (1959), *Review of Skinner's «Verbal Behavior»*, *Language* 35, 26–58.
- Chomsky, Noam (1965), *Aspects of the Theory of Syntax*, Cambridge MA, MIT Press.
- Chomsky, Noam (1981), *Lectures on Government and Binding*, Dordrecht, Foris.
- Chomsky, Noam (1988), *Language and Problems of Knowledge: The Managua Lectures*, Cambridge MA, MIT Press.
- Chomsky, Noam (2005), *Universals of Human Nature*, *Psychotherapy and Psychosomatics* 74, 263–268.
- Chomsky, Noam (2007), *Of Minds and Language*, *Biolinguistics* 1, 9–27, <http://www.biolinguistics.eu>.
- Chomsky, Noam (2015), *Some Core Contested Concepts*, *Journal of Psycholinguistic Research* 44, 91–104.
- Clifton, Charles (2000), *Evaluating models of human sentence processing*, in: Matthew W. Crocker/Martin Pickering/Charles Clifton (edd.), *Architectures and mechanisms for language processing*, Cambridge, Cambridge University Press.
- Costa, Armanda (2005), *Processamento de frases em Português Europeu, Aspectos cognitivos e linguísticos implicados na compreensão da língua escrita*, Lisboa, Fundação Calouste Gulbenkian.

- Costa, Armanda (2012), *Investigação em Leitura – estudos psicolinguísticos*, in: Armanda Costa/Inês Duarte (edd.), *Nada na linguagem lhe é estranho. Homenagem a Isabel Hub Faria*, Porto, Afrontamento, 149–165.
- Costa, Armanda/Faria, Isabel Hub/Matos, Gabriela (1998), *Ambiguidade referencial na identificação do sujeito em estruturas coordenadas*, in: *Actas do XIII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 173–188.
- Crain, Stephen/Steedman, Mark (1985), *On not being led up the garden path: The use of context by the psychological syntax processor*, in: David R. Dowty/Lauri Karttunen/Arnold M. Zwicky (edd.), *Natural language parsing: Psychological, computational and theoretical perspectives*, Cambridge, Cambridge University Press, 320–358.
- Cuetos, Fernando/Mitchell, Don C. (1988), *Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish*, *Cognition* 30, 73–105.
- Cummings, Louise (2008), *Clinical Linguistics*, Edinburgh, Edinburgh University Press Ltd.
- Cutler, Anne/Clifton Jr., Charles (1999), *Comprehending Spoken Language: a blueprint of the listener*, in: Colin M. Brown/Peter Hagoort (edd.), *The Neurocognition of Language*, Oxford, Oxford University Press, 123–166.
- De Vincenzi, Marica (1991), *Syntactic parsing strategies in Italian*, Dordrecht, Kluwer.
- Dehaene, Stanislas/Cohen, Laurent (2011), *The unique role of the visual word form area in reading*, *Trends in Cognitive Sciences* 15:6, 254–262.
- Desmet, Timothy, et al. (2006), *Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account*, *Language and Cognitive Processes* 21:4, 453–485.
- Falé, Isabel/Costa, Armanda/Luegi, Paula (2015), *Leitura em voz alta, movimentos oculares e prosódia. Integração de informação sintática e discursiva*, in: *Textos Seleccionados do XXX Encontro Nacional da Associação Portuguesa de Linguística*, Braga, APL, 261–275.
- Fernandes, Eunice (2015), *Syntactic Priming as a Window into the Representational and Experiential Basis of Syntactic Processing in Comprehension*, tese de Doutoramento em Ciência Cognitiva, Universidade de Lisboa, <http://repositorio.ul.pt/handle/10451/20606>.
- Fernandes, Eunice/Costa, Armanda/Coco, Moreno (2015), *Bridging Mechanisms of Reading, Viewing and Working Memory during Attachment Resolution of Ambiguous Relative Clauses*, in: Gabriella Airenti/Bruno Bara/Giulio Sandini (edd.), *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science*, 264–269, <http://ceur-ws.org/Vol-1419/> (30.01.2016).
- Ferreira, Fernanda (2005), *Psycholinguistics, formal grammars, and cognitive science*, *The Linguistic Review* 22:2–4, 365–380.
- Fodor, Janet Dean (2002), *Psycholinguistics cannot escape prosody*, in: *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence, Université de Provence, 83–88.
- Fodor, Jerry (1983), *Modularity of mind*, Cambridge MA, MIT Press.
- Forster, Ken (1979), *Levels of processing and the structure of the language processor*, in: William Cooper/Edward Walker (edd.), *Sentence Processing*, Hillsdale, NJ, Erlbaum, 27–86.
- Frazier, Lyn (1985), *Syntactic complexity*, in: David Dowty/Lauri Karttunen/Arnold Zwicky (edd.), *Natural Language Parsing. Psychological, computational, and theoretical perspectives*, Cambridge, Cambridge University Press, 129–189.
- Frazier, Lyn (1988), *Grammar and language processing*, in: Frederick Newmeyer (ed.), *Linguistics: The Cambridge Survey. II Linguistic Theory: Extensions and Implications*, Cambridge, Cambridge University Press, 15–34.
- Frazier, Lyn (1990), *Exploring the architecture of the language system*, in: Gerry Altmann (ed.), *Cognitive models of speech processing, Psycholinguistic and Computational Perspectives*, Cambridge MA, MIT Press, 409–433.
- Frazier, Lyn/Clifton Jr., Charles (1996), *Construal*, Cambridge MA, MIT Press.

- Frazier, Lyn/Fodor, Janet Dean (1978), *The sausage machine: A new two-stage parsing model*, *Cognition* 6:4, 291–325.
- Frazier, Lyn/Rayner, Keith (1982), *Making and correcting errors during sentence comprehension. Eye movements in the analysis of structurally ambiguous sentences*, *Cognitive Psychology* 14:2, 178–210.
- Gardner, Howard (1985), *The mind's new science: A history of the cognitive revolution*, New York, Basic Books.
- Garrod, Simon (2006), *Psycholinguistic Research Methods*, in: Keith Brown (ed.), *Encyclopedia of Language & Linguistics*, second edition, vol. 10, New York, Elsevier, 251–257.
- Gernsbacher, Morton Ann/Hargreaves, David J. (1988), *Accessing sentence participants: The advantage of first mention*, *Journal of Memory and Language* 27:6, 699–717.
- Gibson, Edward, et al. (1996), *Recency preference in the human sentence processing mechanism*, *Cognition* 59, 23–59.
- Gilboy, Elizabeth/Sopena, Josep (1996), *Segmentation effects in the processing of complex NPs with relative clauses*, in: Manuel Carreiras/José Garcia-Albea/Núria Sabastián-Gallés (edd.), *Language processing in Spanish*, Mahwah, NJ, Erlbaum, 145–187.
- Grillo, Nino/Costa, João (2014), *A novel argument for the universality of parsing principles*, *Cognition* 133:1, 156–187.
- Grosz, Barbara/Joshi, Aravind K./Weinstein, Scott (1995), *Centering: A framework for modeling the local coherence of discourse*, *Computational Linguistics* 21, 203–225.
- Harris, Zellig S. (1951), *Methods of Structural Linguistics*, Chicago, University of Chicago Press.
- Hauser, Mark/Chomsky, Noam/Fitch, Tecumseh (2002), *The faculty of language: what is it, who has it, and how did it evolve?*, *Science* 298, 1569–1570.
- Hemforth, Barbara/Konieczny, Lars/Scheepers, Christoph (2000), *Syntactic Attachment and Anaphor Resolution: The Two Sides of Relative Clause Attachment*, in: Matthew Crocker/Martin Pickering/Charles Clifton (edd.), *Architectures and Mechanisms for Language Processing*, Cambridge, Cambridge University Press, 259–281.
- Ingram, John C. L. (2007), *Neurolinguistics. An Introduction to Spoken Language Processing and its Disorders*, Cambridge, Cambridge University Press.
- Just, Marcel/Carpenter, Patricia (1992), *Working memory constraints on the processing of syntactic ambiguity*, *Cognitive Psychology* 24, 56–98.
- Just, Marcel/Carpenter, Patricia/Woolley, Jacqueline (1982), *Paradigms and processes in reading comprehension*, *Journal of Experimental Psychology General* 111, 228–238.
- Kail, Michele (1999), *Linguistic variation and cognitive constraints in the processing and the acquisition of language*, in: Catherine Fuchs/Stéphane Robert (edd.), *Language diversity and cognitive representations*, Amsterdam/Philadelphia, Benjamins, 179–194.
- Kail, Michele/Costa, Armanda/Faria, Isabel Hub (2010), *On-line grammaticality judgments: A comparative study of French and Portuguese*, in: Michele Kail/Maya Hickmann (edd.), *Language acquisition across linguistic and cognitive systems*, Amsterdam/Philadelphia, Benjamins, 179–203.
- Kamide, Yuki (2012), *Learning individual talkers' structural preferences*, *Cognition* 124:1, 66–71.
- Keller, Frank (2010), *Cognitively Plausible Models of Human Language Processing*, in: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Short Papers*, Uppsala, Uppsala Universitet, 60–67.
- Levelt, Willem (1999), *Producing spoken language: a blueprint of the speaker*, in: Colin M. Brown/Peter Hagoort (edd.), *The Neurocognition of Language*, Oxford, Oxford University Press, 83–122.
- Lewis, Shevaun/Phillips, Colin (2015), *Aligning grammatical theories and language processing models*, *Journal of Psycholinguist Research* 44:1, 27–46.
- Lourenço-Gomes, Maria do Carmo/Lindemann, Verena (2012), *Interações entre estratégia de desambiguação e modo de segmentação de sentenças em self-paced reading*, in: Armanda Costa/Pilar

- Barbosa/Isabel Falé (orgs.), *Textos Selecionados do XXVII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 360–375.
- Luegi, Paula (2012a), *Processamento de sujeitos pronominais em Português: efeito da posição estrutural dos antecedentes*, tese de Doutorado, Universidade de Lisboa, <http://repositorio.ul.pt/handle/10451/7531>.
- Luegi, Paula (2012b), *Experiências psicolinguísticas: registar tempos e comportamentos no processamento de pronomes*, in: Armanda Costa/Inês Duarte (orgs.), *Nada na linguagem lhe é estranho. Homenagem a Isabel Hub Faria*, Porto, Afrontamento, 673–688.
- Luegi, Paula/Costa, Armanda/Faria, Isabel Hub (2010), *Analizando os comportamentos oculares durante a leitura*, *Revista Linguística* 5:1, 62–80.
- Luegi, Paula/Costa, Armanda/Maia, Marcus (2014), *Processamento e interpretação de sujeitos nulos e plenos em Português Europeu e em Português do Brasil*, *Cadernos de Letras da UFF* 24:49, 67–88.
- Maia, Marcus, et al. (2005), *A compreensão de orações relativas ambíguas em Português Brasileiro e Europeu – um estudo comparativo*, *Revista da Abralín* 3:1–2, 11–39.
- Maia, Marcus, et al. (2007), *Early and late preferences in relative clause attachment in Portuguese and Spanish*, *Journal of Portuguese Linguistics* 5–6, 227–250.
- Marinkovic, Ksenija (2004), *Spatiotemporal Dynamics of Word Processing in the Human Cortex*, *The Neuroscientist* 10:2, 142–152.
- Miller, George A. (1956), *The Magical Number Seven, Plus or Minus Two. Some Limits on Our Capacity for Processing Information*, *Psychological Review* 101:2, 343–352.
- Miller, George A. (1962), *Some psychological studies of grammar*, *American Psychologist* 7, 748–762.
- Miller, George A. (2003), *The cognitive revolution: a historical perspective*, *Trends in Cognitive Sciences* 7:3, 141–144.
- Minsky, Marvin (1961), *Steps toward artificial intelligence*, *Proceedings of IRE* 49, 8–29.
- Mitchell, Don C./Brysbaert, Marc (1998), *Challenges to recent theories of language differences in parsing: Evidence from Dutch*, in: Dieter Hillert (ed.), *Syntax and Semantics. Sentence processing: A Crosslinguistic Perspective*, San Diego, Academic Press, 313–335.
- Mitchell, Don C./Cuetos, Fernando (1991), *The origins of parsing strategies*, in: Carlota S. Smith (ed.), *Current Issues in Natural Language Processing*, Center for Cognitive Science, Austin, University of Texas Press, 1–12.
- Morgado, Sara (2012), *Qual a importância da informação semântica na resolução de sujeitos pronominais no processamento «offline»?.*, in: Fátima Silva/Isabel Falé/Isabel Pereira (edd.), *Textos Selecionados. XXVIII Encontro Nacional da APL*, Coimbra, APL, 493–507.
- Newell, Allen/Shaw, John Calman/Simon, Herbert A. (1958), *Elements of a theory of human problem solving*, *Psychological Review* 65, 151–166.
- Norcliffe, Elisabeth/Harris, Alice/Jaeger, T. Florian (2015), *Cross-linguistic psycholinguistics and its critical role in theory development: early beginnings and recent advances*, *Language, Cognition and Neuroscience* 30:9, 1009–1032.
- Osgood, Charles/Sebeok, Thomas (edd.) (1954, 1965), *Psycholinguistics: A Survey of Theory and Research Problems*, supplement to *International Journal of American Linguistics* 20, Bloomington, Indiana University Press.
- Osterhout, Lee/Swinney, David (1989), *On the Role of the Simplicity Heuristic in Language Processing: Evidence from Structural and Inferential Processing*, *Journal of Psycholinguistic Research* 18:6, 553–562.
- Pickering, Martin (1999), *Sentence comprehension*, in: Simon Garrod/Martin Pickering (edd.), *Language comprehension*, Hove, Psychology Press, 123–153.
- Pickering, Martin J./Branigan, Holly P. (1998), *The representation of verbs: evidence from syntactic priming in language production*, *Journal of Memory and Language* 39:4, 633–651.

- Rayner, Keith (1998), *Eye Movements in Reading and Information Processing: 20 Years of Research*, Psychological Bulletin 124:3, 372–422.
- Rayner, Keith/Slattery, Timothy (2009), *Eye Movements and Moment-to-Moment Comprehension Processes in Reading*, in: Richard Wagner/Christopher Schatschneider/Caroline Phythian-Sence (edd.), *Beyond Decoding: The behavioral and biological foundations of reading comprehension*, New York, Guilford Press, 241–255.
- Rumelhart, David E./McClelland, James L. (1986), *On learning the past tenses of English verbs*, in: James L. McClelland/David E. Rumelhart/PDP Research Group, *Parallel distributed processing: Explorations in the microstructures of cognition*, vol. 2: *Psychological and biological models*, Cambridge MA, MIT Press, 216–271.
- Sandra, Dominiek (2009), *Psycholinguistics*, in: Dominiek Sandra/Jan-Ola Östman/Jef Verschueren (edd.), *Cognition and Pragmatics*, Amsterdam/Philadelphia, Benjamins, 288–368.
- Sekerina, Irina/Fernández, Eva/Clahsen, Harald (2008), *Developmental psycholinguistics: on-line methods in children's language processing*, Amsterdam/Philadelphia, Benjamins.
- Shannon, Claude E. (1948), *A mathematical theory of communication*, Bells Systems Technical Journal 27, 379–423, 623–656.
- Shannon, Claude E./McCarthy, John (edd.) (1956), *Automata Studies*, Princeton, Princeton University Press.
- Skinner, Burrus F. (1957), *Verbal Behavior*, New York, Appleton-Century-Crofts.
- Tanenhaus, Michael K./Carlson, Greg/Trueswell, John C. (1989), *The role of thematic structures in interpretation and parsing*, Language and Cognitive Processes 4:3–4, 211–234.
- Tanenhaus, Michael K./Garnsey, Susan M./Boland, Julie (1990), *Combinatory lexical information and language comprehension*, in: Gerry T. M. Altmann (ed.), *Cognitive models of speech processing. Psycholinguistic and computational perspectives*, Cambridge MA, MIT Press.
- Tanenhaus, Michael K., et al. (1995), *Integration of visual and linguistic information in spoken language comprehension*, Science 268, 1632–1634.
- Townsend, David J./Bever, Thomas G. (2001), *Sentence interpretation. The integration of habits and rules*, Cambridge MA, MIT Press.
- Turing, Alan M. (1950), *Computing machinery and intelligence*, Mind 59, 433–460.
- Ulmann, Michael (2006), *Language and the brain*, in: Ralph Fasold/Jeffrey Connor-Linton (edd.), *An Introduction to Language and Linguistics*, Cambridge, Cambridge University Press, 235–274.

Parte 2: Tópicos de sintaxe, semântica e fonologia

Ana Maria Martins e João Costa

14 Ordem dos constituintes frásicos: sujeitos invertidos; objetos antepostos

Abstract: Partindo da constatação de que o português é uma língua de ordem básica SVO (i.e. Sujeito-Verbo-Objeto), identificam-se neste capítulo diferentes fatores sintáticos, semânticos e pragmático-discursivos que motivam a alteração da ordem básica nas frases declarativas. Considerando primeiro os casos de inversão do sujeito, distinguem-se três tipos de estruturas frásicas: as frases com ordem VSO que correspondem à expressão de juízos téticos e apresentam destopicalização do sujeito; as frases com ordem VOS em que o sujeito em posição final é o foco informacional da frase (ou tem proeminência focal numa frase de foco largo); as frases com ordem VSO em que o sujeito é simultaneamente um foco contrastivo e informacional. Quanto à anteposição do objeto, o capítulo centra-se na clarificação da distinção entre estruturas de topicalização (OSV/OVS) e de focalização contrastiva (OVS), identificando propriedades gramaticais caracterizadoras de cada uma das construções por oposição à outra.

Keywords: inversão do sujeito, teticidade, foco informacional, topicalização, focalização contrastiva

1 Introdução

Tal como a generalidade das línguas românicas contemporâneas, o português é uma língua de ordem básica SVO, ou seja, normalmente o sujeito (S) precede o predicado e este é encabeçado pelo verbo (V), seguindo-se-lhe o objeto (O), sempre que o verbo não é monoargumental.¹ A ordem básica dos constituintes frásicos (também dita canónica, regular ou não marcada) é a que ocorre com mais frequência em frases declarativas simples e, em geral, é sentida pelos falantes como natural em frases descontextualizadas, dependendo menos que outras ordens dos constituintes frásicos de contextos pragmático-discursivos ou fatores lexicais e morfossintáticos particulares. Também não resulta da expressão de atitudes emotivas do falante que criem um nível interpretativo adicional relativamente ao conteúdo denotativo básico da frase, como acontece nas exclamativas.

¹ Usa-se aqui o termo *objeto* (O) no sentido amplo de Larson (1988; 1990), cobrindo também o que para outros autores seriam modificadores do sintagma verbal e não complementos verbais *stricto sensu*. Por isso, o constituinte frásico que aqui identificamos como «O» corresponde a «X» para outros autores (cf., por exemplo, Leonetti 2014).

Embora a ordem básica do português seja SVO, todas as linearizações dos elementos S, V e O são possíveis, mesmo em frases descontextualizadas, como se mostra em (1a–f).

- | | | | |
|-----|----|--|-----|
| (1) | a. | Esse tipo de notícias pouco interessa ao cidadão comum. | SVO |
| | b. | Ao cidadão comum pouco interessa esse tipo de notícias. | OVS |
| | c. | Pouco interessa ao cidadão comum esse tipo de notícias. | VOS |
| | d. | Pouco interessa esse tipo de notícias ao cidadão comum. | VSO |
| | e. | Ao cidadão comum, esse tipo de notícias pouco interessa. | OSV |
| | f. | Esse tipo de notícias, ao cidadão comum pouco interessa. | SOV |

A variação sintática exemplificada em (1) não significa, no entanto, que a ordem dos constituintes frásicos seja livre em português. Por um lado, as alterações da posição dos constituintes frásicos ensaiadas em (1a–f) não produzem generalizadamente, a partir de qualquer frase, resultados gramaticais. Por outro lado, as frases (1a) a (1f) só superficialmente são equivalentes. Destas duas questões (restrições e efeitos da variação na ordem dos constituintes frásicos) nos ocuparemos neste capítulo, selecionando alguns aspetos centrais do tema relativamente à gramática do português europeu.

No exemplo (1), o facto de o objeto ser um complemento preposicional (concretamente, um objeto indireto), e não nominal, é um fator favorecedor da flexibilidade da ordem dos constituintes frásicos, mas é-o também a presença do quantificador adverbial *pouco* como modificador do verbo. Se compararmos as frases em (1a–e) com as frases (2a–e), que apresentam igualmente um complemento preposicional (agora locativo), verificamos que a ordem OSV deixa de ser possível, como mostra a agramaticalidade de (2e). A razão para este contraste decorre de em (1e), mas não em (2e), termos uma estrutura pragmática e informacionalmente adequada no que diz respeito à articulação entre tópico (i.e. *ao cidadão comum, esse tipo de notícias*) e comentário (i.e. *pouco interessa*). Em ambas as frases, (1e) e (2e), o objeto topicalizado e o sujeito correspondem conjuntamente ao «tópico», que equivale ao Sujeito da Predicação, no sentido lógico destes termos. Em (1e) a presença do modificador *pouco* torna o comentário expresso pelo Predicado *pouco interessa* suficientemente relevante do ponto de vista informacional. O mesmo não acontece em (2e) onde o Predicado *estarão* é despojado de conteúdo semântico e relevância informacional. Que é esta inadequação pragmática e semântica da frase (2e) que a torna inaceitável pode confirmar-se pelo contraste com a frase (3). O enriquecimento do Predicado através da adição do modificador *como peixe na água* torna a frase perfeitamente aceitável e natural. Este é só um exemplo de como nem todas as frases permitem todas as ordens dos constituintes frásicos e de como a sintaxe da ordem é condicionada por fatores não limitados à sintaxe.

- | | | |
|-----|----|---|
| (2) | a. | Parlamentares como Mariana Mortágua, do BE, ou João Galamba, do PS, estarão na Comissão Parlamentar de Inquérito. [SVO] |
|-----|----|---|

- b. Na Comissão Parlamentar de Inquérito estarão parlamentares como Mariana Mortágua, do BE, ou João Galamba, do PS. [OVS]
 - c. Estarão na Comissão Parlamentar de Inquérito parlamentares como Mariana Mortágua, do BE, ou João Galamba, do PS. [VOS]
 - d. Estarão parlamentares como Mariana Mortágua, do BE, ou João Galamba, do PS, na Comissão Parlamentar de Inquérito. [VSO]
 - e. *?Na Comissão Parlamentar de Inquérito, parlamentares como Mariana Mortágua, do BE, ou João Galamba, do PS, estarão. [OSV]
- (3) Na Comissão Parlamentar de Inquérito, parlamentares como Mariana Mortágua, do BE, ou João Galamba, do PS, estarão como peixe na água.

Não é objetivo deste capítulo descrever globalmente os múltiplos fatores que estão por detrás da variação na ordem dos constituintes frásicos. O capítulo centra-se nas frases declarativas, excluindo outros tipos de frases. Ficam portanto fora deste capítulo as interrogativas e exclamativas QU- que envolvem inversão do sujeito (*Como conseguiu o João o passaporte?/O que não faz um dono pelo seu cão!*; cf. Ambar 1992; Ambar/Veloso 2001; Barbosa 2001; Kato/Martins 2016), as exclamativas avaliativas que envolvem anteposição do objeto (*Muita cerveja bebeu o João!/Muita cerveja o João bebeu!*; cf. Raposo 1995; 2000; Ambar 1999; Costa/Martins 2011), e as exclamativas sem expressão de grau que derivam a ordem VSO (*Dizes tu que eu vivo em Marte! (Vê lá se não acertei em tudo.)/Quer ele ser escritor! (E escreve desta maneira.)/Trabalhou o miúdo tanto e o professor dá-lhe esta nota!*; cf. Barbosa 2006, 385, ex. 112a; Martins 2013a). Feito este recorte do objeto de estudo, a secção 2 do presente capítulo será dedicada à inversão do sujeito (ordens VSO/VOS) e a secção 3 à anteposição do objeto (ordens OVS/OSV). Na secção 4 apresenta-se uma breve síntese do capítulo.

Na secção 2, sobre a inversão do sujeito nas frases declarativas, identificam-se três tipos de motivação para a inversão nas frases declarativas: 2.1 a expressão de juízos téticos (por oposição a juízos categóricos) através de frases sem a estrutura própria das predicções (i.e. Sujeito-Predicado), o que implica que o sujeito não ocupe a posição inicial de que decorreria a sua interpretação como tópico da frase; 2.2 a marcação do sujeito como foco informacional, quer seja um foco estreito, quer seja o constituinte informacionalmente proeminente numa frase de foco informacional largo; 2.3 a focalização simultaneamente informacional e contrastiva do sujeito. Enquanto a marcação do sujeito como foco informacional determina a sua colocação no final da frase, originando a ordem VOS (secção 2.2), os outros tipos de inversão do sujeito (secções 2.1, 2.3) correspondem à ordem VSO.

Na secção 3, a discussão centra-se na identificação das propriedades gramaticais que caracterizam e diferenciam entre si as estruturas de topicalização (OSV/OVS) e as estruturas de focalização contrastiva (OVS) do objeto. Na esfera das línguas românicas, a construção de topicalização (cf. Duarte 1987) é uma especificidade do português que tem paralelo em línguas como o inglês (cf. Lasnik/Saito 1992). A sua inexistência em línguas românicas muito estudadas, como o espanhol e o italiano, motivou não raras confusões na literatura entre os termos *topicalization* e *focalization*, que importa

dilucidar (cf. Torrego 1984, 110; Cinque 1990, 180; Rizzi 1997, 285–286), ao mesmo tempo que se clarifica a descrição da pouco estudada construção de focalização contrastiva do português europeu (cf. Costa/Martins 2011).

2 Sujeitos invertidos

O termo «sujeito invertido» (ou «inversão do sujeito») faz referência à substituição da posição pré-verbal do sujeito, tida por posição canónica numa língua SVO, pela posição pós-verbal. Os sujeitos invertidos de frases com predicados transitivos correspondem a dois tipos de ordem dos constituintes frásicos: a ordem VSO, como exemplificado em (4a) e (5a), e a ordem VOS, como exemplificado em (4b) e (5b). As frases VSO, como (4a) e (5a), exprimem juízos téticos, no sentido de Kuroda (1972; 1992; 2005). Nestas frases «descritivas», que serão o objeto da secção 2.1, não se põe em destaque nenhuma das entidades que são parte da situação descrita (por isso o sujeito é afastado da posição de tópico). As frases VOS, por seu lado, apresentam o sujeito em posição final como estratégia para atribuir a este constituinte proeminência prosódica e informacional. Estes casos de saliência focal do sujeito serão tratados na secção 2.2.

- | | | | |
|-----|----|--|-----|
| (4) | a. | Entrou <i>um gato</i> no jardim. (Não deixes sair o Rex.) | VSO |
| | b. | Entrou no jardim o <i>gato cinzento que o Rex odeia</i> . (Não o deixes sair.) | VOS |
| (5) | a. | Pousaram <i>dois rabirruivos</i> na ameixeira. (Vou tentar fotografá-los.) | VSO |
| | b. | Pousou na ameixeira <i>um pássaro lindo</i> . (Anda cá ver. Não sei o que é.) | VOS |

Embora em (4a) o sujeito indefinido (*um gato*) torne a frase particularmente natural, seria igualmente possível um sujeito definido (*Entrou o gato cinzento no jardim*). Por outro lado, o sujeito em posição de foco informacional nas frases VOS não tem de ser sintaticamente mais complexo do que o sujeito das frases VSO, ainda que a comparação entres as frases (a) e as frases (b) de (4) e (5) o possa sugerir. Comprova-o o par de frases em (6). Nestes exemplos, a proeminência informacional sobre o sujeito em (6b), mas não em (6a), transmite uma diferença de atitude do falante. A ordem VOS sugere surpresa relativamente à presença de uma coruja no jardim (em cuja ameixeira pousou), enquanto a ordem VSO é compatível com uma situação de familiaridade com a presença de corujas no jardim.

- | | | |
|-----|----|---|
| (6) | a. | Pousou <i>uma coruja</i> na ameixeira. / Está <i>uma coruja</i> pousada na ameixeira. |
| | b. | Pousou na ameixeira <i>uma coruja</i> . / Está pousada na ameixeira <i>uma coruja</i> . |

O terceiro tipo de inversão do sujeito será abordado, de forma mais breve, na secção 2.3, onde se mostrará que a ordem VSO pode marcar o sujeito como foco simultaneamente contrastivo e informacional (*Fiz eu tudo* = ‘Fui eu quem fez tudo’; *Comeu ele as gomas* = ‘Foi ele que comeu as gomas’).

2.1 Ordem VSO como expressão gramatical da teticidade (ingl. *theticity*)

As línguas naturais têm diferentes formas de expressar gramaticalmente a distinção entre juízos categóricos e juízos téticos (cf. Kuroda 1972; 1992; 2005; Sasse 1987; 1995; 1996; 2006; Lambrecht 1988; 1994; 2000; Matras/Sasse 1995). Uma das estratégias usadas por diferentes línguas românicas para sinalizar a referida distinção é a alternância entre as posições pré-verbal e pós-verbal do sujeito (cf. Duarte 1987; 1997; 2013; Martins 1994, 385–429; 2010; Leonetti 2014; Lobo/Martins no prelo-a). Nas línguas em que esta estratégia é observável, as frases que expressam juízos categóricos (referidas na literatura como frases categóricas, predicacionais ou topicalizadas) apresentam a ordem SV, sendo o sujeito interpretado como tópico (especificamente, *aboutness topic*, usando o termo inglês), ou seja Sujeito de uma estrutura de Predicação (no sentido da lógica aristotélica). Por sua vez, as frases que expressam juízos téticos (chamadas frases téticas, descritivas, apresentativas, não predicacionais ou não topicalizadas) podem apresentar a ordem VS, assim indicando que a frase não se estrutura através da articulação Sujeito-Predicado, própria das predicções. A inversão do sujeito funciona assim, neste caso, como forma de bloquear a sua interpretação como tópico da frase (i.e. *aboutness topic*) e permitir descrever uma situação sem pôr em evidência qualquer das entidades que nela participam.

«The function of a non-topicalized sentence is characterized as a description of a situation. A description affirms the cognitive presence of a conceptually or perceptually given situation. «Description» in this sense is meant to be a type of judgment. In order to better keep this in mind, it might be advisable to extend the use of the term *thetic judgment* and use it as a synonym of this sense of *description*.» (Kuroda 2005, 37–38)

«A thetic judgment is a representation of a perceptually apprehended real, recalled, or imagined situation. A thetic judgment is thus by its nature dependent on another form of cognitive act, the perception of a real or imagined situation. By making a thetic judgment, one commits oneself to the truth of a proposition by describing a perceived situation, real or imagined.» (Kuroda 2005, 29–30)

Na sequência textual em (7), que corresponde à descrição de uma fotografia,² estão marcados em itálico os sujeitos pré-verbais de frases que expressam juízos categóricos. O contexto textual torna clara a sua natureza de *aboutness topics* (i.e. sujeitos lógicos de predicções). A frase inicial, no entanto, embora apresente também um sujeito pré-verbal parece dever ser interpretada como expressão de um juízo tético, correspondendo à intenção de descrever uma cena/imagem à qual pertence *um*

² Trata-se de uma fotografia do fotojornalista Rui Oliveira, pertencente à exposição «Aleixo» (Galeria-Atelier Geraldês da Silva, Porto), comentada pelo jornalista Valdemar Cruz. «Aleixo» é o nome de um bairro do Porto.

homem e não de dizer alguma coisa acerca do homem que está à janela. Portanto, embora a interpretação preferencial da ordem SV, em frases descontextualizadas, seja a interpretação predicacional/categórica, a mesma ordem não impede uma interpretação descritiva/tética. Só a ordem VS é não ambígua relativamente à distinção relevante (distinção que é semântica para Kuroda 2005, mas caracterizada como informacional por outros autores, como, por exemplo, Lambrecht 2000 ou Leonetti 2014).

- (7) Um homem descansa o corpo junto a uma janela. Cigarro na boca, barba de vários dias, os olhos a espelharem um cansaço sem nome, sem data, sem fim. Há uma tristeza inominável a ensombrar aquele quadro tão igual, tão decalcado das mais poderosas cenas de um qualquer filme neorrealista italiano. Com uma diferença crucial: *aquele homem* não é uma personagem inventada para ilustrar uma história. *Aquela janela* não integra um cenário imaginário apostado em proporcionar tons de realidade a uma ficção. *Aquele olhar* não obedece a nenhum guião apostado em suscitar veracidade à narrativa ficcionada. *Aquele quadro* não é um quadro. (Expresso online, 30.01.2016, Valdemar Cruz)

Comentaremos a seguir exemplos de frases VSO e VS (com predicados monoargumentais) em que a inversão do sujeito é usada como recurso gramatical para a expressão da teticidade (ingl. *theticity*). A noção de «evidencialidade» (entendida como a fonte da evidência para uma asserção; cf. Aikhenvald 2004; De Haan 2005; Peterson/Sauerland 2010) é relevante para compreendermos alguns dos dados que serão apresentados. De facto, a descrição de situações percebidas pelo falante, visualmente ou não (o tipo particular de evidencialidade designado por (*non-*)*visual sensory evidentiality*), envolve o tipo de ato cognitivo que subjaz aos juízos téticos e à produção das frases não topicalizadas (i.e. com ordem VS) que os expressam gramaticalmente (cf., acima, a citação de Kuroda 2005).

A frase em (8) apresenta inversão do sujeito. Nesta frase coexistem vários ingredientes que favorecem a ocorrência da ordem VSO, entre eles: o verbo *estar*, como nos exemplos em (9), o imperfeito do indicativo e a oração temporal introduzida por *quando*, como nos exemplos em (10) (cf. Barbosa 2006, 385, ex. 112b; Martins 2010; Leonetti 2015). Estes elementos podem funcionar como marcadores de evidencialidade direta, articulando-se assim facilmente com a ordem VSO para descrever situações observadas pelo falante sensorialmente (sobretudo visualmente). Na ficção narrativa (exemplo (10a)) ou na narrativa jornalística (exemplo (8)), a ordem VSO cria um efeito sugestivo de evidencialidade sensorial.

- (8) Estava a *primeira-dama francesa* a fazer furor em Espanha quando o passado rebelde a voltou a assombrar. (Flash, 12.05.2009)
- (9) a. Está *um alarme* a tocar há horas.
 b. Está *um bebé* a chorar.
 c. Estão *uns sapatos* teus no quarto do Pedrinho.
 d. A: Vou fazer a barba.
 B: Está *o teu pai* na casa de banho.

- (10) a. Descia *o nosso herói* a Rua do Alecrim quando uma onda gigante se formou no Tejo.
 b. Diz que não tem apetite, mas ontem quando cheguei a casa, comia *ele* um hamburger duplo e um prato de batatas fritas.

Com predicados transitivos, os objetos preposicionados facilitam enquanto os objetos nominais dificultam a ordem VSO. Imaginemos uma situação em que, estando a falar ao telefone, vemos o nosso cão ser picado por uma abelha. Uma frase possível para justificarmos a necessidade de desligar imediatamente o telefone poderá ser (11a), com ordem VSO, que, na situação descrita, é mais natural do que a ordem SVO.³ A ordem VSO deixa, no entanto, de ser uma opção se a frase for construída com o verbo *picar* como transitivo direto, como mostra a agramaticalidade de (11b). A cliticização do objeto permite eliminar a diferença entre a estrutura transitiva direta e indireta, passando a permitir nos dois casos a inversão do sujeito, como se vê em (12).

- (11) a. (Tenho de desligar.) Picou uma abelha ao meu cão.
 b. (Tenho de desligar.) *Picou uma abelha o meu cão.
- (12) A: O que é que aconteceu para ele estar a ganir?
 B: a. Picou-lhe uma abelha.
 b. Picou-o uma abelha.

Com verbos transitivos declarativos, como *dizer*, *contar*, *escrever*, a ordem VSO atesta-se no relato jornalístico como forma de identificar a fonte da informação (apresentada como evidência indireta fidedigna), indicando ao mesmo tempo, através da inversão do sujeito, que o constituinte com a função de sujeito não é o tópico frásico e discursivo, conforme se exemplifica em (13).

- (13) a. Diz *o jornal «The Sun»* que o inglês tinha uma álibi consistente, apoiado no depoimento de uma jovem de 15 anos, e foi logo descartado. (Expresso, 23.05.2009)
 b. Escrevia *o Le Monde* há poucos dias, que há Boulogne – a bancada histórica, de maioria branca, sem pudor em exibir uma simpatia pela extrema-direita – e no topo oposto do estádio há Auteuil, mais heterogénea, mais recente. Entre elas, um ódio. Referem-se uns aos outros como «nazis» e «escumalha». (Público, 19.03.2010)
 c. Conta *a historiografia oficial* que Paul Gauguin foi para Arles em Outubro de 1888, a convite do seu amigo Vincent Van Gogh, que por essa altura andava entusiasmado com a luminosidade e as cores da Provença. (Diário de Notícias, 26.05.2009)

3 A frase com ordem SVO (*uma abelha picou ao meu cão*) parece menos natural na situação descrita porque seria ambígua entre a expressão de um juízo tético e a expressão de um juízo categórico. Neste último caso, teríamos uma estrutura de predicação com o constituinte *uma abelha* posto em evidência por ser o *aboutness topic* da frase. Pragmaticamente, a escolha do constituinte *o cão* como tópico da frase seria mais adequada, originando uma frase passiva: *o meu cão foi mordido por uma abelha*. Curiosamente, esta frase difere da frase *Picou uma abelha ao meu cão* em termos de evidencialidade pois não tem implícita a informação de que o falante está a descrever uma situação que acabou de presenciar.

Com predicados monoargumentais, a inversão do sujeito ocorre facilmente com verbos inacusativos, embora não esteja vedada a outros verbos intransitivos (cf. Ambar 1992; Martins 1994, 385–429; 2010; Duarte 1997; Costa 1998; 2004; Costa/Figueiredo Silva 2006; Barbosa/Duarte/Kato 2005; Barbosa 2006; 2009; Kato/Martins 2016). Em frases descontextualizadas, as ordens SV e VS podem parecer indistintas nos planos semântico e pragmático-discursivo, como se exemplifica em (14).

- (14) a. Caiu um avião. / Um avião caiu.
 b. O Presidente morreu. / Morreu o Presidente.
 c. Os gêmeos já nasceram. / Já nasceram os gêmeos.
 d. Um dos atletas desmaiou. / Desmaiou um dos atletas.
 d. A Maria telefonou. / Telefonou a Maria.
 e. Já só dois moinhos funcionam. / Já só funcionam dois moinhos.

A observação de dados contextualizados torna-se necessária para perceber as diferenças relevantes, incluindo determinar se a inversão do sujeito está associada à expressão de juízos téticos ou antes à focalização do sujeito, já que nas estruturas monoargumentais a inexistência do terceiro elemento (O) não permite, à partida, separar os dois casos.

O fragmento textual em (15) é parte de uma entrevista dada pelo escritor António Lobo Antunes à revista *Visão* e vem na sequência de uma conversa sobre a vivência da guerra em África pelo escritor. Neste contexto, o constituinte *a guerra* não é, claramente, o foco informacional da frase, nem tem qualquer tipo de proeminência focal. Por outro lado, a frase *parava a guerra* não é acerca do que caracterizava a guerra, antes descreve uma situação excepcional (e surpreendente) observada, no contexto da guerra, por quem a relata. A ordem VS traduz assim perfeitamente a não topicalidade do sujeito frásico, ao mesmo tempo que marca a evidencialidade direta.

- (15) Quando o Benfica jogava, púnhamos os altifalantes virados para a mata e, assim, não havia ataques.
 Visão: Parava a guerra?
 Lobo Antunes: Parava a guerra. Até o MPLA era do Benfica. (Visão, 27.11.2003)

Um traço caracterizador das frases que exprimem juízos téticos é a sua difícil compatibilidade com a negação. Como se vê nos exemplos (16) a (18), as frases afirmativas são totalmente aceitáveis quer com ordem SV quer com ordem VS mesmo que descontextualizadas (abstraindo agora dos contextos discursivos ou pragmáticos particulares em que cada uma seria mais provável ou adequada). Já nas frases negativas há um claro contraste de aceitabilidade entre as que têm ordem SV e as que apresentam inversão do sujeito. Tomadas isoladamente, as frases negativas com ordem VS são estranhas. Ao criarmos um contexto que as torne naturais, verificamos que se tornam, de forma não ambígua, casos de focalização do sujeito, como se exemplifica em (19) e (20). Se admitirmos que há uma relação forte entre teticidade e evidencialidade sensorial, podemos motivar a difícil compatibilidade entre as frases

que exprimem juízos téticos e a negação (mas cf. (23c), onde *uma brisa* parece funcionar como um minimizador). Ou seja, tendencialmente as frases VS descrevem situações que ocorrem e são sensorialmente apreendidas.

- (16) a. O Presidente morreu.
 b. Morreu o Presidente.
 c. O Presidente não morreu. [situação: o Presidente estava em risco de vida]
 d. #Não morreu o Presidente.
- (17) a. O avião caiu.
 b. Caiu o avião.
 c. O avião não caiu. [situação: o avião estava em risco de queda]
 d. #Não caiu o avião.
- (18) a. A Maria telefonou.
 b. Telefonou a Maria.
 c. A Maria não telefonou. [situação: o telefonema da Maria era esperado]
 d. #Não telefonou a Maria.
- (19) Situação: sabe-se que ocorreu um acidente de viação e que há poucos sobreviventes.
 A: Quem é que não morreu?
 B: Não morreu [_{Foco} o Presidente].
- (20) Situação: São esperados telefonemas de várias pessoas.
 A: Já telefonaram todos? / Quem é que não telefonou?
 B: Não telefonou [_{Foco} a Maria].

Com predicados que, semanticamente, envolvem direcionalidade, as frases com ordem VS traduzem obrigatoriamente contacto ou aproximação com a localização do enunciador, como mostram os exemplos (21) e (22), o que de novo remete para a relação entre teticidade e evidencialidade sensorial. Ou seja, as frases com os verbos *chegar* e *telefonar* admitem a ordem VS como expressão de juízos téticos quando o falante está diretamente envolvido na situação que descreve e a apreende visual ou auditivamente. As frases (21d) e (22-B-b) são agramaticais porque essa condição de evidencialidade direta não se verifica. Na interpretação ‘chegou lá, a casa dele’ (com *chegar*) ou ‘telefonou para lá’ (e não ‘telefonou-me ou telefonou-nos’, com *telefonar*) só está disponível a ordem SV, própria das frases categóricas/predicacionais. A frase (22-B-b) parece ser um pouco mais aceitável do que (21d) porque permite, um tanto marginalmente, a interpretação ‘Foi a Maria que telefonou para lá e disseram-lhe que ele saía’ (cf. a secção 2.3).

- (21) a. O teu pai já chegou.
 interpretação (i): ‘chegou cá, a nossa casa (ou ao lugar onde estamos)’
 interpretação (ii): ‘chegou lá, a casa dele’
 b. Já chegou o teu pai.
 interpretação (i): ‘chegou cá, a nossa casa (ou ao lugar onde estamos)’
 [não é possível a interpretação (ii)]

- c. O teu pai já chegou a casa. Acabou de ligar a dizer que chegou bem.
interpretação (ii): ‘chegou lá, a casa dele’
- d. *?Já chegou o teu pai a casa. Acabou de ligar a dizer que chegou bem.
[agramaticalidade decorrente da impossibilidade da interpretação (ii)]
- (22) A: Como é que sabes que o João sai hoje do hospital?
B: a. A Maria telefonou e disseram-lhe que ele saía. (i.e. ‘telefonou para o hospital’)
b. ??Telefonou a Maria e disseram-lhe que ele saía.

Por fim, note-se como certos predicados monoargumentais que descrevem condições atmosféricas ou climáticas favorecem a ordem VS, chegando a torná-la obrigatória:

- (23) a. Corria uma agradável brisa de fim da tarde.
b. *Uma agradável brisa de fim da tarde corria.
c. Não corre uma brisa.
d. *Uma brisa não corre.
- (24) a. Sopravam ventos de 100km/h.
b. *Ventos de 100km/h sopravam.

Nas frases (23b) e (24b), o verbo não tem a força semântica e informacional necessária para que a ordem SV corresponda a uma estrutura tópico-comentário aceitável. Que o problema não está no constituinte sujeito, demonstra-o a gramaticalidade das frases em (25).

- (25) a. Ventos de 100km/h sopravam incessantemente em toda a zona costeira.
b. Uma agradável brisa de fim da tarde corria na praia, enquanto na cidade se sufocava.

2.2 Ordem VOS e foco informacional (estreito / largo)

Os contextos de pergunta/resposta em que a pergunta tem a forma de uma interrogativa QU- são um bom teste para determinar se a frase-resposta exhibe um foco informacional largo ou um foco informacional estreito e, no segundo caso, para identificar o constituinte que, introduzindo informação nova, corresponde ao foco informacional da frase. Uma pergunta do tipo «O que aconteceu?», como em (26), determina que a frase-resposta introduza globalmente informação nova. A frase (26-B) exemplifica, portanto, um foco informacional largo. Já no par pergunta/resposta apresentado em (27) parte da informação presente na frase-resposta é pressuposta (‘X não entrou no avião’); a informação nova é apenas o constituinte que, na resposta, corresponde ao pronome interrogativo (i.e. o sintagma QU-) da pergunta. Este foco informacional estreito estabelece, assim, o valor da variável representada pelo sintagma QU- (em (27-B), *Quem*).

- (26) A: O que aconteceu?
 B: Não nos deixaram embarcar porque não tínhamos o ESTA.
- (27) A: Quem é que não deixaram entrar no avião?
 B: Não deixaram entrar no avião {o professor / dois colegas / nenhum de nós}

Como mostra (28), o sujeito ocorre na posição final da frase quando corresponde ao foco informacional estreito. A estrutura informacional motiva, portanto, a ordem VOS em vez da ordem canônica SVO, que levaria à interpretação do sujeito como tópico da frase e não como foco informacional. Por isso a frase (28-B-b) não constitui uma resposta adequada à pergunta *Quem é que pediu sardinhas assadas?* A sua estrutura informacional não assegura a identificação completa/exaustiva do valor da variável representada por *Quem*, deixando em aberto a possibilidade de outras pessoas além de «a minha mãe, a Rita e eu» terem pedido sardinhas assadas.⁴

- (28) A: Quem é que pediu sardinhas assadas?
 B: a. Pedimos sardinhas assadas [*a minha mãe, a Rita e eu*]_{Foco}.
 b. #[*A minha mãe, a Rita e eu*]_{Tópico} pedimos sardinhas assadas.

Mesmo em frases com foco informacional largo (que transmitem globalmente informação nova), a posição final da frase pode ser usada para pôr em destaque constituintes informacionalmente fortes. Deste modo associa-se a saliência prosódica, que no português recai regularmente sobre o constituinte em posição final, à proeminência discursiva e informacional. Um constituinte informacionalmente forte não tem de ser sintaticamente complexo e prosodicamente pesado, como demonstra o exemplo (29), no qual o sujeito em posição final é o sintagma nominal *os offshores*, ou o exemplo (30), outra frase de foco largo em que a ordem VOS atribui saliência focal ao sujeito nominal *uma águia/orangotangos/neve*.

4 Em contextos situacionais ou pragmáticos que tornam aceitável uma resposta não exaustiva, e só nesses, a ordem SV é possível (com o constituinte sujeito como informação nova). Assim (i-A), por exemplo, permitiria a resposta «Nós os três queremos», resposta que deixa em aberto a possibilidade de outros, além de «nós os três», também quererem.

- (i) A: Quem quer uma fatia de bolo?
 B: a. Queremos [_{Foco} *nós os três*].
 b. [_{Tópico} *Nós os três*] queremos.

As respostas com elipse dos constituintes correspondentes à informação pressuposta, ou parte deles, são sempre possíveis e naturais, podendo realizar-se apenas o constituinte interpretado como foco informacional estreito, como em (ii-Be).

- (ii) A: Quem ganhou a meia maratona de Lisboa este ano?
 B: a. Ganhou a meia maratona de Lisboa este ano *um atleta australiano*.
 b. Ganhou a meia maratona de Lisboa *um atleta australiano*.
 c. Ganhou a meia maratona *um atleta australiano*.
 d. Ganhou *um atleta australiano*.
 e. *Um atleta australiano*.

(29) Os setores imobiliário e de hotelaria, mas também os bancos, funcionam como plataformas cada vez mais vulneráveis para ações de branqueamento de capitais. Depois entram em cena os «*offshores*». (Expresso online, 05.01.2016, Valdemar Cruz)

(30) A: O que é que foi? O que é que estás a ver?

- B: a. Não vais acreditar, mas acabou de pousar no limoeiro do quintal do prédio aqui ao lado *uma águia*.
 b. Não vais acreditar, mas estão a aparecer de todo o lado no quintal do prédio aqui ao lado *orangotangos*.
 c. Não vais acreditar, mas está a cair só mesmo no quintal do prédio aqui ao lado *neve*.

A colocação do sujeito em posição final atesta-se em frases de foco informacional largo na produção escrita contemporânea, como mostram os exemplos (31) a (39). Estes exemplos mostram também que vários fatores podem facilitar a colocação do sujeito na posição final de frases que transmitem, na sua totalidade, informação nova, embora nenhum deles seja só por si condição necessária ou suficiente. São fatores favorecedores da ordem V(O)S como estratégia de atribuição de saliência focal ao sujeito, entre outros: um sujeito sintaticamente complexo, prosodicamente pesado e informacionalmente forte; um predicado informacionalmente fraco; verbos monoargumentais; estruturas passivas; complementos verbais representados por pronomes clíticos; a presença de tópicos de localização temporal ou espacial; certas orações subordinadas, como as relativas ou as orações introduzidas por *como*. A mesma frase pode exibir uma combinação de vários destes fatores.

(31) Com o centro da Europa em estado de sítio, numa deriva securitária de complexos contornos, vão tornar-se inevitáveis *confrontos como os de ontem em Paris*, dos quais resultaram *quase 300 detidos*. (Expresso online, 30.11.2015, Valdemar Cruz)

(32) Nem por isso se encontram *metáforas bastantes para explicar o despudor contido nas notícias sobre a facilidade de circulação de dinheiro*. (Expresso online, 05.01.2016, Valdemar Cruz)

(33) Admite-se que nos inúmeros «*offshores*» espalhados pelo Mundo, inclusive na Madeira, circulem *verbas entre os 8 e os 32 biliões de dólares da riqueza privada*, ou seja, perto de metade do PIB Mundial. Sabem-no *os dirigentes dos principais países*. Sabem-no *os responsáveis pelos principais bancos centrais*. Sabem-no *as principais instituições judiciais e policiais*. (Expresso online, 05.01.2016, Valdemar Cruz)

(34) Vivemos, desse ponto de vista, tempos cruéis, marcados por lavandarias várias e hipocrisias sem fim. Como o revela *a crise dos refugiados e o modo como vários países europeus estão a reagir a esta emergência social*. Fecham-se *fronteiras*, como o fez ontem *a Dinamarca* ao fim de meio século de ligações abertas com a Suécia. Rasgam-se *acordos*. (Expresso online, 05.01.2016, Valdemar Cruz)

(35) Confirma-o *a divulgação, ontem, dos resultados da iniciativa anualmente lançada pela Porto Editora para a escolha da palavra do ano*. (Expresso online, 05.01.2016, Valdemar Cruz)

- (36) Ao nível dos Estados, vão por certo prosseguir *as pequenas e grandes cimeiras com sonoras e piás declarações*. No terreno, persistirá *a certeza de que o fluxo migratório, nem se combate com armas, nem com manifestações de intenções*. (Expresso, 05.01.2016, Valdemar Cruz)
- (37) Está por demonstrar *a eficácia do modelo, com acumulação de debates no mesmo dia*, como está por provar *a bondade democrática de uma alternativa minimalista, reduzida aos frente a frente entre os nomes mais mediáticos*. (Expresso online, 05.01.2016, Valdemar Cruz)
- (38) Na próxima quinta-feira poderá sair da Assembleia da República *uma boa notícia para quem depara com dificuldades económicas*. Serão votados e aprovados, presume-se, *os projetos-lei do PS, BE e PCP que impedem a penhora das casas aos cidadãos com dívidas ao fisco ou à Segurança Social*. (Expresso online, 05.01.2016, Valdemar Cruz)
- (39) Puseram-se a jeito, é verdade. Ou colocou-os a jeito *o próprio programa de arquitetura, materializado num conjunto de torres enormes, com elevados custos de manutenção e para onde foram despejadas dezenas e dezenas de famílias de fracos recursos*. (Expresso online, 30.01.2016, Valdemar Cruz)

Para concluir esta secção, apresenta-se em (40) um par de frases que exemplificam a distinção entre a ordem VOS com proeminência focal sobre o sujeito e a ordem VSO como expressão da teticidade. Tanto (40a) como (40b) são frases de foco largo, não sendo portanto a esse nível que se distinguem. Mas enquanto a ordem VOS de (40a) marca o constituinte em posição final (i.e. *o debate do orçamento de estado*) como a informação em foco, a ordem VSO de (40b) descreve a atmosfera, cenário ou pano de fundo em que ocorre aquilo que é notícia (i.e. *um deputado sentiu-se mal e desmaiou*) e não recai sobre o sujeito invertido qualquer saliência prosódica ou informacional.

- (40) a. Está ao rubro *o debate do orçamento de estado*.
 b. Estava *o debate do orçamento de estado* ao rubro quando um deputado se sentiu mal e desmaiou.

2.3 Ordem VSO com foco contrastivo e informacional sobre o sujeito

A alteração da ordem básica SVO pode corresponder a uma estratégia de marcação do foco informacional da frase, como vimos na secção anterior, ou a uma estratégia de marcação de um foco contrastivo, como veremos detidamente na secção 3. Sintaticamente, focos contrastivos e focos informacionais pertencem, normalmente, a limites opostos da frase, sendo os primeiros iniciais e os segundos finais. Os focos contrastivos ou se mantêm *in situ* e são apenas marcados prosodicamente ou se movem para a margem esquerda da frase. O foco informacional, pelo contrário, pertence caracteristicamente à margem direita da frase, onde recebe um acento prosódico neutro. O contraste posicional entre um constituinte marcado como foco contrastivo e um constituinte marcado como foco informacional exemplifica-se em (41) e (42), respetivamente.

(41) A retórica é a maior arma dos políticos. *Com ela se elevam, com ela se desgraçam.* (Expresso, 11.09.2010, Filipe Santos Costa)

(42) A: Como é que os políticos se elevam?

B: Os políticos elevam-se *com a retórica*.

Do ponto de vista semântico e pragmático, os focos contrastivos introduzem um valor de oposição relativamente a uma asserção, pressuposição ou expectativa, explícita ou implicitamente presentes no domínio discursivo. A propriedade central de um foco contrastivo é a de adicionar ao valor denotativo básico da frase a expressão da atitude (discordante) do falante relativamente ao que sabe ou supõe serem as expectativas ou convicções do ouvinte. A focalização contrastiva assinala o contraste entre a informação transmitida pelo falante e a informação que, do seu ponto de vista, estaria a ser previamente assumida pelo ouvinte (Zimmermann 2007; Zimmermann/Onea 2011). O conceito de informação nova, que está na base da identificação do foco informacional, não é relevante para identificar um foco contrastivo.

O foco informacional e o foco contrastivo correspondem assim a dois planos interpretativos diferentes, pelo que os dois tipos de foco podem ser identificados na mesma frase e um mesmo constituinte pode estar associado a ambos. Relevante para o tema em discussão neste capítulo é o caso de frases com ordem VSO em que o sujeito tem um valor de foco contrastivo mas é, ao mesmo tempo, um foco informacional estreito, como no exemplo (43). Este tipo de sujeito invertido pode também pertencer a uma frase de foco largo mas ser o foco central da informação, como nos exemplos (44) a (47), este último um exemplo atestado. O facto de o sujeito ser simultaneamente marcado como foco contrastivo e foco informacional parece determinar que não ocupe nem a posição final característica do foco informacional nem a posição inicial característica do foco contrastivo anteposto. Ao mesmo tempo, o constituinte sujeito não recebe o acento prosódico contrastivo característico da marcação de foco *in situ*, o que confirma que a estratégia de marcação de foco seguida é sintática e opera com a possibilidade de variação na ordem dos constituintes frásicos (cf. Barbosa 2009).

(43) A: Quem põe a loiça na máquina?

B: Hoje põe o *João* a loiça na máquina. (Não posso ser eu todos os dias!)
[= 'Hoje é o João que põe a loiça na máquina.']

(44) A: Obrigada por terem feito o jantar!

B: Fiz *eu* tudo. (Ele não fez nada.)
[= 'Fui eu que fiz tudo.']

(45) Compra o *João* o gelado para a sobremesa?

[= 'É o João que compra o gelado para a sobremesa?']

(46) Não curaste *tu* o Rex, curou-o o veterinário!

[= 'Não foste tu que curaste o Rex, foi o veterinário.']

- (47) Quando o meu herói, Ernest Shackleton, estava a 97 milhas do Pólo Sul na manhã de 9 de Janeiro de 1909 disse que dera tudo quanto tinha a dar – ora bem, hoje tenho de vos informar que dei *eu* tudo quanto tinha a dar: Já não tenho forças para arrastar um esqui adiante do outro. (Expresso, 06.02.2016, José Cutileiro, citando Henry Worsley)
[= ‘hoje tenho de vos informar que fui eu quem deu tudo quanto tinha a dar’]

Nas três frases retiradas de textos de imprensa que se apresentam a seguir, a ordem VSO parece ser igualmente motivada pela proeminência focal simultaneamente contrastiva e informacional do sujeito. Note-se que nas frases de (49a–49b) o sujeito (destacado a itálico) poderia ser precedido de um advérbio marcador de foco como *precisamente* ou *exatamente*. Quanto à frase (48), a clivada apresentada entre parêntesis retos parece constituir uma boa paráfrase.

- (48) As primárias de New Hampshire revelaram o poder dos eleitores de abanar as elites que controlam a política norte-americana, escreve *a CNN* esta madrugada. (Expresso online, 10.02.2016, Cristina Peres)
[‘é a CNN que o escreve esta madrugada’]
- (49) a. A escritora defende que a proliferação de bancos alimentares, a falta de apoio do Estado aos serviços destinados a crianças, o ataque ao Serviço Nacional de Saúde, a comercialização do ensino nas universidades, fazem de Dickens um autor mais essencial do que nunca, dado serem *muitos destes temas* aqueles que o escritor defendeu e pelos quais lutou. (Expresso, 29.01.2016, Valdemar Cruz)
- b. É a dolorosa verdade do dia a dia daqueles a quem coube o *portuense bairro social do Aleixo* em sorte. (Expresso, 30.01.2016, Valdemar Cruz)

3 Objetos antepostos: focalização vs. topicalização

No português existem vários tipos de construções que marcam o objeto como tópico ao colocá-lo na periferia esquerda da frase. Estas construções envolvem diferentes graus de coesão sintática entre o objeto anteposto e o resto da frase, como se mostra em (50), onde se identificam também os termos usados na literatura sobre o português para designar cada uma das construções. Para uma caracterização detalhada das diferentes construções de tópicos marcados do português europeu, veja-se Duarte (1987; 2003; 2013).⁵

- (50) a. {Livros/Quanto a livros}, não posso dizer que tenho um livro preferido.
(*Tópico Pendente*)
- b. Esse segurança, eu detesto pessoas alarmistas.
(*Tópico Pendente*)

⁵ Considera-se que o tópico não marcado é o sujeito das frases com ordem SVO, daí falar-se de tópico marcado quando o objeto ocupa a posição de tópico no início da frase. Na mesma frase pode existir mais do que um constituinte com função de tópico.

- c. [Esse segurança]_i, dá gosto conversar com [o rapaz/ele]_i.
(*Deslocação à Esquerda de Tópico Pendente/Deslocação à Esquerda*)
- d. [A esse segurança]_i, já [lhe]_i pedi ajuda várias vezes.
(*Deslocação à Esquerda Clítica*)
- e. [A esse segurança]_i, já pedi ajuda [-]_i várias vezes.
(*Topicalização*)

A existência da construção de topicalização, exemplificada em (50e), é uma especificidade do português ausente de outras línguas românicas, como se mostra em (51) a (53), comparando o português europeu, em (53), com o espanhol e o italiano, em (51) e (52) respetivamente. Nestas línguas, a marcação sintática de um objeto acusativo ou dativo como tópico, movendo-o para a periferia esquerda da frase, implica que esteja ligado a um pronome clítico em posição interna à frase. Ou seja, o espanhol e o italiano, como a maioria das línguas românicas, dispõem da construção de deslocação à esquerda clítica (ingl. *clitic left dislocation*) mas não da construção de topicalização (ingl. *topicalization*), enquanto o português permite ambas. Neste aspeto, o português assemelha-se ao inglês, que tem a construção de topicalização e a construção de deslocação à esquerda (sendo uma língua sem clíticos), como se exemplifica em (54).⁶

- (51) a. El diário, Pedro lo compró. (*deslocação à esquerda clítica: √*)
b. *El diário, Pedro compró. (*topicalização: **)
(espanhol; cf. Zubizarreta 1999)
- (52) a. Il tuo libro, lo ho comprato. (*deslocação à esquerda clítica: √*)
b. *Il tuo libro, ho comprato. (*topicalização: **)
(italiano; cf. Rizzi 1997)
- (53) a. O jornal, o Pedro comprou-o. (*deslocação à esquerda clítica: √*)
b. O jornal, o Pedro comprou. (*topicalização: √*)
c. O teu livro, já o comprei. (*deslocação à esquerda clítica: √*)
d. O teu livro, já comprei. (*topicalização: √*)
- (54) a. The newspaper, I have already bought it. (*deslocação à esquerda: √*)
b. The newspaper, I have already bought. (*topicalização: √*)
(inglês; cf. Lasnik/Saito 1992)

O facto de na construção de topicalização o objeto deslocado à esquerda estar ligado a uma posição vazia no interior da frase aproxima superficialmente (no que diz respeito à ordem dos constituintes frásicos) a construção de topicalização da construção de focalização e criou bastante ruído na literatura relativamente ao uso de cada um dos termos.

⁶ Historicamente, a construção de topicalização parece ser no português uma inovação tardia (Martins 1994; ↗1 Introdução: O português numa perspetiva diacrónica e comparativa), o que explicará o contraste com as línguas da mesma família, como o espanhol e o italiano.

Torrego (1984, 110) é talvez a primeira autora que, na literatura sobre as línguas românicas, chama a atenção para a confusão terminológica entre *topicalization* e *focalization*. Mas, alguns anos mais tarde, Cinque (1990) ainda usa o termo *topicalization* para designar a construção de focalização contrastiva:

«As mentioned in chapter 1, topicalization could more appropriately be termed *Focus Movement* in Italian, since its left-peripheral phrase obligatorily bears heavy stress, its pragmatic function being to contrast the «topicalized» constituent with some other constituent. I nonetheless retain the term *topicalization* here to emphasize its syntactic identity to the English construction (though the pragmatics of the latter is indeed closer to that of CLLD [Clitic Left Dislocation] in Italian (...)). For the sake of clarity, I will continue to capitalize topicalized constituents.» (Cinque 1990, 180)

É só a partir do trabalho de Rizzi (1997) que parece criar-se um consenso em torno do significado de cada um dos termos, pelo menos no domínio das línguas românicas:

«A traditional articulation of the clause that typically involves the left periphery is the articulation in topic and comment, as expressed by the English construction referred to as Topicalization [*Your book, you should give t to Paul (not to Bill)*] [...]. Formally similar but interpretatively very different is the focus-presupposition articulation [*YOUR BOOK you should give t to Paul (not mine)*] [...]. In Italian and more generally in Romance, the topic-comment articulation is typically expressed by the construction that Cinque (1990) has called Clitic Left Dislocation (CLLD), involving a resumptive clitic coreferential to the topic [...]. The focus-presupposition articulation can be expressed in Italian by preposing the focal element (focalization) and assigning it special focal stress [...]. In Italian this option is restricted to contrastive focus.» (Rizzi 1997, 285–286)

O tema central desta secção será a identificação das propriedades que distinguem, no português europeu, a construção de focalização contrastiva da construção de topicalização. Como a primeira tem sido menos estudada do que a segunda, os dados que serão apresentados incidem sobretudo na construção de focalização e a descrição será feita na perspetiva de caracterização desta construção, tomando a topicalização como elemento de comparação e contraste. Para uma caracterização exaustiva da construção de topicalização, veja-se Duarte (1987; 2013); sobre a distinção entre foco contrastivo e foco informacional, veja-se acima a secção 2.1.3 e Costa/Martins (2011).

3.1 Equivalência interpretativa com estruturas clivadas

As estruturas de focalização são interpretativamente equivalentes a estruturas clivadas, o que não acontece com as estruturas de topicalização (cf. Hernanz/Brucart 1987; Zubizarreta 1998; 1999). Aos constituintes clivados ou focalizados corresponde, em geral, uma interpretação exaustiva, que os distingue dos constituintes topicalizados.

A frase (55), com topicalização de *a esse país*, não encontra uma boa paráfrase numa clivada canónica, como demonstra a possível continuação *mas não é o único*, que indica que o objeto anteposto não tem uma interpretação exaustiva.

- (55) *A esse país, não tenho vontade de ir. Mas não é o único.*
 [≠ ‘É a esse país que não tenho vontade de ir’]

Ao contrário das frases com topicalização, as frases com focalização do objeto são sempre parafraseáveis por estruturas clivadas (cf. Lobo/Martins no prelo-b), como se exemplifica em (56) e (57). Os constituintes *com ela*, em (56), e *de pés de veludo*, em (57), têm interpretação exaustiva nas frases em que ocorrem como focos contrastivos antepostos, tal como nas frases em que são o constituinte clivado.

- (56) A retórica é a maior arma dos políticos. *Com ela* se elevam, *com ela* se desgraçam. (Expresso, 11.09.2010, Filipe Santos Costa)
 [= ‘é com ela que se elevam e é com ela que se desgraçam’]
- (57) Zeca Afonso era antes de mais um poeta. Num tempo de fala curta, construía metáforas de alcance longo e falava de vampiros para evocar quem *de pés de veludo* chegava para sugar o sangue fresco da manada. (Expresso online, 05.01.2016, Valdemar Cruz)
 [= ‘para evocar quem era de pés de veludo que chegava’]

A proximidade interpretativa entre a focalização e as clivadas faz com que seja possível não só parafrasear uma frase com focalização por outra com uma estrutura clivada, como fazer o exercício contrário, conforme se mostra em (58) e (59). Tenha-se em atenção que a equivalência é apenas semântica, e não sintática. O exemplo (60), que tem como ponto de partida uma estrutura pseudoclivada (em vez de uma clivada canónica), mostra como a equivalência semântica se mantém mesmo quando o constituinte clivado se situa na margem direita da frase, enquanto o constituinte focalizado pertence à margem esquerda da frase.

- (58) Pois é *de lá* que vem a mais escandalosa das notícias: o confisco dos bens dos refugiados. (Expresso online, 28.01.2016, Cecília Meireles)
 [= ‘Pois de lá vem a mais escandalosa das notícias’; o antecedente do locativo *lá* é «a Dinamarca», um país de onde não se esperaria *a mais escandalosa das notícias*]
- (59) É *no final de «Another One»*, o álbum que Mac DeMarco acaba de lançar, que se encontra a chave não só da postura do canadiano, como da sua casa. (Expresso 08.08.2015, Lia Pereira)
 [= ‘No final de «Another One», o álbum que Mac DeMarco acaba de lançar, se encontra a chave...’]
- (60) O que eu transmiti aos meus jogadores foi *isto*: «Ou matamos ou morremos.» (Expresso, 08.08.2015, entrevista a Rui Vitória)
 [= ‘Isto transmiti eu aos meus jogadores: ...’]

As duas frases que se seguem, ambas atestadas, constituem evidência adicional relativamente à proximidade, nos planos semântico e pragmático-discursivo (ainda que não em produtividade de uso), entre a construção de focalização e as estruturas clivadas.

(61) Não lhe respondi, e pareceu-me que aquela cena tinha sido comicamente absurda, como que representada por homens que tinham voltado à adolescência, ou, como seria o caso do Rodrigues e do Luís, não tinham saído dela ou não sairiam nunca. *Isto* me aproximou do meu primo. (Jorge de Sena, *Sinais de Fogo*, in: CRPC, 79 Linguística de corpus e outros usos dos corpora em linguística)

[= 'foi isto que me aproximou do meu primo']

(62) Todos fazemos coisas de sucesso fácil, mas raramente pomos a fasquia acima do que julgamos estar ao nosso alcance. É *isso* que me motiva aqui. (Expresso, 06.02.2016, José Cutileiro, citando Henry Worsley)

3.2 Colocação dos pronomes clíticos

A topicalização não afeta a colocação dos pronomes clíticos. Assim, nas frases finitas, afirmativas, que não incluem elementos desencadeadores de próclise, os pronomes átonos têm a colocação enclítica que caracteriza o português europeu. Pelo contrário, a focalização torna obrigatória a colocação pré-verbal dos pronomes clíticos pois os focos contrastivos antepostos são, eles próprios, desencadeadores de próclise (715 A colocação dos pronomes clíticos em sincronia e diacronia; cf. Martins 2013b para uma identificação completa dos itens proclisadores no português europeu).

No exemplo (63) podemos ver o contraste entre a frase (63a), que apresenta focalização e próclise, e o exemplo (63b), que apresenta topicalização e ênclise. A possibilidade de parafrasear (63a) com uma frase clivada mostra que se trata de focalização. A frase (63b), em que o objeto anteposto está topicalizado, não tem correspondência numa clivada (seria agramatical a frase **se o livro saiu bem, é a si que se deve pouco*).

(63) a. Se o livro saiu bem, *a si* se deve (mensagem de sms de editor para autor)

[= 'Se o livro saiu bem é a si que se deve']

b. Se o livro saiu bem, a si, deve-se pouco.

Em (64) a (66) dão-se exemplos adicionais da colocação proclítica dos pronomes átonos na construção de focalização (em frases em que não há outro proclisador senão o próprio constituinte focalizado). Entre parêntesis retos, a seguir aos exemplos (66) e (67), mostra-se que as frases se tornariam agramaticais com o pronome clítico colocado depois do verbo, o que decorre da impossibilidade de, nestes casos, obter uma estrutura de tópico-comentário bem formada (cf. secção 3.5).

(64) Justamente, meu caro amigo, por vir para as empresas da família *lhe* exijo muito mais do que aos outros. (CRPC, Lobo Antunes, *Fado*)

[= 'por vir para as empresas da família é que lhe exijo muito mais do que aos outros']

- (65) O que aqui interessa é que há vontade de pôr as coisas pela positiva. Sem ponta de ironia o escrevo: todos queremos que as «instituições» melhorem. (Expresso, 09.05.2015, Pedro Santos Guerreiro)
 [= ‘É sem ponta de ironia que o escrevo’]
 [*Sem ponta de ironia, escrevo-o. Cf.: Escrevo-o sem ponta de ironia.]
- (66) Infelizmente, e *com tristeza* o digo, pela consideração que tenho por Vital Moreira, o apontamento dele não é menos abusivo do que o discurso de Paulo Rangel no Parlamento Europeu acerca da minha entrevista. (Expresso online, 05.02.2016)
 [*Com tristeza, digo-o. Cf.: Digo-o com tristeza]

3.3 Anteposição de expressões não referenciais

A topicalização é condicionada pelas propriedades referenciais das expressões nominais, não sendo topicalizáveis as palavras negativas nem os quantificadores puros (cf. Duarte 1987; 1997; Martins 1997; Barbosa 2006). As estruturas de focalização não estão sujeitas a este tipo de restrições, como mostram os exemplos (67) a (71), em que ocorrem focalizadas as palavras negativas *nada*, *ninguém* e os quantificadores *pouco*, *tudo*, *poucas*. As palavras negativas comportam-se, no entanto, de forma particular relativamente à inversão do sujeito (cf. secção 3.5) e a outras propriedades (cf. secção 3.8), constituindo-se num caso especial de focalização.

- (67) As investigações duraram anos e no fim *pouco ou nada* se ficou a saber.
- (68) São declarações que *em nada* contribuem para o bom andamento das negociações.
- (69) Só a diplomacia resolveria este conflito, mas parece que *a ninguém* interessa esta solução. (Diário de Notícias, 06.09.2013, Maria João Tomás)
- (70) Comovente foi ver as igrejas cheias com um público para quem tanto Vivaldi como Feldman são novidade e que *tudo* escutam... religiosamente. (Expresso, 23.07.2011, Jorge Calado)
- (71) *Poucas* são as mulheres que precisam de um homem para as sustentar. (Expresso, 06.02.2016, Maria Filomena Mónica)

3.4 Inversão do sujeito

A focalização de expressões referenciais torna obrigatória a inversão do sujeito, originando a ordem OVS quando o objeto está focalizado. A topicalização, pelo contrário, não condiciona, por si, a posição do sujeito e é perfeitamente compatível com a ordem OSV, como mostram os exemplos em (72).

- (72) a. A esse espetáculo, os meus amigos aconselharam-me a não ir.
 b. Espinafres, o João odeia.

As frases com ordem OVS (73) a (75) exemplificam a inversão do sujeito na construção de focalização. As correspondentes frases agramaticais com ordem OSV (apresentadas entre parêntesis retos a seguir a cada exemplo) não podem ser salvas pela focalização simultânea dos dois constituintes pré-verbais porque a focalização contrastiva (diferentemente da topicalização) apenas pode marcar um constituinte por frase (cf. secção 3.8).

(73) No esquecimento ficaria outra história: (...) (Expresso, 07.12.2013)
[*No esquecimento outra história ficaria.]

(74) De lá vem a mais escandalosa das notícias.
[*De lá a mais escandalosa das notícias vem.]

(75) A grande notícia te dou eu agora.
[*A grande notícia eu te dou agora.]

No exemplo (76), a alteração da ordem OVS para OSV é possível mas altera completamente a interpretação da frase inicial, como evidenciam as paráfrases entre parêntesis retos. A explicação para a diferença semântica entre (76a) e (76b) é que na primeira frase o objeto anteposto está focalizado, enquanto na segunda está topicalizado.

(76) a. Assim trabalho eu. (Expresso, 11.01.2016, Alexandra Carita, citando Júlio Pomar)
[= ‘é assim que eu trabalho’]
b. Assim, eu trabalho.
[= ‘assim/dessa maneira, eu aceito trabalhar’]

A inversão do sujeito torna-se opcional na construção de focalização, podendo mesmo deixar de ser a opção preferencial, quando a expressão referencial focalizada está associada a um marcador de foco como *só* ou *até*, o que é exemplificado em (77).

(77) a. Só na Bélgica, Holanda e Luxemburgo é legal a eutanásia.
b. Só na Bélgica, Holanda e Luxemburgo a eutanásia é legal.
c. Até ao filho pede o *cretino* dinheiro emprestado.
d. Até ao filho o *cretino* pede dinheiro emprestado.

A inversão do sujeito também não é obrigatória quando estão focalizadas palavras negativas. Assim, em (78), podemos ver como a mesma frase admite tanto a ordem OVS como a ordem OSV. Veja-se Costa/Martins (2011) para uma explicação da opcionalidade da inversão do sujeito quando estão focalizados constituintes que integram palavras negativas ou marcadores de foco como *só* e *até*.

(78) a. Parece que a *ninguém* interessa essa solução.
b. Parece que a *ninguém* essa solução interessa.

3.5 Peso informacional do predicado verbal

A topicalização não é permitida quando o comentário (dentro da estrutura tópico-comentário em que a construção se articula) não tem suficiente peso/relevância informacional. A focalização não apresenta este tipo de restrição como demonstra a boa formação das frases em (79) a (82), as quais não encontram correspondência possível em estruturas de topicalização. As frases apresentadas entre parêntesis retos e marcadas como agramaticais podem ser claramente identificadas como casos de topicalização porque exibem ordem OSV ou ênclise (cf. secções 3.2 e 3.4).

- (79) *De notícias se faz o nosso mundo.* (Slogan da RTP N)
 [*De notícias, o nosso mundo faz-se. Cf.: O nosso mundo faz-se de notícias.]
- (80) *Na crise financeira reside o problema.*
 [*Na crise financeira, o problema reside. Cf.: O problema reside na crise financeira.]
- (81) *Destas quatro partes consta o relatório.*
 [*Destas quatro partes, o relatório consta. Cf.: O relatório consta destas quatro partes.]
- (82) *Sem ponta de ironia o escrevo.* (Expresso, 09.05.2015, Pedro Santos Guerreiro)
 [*Sem ponta de ironia, escrevo-o. Cf.: Escrevo-o sem ponta de ironia.]

Alguns autores têm considerado que é uma restrição de natureza prosódica que está por detrás da agramaticalidade das frases com topicalização em que o comentário não tem o peso necessário (cf. Frota/Vigário 1996; Duarte 2013, 420). Ou seja, para estes autores estaria em causa o peso prosódico e não o peso/relevância informacional do comentário como condição de boa formação das estruturas de topicalização. Contudo, diferenças de aceitabilidade como as exemplificadas em (83) e (84) mostram que é a relevância informacional do comentário relativamente ao tópico, e não o peso prosódico do comentário, que condiciona a aceitabilidade da topicalização. Tenha-se em atenção que as formas verbais de (83b/d) e (84b) podem ser produzidas com acento prosódico neutro (ou seja, não necessitam de um acento contrastivo para que as frases sejam bem formadas). O mesmo acontece com a forma verbal *vi* em (85).

- (83) a. ??Esse livro, escrevi.
 b. Esse livro, {odiei / adorei}.
 c. ?Esse livro, lemos. (Duarte 2013, 420)
 d. Esse livro, estudámos.
- (84) a. *Sem ironia, escrevo-o.
 b. Sem ironia, {vegeta-se/definha-se}.
- (85) Esse filme, *vi*. (O do Nanni Moretti é que não *vi*.)

3.6 Relação com o quantificador flutuante *todos*

Os constituintes antepostos por topicalização podem desprender-se do quantificador flutuante *todos*, o que não acontece com os constituintes focalizados (cf. Duarte 1997), como demonstra o contraste de gramaticalidade entre as frases (86b) e (87b). As frases (87a-87b) representam a construção de topicalização, como evidencia a colocação do clítico em ênclise; as frases (88a-88b) representam a construção de focalização, o que é confirmado pela colocação do clítico em próclise (cf. secção 3.2).

- (86) a. Estes livros *todos*, ofereceu-me a Maria.
 b. Estes livros, ofereceu-me a Maria *todos*.
- (87) a. **ESTES LIVROS TODOS** me ofereceu a Maria.
 b. ***ESTES LIVROS** me ofereceu a Maria *todos*.

3.7 Extraposição de orações relativas

A focalização torna possível a extraposição de orações relativas, a topicalização não (cf. Cardoso 2010, de onde são extraídos os exemplos (88) e (89)). A comparação entre os exemplos (88a) e (88b) mostra que o constituinte *crianças que não gostam de chocolate* pode ser topicalizado. Por sua vez, o contraste de gramaticalidade entre (88c) e (88d) demonstra que o nome *crianças* não pode ser movido para a posição de tópico, desligando-se do modificador relativo *que não gostam de chocolate*, mas pode ser movido para a posição de foco contrastivo nas mesmas circunstâncias. A quantificação por *poucas* desambigua a estrutura da frase (88d), tornando-a um exemplo claro de focalização (cf. secção 3.3). As frases (89) a (91) são exemplos adicionais da compatibilidade entre a construção de focalização e a extraposição de orações relativas. O contraste de gramaticalidade entre (89a) e (89b) é mais uma vez revelador de que a focalização e a topicalização interagem de forma diversa com a extraposição de orações relativas.

- (88) a. Também conheço *crianças que não gostam de chocolate*.
 b. *Crianças que não gostam de chocolate*, também conheço.
 c. **Crianças*, também conheço *que não gostam de chocolate*.
 d. *Poucas crianças* conheço *que não gostam de chocolate*.
- (89) a. *Uma estranha doença* lhe diagnosticaram *que lhe retirou toda a alegria*.
 b. **Uma estranha doença*, diagnosticaram-lhe (então) *que lhe retirou toda a alegria*.
- (90) *Poucos colegas* consultei *que tenham contribuído com qualquer sugestão interessante*.
- (91) Parece certo que *novas orientações* haverá que tornam inútil todo o trabalho que fizemos.

3.8 Recursividade e ordem relativa de tópicos e focos contrastivos

Na mesma frase pode ocorrer mais do que um constituinte topicalizado. À recursividade da topicalização opõe-se a limitação da focalização a um constituinte por frase (cf. Zubizarreta 1999). A frase (92a) apresenta o constituinte *em Itália* na posição de tópico, mas é possível modificá-la de modo a trazer para uma segunda posição de tópico outros constituintes, como se mostra em (92b–92c).⁷ A ocorrência de mais do que um tópico por frase não altera as propriedades da topicalização descritas nas secções anteriores. Assim, (92d), por comparação com (92b–92c), mostra, por exemplo, que a opcionalidade de inversão do sujeito se mantém.

- (92) a. Em Itália, o assunto será discutido no Parlamento já em março. (Expresso online, 10.02.2016, Cristina Peres)
 b. Em Itália, já em março, o assunto será discutido no Parlamento.
 c. Em Itália, no Parlamento, o assunto será discutido já em março.
 d. Em Portugal, no Parlamento, não está ainda agendada a discussão sobre a legalização da eutanásia.

Coexistindo na mesma frase dois constituintes topicalizados, a sua posição relativa pode alternar, como se exemplifica em (93). Esta propriedade, tal como a recursividade, estende-se a outras construções de tópicos marcados. Os pronomes clíticos apresentados entre parêntesis nas frases de (93) confirmam que, no aspeto em análise, a topicalização e a deslocação à esquerda clítica são semelhantes.

- (93) a. Um cão, ao meu filho, dou(-lho) assim que ele me pedir.
 b. Ao meu filho, um cão, dou(-lho) assim que ele me pedir.
 c. Os títulos dos livros, nessa estante tão alta, não (os) consigo ler.
 d. Nessa estante tão alta, os títulos dos livros, não (os) consigo ler.
 e. Aos animais, no jardim zoológico, não se pode dar(-lhes) comida.
 f. No jardim zoológico, aos animais, não se pode dar(-lhes) comida.

As palavras negativas constituem uma exceção à limitação dos focos contrastivos antepostos a um constituinte por frase (recorde-se que as palavras negativas não podem ser tópicos). Diferentemente das expressões referenciais, as palavras negativas

⁷ O conjunto de frases em (i) mostra a possibilidade de diferentes constituintes alternarem na posição de tópico (compare-se (ia) com (ib)) sem, no entanto, poderem ser topicalizados simultaneamente, como mostra (ic). A agramaticalidade de (ic) é consequência da condição relativa ao peso informacional do comentário que se discutiu na secção 3.4. O comentário tem de ser suficientemente relevante enquanto predicado acerca do tópico.

- (i) a. Amanhã, quinta-feira, o Eurogrupo discute o plano orçamental português. (Expresso online, 10.02.2016, Cristina Peres)
 b. O plano orçamental português, o Eurogrupo discute amanhã, quinta-feira.
 c. *?Amanhã, quinta-feira, o plano orçamental português, o Eurogrupo discute.

tornam possível a focalização simultânea de diferentes constituintes, mais uma vez indicando serem um caso particular dentro das estruturas de focalização (cf. secções 3.3 e 3.4). Na frase (94a) coocorrem em posição pré-verbal os constituintes *nunca*, *nada* e *de ninguém*. As frases (94b-94d) evidenciam que qualquer deles poderia ocorrer em posição pós-verbal. Diferentemente dos tópicos, as palavras negativas focalizadas parecem ter uma ordenação linear fixa.

- (94) a. Nunca nada de ninguém se saberá.
 b. Nunca nada se saberá de ninguém.
 c. Nada de ninguém se saberá nunca.
 d. Nunca de ninguém se saberá nada.

Sendo topicalização e focalização duas construções distintas, podem coocorrer na mesma frase. Neste caso, como mostram os exemplos (95) a (98), o constituinte topicalizado precede o constituinte focalizado. Note-se que nas frases (97) e (98) não há recursividade da focalização, já que cada um dos focos contrastivos antepostos pertence a uma oração diferente dentro da estrutura de coordenação.

- (95) a. [Dele,]_{Tópico} [pouco ou nada]_{Foco} sei.
 b. [A vocês]_{Tópico} [com alegria]_{Foco} o digo.

- (96) O Muro tinha ruído, a União Soviética tinha implodido, o mundo banhava-se na luz astral do capitalismo como sistema perfeito. [Nos desertos de África e do Afeganistão,]_{Tópico} [um admirável mundo novo]_{Foco} se preparava. (Expresso, 25.08.2012, Clara Ferreira Alves)
 [= ‘Nos desertos de África e do Afeganistão, era um admirável mundo novo o que se preparava’]

- (97) [Na Administração Pública Portuguesa,]_{Tópico} [tudo]_{Foco} se cria, [nada]_{Foco} se perde e [nada]_{Foco} se transforma. (Expresso, 31.12.2015, Miguel Sousa Tavares)

- (98) [Ali,]_{Tópico} [nada]_{Foco} se perdeu e [nada]_{Foco} se transformou. (Expresso, 31.12.2015, Miguel Sousa Tavares)

A partir da frase apresentada em (99a), com o constituinte *na Europa* na posição de tópico, é possível construir (99b), com focalização sintática de *só na Bélgica, Holanda e Luxemburgo* (em vez da focalização *in situ*), mas não é possível (99c). Nesta última frase, o constituinte *na Europa* ocorre à direita do constituinte focalizado, não sendo portanto um tópico. Também não pode ser um foco anteposto, dada a limitação destes a um constituinte por frase. Daqui decorre a agramaticalidade de (99c), tal como de (99d). Na frase (99e) parece ser a incompatibilidade entre a colocação no final da frase do constituinte *na Europa* e a sua interpretação como tópico que provoca a agramaticalidade. Curiosamente, uma estrutura clivada melhoraria a aceitabilidade da frase, enquanto a supressão do marcador de foco *só* a tornaria de todo inaceitável (cf. as frases dadas entre parêntesis retos). Isto mostra que muito fica por explorar relativa-

mente à interface sintaxe-semântica nas estruturas de focalização, topicalização e clivagem e ao modo como estas estruturas interagem entre si.

- (99) a. Na Europa, a eutanásia só é legal na Bélgica, Holanda e Luxemburgo. (Expresso online, 10.02.2016, Cristina Peres)
 b. Na Europa, só na Bélgica, Holanda e Luxemburgo é legal a eutanásia.
 c. *Só na Bélgica, Holanda e Luxemburgo na Europa é legal a eutanásia.
 d. *?Só na Bélgica, Holanda e Luxemburgo SEM RESTRIÇÕES é legal a eutanásia.
 e. ??Só na Bélgica, Holanda e Luxemburgo é legal a eutanásia na Europa.
 [É só na Bélgica, Holanda e Luxemburgo que é legal a eutanásia na Europa.]
 [*Na Bélgica, Holanda e Luxemburgo é legal a eutanásia na Europa.]

4 Conclusão

Descrevem-se neste capítulo ordens não canónicas dos constituintes frásicos que correspondem ou à anteposição do objeto (OSV/OVS) ou à inversão do sujeito (VSO/VOS), relativamente à ordem básica SVO. Procura-se esclarecer a que estratégias gramaticais correspondem as ordens não canónicas (marcação de tópico, marcação de foco contrastivo, marcação de foco/proeminência informacional, marcação de foco simultaneamente contrastivo e informacional, marcação de teticidade), que fatores, internos à frase ou contextuais, favorecem ou limitam cada um dos tipos de anteposição do objeto e de inversão do sujeito, que propriedades sintáticas caracterizam as diferentes estruturas com alteração da ordem básica e que efeitos interpretativos delas decorrem.

A partir da frase com ordem básica «A princesa apaixonou-se pelo ogre», as frases com anteposição do objeto, em (i)–(ii), e com inversão do sujeito, em (iii)–(vi), ilustram as principais estruturas estudadas ao longo do capítulo.

Objetos antepostos

- | | | |
|------|---|--------------------------------|
| (i) | <u>Pelo ogre</u> , a princesa apaixonou-se.
(Mas também se apaixonou pelo príncipe.) | Tópico (topicalização) |
| (ii) | <u>Pelo ogre</u> se apaixonou a princesa.
[= 'Foi pelo ogre que a princesa se apaixonou.'] | Foco contrastivo (focalização) |

Sujeitos invertidos

- | | | |
|-------|--|---------------------------------------|
| (iii) | Quem se apaixonou pelo ogre?
Apaixonou-se pelo ogre <u>a princesa</u> . | Foco informacional (foco estreito) |
| (iv) | Esta história conta como se apaixonou pelo ogre <u>a princesa Branca de Neve</u> . | Saliência informacional (foco largo) |
| (v) | Estava <u>a princesa</u> apaixonada pelo ogre, quando chegou um príncipe diferente de todos os outros. | Frase tética/descriptiva (teticidade) |

Sujeitos invertidos

- | | | |
|------|---|----------------------------------|
| (vi) | Estás a contar mal a história. Apaixonou-se a <u>princesa</u> pelo ogre, não foi a rainha.

[= 'Foi a princesa quem se apaixonou pelo ogre, não foi a rainha.'] | Foco contrastivo e informacional |
|------|---|----------------------------------|

O exemplo (vi) mostra que um foco contrastivo sobre o sujeito numa frase de ordem VSO pode ser corretivo. Pelo contrário, o foco contrastivo anteposto, exemplificado em (ii), não é natural em contextos corretivos no português europeu.⁸ O foco contrastivo associado à ordem VSO tipicamente evoca ou traz para o domínio do discurso as entidades com as quais se estabelece o contraste (i.e. compara explícita ou implicitamente alternativas particulares; cf. os exemplos da secção 2.3), por isso são muito naturais as continuações corretivas.

As diferentes opções de posicionamento do sujeito frásico podem traduzir diferenças subtis de atitude do falante sem alterar o conteúdo proposicional básico da frase, como se exemplifica em (vii)–(ix) fazendo o constituinte sujeito alternar entre a posição pré-verbal de tópico numa frase categórica, a posição imediatamente pós-verbal de uma frase tética e a posição final associada à marcação de foco informacional.

- (vii) Morreram dez pessoas no acidente, 18 encontram-se em estado grave e *cerca de 90* sofreram ferimentos ligeiros. (Expresso online, 10.02.2016, Cristina Peres)
- (viii) Morreram dez pessoas no acidente, encontram-se 18 em estado grave e sofreram *cerca de 90* ferimentos ligeiros.
- (ix) Morreram dez pessoas no acidente, encontram-se em estado grave 18 e sofreram ferimentos ligeiros *cerca de 90*.

5 Referências

- Aikhenvald, Alexandra (2004), *Evidentiality*, Oxford/New York, Oxford University Press.
- Ambar, Manuela (1992), *Para Uma Sintaxe da Inversão Sujeito-Verbo em Português*, Lisboa, Colibri.
- Ambar, Manuela (1999), *Aspects of the Syntax of Focus in Portuguese*, in: George Rebuschi/Laurice Tuller (edd.), *The Grammar of Focus*, Amsterdam/Philadelphia, Benjamins, 23–53.
- Ambar, Manuela/Veloso, Rita (2001), *On the nature of wh-phrases, word order and wh-in-situ: Evidence from Portuguese, French, Hungarian and Tetum*, in: Yves D'Hulst/Johan Rooryck/Jan Schroten (edd.), *Romance Languages and Linguistic Theory 1999: Selected papers from «Going Romance» Leiden 1999*, Amsterdam/Philadelphia, Benjamins, 1–38.

⁸ O português europeu parece comportar-se neste aspeto diferentemente de outras línguas românicas (cf. Rizzi 1997; Zubizarreta 1999; Bianchi/Bocci/Cruschina 2015). Note-se que o foco contrastivo anteposto pode incidir sobre o sujeito, em vez do objeto, caso em que a ordem dos constituintes frásicos será SVO mas se mantêm, genericamente, as propriedades descritas na secção 3 para a construção de focalização (cf. Lobo/Martins no prelo-b).

- Barbosa, Pilar (2001), *On inversion in wh-questions in Romance*, in: Aafke Hulk/Jean-Yves Pollock (edd.), *Subject Inversion in Romance and the Theory of Universal Grammar*, Oxford/New York, Oxford University Press, 20–59.
- Barbosa, Pilar (2006), *Ainda a questão dos sujeitos pré-verbais em Português Europeu: uma resposta a Costa (2001)*, D.E.L.T.A. 22:2, 345–402.
- Barbosa, Pilar (2009), *Two kinds of subject pro*, *Studia Linguistica* 63:1, 2–58.
- Barbosa, Pilar/Duarte, Maria Eugénia L./Kato, Mary A. (2005), *Null subjects in European and Brazilian Portuguese*, *Journal of Portuguese Linguistics* 4:2, 11–52.
- Bianchi, Valentina/Bocci, Giuliano/Cruschina, Silvio (2015), *Focus fronting and its implicatures*, in: Enoch O. Aboh/Jeanette C. Schaeffer/Petra Sleeman (edd.), *Romance Languages and Linguistic Theory 2013: Selected papers from «Going Romance» Amsterdam 2013*, Amsterdam/Philadelphia, Benjamins, 1–20.
- Cardoso, Adriana (2010), *Variation and Change in the Syntax of Relative Clauses*, tese de Doutorado, Universidade de Lisboa.
- Cinque, Guglielmo (1990), *Types of \bar{A} -Dependencies*, Cambridge MA, MIT Press.
- Costa, João (1998), *Word Order Variation: A constraint-based approach*, The Hague, Holland Academic Graphics.
- Costa, João (2004), *Subject Positions and the Interfaces: The Case of European Portuguese*, Berlin/New York, Mouton de Gruyter.
- Costa, João/Figueiredo Silva, Maria Cristina (2006), *On the (in)dependence relations between syntax and pragmatics*, in: Valéria Molnár/Susanne Winkler (edd.), *The Architecture of Focus*, Berlin/New York, Mouton de Gruyter, 83–104.
- Costa, João/Martins, Ana Maria (2011), *On Focus Movement in European Portuguese*, *Probus* 23:2, 217–245.
- CRPC – *Corpus de Referência do Português Contemporâneo*, Lisboa, CLUL, <http://alfclul.clul.ul.pt/CQPweb> (06.02.2016).
- De Haan, Ferdinand (2005), *Encoding speaker perspective: evidential*, in: Zygmunt Frajzyngier/Adam Hodges/David S. Rood (edd.), *Linguistic Diversity and Language Theories*, Amsterdam/Philadelphia, Benjamins, 379–397.
- Duarte, Inês (1987), *A Construção de Topicalização na Gramática do Português*, tese de Doutorado, Lisboa, Universidade de Lisboa.
- Duarte, Inês (1997), *Ordem de palavras: sintaxe e estrutura discursiva*, in: Ana Maria Brito et al. (edd.), *Sentido que a Vida Faz. Estudos para Óscar Lopes*, Porto, Campo das Letras, 581–592.
- Duarte, Inês (2003), *Frases com tópicos marcados*, in: Maria Helena Mira Mateus et al., *Gramática da Língua Portuguesa*, Lisboa, Caminho, 489–506.
- Duarte, Inês (2013), *Construções de Topicalização*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 12, Lisboa, Fundação Calouste Gulbenkian, 401–426.
- Frota, Sónia/Vigário, Marina (1996), *On Weight Effects in European Portuguese*, comunicação apresentada no *GLOW Workshop on Weight Effects*, Atenas.
- Hernanz, M^a. Lluïsa/Brucart, José M^a. (1987), *La Sintaxis: Principios teóricos. La oración simple*, Barcelona, Editorial Crítica.
- Kato, Mary A./Martins, Ana Maria (2016), *The Main Varieties of Portuguese: an overview on word order*, in: Leo Wetzels/Sergio Menuzzi/João Costa (edd.), *Handbook of Portuguese Linguistics*, Hoboken, NJ, Wiley-Blackwell.
- Kuroda, Sige-Yuki (1972), *The Categorical and the Thetic Judgments*, *Foundations of Language* 9, 153–185.
- Kuroda, Sige-Yuki (1992), *Japanese Syntax and Semantics*, Dordrecht, Kluwer.
- Kuroda, Sige-Yuki (2005), *Focusing on the matter of Topic: a study on «wa» and «ga» in Japanese*, *Journal of East Asian Linguistics* 14, 1–58.

- Lambrecht, Knud (1988), *Presentational cleft constructions in spoken French*, in: John Haiman/Sandra A. Thompson (edd.), *Clause Combining in Language and Discourse*, Amsterdam/Philadelphia, Benjamins, 135–179.
- Lambrecht, Knud (1994), *Information Structure and Sentence Form: Topic, focus, and the mental representation of discourse referents*, Cambridge, Cambridge University Press.
- Lambrecht, Knud (2000), *When subjects behave like objects: An analysis of the merging of S and O in Sentence-Focus constructions across languages*, *Studies in Language* 24:3, 611–682.
- Larson, Richard (1988), *On the Double Object Construction*, *Linguistic Inquiry* 19, 335–391.
- Larson, Richard (1990), *Double Object Revisited: Reply to Jackendoff*, *Linguistic Inquiry* 21, 586–632.
- Lasnik, Howard/Saito, Mamoru (1992), *Move α : Conditions on Its Application and Output*, Cambridge MA, MIT Press.
- Leonetti, Manuel (2014), *Spanish VSX*, in: Karen Lahousse/Stefania Marzo (edd.), *Romance Languages and Linguistic Theory 2012: Selected papers from «Going Romance» Leuven 2012*, Amsterdam/Philadelphia, Benjamins, 37–64.
- Leonetti, Manuel (2015), *On word order in Spanish copular sentences*, in: Isabel Pérez-Jiménez/Manuel Leonetti/Silvia Gumiel-Molina (edd.), *New Perspectives on the Study of «Ser» and «Estar»*, Amsterdam/Philadelphia, Benjamins, 203–236.
- Lobo, Maria/Martins, Ana Maria (no prelo-a), *Subjects*, in: Elisabeth Stark/Andreas Dufter (edd.), *Manual of Romance Morphosyntax and Syntax*, Berlin/Boston, De Gruyter.
- Lobo, Maria/Martins, Ana Maria (no prelo-b), *Estratégias de marcação de Foco: ordem dos constituintes frásicos e estruturas clivadas*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 3, Lisboa, Fundação Calouste Gulbenkian.
- Martins, Ana Maria (1994), *Clíticos na História do Português*, tese de Doutorado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Martins, Ana Maria (1997), «*Alguns*», «*poucos*», «*muitos*», «*todos*» e a relação Sintaxe-Semântica, in: Ana Maria Brito et al. (edd.), *Sentido que a Vida Faz: Estudos para Óscar Lopes*, Porto, Campo das Letras, 679–692.
- Martins, Ana Maria (2010), *Constituent order in simple (and root) declarative clauses*, Manuscrito, Universidade de Lisboa, acessível online em: <http://alfclul.clul.ul.pt/wochwel/documents/Martins%20Constituent%20order.pdf> (09.02.2016).
- Martins, Ana Maria (2013a), *The interplay between VSO and coordination in two types of non-degree exclamatives*, *Catalan Journal of Linguistics* 12, 1–27.
- Martins, Ana Maria (2013b), *A colocação dos pronomes pessoais clíticos*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 2, cap. 42, Lisboa, Fundação Calouste Gulbenkian, 2231–2302.
- Matras, Yaron/Sasse, Hans-Jürgen (edd.) (1995), *Verb-Subject Order and Theticity in European Languages*, Berlin, Akademie-Verlag.
- Peterson, Tyler/Sauerland, Uli (edd.) (2010), *Evidence from Evidentials*, The University of British Columbia Working Papers in Linguistics 28.
- Raposo, Eduardo (1995), *Próclise, ênclise e posição do verbo em português europeu*, in: João Camilo dos Santos/Frederick G. Williams (edd.), *O Amor das Letras e das Gentes, in honor of Maria de Lourdes Belchior Pontes*, Santa Barbara, CA, Center for Portuguese Studies, University of California at Santa Barbara, 455–481.
- Raposo, Eduardo (2000), *Clitic Positions and Verb Movement*, in: João Costa (ed.), *Portuguese Syntax. New Comparative Studies*, Oxford/New York, Oxford University Press, 266–297.
- Rizzi, Luigi (1997), *On the fine structure of the left periphery*, in: Liliane Haegeman (ed.), *Elements of Grammar*, Dordrecht, Kluwer, 281–337.
- Sasse, Hans-Jürgen (1987), *The thematic/categorical distinction revisited*, *Linguistics* 25, 511–580.

- Sasse, Hans-Jürgen (1995), «*Theticity*» and VS order: a case study, in: Yaron Matras/Hans-Jürgen Sasse (edd.), *Verb-Subject Order and Theticity in European Languages*, Berlin, Akademie-Verlag, 3–31.
- Sasse, Hans-Jürgen (1996), *Theticity*, Arbeitspapier 27 (Neue Folge), Universität zu Köln, Institut für Sprachwissenschaft.
- Sasse, Hans-Jürgen (2006), *Theticity*, in: Giuliano Bernini/Marcia L. Schwartz (edd.), *Pragmatic Organization of Discourse in the Languages of Europe*, Berlin/New York, Mouton de Gruyter, 257–308.
- Torrego, Esther (1984), *On Inversion in Spanish and Some of its Effects*, *Linguistic Inquiry* 15, 103–127.
- Zimmermann, Malte (2007), *Contrastive Focus*, in: Caroline Féry/Gisbert Fanselow/Manfred Krifka (edd.), *The Notions of Information Structure*, Potsdam, Universitätsverlag Potsdam, 147–159.
- Zimmermann, Malte/Onea, Edgar (2011), *Focus marking and focus interpretation*, *Lingua* 121:11, 1651–1670.
- Zubizarreta, María Luisa (1998), *Prosody, Focus, and Word Order*, Cambridge MA, MIT Press.
- Zubizarreta, María Luisa (1999), *Las funciones informativas: tema y foco*, in: Ignacio Bosque/Violeta Demonte (orgs.), *Gramática Descriptiva de la Lengua Española*, vol. 3, Madrid, Espasa-Calpe, 4215–4244.

Ana Maria Martins

15 A colocação dos pronomes clíticos em sincronia e diacronia

Abstract: Neste capítulo descreve-se a colocação dos pronomes clíticos no português europeu contemporâneo, pondo em destaque as especificidades do português europeu no quadro das línguas românicas. Depois oferece-se uma perspectiva diacrónica que identifica os pontos de estabilidade e de mudança no sistema de colocação dos pronomes clíticos e mostra como, ao longo do tempo, o português divergiu sintaticamente de outras línguas ibéricas com as quais partilhava, no período medieval, um padrão idêntico de distribuição da próclise e da ênclise. Neste percurso de divergência, o português preserva aspetos centrais do sistema original, em contraste com línguas como o espanhol e o catalão. O capítulo aborda ainda a questão da interpolação (i.e. a descontinuidade entre clítico e verbo) e esclarece que a interpolação dialetal observável no português europeu contemporâneo não é a continuação da interpolação medieval. Ao longo do capítulo identificam-se questões em aberto relativamente ao tópico em análise.

Keywords: padrões de colocação dos pronomes clíticos, ênclise/próclise, interpolação, mudança sintática, variedades do português

1 Introdução

O padrão de colocação dos pronomes clíticos apresenta, no português europeu, especificidades que o afastam bastante dos padrões de colocação, mais simples, da maior parte das línguas românicas. Com o português europeu alinham apenas o galego (Álvarez/Xove 2002) e, parcialmente, o asturiano (González i Planas 2007; Fernández-Rubiera 2006; 2009; 2010). Numa perspectiva diacrónica, o português europeu mantém, no entanto, um sistema de colocação dos pronomes clíticos que, em aspetos essenciais, era partilhado com as outras línguas românicas medievais. A sua complexidade poderá estar na origem dos processos evolutivos, no sentido da simplificação, que virão a ocorrer no seio da família românica e que levarão à diversidade de padrões de colocação dos pronomes clíticos que hoje conhecemos. A complexidade do sistema de colocação dos pronomes clíticos está também na base da aquisição mais tardia deste aspeto da sintaxe pelas crianças portuguesas, comparativamente às crianças francesas, italianas ou espanholas (cf. Costa/Fiéis/Lobo 2014) e não é menos desafiante para quem aprende o português europeu como segunda língua (cf. Madeira/Xavier 2009).

No quadro românico, o português brasileiro falado tem o sistema mais simples, pois apresenta próclise generalizada, com o clítico a ocorrer sempre imediatamente

antes do verbo de que é complemento, quer com as formas finitas quer com as formas não finitas do verbo, incluindo, além do infinitivo e do gerúndio, o particípio passado (cf. Kato/Martins 2016).

Português brasileiro

- | | | |
|--------|---|---------------------------------------|
| (1) a. | O chefe <i>me</i> despediu. | <i>próclise ao verbo finito</i> |
| b. | <i>Me</i> dá um beijo. | <i>próclise ao imperativo</i> |
| c. | Você não pode <i>me</i> despedir. | <i>próclise ao infinitivo</i> |
| d. | Ele está sempre <i>me</i> provocando. | <i>próclise ao gerúndio</i> |
| e. | Você não tinha ainda <i>me</i> contado. | <i>próclise ao particípio passado</i> |

O francês tem o sistema que mais se aproxima do português brasileiro, mas deixando fora da próclise generalizada o imperativo morfológico, com o qual os pronomes clíticos ocorrem imediatamente depois do verbo, ou seja, em ênclise. Além disso, o francês, como a maior parte das línguas românicas, não admite cliticização ao particípio passado.

Francês

- | | | |
|--------|--|---------------------------------|
| (2) a. | Je <i>le</i> rencontre chaque jour. | <i>próclise ao verbo finito</i> |
| b. | J'ai tout fait pour <i>te</i> voir. | <i>próclise ao infinitivo</i> |
| c. | Je ne peux <i>m'</i> empêcher de rire. | <i>idem</i> |
| d. | En <i>le</i> voyant, j'ai pensé à toi. | <i>próclise ao gerúndio</i> |
| e. | Donne- <i>lui</i> son cadeau. | <i>ênclise ao imperativo</i> |

O espanhol, o catalão e o italiano apresentam um sistema de colocação dos pronomes clíticos condicionado pela morfologia verbal, tal como o francês, mas a ênclise ocorre com as formas não finitas do verbo (infinitivo e gerúndio) e com o imperativo, enquanto a próclise ocorre regularmente com as formas finitas do verbo.

Espanhol

- | | | |
|--------|--|---------------------------------|
| (3) a. | <i>Le</i> pedí que <i>me lo</i> contara. | <i>próclise ao verbo finito</i> |
| b. | Tengo que contarte algo. | <i>ênclise ao infinitivo</i> |
| c. | Queremos seguir viéndolos juntos. | <i>ênclise ao gerúndio</i> |
| d. | Dáselo a tu hijo. | <i>ênclise ao imperativo</i> |

Embora o fator morfológico (particularmente a oposição entre formas verbais finitas e não finitas) não deixe de ter relevância para a colocação dos pronomes clíticos no português europeu, são fatores de natureza sintática os que se revelam dominantes.¹

¹ A morfologia verbal é um fator determinante no que diz respeito à mesóclise (i.e. a colocação do pronome clítico numa posição interna à forma verbal, como em: *contar-te-ia*, *contar-te-ei*). Este tipo de colocação dos pronomes clíticos é uma variante da ênclise associada à especificidade da morfologia do futuro e condicional (cf. Roberts 1992; Duarte/Matos 2000; Roberts/Roussou 2002; 2003; Oltra-Masuet/Arregi 2005). A mesóclise é, no português europeu contemporâneo, própria da língua escrita e de registos orais formais. Foi comum às línguas ibéricas medievais mas sobrevive hoje apenas no

De facto, em contraste com o francês, o italiano, o espanhol, o catalão e o português brasileiro, o sistema do português europeu apresenta próclise e ênclise quer em frases finitas quer em frases não finitas, com as duas ordens em distribuição complementar nas frases finitas, mas em variação livre em alguns tipos de orações infinitivas. Nas frases (4a) e (4c), a colocação proclítica seria agramatical e na frase (4b) seria agramatical a ênclise. As frases (4d–e) ilustram, pelo contrário, um contexto que torna possível a variação entre próclise e ênclise com o infinitivo simples.

Português europeu

- (4) a. Ela contou-me tudo. *ênclise ao verbo finito*
 b. Eu também *lhe* disse a verdade. *próclise ao verbo finito*
 c. Não pude impedir-me de rir. *ênclise ao infinitivo*
 d. Não parou de *se* queixar o tempo todo. *próclise ao infinitivo*
 e. Não parou de queixar-se o tempo todo. *ênclise ao infinitivo*

Alguns dos fatores que determinam a colocação enclítica ou proclítica no português europeu são fáceis de identificar, em particular a oposição entre orações principais e subordinadas finitas e a oposição de polaridade entre frases afirmativas e negativas.

Português europeu

- (5) a. Ao sábado, geralmente vejo-os. *ênclise em oração principal*
 b. É possível que os veja no sábado. *próclise em oração subordinada finita*
 c. As laranjeiras regam-se. *ênclise em frase afirmativa*
 d. As oliveiras não se regam. *próclise em frase negativa*

As frases negativas e as orações completivas finitas são os dois domínios em que, desde mais cedo, as crianças adquirem o padrão de colocação proclítico (≈16 A aquisição dos pronomes clíticos no português L1). As oposições afirmação/negação, principal/subordinada (finita) são totalmente estranhas à colocação dos pronomes clíticos na maior parte das línguas românicas contemporâneas, mas eram relevantes tanto no português como noutras línguas românicas medievais, conforme se ilustra em (6) e (7) com exemplos do catalão e do espanhol (cf. Granberg 1988; Castillo Lluch 1996; Fischer 2002; Martins 1994; Batllori/Iglésias/Martins 2005; Nieuwenhuijsen 2006; Bouzouita 2008a; 2008b).

Catalão medieval

- (6) a. Los cathalans vengueren a Curial e digueren-li quinys paraments volia
 b. Melchior respòs que *li* playa.
 c. Vós no *m'*avets fet res.

(*Curial e Güelfa*; Ferrando 2007, 105, 134)

português europeu, que neste aspeto se distingue do galego (Álvarez/Xove 2002). Cf. final da secção 3.1.

Espanhol medieval

- (7) a. E en aquel tiempo el rey pagava *se* mucho de caçar con aves
 b. E el rrey dixo que *les* perdonava todos los enojos que *le* avian fecho fasta alli
 c. E don Joan non *lo* quiso fazer

(*Gran Crónica de Alfonso XI*; Catalán 1977, 18, 19)

A secção 2 deste capítulo aborda o tema da colocação dos pronomes clíticos no português europeu numa perspetiva sincrónica, enquanto a secção 3 o faz numa perspetiva diacrónica (que integra dados de variação sincrónica). O tópico principal das secções 2 e 3.1 é a distribuição da próclise e da ênclise, enquanto a secção 3.2 é dedicada ao fenómeno da interpolação (Chenery 1905; Ramsden 1963), i.e. a possibilidade de um clítico pré-verbal ocorrer separado do verbo por outros constituintes. A secção 4 oferece algumas observações conclusivas.

2 Ênclise e próclise no português europeu contemporâneo

Nesta secção traça-se um quadro geral, sumário, da colocação dos pronomes clíticos no português europeu contemporâneo (2.1). De seguida, discutem-se alguns casos de aparente variação entre próclise e ênclise, em frases finitas, com o objetivo de mostrar que diferentes linearizações correspondem a diferenças de estrutura e de interpretação (2.2). Deixa-se fora do âmbito desta secção a questão da subida dos pronomes clíticos, que será brevemente abordada, numa perspetiva diacrónica, no final da secção 3.1.²

2.1 Quadro geral da distribuição da ênclise e da próclise

A ênclise e a próclise estão em distribuição complementar nas frases que apresentam o verbo em forma finita. Na sequência textual em (8), o primeiro pronome clítico é obrigatoriamente próclítico porque ocorre dentro de uma oração subordinada e o segundo é necessariamente enclítico por pertencer a uma oração principal.

2 O termo *subida do clítico* designa os casos em que o pronome clítico é extraído da oração infinitiva e cliticiza ao verbo que a seleciona como complemento. As ordens com e sem subida do clítico, exemplificadas respetivamente em (i) e (ii), são ambas admitidas no padrão contemporâneo.

- (i) Este projeto de investigação é do mais avançado que se pode encontrar no jornalismo mundial. (Expresso online, 04.04.2016, Pedro Santos Guerreiro)
- (ii) Não se trata de gerir informação, mas de poder confirmá-la, estudá-la, aprofundá-la, discernir entre o que é lícito e o que é duvidoso. (Expresso online, 04.04.2016, Pedro Santos Guerreiro).

- (8) Ela é tecedora de teias que outros usam. É uma aranha vegetariana, não come carne. Os outros que *a* comam, isso é com eles. Os outros comem-*na*, isso é conosco. (Expresso online, 04.04.2016, Pedro Santos Guerreiro)

A sequência textual em (9), atesta além da oposição principal/subordinada (*Quando os encontro/pergunta-lhes*), as oposições entre frase afirmativa/frase negativa (*pergunta-lhes/não se frequentam*) e frase afirmativa sem/com desencadeadores de próclise (*pergunta-lhes/talvez até se detestem*).

- (9) Quando *os* encontro *pergunta-lhes* «olha lá, o que está a fazer o fulano de tal?» e nunca sabem dos colegas, não *se* frequentam, talvez até *se* detestem. (Expresso, 28.03.2015)

Do mesmo modo, nas frases afirmativas em (10a–b) só a ênclise é possível, seja em oração principal, como (10a), ou coordenada, como (10b); mas nas orações subordinadas finitas em (11a–d) só a próclise é possível. A próclise é também a única opção nas frases negativas em (12a–e) e nas frases (13a–l), nas quais ocorrem antes do verbo quantificadores (*muito, pouco, todos*), sintagmas QU- (*como, o que*) e advérbios que, tal como os quantificadores e as palavras QU-, não são compatíveis com a colocação enclítica (*até, só, lá* enfático, *também, sempre, já, talvez*).

- (10) a. Sim, pedi-*lhe* que ele fizesse de si próprio. (Expresso, 28.03.2015)
 b. O crime é inextirpável mas o combate dissuade-*o* e pune-*o*. (Expresso online, 04.04.2016, Pedro Santos Guerreiro)
- (11) a. Bem sei que *se* trata de empresas privadas (Expresso online, 04.04.2016, Nicolau Santos)
 b. Empresas fantasma permitem perder o rasto ao dinheiro, ao património e a quem os detém. (Expresso online, 04.04.2016, Pedro Santos Guerreiro)
 c. Quando *se* olha para um tubarão, ninguém repara nas rémoras. (Expresso online, 04.04.2016, Pedro Santos Guerreiro)
 d. O jornalismo de informação está a cumprir uma função social, como *lhe* compete. (Expresso online, 04.04.2016, Pedro Santos Guerreiro)
- (12) a. Não *se* sai incólume de uma viagem destas. (Expresso online, 04.04.2016, Pedro Santos Guerreiro)
 b. Nas restantes ditaduras nem *se* falou do assunto. (Expresso online, 15.04.2016, Daniel Oliveira)
 c. Nunca *me* identifico com a personagem. (Expresso, 13.10.2012)
 d. Mas nada daquilo *se* cola à nossa vida. (Expresso, 07.12.2013)
 e. Quando for para o céu – ou para o inferno – ninguém *me* vai dar uma medalha por ter tocado em 100 países. (Expresso, 05.04.2014)
- (13) a. Muito *se* falou, pouco *se* disse e nada *se* resolveu. (CETEMPúblico)
 b. Queria fazer da segurança o tema do Outono, mas todos *lhe* trocaram as voltas. (CETEMPúblico)
 c. Como *te* atreves?! (Expresso, 01.02.2014)

- d. O que *o* fez voltar ao cinema agora? (Expresso, 28.03.2015)
- e. Gosto de tudo e até *me* descontraíu. (Expresso, 05.04.2014)
- f. A função superior de uma investigação jornalística como esta só *se* completa se o sistema mudar. (Expresso online, 15.04.2016, Pedro Santos Guerreiro)
- g. Mas eu insisti, lá *o* convenci. (Expresso, 28.03.2015)
- h. Garcia Lorca também *se* chamava Frederico. (Expresso, 28.03.2015)
- i. O *star system* sempre *me* desagradou. (Expresso, 13.10.2012)
- j. Muito antes as duas tribos já *se* matavam. (Expresso, 29.11.2014)
- l. Enfim, talvez os métodos da ficção *me* pudessem levar a uma espécie de autobiografia dirigida a Fellini. (Expresso, 28.03.2015)

A caracterização do conjunto de itens que no português europeu tornam a próclise obrigatória em frases afirmativas (não subordinadas) é a parte do sistema de colocação dos pronomes clíticos que mais resiste a uma descrição simples e integradora. Se a classe das palavras QU- não admite exceções (todos os itens desta classe são desencadeadores de próclise), já para os quantificadores é necessário distinguir dois grupos, o dos determinantes *relativos/proporcionais*, que impõem a próclise, e o dos determinantes *absolutos/cardinais* (Fodor/Sag 1982; Higginbotham 1987; Martins 1997), que não têm o mesmo efeito, sendo compatíveis com a ênclise. Além disso, o conjunto dos advérbios que desencadeiam a próclise não corresponde a uma classe semântica particular. Por fim, podem ser fatores puramente estruturais que determinam a colocação dos pronomes clíticos em ênclise ou em próclise. É o caso da oposição entre topicalização (ou outras estruturas de marcação de tópicos) e focalização (≈14 Ordem dos constituintes frásicos: sujeitos invertidos; objetos antepostos). A mesma palavra, ou constituinte pré-verbal, pode assim aparecer associada ora à ênclise ora à próclise, mas a alternância entre uma e a outra colocação não é livre, como ilustraremos na secção 2.2. (Para uma descrição detalhada da colocação dos pronomes clíticos no português europeu contemporâneo, cf. Martins 2013).

O Quadro 1 identifica os itens lexicais e fatores estruturais que fazem emergir o padrão proclítico nas frases afirmativas, não subordinadas. Os itens que estão marcados com asterisco (*) são os desencadeadores de próclise categóricos, ou seja, os que excluem sempre a ênclise. Quanto aos restantes, os contrastes semânticos que exibem conforme se associam à próclise ou à ênclise decorrem ou de ambiguidade lexical ou do tipo de estrutura sintática em que se integram (podendo ser tópicos, focos contrastivos, marcadores de foco ou marcadores de ênfase). A focalização é um fator estrutural associado à próclise, não estando portanto limitada a itens lexicais particulares, pelo que na penúltima linha do Quadro 1 não há lugar à apresentação de uma lista.³

³ Para além dos casos identificados no Quadro 1, a próclise pode ainda ocorrer associada à expressão da afirmação enfática, como se exemplifica em (i)–(ii) abaixo. Também neste caso não há, em princípio, restrições quanto à natureza do constituinte que ocorre à esquerda do verbo e do clítico, mas este tipo de estrutura é infrequente no português contemporâneo. Nas estruturas de afirmação enfática

Quadro 1: Indutores de próclise em frases simples afirmativas (cf. Martins 2013).

Proclisadores: categóricos* não categóricos	Exemplos
Quantificadores <i>algo, alguém, algum, alguns, ambos*, bastante*, bastantes, cada, demasiado*, demasiados*, mais*, menos*, muito*, muitos, pouco*, poucos*, quase, qualquer, raramente*, raros, suficientes*, tal*, tais*, tamanho, tamanhos, tão*, tanto, tantos, todo*, todos, tudo*, vários</i>	<i>Algo <u>me</u> diz que estás a mentir. Ambos <u>me</u> garantiram que é verdade. Poucos <u>se</u> mostraram surpreendidos. Por todo o lado <u>se</u> festeja. Qualquer um <u>te</u> diria o mesmo. Bastantes vezes <u>nos</u> vimos sem dinheiro. Mais <u>o</u> assustas e menos <u>te</u> diz.</i>
Marcadores de foco inclusivos: <i>também*, até*, mesmo*</i> exclusivos: <i>apenas*, só*, somente*, logo, antes*</i> aspetuais: <i>ainda*, já*, quase*, mal*, talvez*</i>	<i>Apenas um turista <u>o</u> confrontou. Até o melro <u>se</u> pôs a cantar. Logo hoje <u>te</u> tens de ir embora mais cedo. Ainda <u>te</u> digo mais. Mal <u>o</u> vi mas quase <u>me</u> fez cair.</i>
Marcadores de ênfase <i>bem, até*, sempre*, lá, aí, aqui, logo, já*</i>	<i>Bem <u>o</u> quisemos salvar, mas não conseguimos. Eu até <u>te</u> contava, mas não posso. Se fores de comboio, sempre <u>te</u> sai mais barato.</i>
Advérbios focalizados Por exemplo: dêiticos locativos (<i>aí, ali, aqui, cá, lá</i>), <i>assim, sempre*, logo, agora, depois, melhor, pior</i> ; alguns advérbios em <i>-mente</i>	<i>Assim <u>se</u> fará. Ali <u>se</u> instalaram e não saíram mais. Rapidamente <u>se</u> afastou. Agora <u>te</u> exijo uma resposta.</i>
Constituintes não-adverbiais focalizados	<i>De notícias <u>se</u> faz o nosso mundo. Se o livro saiu bem, a si <u>se</u> deve. Com a verdade <u>me</u> enganas. Não fico em casa, isso <u>te</u> garanto.</i>
Interrogativas e exclamativas QU- <i>que*, o que*, quem*, onde*, quanto*, como*, quando*, por que*</i>	<i>Quem <u>te</u> contou? Como ele <u>me</u> irrita!</i>

não se põe em destaque o constituinte pré-verbal (como na focalização), mas sim o valor de verdade da asserção.

- (i) Isabel dos Santos chamou «diploma BPI» a um decreto, aprovado pelo Governo e promulgado pelo Presidente da República, e que transcreve orientações europeias no sentido de serem eliminados os limites aos direitos de voto nas instituições financeiras. Ela o disse e não houve uma voz autorizada que a viesse contestar. (Expresso online, 22.04.2016, Nicolau Santos)
- (ii) Passos Coelho tem nas mãos muito mais que o destino da PT. E já fez de Pilatos uma vez: não pode fazer duas. Desta vez ou escolhe defender o país ou não. A história o julgará. (Expresso, 11.10.2014, Nicolau Santos).

Nas orações gerundivas, o fator mais importante relativamente à distribuição da próclise e da ênclise é a polaridade da frase. As orações gerundivas afirmativas normalmente apresentam ênclise, enquanto as orações gerundivas negativas apresentam próclise.

- (14) a. Desde então, o grupo de jornalistas tem investigado milhões de documentos, juntando-*lhes* ainda mais informação, para criar aquilo a que chamamos os «Panama Papers». (Expresso online, 04.04.2016, Pedro Santos Guerreiro)
- b. Mas a Associação de Defesa dos Utentes de Sangue está determinada a levar o problema até às mais altas instâncias, não *se* importando «de perder batalhas para ganhar guerras». (CETEMPúblico)

A presença de um dos itens identificados no Quadro 1 como proclisador tem numa gerundiva afirmativa o mesmo efeito que numa frase finita afirmativa.

- (15) a. Afirmam que, já tendo feito uma retirada no terreno e já se tendo comprometido a novas retiradas, o lado «pragmático» de Netanyahu florescerá agora e as suas inibições «ideológicas» desvanecer-se-ão. (CETEMPúblico)
- b. Entretanto, a questão dos deputados honorários parece estar em banho-maria, só se prevenendo que seja retomada na próxima conferência de líderes. (CETEMPúblico)

Nas orações infinitivas interagem diferentes fatores para determinar quando ocorre a ênclise ou a próclise. Considerando apenas os que são específicos dos domínios infinitivos, será relevante se o infinitivo é simples ou flexionado, se a oração infinitiva é ou não introduzida por preposição e qual a preposição particular que a introduz, como sintetiza o Quadro 2. Embora o infinitivo simples e o infinitivo flexionado tenham um comportamento semelhante em alguns aspetos (por exemplo, a ênclise com as preposições *a* e *com*, e, noutros casos, a possibilidade de variação entre próclise e ênclise), o infinitivo flexionado aproxima-se mais das frases finitas ao reduzir a variação próclise/ênclise ao contexto particular das orações introduzidas pela preposição *em*. De resto, a próclise e a ênclise estão em distribuição complementar nas orações de infinitivo flexionado, tal como nas frases finitas. Pelo contrário, a variação próclise/ênclise é muito característica das orações de infinitivo simples, ainda que seja contextualmente condicionada.

Quadro 2: Distribuição da próclise e da ênclise em orações infinitivas (Martins 2013).

Infinitivo Simples			Infinitivo flexionado		
Ênclise	Variação Pró-clise/Ênclise	Próclise	Ênclise	Variação Pró-clise/Ênclise	Próclise
Orações não introduzidas por preposição, ou introduzidas pelas preposições <i>a</i> e <i>com</i> (sem proclisadores)	Orações introduzidas pelas preposições <i>de</i> , <i>para</i> , <i>por</i> , <i>em</i> , <i>sem</i> (sem proclisadores)	Orações que incluem proclisadores (exceto <i>não</i>)	Orações não introduzidas por preposição, ou introduzidas pelas preposições <i>a</i> e <i>com</i> , (sem proclisadores)	Orações introduzidas pela preposição <i>em</i>	Orações introduzidas pelas preposições <i>de</i> , <i>para</i> , <i>após</i> , <i>até</i> , <i>sem</i>
Interrogativas indiretas (sem proclisadores)	Orações infinitivas negativas (com a negação expressa por <i>não</i>)		Frases simples exclamativas (sem proclisadores)		Orações que incluem proclisadores
	Orações introduzidas pelos pronomes e advérbios relativos ou interrogativos (sem proclisadores)				
	Orações dependentes de <i>ter</i> <i>que</i> , <i>haver</i> <i>que</i>				

Proclisadores = itens identificados no Quadro 1 e itens que expressam negação

Em (16) e (17) exemplifica-se a variação entre próclise e ênclise que o português europeu permite quando o infinitivo é simples e introduzido por uma preposição diferente de *a* ou *com*. Embora as duas ordens sejam consideradas padrão, os falantes têm normalmente preferência por uma ou outra em função de cada preposição particular.⁴

⁴ Gillier (2009) analisou um pequeno corpus representativo de dois autores literários, Jorge de Sena e José Cardoso Pires, e mostrou que têm comportamentos distintos relativamente às orações de infinitivo simples introduzidas por preposição. Enquanto em José Cardoso Pires ocorre sempre a próclise com as preposições *de*, *para* e *sem*, registando-se ênclise a par da próclise apenas com a preposição *por*, Jorge de Sena apresenta variação entre ênclise e próclise com todas as preposições (i.e. *de*, *para*, *sem* e *por*), sendo a ênclise a opção mais frequente. Cf. nota 6, adiante.

- (16) a. Não fiz tal comparação nem tinha de *a* fazer, pois não era esse o objecto do meu texto. (CETEMPúblico)
 b. Essa pergunta terá de *fazê-la* ao senhor Américo Amorim. (CETEMPúblico)
- (17) a. Com salários em atraso, as 250 operárias haviam contratado autocarros para deslocar-se em protesto até à presidência do Governo Regional, mas os motoristas não compareceram. (CETEMPúblico)
 b. É o caso de doze alunos da universidade privada Jayabaya, de Jacarta, que, relata a agência Lusa, estão a encontrar dificuldades para *se* deslocar a Manila, onde deveriam tomar parte na Conferência de Estudantes da Ásia-Pacífico. (CETEMPúblico)

Diferentemente do português brasileiro, as variedades africanas do português têm essencialmente o sistema de colocação dos pronomes clíticos descrito nesta secção, mas com uma pequena margem de variação próclise/ênclise em contextos que no padrão europeu apenas permitem uma das ordens. Essa pequena margem de variação também se encontra nos dialetos do português europeu. Dou aqui exemplos de próclise em contextos de ênclise, em frases finitas. Mas também ocorre a situação inversa (i.e. ênclise em frases com proclisadores), quer no português dialetal quer nas variedades africanas do português (↗ O português em contacto em África; e veja-se a observação sobre o «grupo de controlo» em ↗16 A aquisição dos pronomes clíticos no português L1).⁵

- (18) a. Mano, ela *me* cansou. (Português de Moçambique; Justino 2010)
 b. Eles *se* encontram marginalizados. (Português de Moçambique; Mapasse 2005, 67)
 c. *Me* disseste que era segredo, não meterias a foto dele no facebook, *me* mentiste. (Português de Angola; Domingos 2010)
 d. Fuguh, vi essa foto e *te* identifiquei bem rápido. (Português de Angola; Domingos 2010)
 e. Eu consegui conviver com a população, consegui mergulhar onde há pobreza em Nigéria, as pessoas *me* olhavam com bons olhos (Português de São Tomé; Gonçalves 2009)
 f. No meu caso eu não estou a conseguir envolver com nenhum homem porque homem são-tomense é muito maldoso. Eles acham que nós somos escrava deles, *nos* maltratam muito. (Português de São Tomé; Gonçalves 2009)
 g. *Me* enganaste?! (Português dialetal; CORDIAL-SIN, Melides, Alentejo)
 h. A gente *lhe* chama os miolinhos. (Português dialetal; CORDIAL-SIN, Serpa, Alentejo)
 i. Bem disse o rapaz: «*Se* mete como está este coiso». (Português dialetal; CORDIAL-SIN, Alvor, Algarve)
 j. Às vezes *me* junto com os meus amigos (Português dialetal; CORDIAL-SIN, Santo André, Vila Real)
 l. Depois os foles – *lhe* chamam – enchem-se. (Português dialetal; CORDIAL-SIN, Fiscal, Minho)
 m. Homem, tu *o* viste no outro dia, por que é que queres tornar a ver? (Português dialetal; CORDIAL-SIN, Pico, Açores)

⁵ É possível analisar algumas das frases em (18) como casos de afirmação enfática (cf. nota 3, acima).

Pela sua complexidade e aquisição relativamente tardia (Costa/Fiéis/Lobo 2014), talvez o sistema de colocação dos pronomes clíticos aqui descrito integre de forma equilibrada uma pequena margem de variação estável, que só deixa de ser visível por pressão normativa. Considerando a mudança ocorrida no português brasileiro, pode pensar-se que num contexto de línguas em contacto e de menor pressão normativa se ultrapassou o limite de variação estável e o sistema se tornou demasiado instável para poder ser adquirido na sua configuração original. Quer dizer, a variabilidade no *input* terá ultrapassado um certo limite sensível para a aquisição de L1, conduzindo à simplificação radical que vemos no sistema de colocação dos pronomes clíticos do português brasileiro.

2.2 Ênclise, próclise e interação entre sintaxe e interpretação

Só alguns dos itens listados no Quadro 1 desencadeiam sempre a próclise (i.e. os que estão marcados com asterisco). Relativamente aos que não são proclisadores categóricos, importa esclarecer que não há variação livre entre próclise e ênclise, como veremos nesta secção através de alguns exemplos.

Os dêiticos locativos *aqui, aí, ali, cá, lá* podem ser facilmente topicalizados ou focalizados contrastivamente (↗14 Ordem dos constituintes frásicos: sujeitos invertidos; objetos antepostos). Em ambos os casos são movidos para posições na periferia esquerda da frase, precedendo portanto o verbo. Quando estão focalizados são desencadeadores de próclise, como acontece com os focos contrastivos antepostos em geral. Quando ocupam uma posição de tópico não têm qualquer efeito sobre a colocação dos pronomes clíticos, sendo portanto compatíveis com a ênclise. As frases com dêitico locativo e próclise são interpretativamente diferentes das frases com dêitico locativo e ênclise, o que é independente do dêitico em si, decorrendo antes das diferenças interpretativas entre estruturas de topicalização e estruturas de focalização. Está em causa, portanto, uma diferença sintática, estrutural, e não lexical. A inclusão dos dêiticos locativos *aqui, aí, ali, cá, lá* no Quadro 1, acima, deve-se ao facto de serem itens que ocorrem frequentemente focalizados e atuam, nesta circunstância, como proclisadores. A sequência textual em (19) mostra-nos o dêitico locativo *ali* a ocorrer primeiro como tópico e depois como foco contrastivo. Como tópico, não desencadeia próclise. Por isso, se substituíssemos o verbo *começar* pelo verbo *iniciar-se*, ocorreria a ênclise. Quando o dêitico *ali* está focalizado, a próclise torna-se obrigatória, como se vê na frase: *Ali se sentem de novo mulheres*. As frases em que *ali* está focalizado (mas não a frase em que *ali* é um tópico) são parafraseáveis por uma estrutura clivada, como é próprio das estruturas de focalização contrastiva: ‘É ali que constroem, por instantes, fragmentos de liberdade’; ‘É ali que, por instantes, se sentem de novo mulheres’.

Em (19), os parêntesis retos e elementos associados foram adicionados ao texto original. As etiquetas «Top» e «Foc» correspondem a Tópico e Foco Contrastivo.

- (19) «3 Horas para Amar» é o título do documentário realizado com reclusas do Estabelecimento Prisional de Santa Cruz do Bispo, em Matosinhos, assinado por Patrícia Nogueira (...). Patrícia propõe uma viagem tão fascinante quanto perturbadora na companhia de quatro mulheres disponíveis para falarem da vida que tinham «lá fora» e da que passaram a ter no interior da cadeia. [Ali]_{Top} a rotina começa [inicia-se] às 8h quando se abrem as celas. Uma não trabalhar, outras ficam no pátio. Pela frente têm uma imensidão de tempo. Falam dos que não estão e, afinal, nunca deixam de povoar aqueles quotidianos. Os maridos, os companheiros, os filhos, as mães, os amigos que às vezes já o não são tanto. Uma vez por mês, durante três horas, as reclusas abrangidas pelo regime de Visitas Íntimas, têm à sua disposição um compartimento da prisão. [Ali]_{Foc} constroem, por instantes, fragmentos de liberdade. [Ali]_{Foc} por instantes, se sentem de novo mulheres. (Expresso online, 05.01.2016, Valdemar Cruz)

Outros advérbios, além dos dêiticos locativos, podem ocupar quer uma posição de foco contrastivo quer uma posição de tópico na periferia esquerda da frase. Nalguns casos, daí decorrem contrastes interpretativos ainda mais marcados do que no exemplo (19). Assim, em (20), podemos ver que o advérbio *loucamente* tem a interpretação de advérbio de modo (modificador do sintagma verbal) quando focalizado, como em (20a), mas é interpretado em (20b) como advérbio orientado para o sujeito. A diferente posição do pronome clítico nas duas frases assinala que só na primeira há focalização.

- (20) a. Loucamente *a* amou. (= ‘foi loucamente que *a* amou’)
 b. Loucamente, amou-*a*. (= ‘foi louco ao amá-la’)

O advérbio *bem* precede o verbo em frases com ênclise quando é um tópico, como em (21a), frase que apresenta topicalização do sintagma verbal com apagamento do verbo. Neste caso, *bem* tem a interpretação de advérbio de modo (modificador do sintagma verbal) e pode ser reduplicado, pois a posição de tópico é recursiva. Como marcador de ênfase, por outro lado, *bem*, apresenta-se despojado do seu conteúdo semântico básico, ocorre na periferia esquerda da frase e desencadeia próclise, conforme se vê em (21b). Como marcador de ênfase, *bem* não pode ser reduplicado, o que evidencia que também neste caso há uma diferença estrutural que se associa ao contraste semântico e à oposição entre ênclise e próclise.

- (21) a. Bem, (bem), come-se naquele restaurante. (= Comer bem, (comer bem), come-se naquele restaurante’)
 b. Bem *me* podias oferecer o jantar. (= ‘Podias oferecer-me o jantar!’)

Há advérbios que não podem ser focalizados. É o caso dos evidenciais e alguns avaliativos, como se exemplifica em (22) com o evidencial *obviamente* e em (23) com o avaliativo *surpreendentemente*. Também não sendo marcadores de ênfase nem marcadores de foco, estes advérbios são incompatíveis com a próclise, daí a agramaticalidade de (22b) e (23b). O outro facto interessante em relação a estes advérbios é que a

interpretação das frases em que ocorrem torna claro que não são tópicos (quer dizer, não é possível interpretá-los como *aboutness topic*/Sujeito lógico de uma estrutura de Predicação, numa perspectiva semântica, nem como informação conhecida, partilhada ou familiar, numa perspectiva pragmático-discursiva). A ênclise não depende pois da presença de um tópico na frase nem os constituintes que precedem o verbo numa frase com ênclise têm de ser tópicos.

- (22) a. Obviamente demito-o! (Humberto Delgado, sobre Salazar)
 b. *Obviamente o demito.

- (23) a. Surpreendentemente, amamo-nos.
 b. *Surpreendentemente nos amamos.

No domínio dos quantificadores também se observa a distinção entre proclisadores categóricos e não categóricos (cf. Quadro 1). Se observarmos o comportamento do quantificador *muitos*, que ora se associa à ênclise ora se associa à próclise, voltamos a constatar a interdependência entre sintaxe e interpretação. Ou seja, não são semanticamente equivalentes frases que, à primeira vista, se distinguem apenas pela posição do pronome clítico relativamente ao verbo.

O quantificador *muitos* pode ser um determinante *absoluto/cardinal*, caso em que refere um conjunto de indivíduos ou entidades particulares (identificáveis e avaliados quantitativamente em termos absolutos), ou um determinante *relativo/proporcional*, caso em que é puramente quantificacional, sem a componente referencial, e especifica uma quantidade avaliada relativamente a um certo valor de referência (cf. Fodor/Sag 1982; Higginbotham 1987; Martins 1997). Como determinante absoluto, *muitos* pode ser parte do constituinte tópico das frases categóricas (≈14 Ordem dos constituintes frásicos: sujeitos invertidos; objetos antepostos) e associa-se à ênclise. Como determinante relativo, o quantificador *muitos* associa-se à próclise em estruturas afins às estruturas de focalização. A distinção relevante está exemplificada em (24). Enquanto na frase (a) se diz que há muitas pessoas que nunca falham a meia maratona de Lisboa (pessoas que poderiam ser identificadas consultando a base de dados de inscrições no evento), a frase (b) estabelece apenas que o número de pessoas que se inscreve anualmente na meia maratona de Lisboa é elevado, nada revelando sobre se há ou não repetição dos mesmos participantes ano após ano.

- (24) a. Muitas pessoas inscrevem-se todos os anos na meia maratona de Lisboa.
 (= ‘pessoas há que se inscrevem todos os anos na meia maratona de Lisboa e não são poucas’)
 b. Muitas pessoas se inscrevem todos os anos na meia maratona de Lisboa.
 (= ‘é sempre elevado o número de participantes na meia maratona de Lisboa, relativamente a números de referência para este tipo de evento’)

Esta diferença de interpretação torna-se particularmente clara no par de frases em (25). A frase (25b) é semanticamente estranha porque num mundo de seres mortais o

mesmo indivíduo não pode suicidar-se ano após ano. Num mundo alternativo, a frase (26), com o nome «marcianos» a substituir «pessoas», parece, no entanto, perfeitamente aceitável.

- (25) a. Muitas pessoas se suicidam todos os anos.
 (= 'é elevado o número de pessoas que anualmente comete suicídio')
- b. #Muitas pessoas suicidam-se todos os anos.
 (= 'há pessoas que todos os anos se suicidam e não são poucas')
- (26) Muitos marcianos suicidam-se todos os anos e reaparecem depois rejuvenescidos.

3 Aspetos diacrónicos da colocação dos pronomes clíticos (estabilidade e mudança)

O português dos séculos XIII e XIV não é muito diferente do português contemporâneo no que diz respeito à distribuição da próclise e da ênclise, como mostram os fragmentos da *Crónica Geral de Espanha de 1344* que se apresentam a seguir (cf. Cintra 1990). Continuamos a ver a próclise ocorrer em frases negativas (e.g. *Não vos está bem*), em orações subordinadas (e.g. *E elles disserom que lho tiinham em grande mercee*) e nas frases afirmativas em que precedem o verbo os mesmos itens que são desencadeadores de próclise no português europeu contemporâneo (e.g. *ainda vos ãvya hũa tenda; ambos vos acordaredes no melhor; E assi vos vingaredes*). Fora destes casos, é a ênclise que ocorre maioritariamente nas frases finitas, embora a próclise também seja possível (e.g. *E dom Alvaro Fernandez lhe disse*). Ou seja, os contextos de próclise obrigatória no português contemporâneo também o eram no português medieval (neste aspeto há estabilidade ao longo do tempo), mas nos contextos em que só a ênclise ocorre no português contemporâneo, pode observar-se variação entre ênclise e próclise no português medieval. A frequência relativa de uma ou outra opção muda bastante ao longo do tempo, como se verá na secção 3.1. Nos séculos XIII e XIV a ênclise é, na ausência de proclisadores, o padrão dominante nas frases finitas afirmativas. Como as sequências textuais em (27) a (29) também mostram, uma outra diferença entre o português medieval e o português contemporâneo observa-se nos contextos de próclise obrigatória. No português medieval, os pronomes clíticos pré-verbais podiam estar separados do verbo por diferentes tipos de constituintes (e.g. *despois que me eu de vos parti*). Este tipo de linearização recebe o nome de *interpolação*, fenómeno de que tratará a secção 3.2.

Nos exemplos (27) a (29), todos os clíticos estão destacados a negrito, os constituintes interpolados estão sublinhados, a próclise sem proclisadores está assinalada por maiúsculas pequenas e todos os enclíticos estão em itálico.

(27) E, quando chegarõ ante el rei, dom Alvaro Fernandez e Pero Vermuiz beyjarõ**lhe** as mãaos por o Cide e disseron**lhe**:

– Senhor! Roy Diaz, o Cide, **se** ãvya ãcomendar ã vossa mercee, como de senhor natural, e tem que **lhe** fezeistes muyta mercee em quanta honrra e bẽ mandastes fazer a sua molher e a suas filhas. E, despois que **me eu de vos** parti, vẽceu o Cide hũa batalha de mouros, de que era senhor el rei Unez, filho de Miraamolý de Marrocos, que **o** veeo cercar ã Vallença cõ L mil cavalleiros. E o Cide sayu a elle e venceu**ho** ã cãpo. E, do seu quinto que elle hi guaanhou, ãvy**vos** estes trezentos cavalloos.

E el rei foi muy alegre por a boa vêtuirã do Cide e louvou / [248d] muito tam nobres doas como **lhe** ãvyava e disse:

– Certo nunca foy ã Espanha vassallo que tantas boas doas mandasse a senhor.

E dom Alvaro Fernandez **LHE** disse:

– Senhor, ainda **vos** ãvya hũa tenda que foy del rei Unez, que nunca **a homẽ** vyo melhor.

E el rey mandou**ha** logo armar e entrou dentro em ella com todos seus ricos homẽes. E el rei com todollos outros disserom que nunca virõ tenda tam nobre. E el rey louvou muito o Cide e todollos seus cavalleiros e agradece**o** muyto o grande presente que **lhe** ãviara. E mandou muy bẽ apousentar dom Alvaro Fernandez e Pero Vermuiz e dar**thes** todo o que era mester. (Cap. DCXXX; Miranda 2013, 226)

(28) E el rei mãdou por os messejeiros do Cide e dissel**thes** como / [249a] agradecia muyto ao Cide o grande serviço que **lhe** fazia; e que avya muy grande sabor de **o** veer e que porem **lhe** dissessem que **lhe** mandava elle rogar que **o** vehesse veer a Requena, ca queria com elle fallar algũas cousas de sua prol, ã razom de casamẽtos pera suas filhas, por que os iffantes de Carriõ **lhe** disserom que casariã com ellas, se elle quisesse, e que a elle parecia que seerã muy bem casadas cõ elles. E os messejeiros do Cide disseron**lhe**:

– Senhor, nos bẽ entendemos que o Cide fara todo o que **lhe vós** mandardes. E, quando **vos cõ elle** virdes, ambos **vos** acordaredes no melhor.

E, despois que **lhes el rei** disse todo o que queria mandar dizer ao Cide, beyjarõ**lhe** as mãaos e despediron**se** delle. E el rei mandou logo por os iffantes de Carriõ e dissel**thes** todo o que fallara cõ os messejeiros do Cide. E elles disserom que **lho** tiinham em grande mercee. (Cap. DCXXXI; Miranda 2013, 227)

(29) – Filhos, que foi esto por que mostrades tanta covardice por vista de hũa besta fera? Nõ **vos** está bem. Devyades de seer nẽbrados como sodes de grande sangue e como sodes meus jenros e como **vos** dey duas nobres spadas. E, per todas estas cousas, devyades de perder grande parte do medo.

Destas pallavras ouverom elles mui grande vergonha e firmaron porẽ mais em seus corações a maa entençõ que avyã fallada. Pero meterõ o feito a jogo. E chamarõ seu tio Sueiro Gonçalvez e saïrom cõ elle do paaço. E elle dissel**thes**:

– Que he esto ou como viïdes assy tristes?

E elles disserom:

– Por que **nos** deshonorrou muy mal o Cide ã mandar soltar o leon.

Entom **LHE** cõtãrõ todo o que avyam cuydado. E dissel**he**:

– Calladedev**os** e fazed que nõ dades nada porem, de guisa que nõ entenda desto o Cide nem hũa cousa, e atendamos ataa que seja passado / [251c] este arroydo destes mouros que som viïndos d'aalem mar. E, despois que esto ouver fym, demandaredes vossas molheres pera **as** levardes pera vossa terra. E elle nõ avera scusa de **vollas nõ** dar nem de **vos** tẽer mais conssgo. E, despois que fordes bem alõgados desta terra, podedes em ellas fazer o que quiserdes. E assi **vos** vingaredes da deshorrã que **vos** he feita. (Cap. DCXXXV; Miranda 2013, 231)

A abordagem diacrônica da colocação dos pronomes clíticos que se faz neste capítulo terá por foco as frases finitas. Mas nos excertos da *Crónica Geral de Espanha de 1344* apresentados acima estão destacados todos os clíticos. Nas estruturas com infinitivo, observa-se subida do clítico com verbos de reestruturação (e.g. *E el rey mandouha logo armar; filho de Miraamol̃ de Marrocos, que o veeo cercar*), o que corresponde ao padrão geral do português medieval, e há próclise ao infinitivo em orações introduzidas pelas preposições *de* e *pera* (e.g. *avya muy grande sabor de o veer*).⁶ A ênclise ao infinitivo ocorre nas frases com verbos de reestruturação quando, em estruturas de coordenação, a elipse do verbo superior torna a subida do clítico impossível (*E mandou muy bẽ aposentar dom Alvaro Fernandez e Pero Vermuiz e darlhes todo o que era mester*).

Veremos de seguida, na secção 3.1, que a estabilização da ênclise ao verbo finito nos contextos que começam por ser, historicamente, de variação ênclise/próclise apresenta um percurso surpreendente.

3.1 A linha evolutiva da ênclise nas frases finitas

A variação entre próclise e ênclise exemplificada em (27) a (29) mantém-se nos séculos seguintes, como se mostra em (30) a (33), mas a frequência da próclise relativamente à ênclise vai crescendo gradualmente até a próclise se tornar fortemente dominante no século XVI. Também ao longo do período medieval acontecem outras mudanças ligadas à expansão da próclise, que depois do século XIII começa a aparecer em contextos sintáticos que antes a excluía. A partir do século XIV, um pronome proclítico pode ocorrer na posição inicial de orações coordenadas introduzidas pela conjunção copulativa *e*, como se exemplifica em (31). No século XV, um pronome proclítico pode ocupar a primeira posição da oração principal de uma frase complexa com subordinada anteposta, como em (32). Por fim, já no século XVI, um pronome proclítico pode estar adjacente ao constituinte topicalizado numa estrutura de deslocamento à esquerda clítica, ocupando a primeira posição não periférica da frase, como em (33).

Ao longo de todo o período histórico do português (desde os primeiros textos, no final do séc. XII, até hoje), mantêm-se estáveis os contextos de próclise obrigatória.

⁶ Nos documentos notariais editados por Martins (1994; 2001), que pertencem aos séculos XIII–XVI, há próclise ao infinitivo, sem exceções, com a preposição *de*. Com a preposição *pera* há variação entre próclise e ênclise até meados do século XIV e próclise obrigatória a partir daí. Com a preposição *a* regista-se ênclise até meados do século XIV, após o que só a próclise ocorre. Está por estudar quando, e em que condições, se estabeleceu o atual padrão de variação entre próclise e ênclise em estruturas de infinitivo simples preposicionado, com as preposições *de*, *para*, etc., ao mesmo tempo que as orações infinitivas introduzidas pela preposição *a* parecem ter regredido para o padrão enclítico. Cf. Martins (1994; 2013); Magro (2005).

- (30) a. e eu enuieyuos dizer per mia carta que me prazia
 b. E rey uos *me* enuiastes dizer per uossa carta que uos desembargariades esse castello de Albofeyra
 (Doc. de 1260, Chancelaria de D. Afonso III; Duarte 1986, 87)
- (31) a. E el rey abraçouho e beyjouho e fezlhe muy grande honrra.
 b. o Cide *lhes* fez muyta honrra e *lhes* deu grandes doas em ouro e prata e cavallos e outras cousas
 (Crónica Geral de Espanha de 1344; Miranda 2013, 229, 228)
- (32) a. E quando el-rrei ouvio que elle entrara e que os seus corriam a terra e rroubavom, pesou-*lhe* muito de os leixarem assi entrar
 b. e quando entendeo outra vez de mover esta guerra, *lhe* escreveo que fallasse com o duque e com seu irmão
 (Fernão Lopes, Crónica de D. Fernando; Macchi 1975, 418, 427)
- (33). a. á pessoa que vos tal disse ou espreveo, pergumtelhe vosalteza omde estava symam afomso
 b. aos que imda lá sam, *lhe* tenho dado seguros e *lhe* mando agora notificar ho voso perdam
 (Cartas de Afonso de Albuquerque; Pato 1884, 94, 99)

Entre o século XIII e o século XVI, o português parece estar a evoluir na mesma direção que o espanhol e o catalão relativamente à colocação dos pronomes clíticos, substituindo a ênclise pela próclise nas frases com verbo finito. Mas no século XVII surgem os primeiros sinais de inversão desta tendência e a partir do século seguinte a mudança no sentido da generalização da ênclise torna-se definitiva. O percurso divergente em relação ao espanhol e ao catalão conduzirá ao modelo atual, com eliminação da variação entre próclise e ênclise nas frases finitas (ainda que, fora da variedade padrão europeia, a variação possa ainda observar-se marginalmente, como se mostrou na secção 1.1).

A Figura 1 representa a retração da ênclise a favor da próclise até ao século XVI e, inversamente, a expansão da ênclise e declínio da próclise a partir do século seguinte (cf. Lawton 1966; Ogando 1980; Silva 1989; Salvi 1990; Kaiser 1992; Martins 1994; 2011; Moraes 1995; Ribeiro 1995; Sousa 2004; Galves/Britto/Sousa 2005; Galves/Sousa 2005; Rocha 2009). Os dados relevantes dizem respeito apenas aos contextos que no português europeu contemporâneo são obrigatoriamente de ênclise, mas exibiam variação entre ênclise e próclise em etapas anteriores da história do português.

A linha evolutiva traçada na Figura 1 é surpreendente, quer a interpretemos como evidência de uma mudança gramatical cuja orientação se altera radicalmente a meio do percurso quer como uma sequência de dois processos evolutivos em direções opostas. Uma hipótese alternativa é que a gramática fortemente proclítica, de cariz «pan-ibérico», que está representada na maioria dos textos dos séculos XV a XVII tenha estado limitada sociolinguisticamente à classe dominante e aos setores letrados, através dos quais se tornou o modelo da escrita literária e burocrática (Martins

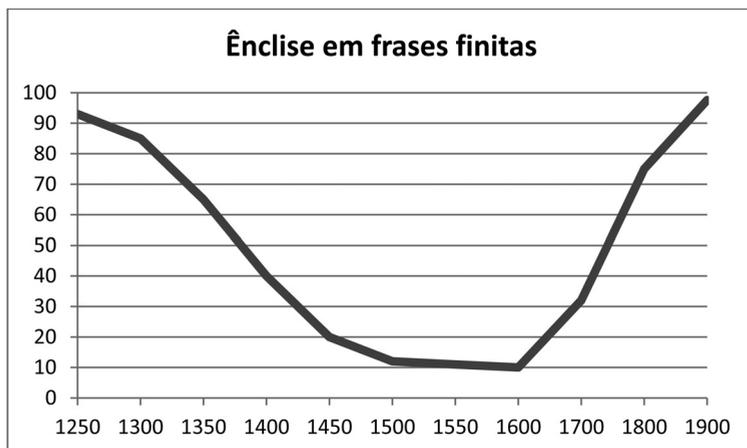


Figura 1: A evolução da ênclise em frases finitas (sem proclisadores), sécs. XIII–XIV.

2011; 2015). O período relevante é de forte influência sociocultural e política de Espanha (iniciada muito antes da perda da independência de Portugal, entre 1580 e 1640) e há outros indicadores linguísticos, além da colocação dos pronomes clíticos, que sugerem um efeito significativo do contacto entre português e espanhol no âmbito da corte, da aristocracia e dos escritores literários (cf. Castro 2002; Döhla 2014). Nesta perspetiva, a linha evolutiva representada na Figura 1 poderá ser enganadora, ao esconder um quadro de mudança linguística mais complexo. Ou seja: entre a elite política, social e cultural ter-se-á desenvolvido uma variedade socioletal que tenderia para a convergência com o espanhol relativamente à generalização da próclise em frases finitas; ao mesmo tempo, desenhas-se e consolidava-se, em outros setores da sociedade, um percurso evolutivo mais direto entre a ênclise (em variação com a próclise) do português medieval e a ênclise (sem variação) do português europeu contemporâneo. A gramática socialmente prestigiada, que favorecia a generalização da próclise, domina de forma esmagadora a produção textual portuguesa entre os séculos XV e XVII, mas acabará por revelar-se uma «mudança falhada» (Potsma 2010), ao perder progressivamente espaço para a gramática mais enclítica.

A hipótese de que as fontes textuais dificultam a perceção do caminho que, historicamente, conduziu ao padrão de colocação dos pronomes clíticos do português europeu pode ser suportada por paralelos com outras situações. Desde logo o caso do português do Brasil, onde é radical a diferença entre língua falada e língua escrita relativamente à colocação dos pronomes clíticos (cf. Kato/Martins 2016). Por outro lado, se nos mantivermos no âmbito do português europeu, as estruturas de subida do clítico e a mesóclise podem ser exemplos paradigmáticos.

No português medieval, a subida dos pronomes clíticos em estruturas infinitivas era, com raras exceções, o padrão geral. A partir do século XVI, a possibilidade de cliticização ao infinitivo vai ganhando terreno pouco a pouco. A subida do clítico (com cliticização ao verbo finito) e a não subida (com cliticização ao infinitivo) estão exemplificadas em (34). O Quadro 3 mostra como a opção por manter o clítico no interior da oração infinitiva se vai estabelecendo ao longo do tempo, com o correspondente decréscimo da frequência da subida do clítico.

- (34) a. ela disse q. não *ma* podia dar aqueles dias
 b. qdo. fui comfesar*me* comt[e]i tudo isto ao confesor
 (Inquisição de Lisboa, ano de 1699; Marquilhas 2000, 301, 303)

Quadro 3: Subida dos pronomes clíticos, séculos XVI–XX (Martins 2006).

	Subida do clítico	Cliticização ao infinitivo
Afonso de Albuquerque (1462?–1515)	95%	5%
Damião de Góis (1502?–1574)	100%	0%
Fernão Mendes Pinto (1510–1583)	97%	3%
Diogo do Couto (1542–1616)	94%	6%
António Vieira (1608–1697)	82%	18%
Luís António Verney (1713–1792)	33%	67%
Almeida Garrett (1799–1854)	55%	45%
Oliveira Martins (1845–1894)	0%	100%

Observando o Quadro 3, poderia pensar-se que a partir do século XIX a subida do clítico deixa de ser uma opção no português europeu e a cliticização ao infinitivo se generaliza (como aconteceu aliás no português do Brasil). Em geral, a produção literária do século XX dá evidência no mesmo sentido. Mas, na verdade, a informação que se extrai dos textos é enganadora porque os dados que nos fornecem são o resultado de uma forte pressão normativa no sentido de considerar a subida do clítico uma opção ilegítima ou incorreta. A escolha normativa de uma das opções permitidas pela gramática criou, no aspeto relevante, uma grande distância entre língua escrita e língua falada. De facto, a análise de dados de língua falada revela que a subida do clítico (por oposição à cliticização ao infinitivo) pode atingir, no português europeu contemporâneo, valores percentuais próximos dos da produção escrita do século XVII (cf. Quadro 3), sem que se registre variação regional ou socioletal significativa (Magro 2005; Andrade 2010; Paiva/Barbosa, submetido).

Por fim, uma referência breve à mesóclise, que na norma padrão do português europeu é obrigatória, nos contextos de ênclise, se o verbo estiver no futuro ou no condicional (quer dizer: sintaticamente a mesóclise é uma variante da ênclise que decorre da especificidade da morfologia de futuro e condicional). A mesóclise existe

hoje apenas no português europeu (tendo desaparecido do português brasileiro e do galego) e é própria da escrita ou da oralidade formal, estando em geral ausente da fala espontânea. Em registos menos monitorados, não é raro ver emergir variantes não normativas, com ênclise em lugar da mesóclise. Poderia pensar-se que está em causa uma mudança recente, de expressão ainda limitada, no sentido da substituição da mesóclise pela ênclise, paralelamente ao que aconteceu noutras variedades românicas. Ora quando lemos textos medievais verificamos que a variação entre ênclise e mesóclise está presente no português desde o século XIII, sugerindo uma linha de continuidade até às produções não normativas do português contemporâneo. De novo, a estabilização da mesóclise na produção escrita é o resultado da pressão normativa e esconde a variação que a língua tem mantido ao longo do tempo.

(35) Eu *te* daria um capão assado e uma regueifa e *faria-te* tudo comer e *dar-te-ia* em cima ãa copa chea de vinho que bebesses (Primeiro Livro de Linhagens; Piel/Mattoso 1980, 48–49)

(36) detremino de o aviar pa [...] ã hũu navyo q estou faltando e asim ordenareivos a vida se poder (*Post Scriptum*; PSCR1145, CLUL 2014)

No exemplo (35), retirado de um texto do século XIII, vemos aparecer a próclise a par da mesóclise e da ênclise, visto que há variação entre ênclise e próclise no português medieval. Surgem assim as três variantes, associadas a formas verbais do condicional, em frases finitas afirmativas. O exemplo (36), com a ênclise a ocorrer em lugar da mesóclise com uma forma verbal do futuro, pertence a uma carta do ano de 1594 (de Jerónimo Leitão, capitão da capitania de São Vicente, para o filho, Francisco Leitão), integrada no arquivo digital *Post Scriptum*.⁷

3.2 A interpolação

Embora as configurações de próclise sejam estáveis ao longo do tempo, o português medieval difere do português contemporâneo por permitir (opcionalmente) que os pronomes clíticos ocorram separados do verbo por outros constituintes, como se observou nas sequências textuais em (27) a (29), de onde se extraem as frases (37a–e). Esta descontinuidade entre clítico e verbo recebe o nome de *interpolação* (cf. Martins

7 É possível encontrar algumas ocorrências de ênclise, em lugar da mesóclise, no corpus CETEMPúblico, mas são provavelmente o resultado de erros de digitalização/formatação do corpus. Por exemplo:

(i) *Irá-se* privilegiar aqueles que tenham maior incidência no índice de preços no consumidor, no sentido de que, pela evolução dos seus preços, se possa aferir das suas margens de lucro. (CETEMPúblico, par=ext206481-eco-96b-2)

1994; Fiéis 1996; Namiuti 2006; 2008; Magro 2006; 2010a; 2010b). Nos exemplos em (37) os constituintes interpolados estão delimitados pelas palavras em itálico, i.e. o clítico, à esquerda, e o verbo, à direita.

- (37) a. E, depois que *me* eu de vos *parti*
 b. hũa tenda que foy del rei Unez, que nunca *a* homẽ *vyo* melhor.
 c. nos bẽ entendemos que o Cide fara todo o que *lhe* vós *mandardes*. E, quando *vos* cõ elle *virdes*, ambos vos acordaredes no melhor.
 d. depois que *lhes* el rei *disse* todo o que queria mandar dizer ao Cide
 e. E elle nõ avera scusa de *vollas* nõ *dar*

A interpolação do português medieval só é permitida nos contextos de próclise obrigatória, estando excluída dos contextos de variação próclise/ênclise, mesmo quando nesses contextos ocorre a próclise. Além disso, o marcador de negação predicativa não pode legitimar a interpolação porque *nom/não* tem um requisito de adjacência ao verbo, do qual só pode ser separado por pronomes clíticos. A dependência entre *nom/não* e o verbo determina, por outro lado, que o marcador de negação predicativa ocorra frequentemente interpolado, desde que haja um outro proclisador à sua esquerda (cf. (37e), onde a preposição *de* é o proclisador que legitima a interpolação de *nõ*). Nas mesmas condições, diferentes tipos de constituintes podem ser interpolados, como ilustrado em (38). Entre eles: o sujeito, modificadores adverbiais, predicados adjetivais e complementos verbais antepostos (próprios das frases com ordem SOV do português medieval), incluindo complementos oracionais com verbo não finito. Na mesma frase pode ocorrer interpolado mais do que um constituinte frásico e não parece haver condicionamentos quanto à sua extensão e complexidade (embora exemplos como (38j-1) sejam infrequentes). Em geral, qualquer constituinte que possa ocorrer à esquerda do verbo e que não esteja topicalizado nem focalizado (i.e. que não pertença à periferia esquerda da frase) pode ser interpolado entre o clítico e o verbo no português medieval.

- (38) a. com os lavradores que *as* ssemeadas *teuerẽ* (ano de 1472, Martins 2001, 516)
 b. E sse *as* nos comprar nõ *quisermos* entõ uẽderdelas uos (ano de 1329; Martins 2001, 414)
 c. como *lhe* entregou as terras que *lhe* de dar *avia* (Fernão Lopes; Costa 2001, 184)
 d. E com todas as pertenças que aos ditos casaaes *lhes* dereitamẽte *pertẽçe* (ano de 1522; Martins 2001, 306)
 e. ssegundo *as* elle sempre *trouxera* (ano de 1472; Martins 2001, 277)
 f. todollos adubynos que *lhes* conpridoiros e neçareos *forem* (ano de 1476; Martins 2001, 459)
 g. E mjlhor *se* o elle Jorge Duarte e pessoa depos elles mjlhor *poder* aver (ano de 1513; Martins 2001, 301)
 h. mas nõca achou quẽ *lhe* delle nouas allgũas *dissese* (Demanda do Santo Graal; WOChWEL, ID DSG168,15)
 i. E sse pela uẽtujra *uos* algẽ a dita vỹa *embargar* (ano de 1296; Martins 2001, 377)

- j. quem quer que *lhe* sobre elle e parte delle e sobre as ditas casas e qualquer cousa delas algũu embarguo ou empedymento *puser* (ano de 1540; Martins 2001, 556)
- l. Por as sobreditas vinha e oliual e câpo e mato que *lhus* assy os sobreditos Prior e Raçoeiros ã nome da dita Eigreja de Sam bartholameu escambhadas e pela guisa que suso dito he e declarado ã escambho dadas e outorgadas *auyã* (ano de 1385; Martins 2001, 462)

Embora a interpolação se registre sobretudo em orações subordinadas, é possível encontrá-la também em frases simples e orações principais, desde que um proclisador preceda o verbo. Exemplo disso são as frases em (39a–c), nas quais a próclise é desencadeada, e a interpolação legitimada, respetivamente pela palavra negativa *nunca*, pelo quantificador *todos* e pelo advérbio *ainda*.

- (39) a. e nunca *se* i homem *assentou* (José de Arimateia; WOChWEL, ID JAR41, .36)
- b. e todos *o* dantes nom *tinhão* por tal cavaleiro (José de Arimateia; WOChWEL, ID JAR49, .41)
- c. Certas, disse ho scudeiro, ajnda *uos* oje mujto pessar *averrá* (Demanda do Santo Graal; WOChWEL, ID DSG131, .12)

No período medieval, a interpolação de *não* distingue-se da interpolação de outros constituintes por apresentar níveis de frequência muito superiores, como se mostra, com dados extraídos de documentos notariais, nos Quadros 4 e 5. Só a interpolação de *não* atinge valores de 90% ou superiores, aparecendo ao longo do período medieval como padrão não marcado nas frases negativas em que há um clítico e um proclisador independente de *não*. Além disso, durante um intervalo de tempo limitado (séculos XV–XVII) a interpolação de *não*, em contraste com a interpolação de outros constituintes, podia ocorrer, ocasionalmente, fora dos contextos de próclise obrigatória (cf. Martins 1994; Namiuti 2006; 2008).

Quadro 4: Interpolação de *não* entre os séculos XIII e XVI (Martins 1994).

	Séc. XIII	Séc. XIV	Séc. XV	Séc. XVI
clítico- <i>não</i> -verbo	94,10%	96,80%	90,70%	90,00%
<i>não</i> -clítico-verbo	5,90%	3,20%	9,30%	10,00%

Quadro 5: Interpolação de outros constituintes entre os séculos XIII e XVI (Martins 1994).

	Séc. XIII	Séc. XIV	Séc. XV	Séc. XVI
clítico-X-verbo	66,70%	69,10%	57,00%	51,70%
X-clítico-verbo	33,30%	30,90%	43,00%	48,30%

A interpolação de *não* mantém-se uma opção gramatical até aos nossos dias, como se exemplifica em (40), e é considerada uma variante padrão (cf. Duarte 2003, 866–867), ainda que não esteja disponível para todos os falantes nem seja a opção mais usada (i.e. a ordem *não*-clítico-verbo é mais comum que a ordem clítico-*não*-verbo). Como as frases em (40) também ilustram, continua a ocorrer apenas nos contextos sintáticos que a permitiam no português medieval, o que se deve à estabilidade diacrónica dos contextos proclíticos (descontada a próclise opcional de etapas passadas da história da língua).

- (40) a. Os maridos, os companheiros, os filhos, as mães, os amigos que às vezes já *o* não são tanto. (Expresso online, 05.01.2016, Valdemar Cruz).
- b. O que parece ser então verdade? Em poucas palavras, os portugueses não confiam no governo e no parlamento, mas também *os* não *receiam*. (Visão, 23.04.2005, António Barreto).
- c. Também se *o* não *comeres* agora, podes dizer-lhe adeus. (Cardoso Pires ⁴1975, 37)

Enquanto a interpolação de *não* persiste ininterruptamente até ao português contemporâneo, a interpolação de outros constituintes, com as características acima descritas, deixa de se registar nas fontes escritas a partir do século XVII.

Nos dialetos do português europeu contemporâneo (de Norte a Sul de Portugal e também nas ilhas dos Açores) é possível encontrar interpolação não limitada à palavra *não*, o que aparentemente indicaria a persistência residual da interpolação medieval fora da variedade padrão (Martins 1994; Barbosa 1996; Hinzelin 2010). Mas Magro (2006; 2010a; 2010b) demonstrou que não é assim. Por um lado, a interpolação dialetal tem propriedades muito diferentes da interpolação medieval, encontrando-se limitada a palavras que expressam dêixis: pessoal (*eu, nós, a gente, ele, ela, eles, elas, a nós*), espacial (*aqui, ali, aí, cá, lá, para lá, para aí; esta, isto, isso*), temporal (*agora, depois, então, hoje, ontem*), aspetual (*ainda, já*) e modal (*assim*). Por outro lado, o tipo de interpolação característico do português dialetal ocorre, de facto, também em autores literários, mas só a partir do século XIX. Ou seja, no século XVII é ainda visível, de forma residual, a interpolação medieval (limitada a autores nascidos na primeira metade do século). A partir daí e ao longo de todo o século XVIII não há interpolação senão do marcador de negação predicativa. No século XIX emerge um outro tipo de interpolação, limitada a pronomes e advérbios dêiticos, que Magro (2006) mostra ser uma inovação. A persistência ao longo do tempo da interpolação de *não* terá sido o fator que permitiu a inovação, tornando-se assim, de alguma maneira, a ponte entre a interpolação generalizada do português medieval e a nova interpolação associada à dêixis (sem que a última seja a continuação da primeira). Enquanto a interpolação medieval depende estritamente de fatores sintáticos (perdendo-se porque se altera a estrutura funcional da frase e, concomitantemente, as possibilidades de linearização dos constituintes frásicos), a interpolação contemporânea é um fenómeno mais superficial, de metátese morfológica (Harris/Halle 2005), sem deixar de ter uma base sintática (Magro 2006).

O Quadro 6 evidencia o desaparecimento da interpolação medieval no século XVII e a emergência da interpolação de dêiticos no século XIX. Do conjunto de doze autores oitocentistas estudados por Magro (2006), apenas três, Eça de Queiroz, Trindade Coelho e Mário de Sá Carneiro, não apresentam exemplos de interpolação de dêiticos.

Quadro 6: Interpolação entre os séculos XVI e XIX (Magro 2006; Martins 1994; Namiuti 2006; 2008).

	autor – data de nascimento		interpolação de <i>não</i>	interpolação de outros constituintes
Século XVI	Damião de Góis	1502	sim	sim
	Fernão Mendes Pinto	1510?	sim	sim
	Diogo do Couto	1542	sim	sim
	Luís de Sousa	1556	sim	sim
	Francisco Rodrigues Lobo	1579	sim	sim
Século XVII	Francisco Manuel de Melo	1608	sim	sim
	António Vieira	1608	sim	não
	António das Chagas	1631	sim	sim
	Manuel Bernardes	1644	sim	não
	José da Cunha Brochado	1651	sim	não
	André Barros	1675	sim	não
Século XVIII	Matias Aires	1705	sim	não
	Luís António Verney	1713	sim	não
	Correia Garção	1724	sim	não
	Marquesa de Alorna	1750	sim	não
Século XIX	Almeida Garrett	1799	sim	sim
	Alexandre Herculano	1819	sim	sim
	Camilo Castelo Branco	1825	sim	sim
	Júlio Dinis (1839)	1839	sim	sim
	Eça de Queiroz	1845	sim	não
	Guerra Junqueiro	1850	sim	sim
	Fialho de Almeida	1875	sim	sim
	Trindade Coelho	1861	sim	não
	Raul Brandão	1867	sim	sim
	Aquilino Ribeiro	1885	sim	sim
	Mário de Sá Carneiro	1890	sim	não
	João de Araújo Correia	1899	sim	sim

Os exemplos apresentados em (41) e (42) são extraídos de Magro (2006, 98, 100–102, 112–113, 178–179, 233) e comprovam que a interpolação encontrada nos autores do século XIX e a interpolação dialetal atestada no corpus CORDIAL-SIN são da mesma natureza. Tanto nos dados dialetais em (41) como nos dados literários em (42), vemos interpolados dêiticos pessoais, locativos e aspetuais. A nova interpolação, tal como a interpolação medieval e a interpolação de *não*, situa-se sempre à esquerda do verbo e depende da presença na frase de um proclisador diferente de *não*, sem implicar adjacência entre o proclisador e o clítico – cf. (41e/41m) e (38d/38i), acima.

- (41) a. Oi! Quando *me* ela *disse* aquilo! (CORDIAL-SIN, Corvo)
 b. Mesmo dentro de uma marmita *se* eles *governam* (CORDIAL-SIN, Alte)
 c. Tem de se pôr um garmel para *se* eles *não forçar*. (CORDIAL-SIN, Ponta Garça)
 d. Também *as* eu ali *tenho*, que já dá para a minha vida (CORDIAL-SIN, Monsanto)
 e. Vamos lá que eu *lhe* lá *vou* ensinar. (CORDIAL-SIN, Monsanto)
 f. Eu a gilete ainda *a* aí *tenho* (CORDIAL-SIN, Unhais da Serra)
 g. Sei que *a* aqui *tenho* por baixo, a maça de amasar o linho (CORDIAL-SIN, Porto de Vacas)
 h. Também *se* cá *usa*, mas agora não sei dizer qual é esta (CORDIAL-SIN, Figueiró)
 i. Todos os dias *se* para lá *ia* achar túbara (CORDIAL-SIN, Lavre)
 j. Ele veio lá a casa mas eu já o lá não *tinha*. (CORDIAL-SIN, Unhais da Serra)
 l. Trigo é que *se* já *semeou* aqui muito mas agora é pouco (CORDIAL-SIN, Alcochete)
 m. aquele, o tal que eu *vos* já *falei* que tinha uns livros muito bons (CORDIAL-SIN, Covo)
- (42) a. Tão bem arrecadado está, que não há meio de *lhe* eu *pôr* a vista. (João Araújo Correia, *Terra Ingrata*)
 b. O que *me* eu *parece* é que nós temos cedo muita pancada rija (Almeida Garrett, *Viagens na minha Terra*)
 c. Que queres mais que *te* ele *dê*? (Raul Brandão, *Os Pobres*)
 d. Nunca duvidei do que *me* elas *diziam* (Almeida Garret, *Viagens na minha Terra*)
 e. Não houve quem fosse indiferente ao sucesso, que o conselheiro narrou ainda sob a opressiva influência que *lhe* êle *deixara* (Júlio Dinis, *A Morgadinha dos Canaviais*)
 f. E voto a Christo, que me parece o mesmo mancebo que *vos* aqui *procurou* há um ano... (Alexandre Herculano, *O Monge de Cisté*)
 g. A receber alguma paga há de ser de quem *me* cá *mandou*. (Camilo Castelo Branco, *Amor de Perdição*)
 h. Quem *me* lá *dera* outra vez! (Camilo Castelo Branco, *O Condenado*)
 i. Sorriu-se o morgado para dentro por que *lhe* já não *ficava* bem indignar-se por dentro nem por fôra. (Camilo Castelo Branco, *A Queda d'um Anjo*)

A julgar pelos exemplos em (42), no século XIX a interpolação de dêiticos não seria tida por fenómeno dialetal a excluir da língua literária. Posteriormente, tornou-se quase invisível na produção escrita, ainda que continue a ser possível atestá-la em textos jornalísticos (e talvez uma investigação sistemática em textos literários venha a revelar que também aí não desapareceu – cf. Cunha/Cintra 1984, 314).

- (43) a. Quem não hesita em fazer uma coisa daquelas não tem com certeza, em privado, princípios que nos possam tranquilizar. Como, de resto, *se já constatou*. (Público, 25.03.2011, Vasco Pulido Valente)
- b. Onde é que *se já ouviu* falar de semelhante esquema? (CETEMPúblico)
- c. Os verdadeiros aristocratas apreciadores de pólo – os únicos que no domingo se mantiveram fiéis ao que os lá *levava*, assistindo à segunda partida – não vão habitualmente para as bancadas. (CETEMPúblico)
- d. Quando *se lá chega*, chuta quem estiver em melhores condições. (Expresso online, 27.06.2016, Nicolau Santos)
- e. E no beijo de gratidão que *lhe* ele deu, bem viu ela que jamais a fome lhes bateria à porta. (CETEMPúblico)

4 Conclusão

Neste capítulo caracterizou-se o sistema de colocação dos pronomes clíticos no português europeu, pondo em destaque as suas especificidades, no quadro das línguas românicas, a sua complexidade, decorrente de depender da articulação entre fatores gramaticais diversos (sintáticos, lexicais, semânticos, pragmáticos, morfológicos), e o percurso diacrónico inusual e intrigante que se desenha entre o português das primeiras fontes escritas e o português contemporâneo, envolvendo variação, mudança e estabilidade. Apesar de o tema da colocação dos pronomes clíticos ter suscitado continuamente muito interesse e atenção (e ter originado uma extensíssima bibliografia que só em parte pôde ser referida neste capítulo), muitas questões permanecem ainda em aberto, são parte essencial do estado da arte e continuarão a alimentar a investigação neste domínio.

5 Referências

- Álvarez, Rosario/Xove, Xosé (2002), *Gramática da Língua Galega*, Vigo, Galaxia.
- Andrade, Aroldo (2010), *A Subida de Clíticos em Português: Um estudo sobre a variedade europeia dos séculos XVI a XX*, tese de Doutoramento, Campinas, SP, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.
- Barbosa, Pilar (1996), *Clitic Placement in European Portuguese and the Position of Subjects*, in: Aaron Halpern/Arnold Zwicky (edd.), *Approaching Second: Second Position Clitics and Related Phenomena*, Stanford, California, CSLI, 1–40.
- Barbosa, Pilar (2000), *Clitics: A Window into the Null Subject Parameter*, in: João Costa (ed.), *Portuguese Syntax: New Comparative Studies*, Oxford/New York, Oxford University Press, 31–93.
- Batlloiri, Montserrat/Iglésias, Narcís/Martins, Ana Maria (2005), *Sintaxi dels clítics pronominals en català medieval*, *Caplletra* 38, 137–177.
- Bouzouita, Miriam (2008a), *The Diachronic Development of Spanish Clitic Placement*, tese de Doutorado, London, King's College London.

- Bouzouita, Miriam (2008b), *At the Syntax-Pragmatics Interface: Clitics in the History of Spanish*, in: Robin Cooper/Ruth Kempson (edd.), *Language in Flux: Dialogue Coordination, Language Variation, Change and Evolution*, London, King's College Publications, 221–263.
- Cardoso Pires, José (⁴1975, ¹1963), *Jogos de Azar*, Lisboa, Moraes.
- Castillo Lluch, Mónica (1996), *La Posición del Pronombre Átono en la Prosa Hispánica Medieval*, tese de Doutoramento, Paris, Université Paris XIII.
- Castro, Ivo (2002), *Sur le bilinguisme littéraire castillan-portugais*, in: Luísa Bras de Oliveira (ed.), *La littérature d'auteurs portugais en langue castillane*, Lisboa/Paris, Fundação Calouste Gulbenkian, 11–23.
- Catalán, Diego (ed.) (1977), *Gran Crónica de Alfonso XI*, vol. 2, Madrid, Gredos.
- CETEMPúblico – *Corpus de Extratos de Textos Electrónicos MCT/Público*, <http://www.linguateca.pt/CETEMPUBLICO/> (13.05.2016).
- Chenery, Winthrop Holt (1905), *Object Pronouns in Dependent Clauses: A Study in Old Spanish Word Order*, Publications of the Modern Language Association of America 20, 1–151.
- Cintra, Luís Filipe Lindley (ed.) (1990), *Crónica Geral de Espanha de 1344, Edição crítica do texto português*, vol. 4, Lisboa, Academia Portuguesa de História.
- CLUL (ed.) (2014), *P.S. Post Scriptum, Arquivo Digital de Escrita Quotidiana em Portugal e Espanha na Época Moderna* [2016], <http://ps.clul.ul.pt> (13.05.2016).
- CORDIAL-SIN – *Syntax-oriented Corpus of Portuguese Dialects*, <http://www.clul.ul.pt> (13.05.2016).
- Costa, João/Fiéis, Alexandra/Lobo, Maria (2014), *Input variability and late acquisition: clitic misplacement in European Portuguese*, *Lingua* 161, 10–26.
- Costa, Maria João (2001), *Contributos para o estudo das perífrases verbais com «aver» e «teer» no português medieval*, in: Clara Nunes Correia/Anabela Gonçalves (edd.), *Actas do XVI Encontro da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística, 179–186.
- Cunha, Celso/Cintra, Luís Filipe Lindley (1984), *Nova Gramática do Português Contemporâneo*, Lisboa, Edições João Sá da Costa.
- Döhla, Hans-Jörg (2014), *Diachronic convergence and divergence in differential object marking between Spanish and Portuguese*, in: Kurt Braunmüller/Steffen Höder/Karoline Köhl (edd.), *Stability and Divergence in Language Contact. Factors and Mechanisms*, Amsterdam/Philadelphia, Benjamins, 265–289.
- Domingos, Manuel (2010), *A colocação dos pronomes clíticos no português (oral) de Angola (POA)*, Trabalho do seminário de Linguística Comparada: Tópicos de Gramática do Português numa perspetiva comparativa, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Duarte, Inês (2003), *Padrões de colocação dos pronomes clíticos*, in: Maria Helena Mira Mateus et al., *Gramática da Língua Portuguesa*, Lisboa, Caminho, 847–867.
- Duarte, Inês/Matos, Gabriela (2000), *Romance Clitics and the Minimalist Program*, in: João Costa (ed.), *Portuguese Syntax: New Comparative Studies*, Oxford/New York, Oxford University Press, 116–142.
- Duarte, Luiz Fagundes (ed.) (1986), *Os Documentos em Português da Chancelaria de D. Afonso III (Edição)*, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Fernández-Rubiera, Francisco J. (2006), *Clitic Placement in Asturian: Evidence for a Syntactic FocusP Interaction*, Georgetown University Working Papers in Linguistics 2006, 89–126.
- Fernández-Rubiera, Francisco J. (2009), *Clitics at the Edge: Clitic Placement in Western Iberian Languages*, tese de Doutoramento, Washington, D.C., Georgetown University.
- Fernández-Rubiera, Francisco J. (2010), *Force⁰, Finiteness⁰ and the placement of clitics in Western Iberian Romance languages*, *Estudos de Linguística Galega* 2, 75–95.
- Ferrando, Antoni (ed.) (2007), *Curial e Güelfa*, Toulouse, Anacharsis.
- Fiéis, Alexandra (1996), *Clíticos num Corpus do Português do Século XIII*, dissertação de Mestrado, Lisboa, Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa.

- Fischer, Susan (2002), *The Catalan Clitic System: A Diachronic Perspective on its Syntax and Phonology*, Berlin/New York, Mouton de Gruyter.
- Fodor, Janet Dean/Sag, Ivan A. (1982), *Referential and Quantificational Indefinites*, *Linguistics and Philosophy* 5, 355–398.
- Galves, Charlotte/Britto, Helena Britto/Sousa, Maria Clara Paixão de (2005), *The change in clitic placement from Classical to Modern European Portuguese: Results from the Tycho Brahe Corpus*, *Journal of Portuguese Linguistics* 4, 39–67.
- Galves, Charlotte/Sousa, Maria Clara Paixão de (2005), *Clitic placement and the position of subjects in the history of European Portuguese*, in: Twan Geerts/Ivo van Ginneken/Haike Jacobs (edd.), *Romance Languages and Linguistic Theory 2003: Selected Papers from «Going Romance»*, Amsterdam/Philadelphia, Benjamins, 97–113.
- Gillier, Raíssa (2009), *A colocação dos pronomes pessoais clíticos em Jorge de Sena e José Cardoso Pires – variação idiolectal*, Trabalho do seminário de Linguística Comparada: Tópicos de Gramática do Português numa perspetiva comparativa, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Gonçalves, Rita (2009), *A colocação dos pronomes clíticos no português oral de São Tomé: Análise e discussão de contextos numa perspetiva comparativa*, Trabalho do seminário de Linguística Comparada: Tópicos de Gramática do Português numa perspetiva comparativa, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- González i Planas, Francesc (2007), *Sintaxis de los clíticos pronominales en asturleonés*, *Ianua Revista Philologica Romanica* 7, 15–35.
- Granberg, Robert Arthur (1988), *Object Pronoun Position in Medieval and Early Modern Spanish*, tese de Doutoramento, Los Angeles, University of California at Los Angeles.
- Harris, James/Halle, Morris (2005), *Unexpected Plural Inflections in Spanish: Reduplication and Metathesis*, *Linguistic Inquiry* 36, 195–222.
- Higginbotham, James (1987), *Indefiniteness and Predication*, in: Eric J. Reuland/Alice G. ter Meulen (edd.), *The Representation of (In)definiteness*, Cambridge, MA, MIT Press, 43–70.
- Hinzelin, Marc-Olivier (2010), *L'interpolation dans les langues romanes: aspects diachroniques*, in Maria Iliescu/Heidi Siller Runggaldier/Paul Danler (edd.), *Actes du XXV^e Congrès International de Linguistique Romane, Innsbruck 2007*, vol. 2, Berlin/Boston, De Gruyter, 331–340.
- Justino, Vítor Mércia (2010), *Estudo Comparativo dos Padrões de Colocação dos Pronomes Clíticos nos Discursos Oral e Escrito do Português de Moçambique*, Trabalho do seminário de Linguística Comparada: O Português em África, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Kaiser, Georg A. (1992), *Zur Grammatik der klitischen Personalpronomina im Französischen und Portugiesischen: Eine synchronische und diachronische Analyse im Rahmen der Prinzipien- und Parametertheorie*, tese de Doutoramento, Hamburg, Universität Hamburg.
- Kato, Mary/Martins, Ana Maria (2016), *European Portuguese and Brazilian Portuguese: An overview on word order*, in: Leo Wetzels/Sergio Menuzzi/João Costa (edd.), *The Handbook of Portuguese Linguistics*, Hoboken, NJ, Wiley/Blackwell, 15–40.
- Lawton, Robert A. (1966), *La syntaxe des pronoms personnels atones en ancien portugais: la phrase principale à sujet initial, et à verbe initial*, in: José Gonçalo Herculano de Carvalho (ed.), *Actas do V Colóquio Internacional de Estudos Luso-Brasileiros*, vol. 3, Coimbra, [Gráfica de Coimbra,] 5–17.
- Luís, Ana R./Kaiser, Georg A. (2016), *Clitic Pronouns: Phonology, Morphology and Syntax*, in: Leo Wetzels/Sergio Menuzzi/João Costa (edd.), *The Handbook of Portuguese Linguistics*, Hoboken NJ, Wiley/Blackwell, 210–233.
- Macchi, Giuliano (ed.) (1975), *Fernão Lopes, Crónica de D. Fernando, Edição crítica*, Lisboa, Imprensa Nacional, Casa da Moeda.
- Madeira, Ana Maria (1993), *Clitic-second in European Portuguese*, *Probus* 5, 155–174.

- Madeira, Ana Maria/Xavier, Maria Francisca (2009), *The Acquisition of Clitic Pronouns in L2 European Portuguese*, in: Acrísio Pires/Jason Rothman (edd.), *Minimalist Inquiries into Child and Adult Language Acquisition: Case Studies across Portuguese*, Berlin/Boston, De Gruyter, 273–299.
- Magro, Catarina (2005), *Introdutores de orações infinitivas – o que diz a sintaxe dos clíticos*, in: Inês Duarte/Isabel Leiria (edd.), *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística, Lisboa 2004*, Lisboa, Associação Portuguesa de Linguística, 649–664, <http://repositorio.ul.pt/handle/10451/6083> (13.05.2016).
- Magro, Catarina (2006), *Clíticos: Variação sobre o Tema*, tese de Doutoramento, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Magro, Catarina (2010a), *Interpolação & Cia. nos dialectos do Português Europeu*, Estudos de Linguística Galega 2, 97–119.
- Magro, Catarina (2010b), *When corpus analysis refutes common beliefs: The case of interpolation in European Portuguese dialects*, Corpus 9, 115–135.
- Mapasse, Ermelinda Lúcia Atanasio (2005), *Clíticos Pronominais em Português de Moçambique*, dissertação de Mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Marquilhas, Rita (2000), *A Faculdade das Letras: Leitura e escrita em Portugal no séc. XVII*, Lisboa, Imprensa Nacional – Casa da Moeda.
- Martins, Ana Maria (1994), *Clíticos na História do Português*, tese de Doutoramento, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Martins, Ana Maria (1997), «Alguns», «muitos», «poucos», «todos» e a relação sintaxe-semântica, in: Ana Maria Brito et al. (edd.), *Sentido que a Vida Faz: Estudos para Óscar Lopes*, Porto, Campo das Letras, 679–692.
- Martins, Ana Maria (2001), *Documentos Portugueses do Noroeste e da Região de Lisboa: Da Produção Primitiva ao Século XVI*, Lisboa, Imprensa Nacional – Casa da Moeda.
- Martins, Ana Maria (2006), *Aspects of infinitival constructions in the history of Portuguese*, in: Randall S. Gess/Deborah Arteaga (edd.), *Historical Romance Linguistics: Retrospective and perspectives*, Amsterdam/Philadelphia, Benjamins, 327–355.
- Martins, Ana Maria (2011), *Clíticos na história do português à luz do teatro vicentino*, Estudos de Linguística Galega 3, 83–109.
- Martins, Ana Maria (2013), *A colocação dos pronomes pessoais clíticos*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 2, cap. 42, Lisboa, Fundação Calouste Gulbenkian, 2231–2302.
- Martins, Ana Maria (2015), *Variação sintática no português quinhentista: a colocação dos pronomes clíticos*, Estudos de Linguística Galega 7, 83–94.
- Miranda, Sílvia (2013), *Reconstituição do ms. L da Crónica Geral de Espanha de 1344 (2.ª Parte)*, Relatório final de estágio de mestrado, Lisboa, Faculdade de Letras da Universidade de Lisboa, <http://repositorio.ul.pt/handle/10451/9403> (13.05.2016).
- Moraes, Maria Aparecida Torres (1995), *Do Português Clássico ao Português Moderno: Um estudo da cliticização e do movimento do verbo*, tese de Doutoramento, Campinas, SP, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.
- Namiuti, Cristiane (2006), *Um Estudo sobre o Fenômeno da Interpolação na História do Português*, Cadernos de Estudos Linguísticos 48, 171–194.
- Namiuti, Cristiane (2008), *Aspectos da História Gramatical do Português. Interpolação, Negação e Mudança*, tese de Doutoramento, Campinas, SP, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.
- Nieuwenhuijsen, Dorine (2006), *Cambios en la colocación de los pronombres átonos*, in: Concepción Company Company (ed.), *Sintaxis Histórica del Español*, México, Universidad Nacional Autónoma de México, Fondo de Cultura Económica, 1337–1404.

- Ogando, Victoria (1980), *A colocación do pronome átono en relación co verbo no galego-portugués medieval*, Verba 7, 251–282.
- Oltra-Massuet, Isabel/Arregi, Karlos (2005), *Stress-by-Structure in Spanish*, Linguistic Inquiry 36, 43–84.
- Paiva, Conceição/Barbosa, Pilar (submetido), *Clitic climbing in the speech of Braga and Lisbon*, in: Pilar Barbosa/Conceição Paiva/Celeste Rodrigues (edd.), *Studies on Variation in varieties of Portuguese*, Amsterdam/Philadelphia, Benjamins.
- Pato, Raymundo A. Bulhão (ed.) (1884), *Cartas de Afonso de Albuquerque seguidas de Documentos que as Elucidam*, vol. 1, Lisboa, Academia Real das Sciencias de Lisboa.
- Piel, Joseph M./Mattoso, José (edd.) (1980), *Livros Velhos de Linhagens*, Lisboa, Academia das Ciências de Lisboa.
- Potsma, Gertjan (2010), *The impact of failed changes*, in: Anne Breitbarth et al. (edd.), *Continuity and Change in Grammar*, Amsterdam/Philadelphia, Benjamins, 269–302.
- Ramsden, Herbert (1963), *Weak-Pronoun Position in the Early Romance Languages*, Manchester, Manchester University Press.
- Ribeiro, Ilza (1995), *A sintaxe da ordem no português arcaico: o efeito V2*, tese de Doutorado, Campinas, SP, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.
- Roberts, Ian (1992), *A formal account of grammaticalization in the history of Romance futures*, Folia Linguistica Historica 13, 219–258.
- Roberts, Ian/Roussou, Anna (2002), *The History of the Future*, in: David W. Lightfoot (ed.), *Syntactic Effects of Morphological Change*, Oxford/New York, Oxford University Press, 23–56.
- Roberts, Ian/Roussou, Anna (2003), *Syntactic Change: A Minimalist Approach to Grammaticalization*, Cambridge, Cambridge University Press.
- Rocha, Nilzete (2009), *Clíticos: Ingrediente na Cozinha Portuguesa do século XVII*, dissertação de Mestrado, Salvador, Bahia, Universidade Federal da Bahia.
- Salvi, Giampaolo (1990), *La sopravvivenza della legge di Wackernagel nei dialetti occidentali della Penisola Iberica*, Medioevo Romanzo 15, 117–210.
- Silva, Rosa Virgínia Mattos e (1989), *Estruturas Trecentistas: Elementos para uma Gramática do Português Arcaico*, [Lisboa,] Imprensa Nacional – Casa da Moeda.
- Sousa, Maria Clara Paixão de (2004), *Língua Barroca: Sintaxe e História do Português de 1600*, tese de Doutorado, Campinas, SP, Instituto de Estudos da Linguagem, Universidade Estadual de Campinas.
- WOChWEL's POS-tagged and Parsed Old Portuguese Texts (Ana Maria Martins, coord., CLUL/FCT), <http://alfclul.clul.ul.pt/wochwel/oldtexts.html> (13.05.2016).

João Costa, Alexandra Fiéis e Maria Lobo

16 A aquisição dos pronomes clíticos no português L1

Abstract: Neste capítulo, mostramos que as propriedades morfosintáticas dos pronomes clíticos são adquiridas a ritmos diferentes no português europeu como língua materna. Os contextos em que o pronome é obrigatório são de desenvolvimento tardio, possivelmente por haver uma sobregeneralização da construção de objeto nulo. Em contrapartida, a interpretação das propriedades referenciais dos clíticos é adquirida cedo, ao contrário do que acontece quando o pronome é uma forma forte.

Keywords: aquisição, L1, omissão de clíticos, colocação de clíticos, compreensão

1 Introdução

Os pronomes clíticos têm propriedades morfofonológicas, distribucionais e interpretativas que os fazem diferir de outros pronomes. Em (1), observamos que os clíticos, enquanto formas átonas, formam uma palavra prosódica com o verbo. Conforme descrito em Vigário (2001), numa sequência de palavras verbo-nome, a vogal [i] átona em posição final é geralmente inaudível. O mesmo não acontece internamente à palavra, como atesta o exemplo em (1b), observando-se que a vogal forma ditongo com a vogal seguinte, semivocalizando. Se o verbo for seguido de um clítico, também se dá semivocalização, conforme se observa em (1c), o que atesta que o clítico não tem autonomia prosódica face à palavra a que se liga. Quando o clítico ocorre em posição proclítica, o mesmo processo fonológico pode ligar o clítico à palavra que o precede (1d):

- (1) a. bebe agora – [bebe'gɔrɐ]
- b. teatro – ['tjatrɐ]
- c. bebe-a – ['bebjɐ]
- d. se a beberes – [sɐbi'berɨ]

Sintaticamente, é possível observar que o clítico tem propriedades diferentes das de outros pronomes. Conforme descrito no capítulo 15 A colocação dos pronomes clíticos em sincronia e diacronia, a posição dos clíticos varia em função de contextos sintáticos. Por exemplo, o clítico é pós-verbal (enclítico) numa frase declarativa afirmativa simples (2a), mas a presença de negação (2b), de um conector subordinativo (2c) ou de um constituinte interrogativo (2d) induz próclise, atraindo o clítico para uma posição pré-verbal:

- (2) a. O Pedro viu-a.
 b. O Pedro não a viu.
 c. O Pedro disse que a viu?
 d. Quem a viu?

Este comportamento contrasta com o de sintagmas nominais complemento ou com o de outros pronomes, cuja posição não varia com o contexto sintático. Por exemplo, em português brasileiro, que quase não utiliza pronomes clíticos nos contextos de (2), mas sim pronomes fortes, observamos que a posição do pronome é invariável:

- (3) a. O Pedro viu ela.
 b. O Pedro não viu ela.
 c. O Pedro disse que viu ela?
 d. Quem viu ela?

Sob o ponto de vista interpretativo, vemos que um pronome clítico em posição de complemento, como em (4a), não pode ser interpretado como correferente com o *Pedro*, enquanto em (4b) o pronome forte, inserido num sintagma preposicional, é ambíguo, admitindo as duas interpretações. Um pronome clítico reflexo, ao contrário de um não reflexo, é sempre correferente com o sujeito da sua oração (5a). Um pronome forte reflexo, como *si*, no entanto, pode, em determinados contextos remeter para o interlocutor (segunda pessoa mais formal) ou ser correferente com um sujeito mais distante como em (5c).

- (4) a. Só o Pedro_i o_{j/*i} viu.
 b. Só o Pedro_i confia nele_{j/i}.
 (5) a. Só o Pedro_i se_{i/*j} viu.
 b. Só o Pedro_i confia em si_i.
 c. O Pedro_i acha que ninguém_j confia em si_i.

Adquirir pronomes clíticos é, assim, adquirir formas fonologicamente deficientes, com propriedades sintático-semânticas que os fazem diferir de outras expressões nominais.

No caso específico do português, a aquisição dos clíticos implica ainda que, em alguns contextos, se domine o facto de os clíticos alternarem livremente com objetos nulos, conforme demonstrado em Raposo (1986) e ilustrado nos exemplos em (6):

- (6) a. Tirei a chávena da mesa e pu-la na máquina de lavar.
 b. Tirei a chávena da mesa e pus \emptyset na máquina de lavar.

Neste capítulo, apresentamos vários dados relativos à aquisição de pronomes clíticos por crianças falantes do português, mostrando que, em determinados aspetos, a aquisição de clíticos em português europeu é mais lenta do que a que se encontra em outras línguas, o que pode ser atribuído a uma maior complexidade no *input*.

O capítulo encontra-se estruturado da seguinte forma: na secção 2, apresentamos dados sobre a produção de clíticos acusativos, reflexos e especificados para primeira e

segunda pessoa, mostrando que as crianças falantes do português europeu omitem clíticos até mais tarde do que as crianças que adquirem outras línguas com clíticos, embora distingam diferentes tipos de clíticos; na secção 3, apresentamos dados da compreensão que atestam que as crianças compreendem clíticos, mas sobregeneralizam a construção de objeto nulo, o que permite supor que a omissão encontrada se deve a um uso massivo da construção de objeto nulo; na secção 4, apresentamos dados sobre a aquisição da posição dos clíticos, mostrando que as crianças sobregeneralizam a ênclise, observando-se que a aquisição da próclise é gradual e sensível aos diferentes contextos sintáticos; finalmente, na secção 5, mostramos que a aquisição de alguns dos princípios que subjazem à compreensão dos clíticos não é problemática para as crianças, existindo desde cedo um contraste claro entre clíticos e pronomes fortes.

2 Produção e omissão de pronomes clíticos

A literatura das últimas duas décadas tem dedicado bastante atenção à aquisição dos pronomes clíticos por se observar que, em línguas em que o preenchimento da posição de objeto é obrigatória, como o francês ou o italiano, as crianças produzem frases com omissão de pronomes. Esta observação tem sido feita para diferentes línguas, destacando-se os estudos para o francês de Jakubowicz et al. (1998) e para o italiano de Schaeffer (1997). Seria legítimo supor que os pronomes são omitidos por se tratar de formas fonologicamente deficientes. A sua baixa saliência prosódica justificaria, assim, que as crianças os omitissem, se se seguir a hipótese segundo a qual a produção poderia ser guiada por princípios de proeminência prosódica.

Jakubowicz et al. (1998) mostram, contudo, que, na aquisição do francês, esta omissão de pronomes é seletiva: afeta apenas pronomes acusativos, como em (7a), e não os dativos, como em (7b):

- (7) a. Il la lave.
 ele a lava
 b. Il lui téléphone.
 ele lhe telefona

Sabendo-se que tanto os pronomes acusativos como os dativos são prosodicamente deficientes, Jakubowicz et al. (1998) mostram, através destes dados, que a omissão de pronomes clíticos não se deve à sua fraca proeminência fonológica. Esta conclusão é corroborada pela comparação entre a produção de pronomes clíticos e determinantes em francês (que são homófonos), através da qual se mostra que as crianças apenas omitem os pronomes clíticos e não os determinantes, o que permite argumentar que a omissão de clíticos não depende da sua atonicidade.

Vários estudos para diferentes línguas têm procurado determinar se as crianças omitem pronomes clíticos nas suas produções iniciais. Observou-se que a omissão não é atestada em todas as línguas. Encontra-se omissão em francês, italiano (Schaeffer

fer 1997), catalão (Wexler/Gavarró/Torrens 2004), mas não em espanhol (Wexler/Gavarró/Torrens 2004) ou grego (Tsakali/Wexler 2004). Alguma da literatura sobre o assunto tem tentado associar a omissão de clíticos a outras propriedades dos sistemas linguísticos, como a existência de concordância nos participios passados, o que não é de todo consensual. Percebe-se, contudo, a relevância de avaliar a robustez da generalização de que os clíticos são adquiridos tardiamente.

Costa/Lobo (2006; 2007a; 2007b), Costa/Lobo/Silva (2012) e Silva (2008) testaram a produção de pronomes clíticos por crianças falantes de português, usando diferentes testes de produção induzida, adaptados para o português. Num dos exemplos usados, mostra-se uma imagem como a de (8) às crianças e induz-se a produção do pronome através de uma interação como a de (9):

- (8) Exemplo de figura em teste de produção de clíticos



- (9) Olha! Um menino e um gato! O que é que o menino está a fazer ao gato?

Assim, (10a) é uma resposta possível para a pergunta em (9). Note-se, contudo, que, porque o português é uma língua de objeto nulo, a resposta em (10b), com omissão do pronome, também é legítima (ao contrário do que acontece em italiano ou francês):

- (10) a. Está a molhá-lo.
b. Está a molhar \emptyset .

Por este motivo, Costa/Lobo (2006; 2007a) testaram também a produção de pronomes em contexto de ilha sintática¹ (neste caso uma oração subordinada adverbial), já que, conforme se mostra em (11), o objeto nulo é agramatical neste contexto (Raposo 1986):

- (11) *Tirei a chávena da mesa quando pus \emptyset na máquina de lavar.

¹ Uma ilha sintática é um domínio sintático que apresenta restrições à extração de constituintes, como por exemplo orações adverbiais e orações relativas.

Assim, para imagens como a de (8), foram criados contextos de elicitção como os de (12):

- (12) Olha! Um menino e um gato. O gato está todo molhado! Porque é que o gato está todo molhado? O gato está molhado porque o menino...

Neste contexto, apenas a continuação «...o molhou», em que o pronome é realizado, é possível.

Os dados da produção obtidos em Costa/Lobo (2006; 2007a) permitiram chegar às seguintes conclusões:

- A. As crianças falantes do português omitem clíticos até mais tarde do que as crianças falantes do inglês ou italiano. Observaram-se taxas de omissão bastante altas em crianças com 6 anos, enquanto nas outras línguas a omissão tende a desaparecer a partir dos 3 anos.
- B. As crianças omitem clíticos quer em frases raiz, quer em ilhas sintáticas, embora as taxas de produção de clítico sejam mais elevadas em contexto de ilha do que nas frases raiz. Isto atesta alguma sensibilidade ao contexto em que os objetos nulos são impossíveis, embora não haja convergência com a performance dos adultos neste contexto.

Nas Tabelas 1 e 2, ilustra-se esta diferença entre produção de clíticos em frases raiz e em ilhas, com os dados de Silva (2008):

Tabela 1: Resultados da produção de clíticos acusativos de 3.ª pessoa em contextos raiz.

	CLÍTICOS ACUSATIVOS DE 3.ª PESSOA				Adultos
	[3,0; 4,0[[4,0; 5,0[[5,0; 6,0[[6,0; 6,5[
Clítico	23/240 9,58%	99/432 22,92%	146/336 43,45%	74/160 46,25%	229/240 95,42%
Objeto nulo	210/240 87,50%	298/432 68,98%	149/336 44,35%	84/160 52,50%	3/240 1,25%
Expressão nominal	6/240 2,50%	33/432 7,64%	41/336 12,20%	2/160 1,25%	8/240 3,33%
Pronome forte	1/240 0,42%	2/432 0,46%	0/336 0%	0/160 0%	0/240 0%

Tabela 2: Resultados da produção de clíticos acusativos de 3.^a pessoa em contextos de ilha.

	CLÍTICOS ACUSATIVOS DE 3. ^a PESSOA EM ILHAS				Adultos
	[3,0; 4,0[[4,0; 5,0[[5,0; 6,0[[6,0; 6,5[
Clíticos	4/120 3,33%	33/216 15,28%	49/168 29,17%	32/80 40%	78/120 65%
Objeto nulo	77/120 64,17%	76/216 35,18%	27/168 16,07%	20/80 25%	0/120 0%
Expressão nominal	38/120 31,67%	107/216 49,54%	92/168 54,76%	28/80 35%	42/120 35%
Pronome forte	1/120 0,83%	0/216 0%	0/168 0%	0/80 0%	0/120 0%

Para se perceber se as crianças estariam a sobregeneralizar a construção de objeto nulo em contextos em que este não é legítimo, conduziram-se testes para avaliar a produção de pronomes noutros contextos em que não há alternância livre entre clíticos e objetos nulos, como nos casos em que os clíticos estão especificados para 1.^a e 2.^a pessoa ou em casos de clíticos reflexos, conforme se ilustra em (13) e (14):

(13) Clíticos de 1.^a e 2.^a pessoa não variam com objeto nulo:

- a. Eu estava no cinema e tu encontraste-me.
- b. *Eu estava no cinema e tu encontraste \emptyset .
- c. Tu estavas no cinema e o Pedro encontrou-te.
- d. *Tu estavas no cinema e o Pedro encontrou \emptyset .

(14) Clíticos reflexos não variam com objeto nulo:

- a. O Pedro olhou para o espelho e penteou-se.
- b. *O Pedro olhou para o espelho e penteou \emptyset .

Em Silva (2008), desenvolveram-se testes para induzir a produção de clíticos reflexos e especificados para primeira e segunda pessoa. Observou-se que, à semelhança do que acontece nos contextos de ilha, as crianças omitem estes pronomes, ao contrário do que fazem os adultos, embora com taxas de omissão inferiores às dos clíticos de terceira pessoa.

Pode, assim, concluir-se que os clíticos são massivamente omitidos na gramática das crianças que adquirem o português e de forma diferenciada face ao que acontece noutras línguas.

Com base nestes resultados, duas hipóteses podem ser formuladas:

- A. A omissão de pronomes clíticos em português é semelhante à das outras línguas, mas a estabilização da produção é mais tardia.
- B. A omissão de pronomes em português é um caso de sobregeneralização no uso de objetos nulos.

A distinção entre estas duas hipóteses implica verificar se a gramática das crianças admite objetos nulos e se a sua compreensão de objetos nulos reflete o que se passa na produção. É disso que tratará a próxima secção.

3 A compreensão de estruturas de objeto nulo: o que é um clítico omitido?

Como vimos no final da secção anterior, para entendermos o fenómeno de omissão de clítico e verificarmos se o podemos interpretar como um uso sobregeneralizado da construção de objeto nulo, é importante percebermos se existe uma gramática de objeto nulo para as crianças. Esta conjectura foi primeiramente formulada em Grüter (2006), que avaliou a hipótese de que a omissão de pronomes em francês se deveria à passagem das crianças falantes de francês por um estágio em que a sua gramática admitiria objetos nulos.

Grüter desenvolveu um teste de compreensão, em que testou a compreensão de verbos com potencial para serem intransitivos ou transitivos, como o verbo *monter* ('subir') em francês:

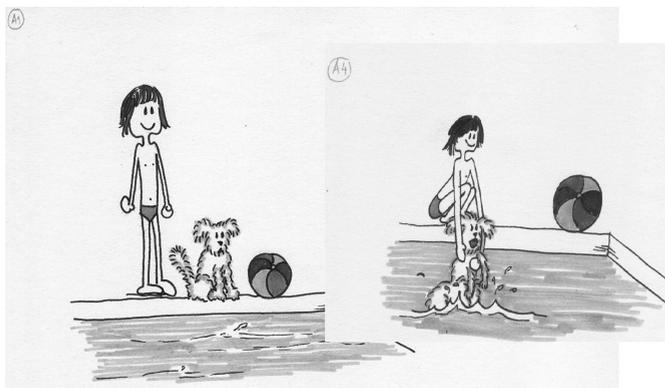
- (15) a. Il monte.
 Ele sobe
 b. Il monte le garçon.
 Ele sobe o rapaz
 «Ele eleva o rapaz.»

Crucialmente, para o adulto falante de francês, uma frase como (15a) não é aceitável como uma variante com objeto nulo de (15b). Se as crianças francesas admitissem uma frase como (15a) num contexto transitivo, isso significaria que a gramática das crianças falantes de francês incluiria objetos nulos. Quando testadas as frases, verificou-se que as crianças que estavam a adquirir o francês rejeitaram objetos nulos de forma categórica, o que permitiu rejeitar a hipótese de que passariam por um estágio de objeto nulo na sua gramática.

Em Costa/Lobo (2009; 2010), replicámos o teste de Grüter (2006) para o português, modificando-o para incluir verbos de alternância transitiva-intransitiva em português e incluindo, para além de frases simples e de frases com ilhas, contextos reflexos, para se poder fazer uma comparação direta com os contextos de produção.

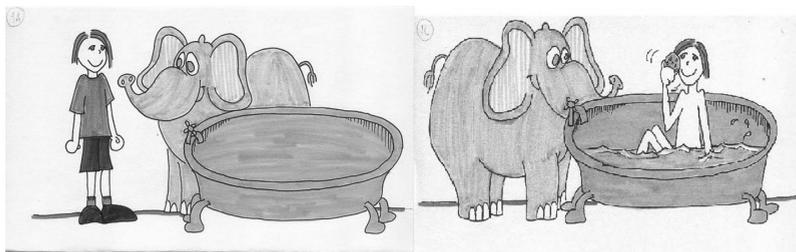
Em (16) apresentamos um dos itens de teste.

(16) Exemplo de item de teste para aceitabilidade de objeto nulo



Num primeiro momento, o experimentador mostra a imagem da esquerda, apresentando as personagens: «Aqui está o Rui, o cão e uma bola junto à piscina. Vamos ver o que o Rui fez ao cão». Mostra-se a imagem da direita e um fantoche diz: «Olha! Mergulhou na piscina!». Neste contexto, a frase é possível em português europeu, alternando livremente com «Olha! Mergulhou-o na piscina!». Para se testar contextos reflexos, foram usados pares de imagens como os de (17):

(17) Exemplo de item de teste para aceitabilidade de objeto nulo em contexto reflexo



O experimentador mostra a primeira imagem, apresentando as personagens: «Aqui está o Rui e o elefante junto à banheira. Vamos ver o que o Rui vai fazer ao elefante». O fantoche, mediante a apresentação da segunda imagem diz: «Olha! O Rui está a lavar». Pede-se à criança que avalie se a frase constitui uma descrição adequada da imagem ou não. Neste contexto, para o adulto, a frase é impossível como uma descrição adequada desta situação, já que o objeto nulo não tem interpretação reflexa.

Os resultados de Costa/Lobo (2009; 2010) neste teste permitiram chegar a três conclusões:

- A. As crianças falantes do português aceitam frases com objeto nulo em contexto raiz, tal como os adultos e diferentemente do que foi encontrado por Grüter (2006) para crianças falantes do francês.
- B. As crianças falantes do português aceitam frases com objeto nulo em ilhas sintáticas, ao contrário dos adultos.
- C. As crianças falantes do português aceitam frases com objeto nulo em contextos reflexos, ao contrário dos adultos.

A conclusão expressa em A permite afirmar que as crianças que estão a adquirir o português, aos três anos, já adquiriram a construção de objeto nulo nesta língua. Por outras palavras, num modelo de princípios e parâmetros como o de Chomsky (1986), as crianças já terão fixado o valor positivo para o parâmetro do objeto nulo. Contudo, não atingiram ainda um estágio adulto. As conclusões registadas em B e C mostram que, apesar de terem uma gramática que contém objetos nulos, o seu conhecimento não é igual ao do adulto, na medida em que admitem objetos nulos num conjunto de contextos mais amplo do que na gramática do adulto. Curiosamente, a sobreaceitação do objeto nulo na compreensão é em tudo compatível com a sobregeneralização do objeto nulo na produção. Vimos que as crianças omitem clíticos em contextos ilha e em contextos reflexos e vemos agora que essa omissão na produção tem um reflexo claro na compreensão.

Perante estes dados, é legítimo supor que, conforme proposto em Costa/Lobo (2006; 2007a; 2007b), a omissão de clíticos pelas crianças falantes do português é um caso de uso generalizado da construção de objeto nulo.

Neste sentido, os dados da compreensão de objeto nulo ajudam-nos a perceber que muito provavelmente não há qualquer problema com a produção de clíticos na aquisição do português. As crianças conhecem os clíticos, produzem-nos, sendo a sua baixa frequência apenas uma consequência da competição com a construção de objeto nulo. É, aliás, esta competição que, por hipótese, leva a que a convergência com as taxas de produção dos adultos demore algum tempo. A tarefa das crianças consistirá na estabilização dos contextos em que o clítico e o objeto nulo estão em competição ou em variação livre. É esta complexidade no *input* que induz, de acordo com esta hipótese, uma estabilização da produção dos clíticos e do objeto nulo mais tardia em português do que em outras línguas.²

Nas próximas secções, defenderemos que a complexidade, entendida como competição entre formas em variação, pode explicar outras propriedades da aquisição dos pronomes clíticos.

² Em Costa/Lobo (2014) e Costa/Grolla/Lobo (2014) mostramos, numa comparação entre a aquisição de propriedades do sujeito nulo e do objeto nulo em português europeu e português brasileiro, contrastadas com as propriedades do japonês e da aquisição do japonês descritas em Saito (2007), Takahashi (2008) e Sugisaki (2007; 2009), que a diferente complexidade inerente a cada uma destas línguas tem poder preditivo sobre ritmos e velocidades de convergência com a gramática do adulto.

4 Aquisição da colocação dos clíticos: variação entre próclise e ênclise

Conforme já foi referido, a posição dos clíticos é sintaticamente condicionada em português europeu. Tal como descrito por vários autores (ver Duarte/Matos 2000; Martins 2013, entre outros), a próclise ocorre nos seguintes contextos:

- i) com negação (18);
- ii) com sujeitos negativos pré-verbais (19);
- iii) com alguns advérbios pré-verbais, como *já*, *ainda*, *sempre*, *também* ou *só* (20);
- iv) com alguns sujeitos quantificados em posição pré-verbal (21);
- v) em orações com palavras interrogativas-qu e exclamativas-qu a preceder o verbo (22);
- vi) em orações subordinadas finitas com conector subordinativo explícito (23);
- vii) em orações com anteposição de foco (24).

(18) O João *não se* lavou.

(19) *Ninguém se* lavou.

- (20) a. O João *já se* lavou.
b. O João lavou-*se já*.

- (21) a. *Todos os meninos se* lavaram.
b. Lavaram-*se* todos os meninos.
c. *Dois meninos* lavaram-*se*.

- (22) a. *Quem se* magoou?
b. *Que bem lhe* respondeste!

- (23) a. O João disse *que se* lavava todos os dias.
b. O João cheia bem *porque se* lava todos os dias.

(24) *Isso te* digo eu!

A ênclise ocorre na ausência de desencadeadores de próclise, quer em frases raiz, quer em frases coordenadas:

(25) O João lavou-*se*.

(26) A mãe abriu a torneira e o João lavou-*se*.

Em vários estudos, com particular destaque para Duarte/Matos/Faria (1995), tem sido reportado que durante o processo de aquisição do português europeu as crianças produzem clíticos com colocações não conformes às da gramática do adulto, observando-se, em dados da fala espontânea, uma tendência para o uso da ênclise em contextos proclíticos, conforme se ilustra no seguinte conjunto de dados:

- (27) a. Foste tu *que* daste-me. (J. 4;8)
 b. Foi a Mariana *que* deu-me este. (S. 3;0.21; in Soares 2006, 375)
 c. foi alguém *que* meteu-me nesta fotografia. (J.G. 3;3; in Duarte/Matos/Faria 1995)
- (28) a. O mano *não* deixa-me dormir. (J. 3;8)
 b. *não* chama-se nada. (M. 20 m.; in Duarte/Matos/Faria 1995)
- (29) a. *Porque* partiu-se, mãe? (J. 3;4)
 b. *Porque* é que foste-me interromper? (R., 2;5; in Duarte/Matos/Faria 1995)

Vemos, nos dados de (27) a (29), casos de ênclise em contextos proclíticos com orações subordinadas (27), com negação (28) e em interrogativas-qu (29). Apesar de estes serem os dados mais comumente descritos na produção espontânea, é possível encontrar atestações de próclise em contexto de ênclise (30), de mesóclise na ausência de morfemas de futuro ou condicional (31) e de repetição de clítico com próclise e ênclise simultâneas (32):

- (30) a. Uma carta me caiu, do pokémon. (J. 4;8)
 b. Se queres levar isto, eu te empresto. (J. 3;6)
 c. Eu te empresto um, pai. (J. 3;7)
- (31) a. Dá-me-s uma moeda no meu porquinho? (J. 3; 4)
 b. Ai, duas pessoas a agarrar-m-em! (J. 3;5)
 c. Pai, deixa-me-s comer a sopa? (J. 3;5)
- (32) a. Eu disse que não se põe-se em pé. (J. 3;4)
 b. não te engasgas-te nada! (R. 2;5; in Duarte/Matos/Faria 1995)

Em Costa/Fiéis/Lobo (2014), pretendemos confirmar se, em testes de produção induzida, em que é possível obter uma amostra maior e um conjunto maior de produções, se observa alguma tendência na colocação dos pronomes clíticos, isto é, se as crianças produzem preferencialmente ênclise ou próclise e se todos os contextos são tratados da mesma forma pelas crianças.

No estudo de elicitación, foram criadas situações para levar as crianças a produzir pronomes clíticos nos seguintes contextos:

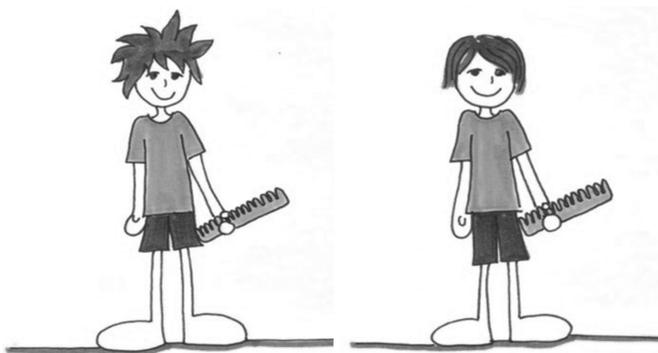
- i) frases simples sem desencadeador de próclise
- ii) orações coordenadas sem desencadeador de próclise
- iii) frases simples com negação
- iv) frases simples com sujeitos negativos
- v) frases simples com sujeitos quantificados
- vi) frases simples com o advérbio pré-verbal *já*
- vii) orações subordinadas completivas com o verbo *querer*
- viii) orações subordinadas adverbiais causais.

Os contextos i) e ii) são contextos de ênclise e os restantes são contextos de próclise.

Para minorar os efeitos da possível omissão de clíticos, apenas foram testados contextos com o clítico *se* por ser aquele que apresenta taxas menores de omissão, conforme foi explicado na secção 2.

Vejam alguns exemplos dos itens de teste:

(33) Exemplo de item de teste para frase simples (ênclise):

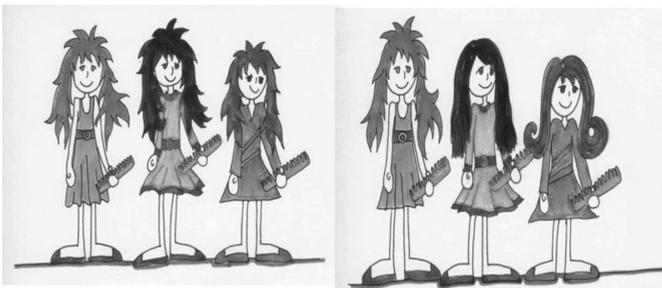


[imagem 1] Este menino está despenteado. Olha para o cabelo dele. Mas ele tem um pente.

[imagem 2] O que é que o menino fez?

Resposta esperada: *Penteou-se*.

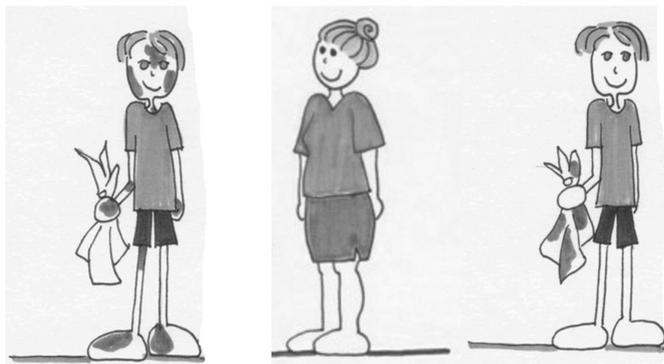
(34) Exemplo de item de teste para contextos negativos (próclise):



[imagem 2] Estas duas meninas usaram o pente e agora estão penteadas. Mas esta menina ainda está despenteada. O que é que ela não fez?

Resposta esperada: *Não se penteou*.

(35) Exemplo de item de teste para orações completivas finitas (próclise):



[imagem 1] Este menino foi para o parque e ficou todo sujo. A avó deu uma toalha ao menino. O que é que a avó quer que o menino faça? A avó quer...

Resposta esperada: ... *que o menino/ele se limpe.*

Foram testadas 60 crianças entre os 5 e os 7 anos e 20 adultos. Os resultados globais confirmaram a observação já feita na literatura de que há uma tendência para a sobregeneralização da ênclise:³

Tabela 3: Respostas obtidas em contextos enclíticos.

	Ênclise				
	Ênclise	Próclise	Redobro	Omissão	Outras respostas
5 anos	69,2%	0,4%	0,4%	5%	25%
6 anos	89,2%	0	0	0	10,8%
7 anos	92,5%	0	0	0	7,5%
Adultos	90,4%	0	0	0,4%	9,2%

³ Nas tabelas, a resposta redobro corresponde a casos em que há produção simultânea do clítico em próclise e ênclise.

Tabela 4: Respostas obtidas em contextos de próclise.

	Próclise				
	Ênclise	Próclise	Redobro	Omissão	Outras respostas
5 anos	49,58%	25,83%	0,42%	11,66%	12,5%
6 anos	42,7%	49,8%	0	3,1%	4,4%
7 anos	41,5%	55,4%	0	1,2%	1,9%
Adultos	10,83%	86,04%	0	0,42%	2,71%

Conforme se observa nas tabelas, em contextos de ênclise praticamente não se registam casos de próclise. Já nos contextos proclíticos, a taxa de ênclise é muito grande, registando-se uma estabilização muito gradual da próclise, mas, ainda aos 7 anos, uma enorme variação entre próclise e ênclise. Curiosamente, mesmo no grupo de controlo adulto, há 10,83% de produção de ênclise em contexto proclítico (a colocação dos pronomes clíticos em sincronia e diacronia).

Considerando os diferentes contextos de próclise, registámos a taxa de próclise e ênclise para cada um dos contextos de eliciação, tendo-se obtido os seguintes resultados:

Tabela 5: Taxas de próclise por condição.

	5 anos	6 anos	7 anos	adultos
Negação	61%	87,5%	84,4%	97,4%
Sujeitos negativos	48,3%	69,2%	70,9%	96,1%
Orações completivas	46,7%	74,6%	65,8%	97,3%
Advérbio <i>já</i>	36%	58,2%	65,8%	93,6%
Orações adverbiais	14%	20%	32,9%	77,6%
Sujeitos quantificados	9,9%	12,2%	21,3%	70,8%

Estes dados são interessantes porque mostram que a sobregeneralização da ênclise é condicionada pelo contexto sintático, sendo diferenciada em função dos diferentes contextos desencadeadores de próclise. Costa/Fiéis/Lobo (2014) estabeleceram a seguinte escala de desenvolvimento da próclise:

1. Negação > 2. Sujeitos negativos/Orações completivas finitas > 3. Advérbio *já* >
4. Orações adverbiais finitas > 5. Sujeitos quantificados

Esta ordem de aquisição não é inesperada se tivermos em conta questões de complexidade e variação no *input*. Na verdade, os contextos mais categóricos são os que são adquiridos mais precocemente. Vejamos caso a caso:

1) *Negação:*

Em português o principal marcador de negação frásica é o advérbio *não*. A criança terá apenas de descobrir que a negação induz próclise.

2) *Sujeitos negativos:*

Todos os sujeitos negativos (com palavras como *ninguém* ou *nenhum*) induzem próclise, mas nem todos os sujeitos são indutores de próclise. Assim, a tarefa da criança será distinguir quais os sujeitos que são indutores de próclise. Terá de fazer uma diferenciação entre tipos de sujeitos.

3) *Orações completivas finitas:*

Estabilizada a diferença entre coordenadas e subordinadas, a criança terá de descobrir que todas as orações subordinadas são contexto de próclise, desde que tenham conector subordinativo expreso e sejam finitas. Terá de fazer uma diferenciação entre subtipos de subordinação.⁴ O facto de o conector subordinativo *que* ser claramente um marcador de subordinação, introduzindo orações completivas finitas de vários tipos, poderá facilitar a aquisição.

4) *Advérbios:*

Com os advérbios, a criança terá de começar a adquirir propriedades de itens lexicais específicos. Note-se que nem todos os advérbios pré-verbais são indutores de próclise, como se ilustra no contraste entre (36a) e (36b):

- (36) a. O João já o leu. / *O João já leu-o.
 b. *O João ontem o leu. / O João ontem leu-o.

A criança terá de recorrer a conhecimento lexical para descobrir quais os advérbios que, em posição pré-verbal, induzem próclise.⁵

5) *Orações adverbiais finitas:*

Conforme demonstrado em Lobo (2003), apenas conhecendo alguns conectores específicos, é possível determinar se estamos perante subordinação adverbial ou coordenação. Por exemplo, em (37), a diferença entre as duas frases prende-se sobretudo com o conhecimento das propriedades das conjunções envolvidas:

- (37) a. O João ficou em casa, pois constipou-se. *Coordenação explicativa*
 b. O João ficou em casa, porque se constipou. *Subordinação causal*

A criança terá, assim, de recorrer a conhecimento item a item, para determinar se está perante um contexto de próclise ou de ênclise. Ao contrário do que acontece com as completivas, na subordinação adverbial encontramos uma multiplicidade

4 Martins (2013) mostra que há alguma variação na colocação de clíticos na gramática adulta em orações subordinadas completivas com o indicativo. Como o contexto que usámos era um contexto de conjuntivo, espera-se que seja um contexto onde ocorra categoricamente próclise.

5 Estes advérbios não constituem uma classe semântica nem sintática claramente delimitada, o que torna a tarefa mais difícil.

de conectores que podem funcionar como marcadores de um domínio de subordinação.

6) *Sujeitos quantificados:*

O mesmo acontece com os sujeitos quantificados. Se, em alguns casos, os quantificadores envolvidos são contextos categóricos de próclise, como em (38a) e (38b), outros quantificadores há que permitem variação entre próclise e ênclise, como se atesta em (38c) e (38d) (cf. Martins 2013):

- (38) a. Todos se viraram.
 b. *Todos viraram-se.
 c. Alguns meninos o viram.
 d. Alguns meninos viram-no.

É, portanto, tarefa da criança determinar quais os quantificadores específicos que funcionam como indutores de próclise ou não, e se induzem próclise obrigatoriamente ou opcionalmente (com consequências interpretativas que aqui não se discutem). Curiosamente, os contextos em que há uma maior dependência de conhecimento lexical são aqueles em que os próprios adultos exibem alguma variação e oscilação entre próclise e ênclise.

Deste padrão de resultados, podemos extrair a conclusão de que quanto menos lexicalmente especificado é o contexto desencadeador de próclise, mais precoce é a sua estabilização. Estes resultados sobre a colocação dos pronomes clíticos são importantes, já que constituem evidência para a interação entre fatores sintáticos e lexicais na estabilização do conhecimento linguístico, mostrando-se que a aquisição de propriedades específicas de itens lexicais é mais complexa do que a estabilização do conhecimento sintático.

5 Aquisição das propriedades referenciais dos pronomes clíticos

Consideremos agora a compreensão das propriedades referenciais dos pronomes clíticos. Como é sabido, os clíticos comportam-se como pronomes ou anáforas, no sentido clássico da Teoria da Ligação de Chomsky (1981). Assim, conforme ilustrado em (39), um clítico anafórico deve encontrar o seu antecedente no domínio em que se encontra, enquanto o clítico pronominal não pode ter um antecedente que o c-comande no mesmo domínio, conforme ilustrado em (40):

- (39) a. A Maria_i lavou-se_{i/*j}.
 b. A Maria_i disse que a Ana se_i lavou.

- (40) a. A Maria_i lavou-a_{i/j}.
 b. A Maria_i disse que a Ana a_i lavou.
 c. A filha d[a Maria]_i lavou-a_i.

Desde Chien/Wexler (1990), a aquisição das formas referencialmente dependentes tem recebido muita atenção. Estes autores, num estudo sobre o inglês, mostraram que as crianças falantes do inglês exibem um contraste na sua performance entre a compreensão de anáforas e a compreensão de pronomes. Por outras palavras, as crianças testadas naquele estudo não exibiram problemas na compreensão de anáforas, em frases como as de (41a), mas não excluíram leituras correferenciais entre o pronome e o sujeito em frases como as de (41b):

- (41) a. Ann washes herself.
 b. Ann washes her.

Este teste foi replicado em várias línguas e, desde McKee (1992), que conduziu este estudo em italiano, tornou-se claro que, em línguas em que o pronome é um clítico, não se encontram problemas na compreensão das propriedades interpretativas do pronome. Em português europeu, este teste foi replicado por Cristóvão (2006), tendo-se confirmado que, à semelhança do que acontece em outras línguas com clíticos, as crianças portuguesas não exibem dificuldades na compreensão de frases como (42):

- (42) A mãe lava-a.

Perante frases como estas, as crianças rejeitam interpretações correferentes para o pronome e o sujeito da frase. Evidência independente de que o estatuto do pronome (enquanto clítico ou não) desempenha um papel importante na previsibilidade de se encontrarem diferenças na compreensão vem da comparação entre o português europeu e o português brasileiro. Grolla (2006) testou a compreensão dos pronomes em português brasileiro e encontrou nas crianças testadas uma performance semelhante à que Chien/Wexler tinham encontrado para o inglês. Crucialmente, as frases utilizadas no teste de Grolla envolviam pronomes fortes e não clíticos:

- (43) A mãe lava ela.

A mesma conclusão pode ser retirada do estudo de Silva (2014; 2015), em que se mostra que, mesmo internamente ao português europeu, a compreensão de um pronome forte é mais difícil do que a de um clítico.

Silva comparou a compreensão de frases com pronomes clíticos e pronomes fortes, como as que se encontram em (44):

- (44) a. A mãe lava-a.
 b. A mãe bate nela.

Em ambos os casos encontramos-nos perante verbos transitivos, mas em (44b), devido à presença da preposição, o pronome utilizado é um pronome forte e não um clítico. O tipo de pronome tem, crucialmente, consequências para a sua interpretação pelas crianças.

Silva (2014) compara a compreensão de frases como as seguintes:

- | | | |
|---------|--------------------------------|---------------------------------|
| (45) a. | A mãe lava-se. | <i>Clítico anafórico</i> |
| b. | A mãe lava-a. | <i>Clítico pronominal</i> |
| c. | A mãe está a apontar para si. | <i>Pronome forte anafórico</i> |
| d. | A mãe está a apontar para ela. | <i>Pronome forte pronominal</i> |

Vejamos a performance das crianças, quando comparada com a dos adultos. O primeiro par de gráficos (de Silva 2014) mostra os resultados das crianças de 3, 4 e 5 anos na compreensão de clíticos, comparando-as com os adultos:

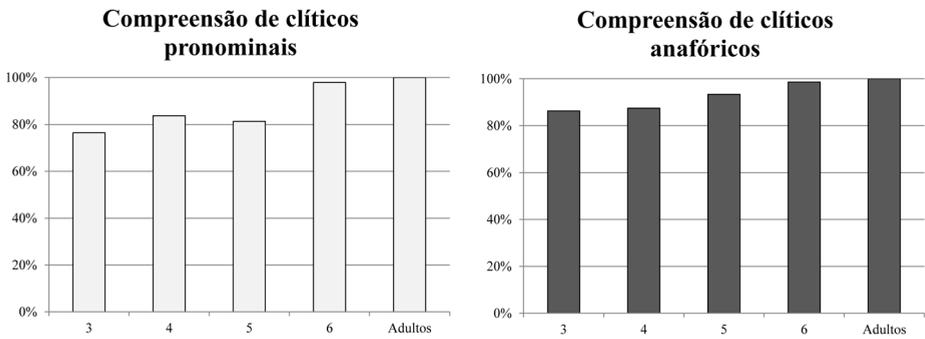


Figura 1: Compreensão de clíticos anafóricos e pronominais (Silva 2014).

Apesar de haver algum desenvolvimento, sobretudo entre os 5 e os 6 anos, nota-se que há, desde cedo, uma clara convergência com a gramática do adulto. Os resultados das crianças portuguesas estão muito longe das taxas de 50% obtidas para outras línguas. O mesmo não acontece na interpretação de pronomes fortes, conforme se mostra nos seguintes gráficos:

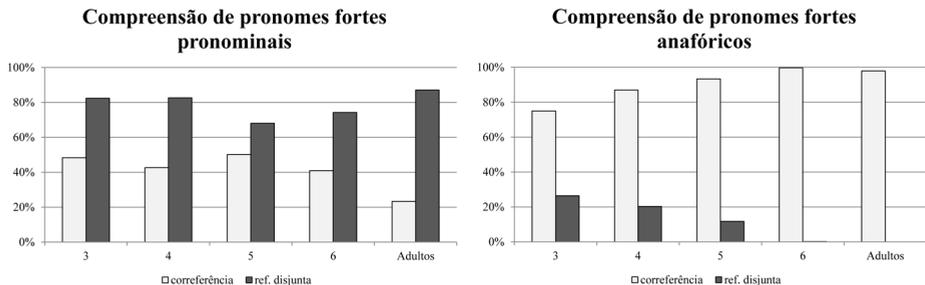


Figura 2: Compreensão de pronomes fortes anafóricos e pronominais (Silva 2014).

Conforme se vê nos gráficos acima, a compreensão de pronomes fortes, sobretudo no contexto pronominal, difere bastante da performance dos adultos. Aos seis anos, a diferença entre a taxa de aceitação de correferência para pronomes em relação aos adultos é ainda de cerca de 20%, o que aponta para uma estabilização bastante mais tardia das formas pronominais fortes.

Estes dados são importantes no contexto deste capítulo por dois motivos. Por um lado, confirmam que as crianças têm um bom conhecimento das formas clíticas. Nas secções anteriores, vimos que as crianças são capazes de produzir clíticos e que a convergência tardia com as taxas de produção dos adultos não se deve a um mau conhecimento dos clíticos, mas sim a uma aquisição tardia das condições que restringem o uso da construção de objeto nulo. Nestes dados, observa-se que o bom conhecimento dos clíticos se confirma numa performance bastante semelhante à dos adultos no que concerne à interpretação da referência dos clíticos.

Por outro lado, a performance das crianças na compreensão de pronomes fortes apoia a hipótese de que a variação e a competição no input pode levar a uma estabilização mais lenta do conhecimento de algumas construções. Uma vez que, conforme demonstrado em Menuzzi (1999) e Estrela (2006), as formas pronominais fortes anafórica e pronominal competem, por vezes, para o mesmo uso e com a mesma interpretação, por exemplo em contextos como os de (46), espera-se que a sua aquisição seja mais lenta do que a dos clíticos, para os quais não se observa o mesmo tipo de competição.

- (46) a. O João_i viu uma cobra atrás de si_i.
 b. O João_i viu uma cobra atrás dele_i.

6 Conclusões

Reportámos, neste capítulo, os resultados de vários estudos sobre a produção e compreensão de pronomes clíticos. Sintetizamos os resultados nos seguintes pontos:

- a) As crianças falantes de português europeu produzem clíticos desde cedo, mas tendem a omiti-los devido aos efeitos de sobregeneralização da construção de objeto nulo.
- b) A hipótese de que as crianças falantes de português europeu sobregeneralizam o objeto nulo é confirmada pelo facto de compreenderem bem estruturas com objeto nulo e por aceitarem objetos nulos nos mesmos contextos em que os produzem.
- c) As crianças falantes de português europeu conhecem as propriedades referenciais dos pronomes, sendo a sua compreensão mais precoce do que a dos pronomes fortes.
- d) As crianças falantes de português europeu tendem a sobregeneralizar a posição enclítica dos pronomes, sendo o tempo para a estabilização da próclise condicio-

nado pela necessidade de determinar aspetos lexicais ou apenas sintáticos dos desencadeadores de próclise. Observou-se que quanto menor é a dependência de aspetos lexicais, mais prematura é a convergência com a performance dos adultos.

7 Referências

- Chien, Yu-Chin/Wexler, Ken (1990), *Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics*, *Language Acquisition* 1, 225–295.
- Chomsky, Noam (1981), *Lectures on Government and Binding*, Dordrecht, Foris.
- Chomsky, Noam (1986), *Knowledge of Language, its Nature, Origin and Use*, New York, Praeger.
- Costa, João/Fiéis, Alexandra/Lobo, Maria (2014), *Input variability and late acquisition: clitic misplacement in European Portuguese*, *Lingua* 161, 10–26, <http://dx.doi.org/10.1016/j.lingua.2014.05.009> (21.10.2015).
- Costa, João/Grolla, Elaine/Lobo, Maria (2014), *The acquisition of microvariation in silent categories*, in: Cornelia Hamann/Esther Ruigendijk (edd.), *Language Acquisition and Development. Proceedings of GALA 2013*, Newcastle upon Tyne, Cambridge Scholars Publishing, 49–60.
- Costa, João/Lobo, Maria (2006), *A aquisição de clíticos em PE: Omissão de Clíticos ou Objectos Nulos?*, in: *XXI Encontro Nacional da APL. Textos seleccionados*, Lisboa, APL, 285–293.
- Costa, João/Lobo, Maria (2007a), *Clitic Omission, null objects or both in the acquisition of European Portuguese?*, in: Sergio Baauw/Frank Drijkoningen/Manuela Pinto (edd.), *Romance Languages and Linguistic Theory 2005*, Amsterdam/Philadelphia, Benjamins, 59–72.
- Costa, João/Lobo, Maria (2007b), *Complexidade e omissão de clíticos: o caso dos reflexos*, in: *XXII Encontro Nacional da APL. Textos seleccionados*, Lisboa, APL, 303–313.
- Costa, João/Lobo, Maria (2009), *Clitic omission in the acquisition of European Portuguese: Data from comprehension*, in: Acrísio Pires/Jason Rothman (edd.), *Minimalist Inquiries into Child Language Acquisition. Case Studies across Portuguese*, Berlin/Boston, De Gruyter, 63–84.
- Costa, João/Lobo, Maria (2010), *Clitic omission is null object: evidence from comprehension*, in: João Costa et al. (edd.), *Language Acquisition and Development: Proceedings of GALA 2009*, Newcastle upon Tyne, Cambridge Scholars Publishing, 96–106.
- Costa, João/Lobo, Maria (2014), *Crosslinguistic variation in the acquisition of null categories: Portuguese vs. Japanese*, in: Kleantes K. Grohmann/Theoni Neokleous (edd.), *Developments in the Acquisition of Clitics*, Newcastle upon Tyne, Cambridge Scholars Publishing, 276–291.
- Costa, João/Lobo, Maria/Silva, Carolina (2012), *Which category replaces an omitted clitic? The case of European Portuguese*, in: Maria Pilar Larrañaga/Pedro Guijarro-Fuentes (edd.), *Pronouns and clitics in early acquisition*, Berlin/Boston, De Gruyter, 105–130.
- Cristóvão, Sandra (2006), *A co-referência nos pronomes objecto directo na aquisição do português europeu*, dissertação de Mestrado, FCSH-Universidade Nova de Lisboa.
- Duarte, Inês/Matos, Gabriela (2000), *Romance Clitics and the Minimalist Program*, in: João Costa (ed.), *Portuguese Syntax. New Comparative Studies*, Oxford, Oxford University Press, 116–142.
- Duarte, Inês/Matos, Gabriela/Faria, Isabel Hub (1995), *Specificity of European Portuguese Clitics in Romance*, in: Isabel Hub Faria/Maria João Freitas (edd.), *Studies on the Acquisition of Portuguese*, Lisboa, APL/Colibri, 129–154.

- Estrela, Antónia (2006), *A Teoria da Ligação: Dados do Português Europeu*, dissertação de Mestrado, FCSH-Universidade Nova de Lisboa.
- Grolla, Elaine (2006), *The Acquisition of A- and A'-Bound Pronouns in Brazilian Portuguese*, in: Vincent Torrens/Linda Escobar (edd.), *The Acquisition of Syntax in Romance Languages*, Amsterdam/Philadelphia, Benjamins, 227–250.
- Grüter, Thérèse (2006), *Object clitics and null objects in the acquisition of French*, tese de Doutorado, Montreal, McGill University.
- Jakubowicz, Celia, et al. (1998), *Determiners and Clitic Pronouns in French-Speaking Children with SLI*, *Language Acquisition* 7:2, 113–160.
- Lobo, Maria (2003), *Aspectos da Sintaxe das Orações Subordinadas Adverbiais do Português*, tese de Doutorado, Universidade Nova de Lisboa.
- Martins, Ana Maria (2013), *Posição dos Pronomes Pessoais Clíticos*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 2, cap. 42, Lisboa, Fundação Calouste Gulbenkian, 2229–2302.
- McKee, Cecile (1992), *A Comparison of Pronouns and Anaphors in Italian and English Acquisition*, *Language Acquisition* 2, 21–54.
- Menuzzi, Sérgio (1999), *Binding Theory and Pronominal Anaphora in Brazilian Portuguese*, The Hague, Holland Academic Graphics.
- Raposo, Eduardo Paiva (1986), *On the null object construction in European Portuguese*, in: Osvaldo Jaeggli/Carmen Silva-Corvalán (edd.), *Studies in Romance Linguistics*, Dordrecht, Foris, 373–390.
- Saito, Mamoru (2007), *Notes on East Asian argument ellipsis*, *Language Research* 43, 203–227.
- Schaeffer, Jeannette (1997), *Direct object scrambling in Dutch and Italian child language*, UCLA dissertations in Linguistics, Los Angeles, University of California.
- Silva, Carolina (2008), *Assimetrias na aquisição de clíticos diferenciados em português europeu*, dissertação de Mestrado, FCSH-Universidade Nova de Lisboa.
- Silva, Carolina (2014), *Interpretation of Clitic, Strong and Null Pronouns by Typically Developing Portuguese Children*, in: João Costa et al. (edd.), *New Directions in the Acquisition of Romance Languages. Selected Proceedings of The Romance Turn V*, Newcastle upon Tyne, Cambridge Scholars Publishing, 302–325.
- Silva, Carolina (2015), *Interpretation of Clitic, Strong and Null Pronouns in the Acquisition of European Portuguese*, tese de Doutorado, Universidade Nova de Lisboa.
- Soares, Carla (2006), *La syntaxe de la périphérie gauche et son acquisition en Portugais Européen*, tese de Doutorado, Université Paris 8.
- Sugisaki, Koji (2007), *The configurationality parameter in the minimalist program: A view from child Japanese*, in: Heather Caunt-Nulton/Samantha Kulatilake/I-hao Woo (edd.), *Proceedings of the Boston University Conference on Language Development 31*, Somerville MA, Cascadia Press, 597–608.
- Sugisaki, Koji (2009), *Argument Ellipsis in Child Japanese: A Preliminary Report*, in: Yukio Otsu (ed.), *The Proceedings of the Tenth Tokyo Conference on Psycholinguistics*, Tokyo, Hituzi Syobo, 291–312.
- Takahashi, Daiko (2008), *Noun phrase ellipsis*, in: Shigeru Miyagawa/Mamoru Saito (edd.), *The Oxford Handbook of Japanese Linguistics*, New York, Oxford University Press, 394–422.
- Tsakali, Vina/Wexler, Ken (2004), *Why children omit clitics in some languages but not in others: new evidence from Greek*, in: Jacqueline van Kampen/Sergio Baauw (edd.), *Proceedings of GALA 2003*, vol. 2, Utrecht, LOT, 493–504.
- Vigário, Marina (2001), *The Prosodic Word in European Portuguese*, tese de Doutorado, Universidade de Lisboa.

Wexler, Ken/Gavarró, Anna/Torrens, Vicenç (2004), *Feature checking and object clitic omission in child Catalan and Spanish*, in: Reineke Bok-Bennema/Bart Hollebrandse/Brigitte Kaspers-Manhe (edd.), *Romance language and linguistic theories: Selected papers from Going Romance 2002*, Amsterdam/Philadelphia, Benjamins, 253–269.

Inês Duarte, Ana Lúcia Santos e Anabela Gonçalves

17 O infinitivo flexionado na gramática do adulto e na aquisição de L1

Abstract: O presente capítulo tem como objeto de estudo o infinitivo flexionado em português, sendo a descrição feita quer na perspectiva da gramática adulta do português padrão quer na da aquisição do português como língua materna (L1). No que diz respeito à gramática do adulto, apresentam-se os contextos que legitimam a ocorrência de infinitivo flexionado, bem como os mecanismos de interpretação dos sujeitos nulos que ocorrem na frase infinitiva. Do ponto de vista da aquisição do infinitivo por crianças que têm o português como L1, apresentam-se resultados de investigação que se centram na emergência do infinitivo flexionado com referência à idade, bem como na identificação de estádios de emergência do mesmo em função de diferentes contextos sintáticos. Os resultados obtidos para o português europeu são comparados com dados do português brasileiro relativos à construção de infinitivo pessoal, de forma a mostrar que as duas construções emergem em idades comparáveis e nos mesmos contextos.

Keywords: infinitivo flexionado, aquisição, completivas, frases adjuntas, infinitivo preposicionado

1 Introdução

O infinitivo flexionado é uma forma verbal rara, característica do português e do galego.¹ Trata-se de uma forma infinitiva com morfemas de pessoa e número, como se pode observar no paradigma apresentado na tabela 1, com verbos regulares de 1.^a, 2.^a e 3.^a conjugações:

1 Alguns autores apresentam dados de outras línguas, como o húngaro e o sardo, que, aparentemente, dispõem também de infinitivo flexionado. No entanto, trata-se de uma forma distinta da que existe em português. Por exemplo, em húngaro, coocorre apenas com predicados impessoais, que não envolvem concordância em pessoa e número, e o sujeito tem caso dativo (Tóth 2000; Szabolcsi 2009); em sardo, a ocorrência de infinitivo flexionado restringe-se a completivas objeto, estando excluído de completivas sujeito (Pountain 1995).

Tabela 1: Infinitivo flexionado.

Pessoa	Número	
	Singular	Plural
1. ^a	Cantar, comer, partir	Cantarmos, comeremos, partirmos
2. ^a	Cantares, comeres, partires	Cantardes, comerdes, partirdes
3. ^a	Cantar, comer, partir	Cantarem, comerem, partirem

No singular, apenas a forma de 2.^a pessoa apresenta morfemas explícitos (-es), que permitem distingui-la da forma do infinitivo não flexionado, sendo as restantes homónimas. Já no plural, todas as pessoas exibem morfemas explícitos. Com exceção da 2.^a pessoa do plural (caída em desuso neste e nos restantes tempos do paradigma verbal na variedade padrão),² as restantes são frequentes na língua oral e escrita, emergindo o infinitivo flexionado (ou, pelo menos, algumas das suas formas) bastante cedo nas produções das crianças (ver secção 5).

Como os exemplos (1) e (2) mostram, esta forma verbal pode ocorrer com sujeitos nominativos (ver (1)) ou com sujeitos sem realização lexical (ver (2)):³

- (1) a. Mas o curioso é *tu* não perceberes que não houve nunca «ilusão» alguma.
(Vergílio Ferreira, *Nítido Nulo*. *apud* Cunha/Cintra ⁷1990, ¹1984, 485)
- b. Soube da novidade ainda antes de *eles* me telefonarem.
- (2) a. Quem te deu, pois, o direito de correres a morte certa? (Herculano 1844, 95).
- b. Lamento terem-te ofendido tão injustamente.

O facto de haver homonímia na 1.^a e na 3.^a pessoa do singular leva a que possa haver ambiguidade entre estas formas e as de infinitivo não flexionado. Como veremos adiante com mais pormenor, há contextos sintáticos que impedem a ocorrência de infinitivo flexionado, pelo que, nesses contextos, não existe ambiguidade (ver (3)); por outro lado, na língua padrão, a presença de um sujeito lexical é um indício de que a forma é de infinitivo flexionado (ver (4)):

2 Vejam-se os seguintes exemplos de Camões e de Herculano (este último mima um diálogo que localiza no início do século XV):

- (i) «Dizei: se sois mandados, como estais/Sem irdes a pregar a santa Fé?» (Camões, *Os Lusíadas*, 1572 X, 119. 3–4)
- (ii) «Beijo-vos as mãos, senhor rei, por vos lembrardes ainda de um velho homem de armas que para nada presta hoje.» (Herculano ¹1851, 282)

As formas verbais de 2.^a pessoa do plural persistem na língua literária pelo menos até final do séc. XVII (Bechara 1991).

3 Nos exemplos (1) e (2), os sujeitos explícitos da oração infinitiva estão em itálico e a forma de infinitivo flexionado está sublinhada.

- (3) a. Nós queremos ir ao cinema.
b. *Nós queremos irmos ao cinema.
- (4) a. Calhou [eu estar em casa a essa hora].
b. Foi um alívio [você ter encontrado a minha agenda].
c. É melhor [ela/a minha prima consultar um advogado].

Nas secções 2, 3 e 4, ocupar-nos-emos das propriedades do infinitivo flexionado na gramática adulta do português padrão. A primeira delas tratará da distribuição do infinitivo flexionado, em particular em contextos de subordinação, aqueles em que ocorre com maior frequência. A ocorrência do infinitivo flexionado na construção denominada infinitivo preposicionado ou gerundivo será considerada na secção 3. A secção 4 tem por tema a interpretação referencial do sujeito das orações de infinitivo flexionado, com especial atenção aos casos em que tal sujeito é foneticamente nulo. Finalmente, a secção 5 ocupar-se-á do infinitivo flexionado na aquisição do português como L1. Uma curta secção de síntese rematará o capítulo.

2 Propriedades do infinitivo flexionado na gramática do adulto

Tal como as formas não finitas em geral, o infinitivo flexionado apresenta uma distribuição restrita, i.e., ocorre, tipicamente, em contextos de frase subordinada. De facto, em contextos de frases-raiz, só em enunciados com valor ilocutório avaliativo e expressivo (exclamativas e optativas) é possível encontrar infinitivo flexionado. Vejam-se os seguintes exemplos – (5c) contém duas fraseologias:

- (5) a. Tu, Hermengarda, recordares-te?! Mentira! (Herculano 1844, 95).
b. E eles a teimarem!
c. Eles a darem-lhe e a burra a fugir!⁴
- (6) a. Poderes fazer o que te apetece sem ninguém a mandar em ti!
b. Vivermos sem medo! Quem nos dera!

Em contextos de subordinação, a distribuição do infinitivo flexionado pode ser descrita através das generalizações descritivas formuladas em A a E:

A. Em contextos distintos dos de controlo obrigatório, o infinitivo flexionado tem uma distribuição livre (ver secção 2.1).⁵

⁴ Esta expressão idiomática utiliza-se a propósito de alguém que defende obstinada e irracionalmente uma opinião ou um comportamento.

⁵ O termo *controlo obrigatório* identifica estruturas em que, quando o infinitivo não flexionado é usado, o sujeito da oração infinitiva tem um antecedente fixo na oração superior: o sujeito, como em

- B. Com verbos de controlo obrigatório de sujeito, o infinitivo flexionado obedece a restrições de natureza temporal: não pode ocorrer em completivas de tempo dependente nem em completivas de tempo independente selecionadas por um verbo superior que tenha uma orientação temporal especificada (ver secção 2.2).
- C. Em contextos de controlo obrigatório de objeto direto, o infinitivo flexionado ocorre na 3.^a pessoa do singular ou do plural (ver secção 2.3).
- D. Em contextos de controlo obrigatório de objeto indireto ou em que se verifica opcionalidade entre controlo de objeto indireto ou de sujeito, o infinitivo flexionado é sempre possível nas orações introduzidas por *para* (ver secção 2.4).
- E. Em orações dependentes de verbos de elevação pode ocorrer o infinitivo flexionado (ver secção 2.5).

Nas secções seguintes ocupar-nos-emos da descrição das propriedades das estruturas com infinitivo flexionado em cada um dos contextos considerados nestas generalizações.

2.1 Contextos previstos pela Generalização A

Os sujeitos frásicos não finitos constituem um dos contextos descritos pela Generalização A, que se repete de seguida:

Generalização A

Em contextos distintos dos de controlo obrigatório, o infinitivo flexionado tem uma distribuição livre.

As orações de infinitivo flexionado com função sintática de sujeito ocorrem tipicamente em posição pós-verbal, denominando-se, neste caso, extrapostas. Podem ser selecionadas por verbos, nomes e adjetivos, como se mostra em (7):

- (7) a. Calhou [(nós/eu e o meu marido) estarmos em casa a essa hora].
- b. Foi um alívio [(você/eles/os vizinhos) terem encontrado a minha agenda].
- c. É melhor [(tu) consultares um advogado].

O sujeito da oração infinitiva pode estar realizado lexicalmente (por exemplo, em (7a), o pronome nominativo *nós* ou o sintagma nominal coordenado *eu e o meu marido*) e, quando tal acontece, a ordem canónica sujeito-verbo é observada na oração infinitiva.

(i), o objeto direto, como em (ii), ou o objeto indireto, como em (iii) – o antecedente está assinalado a itálico:

- (i) *Os meninos* prometeram aos pais [-] chegar a casa a horas.
- (ii) Os meninos obrigaram *os pais* a [-] ir com eles ao concerto.
- (iii) Os meninos disseram *aos pais* para [-] não esperar por eles.

As orações infinitivas sujeito têm uma distribuição mais restrita em posição pré-verbal. Assim, com verbos que introduzem no universo do discurso uma eventualidade,⁶ esta ordem de palavras induz agramaticalidade (cf. (8)), o mesmo acontecendo com verbos que exprimem uma condição necessária ou suficiente para que uma dada eventualidade tenha lugar⁷ (veja-se o contraste entre (8a) e (8b) e entre (9a) e (9b)):

- (8) a. Calhou [estarmos em casa a essa hora].
 b. *[Estarmos em casa a essa hora] calhou.
- (9) a. Basta [pedirem uma dose de caril de gambas].
 b. *[Pedirem uma dose de caril de gambas] basta.

Embora menos forte, esta restrição também se verifica com verbos psicológicos que selecionam um experienciador como objeto direto⁸ (cf. (10a) vs. (10b)) ou indireto⁹ (cf. (11a) vs. (11b))¹⁰ e com predicados nominais e adjetivais (12), oscilando os falantes na avaliação da gramaticalidade das frases:¹¹

- (10) a. Comoveu-as [encontrarem as fotografias do tempo da escola].
 b. (?)[Encontrarem as fotografias do tempo de escola] comoveu-as.
- (11) a. Custou-lhes [deixarem a casa de família abandonada].
 b. ??[Deixarem a casa de família abandonada] custou-lhes.
- (12) a. É melhor [consultares um advogado].
 b. *[Consultares um advogado] é melhor.

Também verbos causativos de dois e três lugares cujo argumento externo é a causa do evento expresso pela frase podem selecionar orações de infinitivo flexionado como sujeito. No primeiro caso, trata-se de verbos que admitem um argumento experienciador em dativo (ver (13)); no segundo caso, trata-se de verbos de controlo de objeto, que selecionam um argumento humano como objeto direto e uma frase completiva como complemento oblíquo (ver (14)):

6 Pertencem a esta classe os verbos *acontecer*, *constar*, *ocorrer*, *suced*.

7 Comportam-se como *bastar* os verbos *chegar* e *faltar*.

8 Pertencem a esta classe verbos como *afligir*, *chatear*, *divertir*, *impressionar*.

9 Pertencem a esta classe verbos como *agradar*, *custar*, *doer*, *importar*, *interessar*.

10 Note-se que os exemplos (a) de (10) e (11) são contextos de controlo obrigatório, de acordo com Landau (2010; 2013): o verbo superior é psicológico, selecionando um argumento com o papel de Causa (a oração infinitiva) e outro com o papel de Experienciador (respetivamente, os clíticos acusativo e dativo em (10) e (11)). Na secção 3 retomaremos este assunto.

11 Em construções que envolvem graduação, com material lexical à direita, as frases (b) tornam-se mais naturais:

- (i) Encontrarem as fotografias do tempo de escola comoveu-as muito mais do que eu esperaria.
 (ii) Deixarem a casa de família abandonada custou-lhes muitíssimo.
 (iii) Consultares um advogado é melhor do que não fazeres nada.

- (13) a. [Beberes tanto] prejudica-te a saúde.
 b. [Terem diminuído os comboios da linha de Cascais] causou-nos imenso transtorno.
- (14) a. [Estudares comigo] ajudou-me a compreender a matéria de álgebra.
 b. [Termos sido assaltados] obrigou-me a ser mais prudente.

Finalmente, com uma subclasse de verbos causativos denominados inferenciais, o argumento externo do verbo superior pode ser uma oração de infinitivo flexionado e o argumento interno pode também ser uma frase (finita ou não finita):

- (15) a. [Acreditares em videntes] mostra que és muito pouco racional.
 b. [Os teus miúdos terem-se oferecido como voluntários] revela que os educaste bem.

Como os exemplos (13)-(15) mostram, com verbos causativos de dois e três lugares e com verbos inferenciais, a oração de infinitivo flexionado com a função de sujeito pode ocorrer na ordem direta (i.e. sujeito-verbo).

Outro dos contextos cujo comportamento é previsto pela Generalização A é o das frases adjuntas.¹² Como referido acima, é nestas que o infinitivo flexionado ocorre com maior frequência, em particular em adjuntas finais e temporais (veja-se (16) e (17)), mas também em causais e em concessivas (veja-se (18) e (19)):

- (16) a. Saímos das aulas a correr [para irmos ao teatro].
 b. Fico com a tua filha [para ires à festa].
- (17) a. Apanha um autocarro [depois de chegares ao Marquês].
 b. [Ao chegarmos a casa], já chovia a potes.
- (18) a. Tivemos um grande aborrecimento [por não termos sido prudentes].
 b. [Dado o governo ter aumentado brutalmente os impostos], a classe média perdeu poder de compra.
- (19) a. [Apesar de não terem estudado o suficiente], os miúdos acabaram por passar.
 b. Ainda conseguimos entrar na sala, [apesar de termos chegado atrasados].

Com exceção da construção temporal exemplificada em (17b), as orações adjuntas de infinitivo flexionado admitem a ordem direta sujeito-verbo quando o sujeito está lexicalmente realizado, como se pode observar em (18b) e nos exemplos (20):

- (20) a. Ficámos com as crianças [para os pais irem ao teatro].
 b. [Apesar de os miúdos não terem estudado o suficiente], acabaram por passar no exame.

¹² Pelo menos as frases adjuntas encaixadas acima do VP, como acontece com as temporais, as concessivas, as causais e as finais.

O infinitivo flexionado ocorre ainda numa subclasse de frase relativas, introduzidas por *para* e com valor final. Tais frases, designadas como relativas finais, encontram-se exemplificadas em (21) – cf. Duarte/Santos/Alexandre (2014):¹³

- (21) a. Os livros [para lermos] encontram-se na biblioteca da escola.
 b. Os livros [para os alunos lerem] encontram-se na biblioteca da escola.

A Generalização A capta igualmente a distribuição do infinitivo flexionado em complementos de nomes e de adjetivos. Assim, nomes formados a partir de verbos que selecionam complementos podem construir-se com orações de infinitivo flexionado, sempre precedidas de preposição. Quando a forma base verbal é transitiva direta, não selecionando, portanto, uma preposição específica, é usada a preposição de ligação *de* (ver (22a)):

- (22) a. Eles sempre confessaram o desejo [de emigrarem].
 b. É óbvio o interesse [em esconderem a verdade sobre o caso].

Também nomes modais (ver (23a)) e nomes que descrevem eventualidades (ver (23b)) podem construir-se com orações de infinitivo flexionado que especificam o seu significado (cf. Barbosa/Raposo 2013):

- (23) a. Não acredito na possibilidade [de seres convidada para a festa].
 b. Agrada-me a ideia [de irmos ao cinema logo].

Como os exemplos acima mostram, o sintagma nominal que contém a oração infinitiva pode desempenhar diferentes funções sintáticas na frase superior: exemplificando, tem a função de complemento direto em (22a), de complemento oblíquo em (23a) e de sujeito nos dois casos restantes.

13 Como mostram Duarte/Santos/Alexandre (2014), algumas frases infinitivas com valor final introduzidas por *para* são ambíguas, podendo ser interpretadas como adjuntas finais ou como relativas finais. Tal acontece, em particular, quando a frase ocorre em posição final, após o objeto direto, como em (i):

- (i) Já comprei peixe [para grelhar] (Duarte/Santos/Alexandre 2014, ex. (1c)).

Baseadas em Jones (1991), as autoras apresentam um conjunto de critérios que permitem distinguir as relativas finais das adjuntas finais. Assim, por exemplo, uma relativa final, como as de (21), é obrigatoriamente clivada conjuntamente com o grupo nominal que a precede, como se mostra em (ii):

- (ii) a. São [os livros para lermos] que se encontram na biblioteca da escola.
 b. *São [os livros] que se encontram na biblioteca da escola [para lermos]
 (agramatical na interpretação de (21a)).

Em (i), porém, existem duas possibilidades de clivagem: a frase infinitiva pode ser clivada conjuntamente com o grupo nominal que a antecede, o que faz dela uma relativa final (iiia), mas também pode manter-se autónoma desse grupo (iiib), o que a aproxima das adjuntas finais.

- (iii) a. Foi [peixe para grelhar] que já comprei.
 b. Foi [peixe] que já comprei [para grelhar].

Orações de infinitivo flexionado podem igualmente ser seleccionadas por adjetivos psicológicos (ver (24a)) ou por adjetivos deverbais (ver (24b)):

- (24) a. Eles estão ansiosos [por irem de férias].
b. Detesto as decisões destinadas [a parecermos solidários].

Nestes casos, a oração infinitiva pode integrar o predicado secundário de uma frase copulativa (como acontece com *ansiosos por irem de férias*) ou ocorrer internamente a um sintagma nominal (como se pode observar em *as decisões destinadas a parecermos solidários*).

Refira-se que é em contextos de controlo não obrigatório, em particular em frases adjuntas, que ocorre em outras línguas românicas o chamado infinitivo pessoal, i.e., uma estrutura infinitiva com sujeito expreso marcado com nominativo e sem flexão de pessoa e número. Os exemplos do espanhol e do português brasileiro apresentados, respetivamente, em (25) e (26) são ilustrativos desta estrutura:

- (25) a. No me sorprendería tener algún tío ricacho en América, [sin yo saberlo].
b. Todo por culpa de cosas que pasaron [antes de yo nacer].
(Exemplos de Mensching 2000, 26)

- (26) a. A Maria ligou antes de nós/*de mim sair. (Exemplo de Pires 2002, 156)
b. pega o vermelho pra nós botar.
(Corpus de Ruth Lopes, s/d, citado por Rothman et al. 2013, 169)

Na gramática padrão adulta do português europeu, as formas de infinitivo pessoal não são aceites, i.e., as frases em (26) são agramaticais. Este facto leva a considerar que são casos de infinitivo flexionado, e não de infinitivo pessoal, estruturas em português europeu com pronomes de 1.^a e 3.^a pessoa no singular, como as exemplificadas em (27):

- (27) a. A Maria ligou [antes de eu sair].
b. O Pedro preparou uma festa [sem ela saber].
c. A festa animou [depois de você sair].

Sobre a ocorrência deste tipo de estrutura na aquisição do português europeu, veja-se a secção 5.

2.2 Contextos previstos pela Generalização B

A distribuição do infinitivo flexionado em complementos de verbos de controlo obrigatório de sujeito é captada pela Generalização B. Repetimos em baixo a primeira parte dessa generalização:

Generalização B

Com verbos de controle obrigatório de sujeito, o infinitivo flexionado obedece a restrições de natureza temporal: não pode ocorrer em completivas de tempo dependente.

Esta condição foi proposta no artigo clássico de Raposo (1987). O conceito de tempo dependente usado nesse texto reduzia-o a uma propriedade lexical do verbo que selecionava a completiva: alguns verbos selecionariam um complemento oracional [+ T(empo)], caso em que a oração teria tempo independente; outros selecionariam completivas [- T], i.e., com tempo dependente do tempo da frase superior. Outros autores preferiram caracterizar o conceito em termos sintáticos: uma completiva infinitiva sem traços temporais autônomos teria tempo dependente (cf. Ambar 1992; 2007).

Mas qualquer das definições acima propostas faz predições erradas. Efetivamente, assumindo que a ocorrência de um advérbio de tempo orientado para o futuro numa completiva infinitiva selecionada por um verbo superior no pretérito perfeito seja um teste seguro de independência temporal, no contexto exemplificado em (28a) as completivas infinitivas têm tempo independente, contrariamente às que ocorrem em (28b):

- (28) a. O João e a Maria decidiram/prometeram [partir amanhã].
 b. *O João e Maria quiseram/pretenderam [partir amanhã].

Assim, a assumir análises como as de Raposo (1987) ou Ambar (1992; 2007), esperar-se-ia que os verbos superiores presentes em (28a) aceitassem complementos oracionais com infinitivo flexionado, contrariamente aos que ocorrem em (28b). Ora, como se mostra em (29), a primeira parte da predição não se confirma:

- (29) a. *O João e a Maria decidiram/prometeram [partirem amanhã].
 b. *O João e Maria quiseram/pretenderam [partirem amanhã].

Adicionalmente, a definição lexical de tempo dependente não permite explicar por que razão em contextos paralelos aos de (29b), apesar de ocorrer na completiva infinitiva uma expressão adverbial temporal orientada para o futuro, obtemos frases gramaticais desde que essa expressão não seja dêitica, como acontece com (30):

- (30) O João e Maria quiseram/pretenderam partir uns dias depois.

Gonçalves/Cunha/Silvano (2010) propuseram uma redefinição semântica do conceito de tempo dependente que resolve os problemas que se colocavam às definições anteriores. Assim, estes autores redefinem tempo dependente a partir dos conceitos de domínio temporal (o intervalo de tempo em que se localizam situações temporalmente relacionadas) e ponto de perspectiva temporal (o ponto a partir do qual uma

situação é considerada):¹⁴ há tempo dependente quando duas situações partilham o mesmo domínio temporal e o ponto de perspetiva temporal da oração infinitiva é o tempo da frase superior. Definido deste modo, tempo dependente permite não só captar a agramaticalidade de (28b) como também dar conta do contraste entre (28b), em que ocorre um advérbio dêitico ancorado no tempo da enunciação (o que é ilegítimo numa oração de tempo dependente), e (30), em que a expressão adverbial temporal está ancorada no tempo da frase superior.

Contudo, ainda persiste um problema, ilustrado pela agramaticalidade dos exemplos em (29a), que mostram que, mesmo quando a completiva não é temporalmente dependente, podem não estar reunidas as condições para a legitimação de um infinitivo flexionado. É esse problema que a segunda parte da Generalização B, aqui repetida na totalidade, pretende resolver, recorrendo ao conceito de orientação temporal:

Generalização B

Com verbos de controlo obrigatório de sujeito, o infinitivo flexionado obedece a restrições de natureza temporal: não pode ocorrer em completivas de tempo dependente *nem em completivas de tempo independente selecionadas por um verbo superior que tenha uma orientação temporal especificada*.

Duarte/Gonçalves/Santos (2012) mostraram que existe uma forte correlação entre a possibilidade de ocorrência de infinitivo flexionado e as propriedades de orientação temporal do verbo que seleciona a completiva. Com efeito, quando os verbos permitem que a situação descrita pela oração infinitiva seja anterior, sobreposta ou posterior ao tempo da frase superior, o infinitivo flexionado pode ocorrer, como acontece com muitos verbos epistémicos e declarativos (cf. (31)) e com verbos avaliativos factivos (cf. (32)); os exemplos (a) ilustram casos em que a situação descrita pela oração infinitiva é posterior à descrita pela frase superior, os exemplos (b) casos em que é anterior e os exemplos (c) casos em que há sobreposição das duas situações:

- (31) a. Os professores declararam [trazerem os testes corrigidos amanhã].
 b. Os professores declararam [já terem trazido os testes corrigidos].
 c. Os professores declararam [confiarem plenamente nos alunos].
- (32) a. Os professores acharam mal [não terem subsídio de férias para o ano].
 b. Os professores acharam mal [não terem tido subsídio de férias].
 c. Os professores acharam mal [viverem em permanente sobressalto].

Pelo contrário, quando o verbo superior exige uma localização temporal específica da oração infinitiva, o infinitivo flexionado não pode ocorrer. É exatamente isso que se passa com os verbos exemplificados em (28a) e (29a), repetidos em (33), que selecionam complementos frásicos de tempo independente:

¹⁴ Sobre os conceitos de domínio temporal e ponto de perspetiva temporal veja-se, respetivamente, Declerck (1991) e Kamp/Reyle (1993).

- (33) a. O João e a Maria decidiram/prometeram [partir amanhã].
 b. *O João e a Maria decidiram/prometeram [ter partido ontem].

Em síntese, as restrições sobre a ocorrência do infinitivo flexionado têm, na configuração em que, com infinitivo não flexionado, se verifica controlo obrigatório de sujeito, uma componente semântica (o requisito de tempo independente da oração infinitiva) e uma componente lexical (um verbo superior que selecione uma orientação temporal não especificada do complemento infinitivo).¹⁵

Consideremos agora o sujeito das orações de infinitivo flexionado. Verbos epistémicos e declarativos aceitam como complemento orações de infinitivo flexionado cujo sujeito, nulo ou realizado, pode ser tanto referencialmente idêntico como disjunto do da frase superior:

- (34) a. [-]_i Afirmámos repetidas vezes [[-]_i estarmos conscientes do problema].
 b. O governo pensa erradamente [estarem os cidadãos subordinados à vontade dos banqueiros].

Contudo, quando o sujeito tem realização fonética, deve ocorrer em posição pós-verbal (veja-se o contraste entre (34b) e (35a)), salvo se for modificado por um operador de foco e o predicado da oração infinitiva contiver um verbo modal (cf. (35b)):

- (35) a. *O governo declarou [os detentores de passaportes nacionais apresentarem-se a concurso].
 b. O governo declarou [só os detentores de passaportes nacionais poderem apresentar-se a concurso].

Este tipo de restrição não se aplica às completivas de infinitivo flexionado selecionadas por verbos avaliativos factivos, estruturas em que tanto a ordem de palavras direta quanto a ordem verbo-sujeito são possíveis, como os exemplos (36) mostram:

- (36) a. Lamento/acho mal [o governo não ter confiado nos professores].
 b. Lamento/acho mal [não ter o governo confiado nos professores].

As completivas de infinitivo flexionado selecionadas por estes verbos admitem uma construção que «nominaliza» a oração, ao fazê-la preceder do artigo definido *o* (veja-se Raposo 1987), como se pode observar em (37):

- (37) O governo lamentou o [terem sido disparados tantos tiros].

Esta construção é considerada por alguns autores uma completiva de nome elítica, em que o nome *facto* e a preposição *de* estariam omitidos (Barbosa/Raposo 2013):

- (38) O governo lamentou o facto de terem sido disparados tantos tiros.

¹⁵ Sobre este assunto, ver Duarte/Gonçalves/Santos (2012) e Gonçalves/Santos/Duarte (2014).

Note-se, contudo, que pesam sobre a construção em (37) restrições que não afetam a completiva nominal, nomeadamente, restrições sobre a ordem de palavras dentro da completiva. Assim, se o constituinte com a função de sujeito na oração infinitiva for um sintagma nominal, a ordem direta é impossível (veja-se o contraste entre (39a) e (39b)):

- (39) a. *O governo lamentou o [tantos tiros terem sido disparados].
 b. O governo lamentou o facto de tantos tiros terem sido disparados.

A ordem direta só é possível quando o sujeito é um pronome, como ilustrado em (40):

- (40) O governo lamentou o [eles terem disparado tantos tiros].

2.3 Contextos previstos pela Generalização C

Esta generalização, repetida abaixo, cobre os contextos em que ocorre na frase superior um verbo que seleciona como objeto direto um argumento humano e como oblíquo uma frase completiva, tipicamente introduzida pela preposição *a*.

Generalização C

Em contextos de controlo obrigatório de objeto direto, o infinitivo flexionado ocorre na 3.^a pessoa do singular ou do plural.

Os verbos em questão pertencem à esfera semântica da obrigação/permmissão e do aconselhamento/persuasão e o seu argumento alinhado com a função sintática de objeto direto é o controlador obrigatório do sujeito da oração infinitiva, quer ela exiba infinitivo flexionado quer não flexionado:¹⁶

- (41) a. O professor forçou-as_i [a [-]_i repetir(em) o teste].
 b. Os amigos convidaram-nos_i (a eles_i) [a [-]_i acampar(em) no Gerês].

Contudo, para muitos falantes e contrariamente ao que caracteriza em geral as construções que admitem infinitivo flexionado, apenas são aceites formas de 3.^a pessoa, sendo as formas com marcação dêitica de pessoa completamente excluídas ou consideradas marginais, como se pode observar em (42):

- (42) a. */?O professor forçou-nos_i [a [-]_i repetirmos o teste].
 b. */?Os meus amigos convidaram-te_i [a [-]_i acampares no Gerês].

O contraste entre (41) e (42) sugere que os infinitivos flexionados que podem ocorrer nestes contextos são deficitários na categoria pessoa. Este comportamento peculiar é

¹⁶ A possibilidade de ocorrência de infinitivo flexionado neste contexto foi notada, para o português brasileiro, por Modesto (2010).

acentuado pela impossibilidade de ocorrência de um sujeito lexical (ver a agramaticidade dos exemplos (43)), o que, mais uma vez, não é esperado em orações de infinitivo flexionado:¹⁷

- (43) a. *O professor forçou-as_i [a [elas]_i repetirem o teste].
 b. *Os amigos convidaram-nos_i outra vez [a [eles]_i acamparem no Gerês].

Assim, os dados acima levam a concluir que os infinitivos flexionados que ocorrem neste contexto não partilham as propriedades sintáticas características desta forma verbal, parecendo a flexão de plural uma mera consequência morfológica do processo sintático em ação no controlo de objeto.

Também verbos da esfera semântica da proibição podem selecionar completivas de infinitivo flexionado. Nestes casos, a oração infinitiva é introduzida pela preposição *de*, sendo o sujeito controlado pelo objeto direto da frase superior, como se mostra em (44):

- (44) a. O professor proibiu-as_i [de [-]_i repetir(em) o teste].
 b. O temporal impediu-os_i [de [-]_i acampar(em) no Gerês].

Com estes verbos, a restrição de pessoa não é tão severa quanto a observada com os verbos de controlo de objeto que selecionam completivas introduzidas por *a*, embora se mantenha a impossibilidade de ocorrência de um sujeito lexical na oração encaixada, como se pode ver em (45) e (46):

- (45) a. ?O professor proibiu-te_i [de [-]_i repeties o teste].
 b. O temporal impediu-nos_i [de [-]_i acamparmos no Gerês].

- (46) a. *O professor proibiu-as_i [de [elas]_i repetirem o teste].
 b. *O temporal impediu-os_i outra vez [de [eles]_i acamparem no Gerês].

A diferença de comportamento entre os dois grupos de verbos, patente no contraste entre (42) e (45), pode dever-se a diferentes estruturas argumentais. De facto, enquanto os primeiros mantêm uma estrutura argumental estável, de dois complementos,

17 A ocorrência do pronome nominativo (*elas*) em posição pós-verbal dá origem a frases gramaticais, como notado em Barbosa (2009):

(i) O professor_i forçou-as [a repetirem (só) [elas]_i o teste].

Neste caso, o pronome é anafórico do objeto direto da frase superior e tem um valor contrastivo. Assim, no exemplo (i), o pronome nominativo *elas* tem o mesmo referente do objeto direto *as* e a interpretação obtida é a de que apenas elas, e não os restantes alunos, tiveram de repetir o teste.

Note-se que a presença deste tipo de pronomes (ou anáforas focalizadas) não é específica dos casos de infinitivo flexionado, podendo surgir, generalizadamente, na construção de controlo. Veja-se a frase (ii), que exemplifica a ocorrência de uma anáfora focalizada na construção de controlo de sujeito, que não admite infinitivo flexionado:

(ii) Os meninos_i querem [descascar [eles]_i a fruta].

quando seleccionam uma completiva finita (cf. (47a)), os últimos apenas exibem um complemento, a frase finita (cf. (47b)):

- (47) a. O professor forçou-as/aconselhou-as a que repetissem o teste.
 b. O professor proibiu/impediu que elas repetissem o teste.

Este comportamento dos verbos de proibição aproxima-os da classe dos verbos de elevação de sujeito para objeto, caso em que o sujeito lógico do verbo da frase encaixada é elevado para a frase superior, sendo realizado sintaticamente como objeto direto do verbo dessa frase (ver secção 2.5).

2.4 Contextos previstos pela Generalização D

Os contextos a que se refere esta generalização, aqui repetida, são os seleccionados por verbos diretivos como *dizer* ou *pedir*:

Generalização D

Em contextos de controlo obrigatório de objeto indireto ou em que se verifica opcionalidade entre controlo de objeto indireto ou de sujeito, o infinitivo flexionado é sempre possível nas orações introduzidas por *para*.

Os contextos previstos nesta generalização encontram-se exemplificados em (48) e (49):

- (48) O professor disse aos alunos [para (eles) saírem mais cedo].
- (49) a. A Ana pediu ao Pedro [para (ele) me telefonar].
 b. A Ana e a Rita pediram aos pais [para [-] chegarem depois da meia-noite].
 c. A Ana pediu-me [para [-] irmos juntas ao concerto].

Contrariamente ao verbo *dizer*, que apenas permite que o sujeito da oração infinitiva tenha como antecedente o objeto indireto da frase superior, no caso do verbo *pedir*, o sujeito da oração infinitiva pode ter como antecedente o objeto indireto da frase superior (cf. (49a)), o sujeito (cf. (49b)) ou ter um antecedente repartido em duas posições sintáticas, a de sujeito e a de objeto indireto (cf. (49c)), dependendo a interpretação de fatores pragmáticos. Nestes contextos, o infinitivo flexionado parece ser geralmente possível, sendo que a sua presença pode ser necessária à desambiguação, tornando-se obrigatória em contextos como o de (49c).

2.5 Contextos previstos pela Generalização E

Considerem-se, finalmente, os contextos em causa na Generalização E:

Generalização E

Em orações dependentes de verbos de elevação pode ocorrer o infinitivo flexionado.

Observem-se, em primeiro lugar, orações dependentes de verbos de elevação de sujeito para sujeito. Neste contexto, o infinitivo flexionado é possível em duas situações distintas, que os exemplos (50) e (51) exemplificam:

- (50) a. Parece [estarmos todos cansados].
 b. Parece [ter tocado o telefone quando estavas na cozinha].

(51) Os miúdos parece [~~os miúdos~~] terem apanhado sol demais].

Em (50), deparamos com uma construção típica de verbo monoargumental. O sujeito da oração infinitiva pode ser nulo ou lexical; neste último caso, ocorre em posição pós-verbal, como ilustrado em (50b).

Construção bastante diferente em português europeu é a que (51) exemplifica. Nesta frase, o constituinte interpretado como sujeito da oração infinitiva, *os miúdos*, desencadeia a concordância do complexo verbal infinitivo, embora ocorra no início da frase superior. Contudo, o verbo da frase superior, *parecer*, não exhibe traços de plural, o que mostra que o constituinte *os miúdos* não desempenha a função de sujeito dessa oração, antes ocupando uma posição de tópico marcado, na periferia esquerda da frase. Note-se que, a haver concordância com o verbo *parecer*, a frase seria agramatical, como se mostra em (52):

(52) *Os miúdos parecem [~~os miúdos~~] terem apanhado sol demais].

Também em orações selecionadas por verbos de elevação de sujeito para objeto pode ocorrer infinitivo flexionado. Pertencem a esta classe verbos que admitem uma grande variedade de construções não finitas, como os verbos perceptivos, os causativos e o verbo de permissão *deixar*. Vejam-se os exemplos (53):

- (53) a. A Joana viu [os assaltantes/eles partirem o vidro com um pontapé].
 b. O professor mandou [pintarmos todos (nós) um balão].
 c. Os pais deixaram [elas irem acampar nas férias].

Nestes exemplos, os constituintes com a função de sujeito da oração infinitiva são legitimados internamente ao domínio encaixado, o que a ocorrência das formas nominativas do pronome pessoal comprova (cf. *eles* em (53a), *nós* em (53b), *elas* em (53c)). Com os mesmos verbos, quando é elevado para a frase superior e nela passa a desempenhar a função de objeto direto, o sujeito do complemento infinitivo assume, como esperado, a forma acusativa. Neste caso, ocorre, por norma, o infinitivo não

flexionado, embora o infinitivo flexionado seja também aceite por muitos falantes, como notado em Hornstein/Martins/Nunes (2008) – cf. (54):

- (54) a. ?A Joana viu-os [partirem o vidro com um pontapé].
 b. ?O professor mandou-nos [pintarmos todos um balão].
 c. ?Os pais deixaram-nas [irem acampar nas férias].

O facto de, para alguns falantes, nas estruturas em (54), poderem ocorrer infinitivos flexionados que não legitimam um sujeito nominativo sugere que, neste contexto, tais infinitivos não partilham as propriedades sintáticas características desta forma verbal, tal como já vimos acontecer no caso de infinitivos flexionados em complementos de verbos de controlo de objeto. A flexão pode ser neste caso, mais uma vez, uma mera consequência morfológica do processo sintático envolvido na elevação do sujeito.

3 Infinitivo Preposicionado

Paralelamente às estruturas de infinitivo flexionado descritas na secção anterior, está disponível no português europeu o chamado infinitivo preposicionado (Raposo 1989), ou infinitivo gerundivo (Sten 1952). Trata-se de uma estrutura com a forma *a* + V no infinitivo, com valor semântico próximo do gerúndio, que pode ocorrer como complemento de verbos aspetuais e de verbos percetivos. Em orações dependentes desta última classe de verbos, infinitivo flexionado e não flexionado encontram-se em distribuição livre. Observem-se os exemplos em (55):

- (55) a. Vi os miúdos [a jogar(em) à bola].
 b. Ouvi-te [a tocar(es) violino a noite inteira].

De acordo com a análise de Barbosa/Cochofel (2005), o constituinte selecionado pelos verbos *ver* e *ouvir* é uma oração pequena, com um núcleo aspetual instanciado pela preposição *a*. Assumindo esta análise, em (55a), por exemplo, o constituinte *os miúdos* é o sujeito da oração pequena e o constituinte à sua direita é um domínio aspetual com uma posição vazia de argumento externo do verbo *jogar*. O sujeito da oração pequena (*os miúdos*, em (55a)) necessita de ser legitimado pelo verbo superior, o qual, sendo um verbo de elevação de sujeito para objeto, admite que ele passe a funcionar como seu objeto direto, como mostra a sua realização sob a forma de um pronome acusativo em (55b).

Há outros contextos em que o infinitivo flexionado também pode ocorrer nesta construção em distribuição livre com o não flexionado. Assim, encontra-se em estruturas de predicação secundária, associado ao sujeito (cf. (56)), ao objeto direto (cf. (57)) ou a complementos oblíquos (cf. (58)):

(56) Os pais telefonaram à polícia [a pedirem auxílio].

(57) A mãe encontrou-nos escondidos [a chorarmos].

(58) É preciso pensarmos nas crianças [a passarem fome em muitos países].

Também em adjuntos introduzidos pelas preposições *com* e *sem* pode ocorrer esta construção com infinitivo flexionado, como mostram os exemplos (59):

(59) a. Com todos [a falarem ao mesmo tempo], ninguém conseguiu entender-se.

b. Tudo correu mal, sem mim e a Ana [a planearmos a festa].

Note-se que, neste caso, o sujeito da oração pequena é legitimado casualmente pela preposição, ocorrendo na forma forte oblíqua. Isso mesmo é empiricamente observável em frases como (59b) e (60):

(60) Sem ti [a pilotares o avião], não vou voar.

Finalmente, o infinitivo flexionado pode igualmente ocorrer nesta construção em frases independentes de valor exclamativo, avaliativo ou imperativo, como se pode observar nos seguintes exemplos:¹⁸

(61) a. Ele [a dar-lhe] e a burra [a fugir]!

b. Tu [a dares-me conselhos]? O melhor é ficares calado!

c. Todos [a saírem da piscina já]!

Note-se que, neste último contexto, o sujeito da oração pequena assume a forma nominativa, i.e., a forma casual por defeito em português europeu.

4 Interpretação do sujeito encaixado

A descrição apresentada nas secções anteriores permite concluir que, em muitos dos contextos em que pode ocorrer o infinitivo flexionado, o sujeito da oração infinitiva pode não ter realização lexical, i.e., pode ser a contrapartida nula de um pronome pessoal. Interessa, portanto, determinar em quais desses contextos tal pronome nulo tem necessariamente a sua referência fixada por algum constituinte da frase superior e em quais deles pode ter um referente fixado discursiva ou situacionalmente. Por outras palavras, é necessário determinar se o pronome nulo das orações de infinitivo flexionado se comporta como o das orações de infinitivo não flexionado em contextos denominados de controlo, obrigatório e não obrigatório.

¹⁸ O exemplo (61a) é uma expressão idiomática, ambígua entre uma estrutura de infinitivo flexionado ou não flexionado.

Nas completivas sujeito (cf. (62)), nas frases adjuntas (cf. (63)) e nas relativas finais (cf. (64)), contextos de controlo não obrigatório, o sujeito nulo não necessita de antecedente na frase superior que lhe fixe a referência, como os exemplos abaixo mostram:

- (62) a. Calhou [-] assistirmos à inauguração da exposição.
 b. Agradou ao júri [-] teres escolhido esse tema para a lição.
 c. [-] Estudares comigo ajudou-me a compreender a matéria de álgebra.
- (63) a. Apanhei um autocarro para Belém [depois de [-] te ires embora].
 b. A classe média perdeu poder de compra [por [-] terem implementado políticas severas de austeridade].
 c. Ficas com as crianças [para [-] irmos à festa]?
- (64) As meias [para [-] coseres] estão no cesto da costura.

Mas quando existe na frase superior um constituinte semanticamente adequado e com traços de pessoa e número idênticos aos da forma infinitiva, a interpretação preferencial do sujeito sem realização lexical é a de correferência com tal constituinte:

- (65) a. Calhou-lhes_i [[-]_i assistirem à inauguração da exposição].
 b. Agradou [aos candidatos]_i [[-]_i poderem escolher o tema para a lição].
 c. [[-]_i Estudarem] ajudou-os_i a compreender a matéria de álgebra.
- (66) a. [-]_i Apanharam um autocarro para Belém [depois de [-]_i saírem da biblioteca].
 b. [A classe média e a média-alta]_i perderam poder de compra [por [-]_i terem sido alvo de políticas severas de austeridade].
 c. [-]_i Ficam agora com as crianças [para [-]_i depois irem à festa].
- (67) [-]_i Compraram livros [para [-]_i lerem durante as férias].¹⁹

Com verbos de controlo de sujeito (como em (68a) e (69a)), o sujeito nulo das orações de infinitivo flexionado tem, em geral, uma referência total ou parcialmente disjunta da do sujeito superior, como se pode observar, respetivamente, em (68b), (69b) e (69c). Isso mostra claramente que os contextos em (68b) e (69b,c) não são contextos de controlo.

- (68) a. O deputado_i afirmou [[-]_i ir ser em breve informado da medida legislativa em estudo].
 b. O deputado afirmou [[-] irmos ser em breve informados da medida legislativa em estudo].
- (69) a. Todos_i lamentaram [[-]_i ter saído].
 b. Todos lamentaram [[-]teres saído].
 c. [-]_i Lamento [[-]_{i+j} não poderemos almoçar amanhã].

¹⁹ Na interpretação de relativa final; ver nota 13.

De facto, nestes contextos, o infinitivo flexionado não é a opção preferencial quando existe identidade referencial total entre sujeito superior e sujeito encaixado, preferindo-se, neste caso, a construção de controlo com infinitivo não flexionado, sobretudo se o sujeito superior for de 1.^a ou 2.^a pessoa – compare-se (70) com (71):

- (70) a. (?) [Nós_i confirmámos [-]_i termos de trabalhar no sábado].
 b. (?) [-]_i lamentas [-]_i teres de ir à faculdade nas férias?
- (71) a. [Nós_i confirmámos [-]_i ter de trabalhar no sábado].
 b. [-]_i lamentas [-]_i ter de ir à faculdade nas férias?

Pelo contrário, em complementos de verbos de controlo de objeto, o sujeito nulo da oração infinitiva tem a sua referência fixada, obrigatoriamente, pelo objeto direto (72a) ou indireto (72b) do verbo superior, independentemente de o infinitivo na encaixada ser ou não flexionado:

- (72) a. [Os pais]_i aconselharam [os miúdos]_j [a [-]_{j/i} ter(em) cuidado com o sol].
 b. [A mãe]_i disse-nos_j [para [-]_{j/i} trazer(mos) leite do supermercado].

No caso de verbos como *pedir*, o sujeito da oração de infinitivo flexionado tem obrigatoriamente como antecedente o objeto indireto ou o sujeito do verbo superior ou ambos, i.e., é obrigatoriamente controlado:

- (73) a. [Os pais]_i pediram [aos miúdos]_j [para [-]_j terem cuidado com o sol].
 b. [Os miúdos]_i pediram [aos pais]_j [para [-]_{i/i+j} irem à final da Taça].

Porém, quando o argumento interno de verbos de controlo de objeto direto ou indireto está omissivo, o sujeito nulo da oração de infinitivo flexionado pode ter uma referência independente, definida ou arbitrária (i.e., indeterminada), como se mostra em (74). No entanto, nestas frases, a referência do sujeito nulo da infinitiva é sempre entendida como coincidente com a referência do objeto direto ou indireto implícito na superior.

- (74) a. A situação aconselha [a [-] sermos/ser-se prudente(s)].
 b. O peregrino disse [para [-] termos/se ter cuidado neste troço da caminhada].
 c. Os professores pediram [para [-] manteres/se manter a calma].

Em orações dependentes de verbo de elevação de sujeito para sujeito como *parecer*, o sujeito nulo tem referência independente (cf. (75a)), a não ser que ocorra na frase superior um argumento experienciador (cf. (75b)):

- (75) a. Parece [-] estarmos a precisar de descanso].
 b. Parece-nos_i [-]_i estarmos a precisar de descanso].

Já em orações de infinitivo flexionado dependentes de verbos de elevação de sujeito para objeto, o sujeito nulo é referencialmente independente do sujeito da frase superior, como se mostra em (76):

- (76) a. [A Joana e o Pedro]_i viram [[-]_j/*_i partirem o vidro com um pontapé].
 b. O professor mandou [[-] pintarmos um balão].

Quanto ao sujeito nulo da oração de infinitivo flexionado na construção de infinitivo preposicionado, tem sempre a sua referência estabelecida pelo sujeito da oração pequena (77), podendo dizer-se que o sujeito nulo da infinitiva é controlado pelo sujeito da oração pequena. A interpretação do sujeito nulo da oração de infinitivo flexionado é, neste caso, indistinta da interpretação do sujeito nulo na contrapartida com infinitivo não flexionado (78).

- (77) [-]_i Viram [[os miúdos]_j [a [-]_j/*_i/*_k jogarem à bola]].
 (78) [-]_i Viram [[os miúdos]_j [a [-]_j/*_i/*_k jogar à bola]].

5 O infinitivo flexionado na aquisição de L1

Apesar de o infinitivo flexionado ser uma característica saliente do português europeu e ser frequente na fala espontânea, não havia, até muito recentemente, estudos sobre a sua aquisição. Nesta secção, sintetizamos os principais resultados de estudos recentes, que se desenvolveram em torno de três questões principais: (i) idade e contextos de emergência de infinitivos flexionados no português europeu; (ii) interpretação dos sujeitos nulos de infinitivos flexionados por crianças em idade pré-escolar; (iii) diferenças e semelhanças entre o português europeu e o português brasileiro, reveladas pelos dados da aquisição.

A observação de contextos de emergência de infinitivos flexionados em fala espontânea de crianças baseia-se na análise de um *corpus* longitudinal de fala espontânea de crianças e de fala dirigida a crianças (*corpus* Santos, Santos 2006; Santos et al. 2014), contendo 27.595 enunciados produzidos por crianças (3 crianças: 1;6–3;11, 1;6–3;10, 1;5–2;9) e 70.736 enunciados produzidos por adultos.²⁰ Este trabalho permitiu chegar a algumas generalizações sobre as primeiras ocorrências de infinitivos flexionados na fala espontânea das crianças. Em primeiro lugar, concluiu-se que as crianças são precocemente sensíveis à distribuição dos infinitivos flexionados, produzindo-os muito cedo (com idades próximas dos dois anos) e apenas nos contextos em que são esperados. Em segundo lugar, a distribuição dos infinitivos flexionados nestas idades permite definir estádios com base na emergência de diferentes contextos.

O contexto mais precoce e consistentemente mais frequente de produção de infinitivos flexionados no discurso espontâneo das crianças é o de oração final introduzida por *para* (vejam-se os resultados de Santos et al. 2013, com base numa subparte do *corpus* alargado que aqui consideramos), como exemplificado em (79) ou

²⁰ Para mais informações sobre o *corpus*, veja-se <http://www.clul.ul.pt/en/resources/546>.

(80). Estas orações com valor final incluem casos claros de adjuntas (ver (79b)) ou de relativas finais (ver (80)).

- (79) a. MAE: vão buscar papa ?
 MAE: para quem ?
 TOM: pó [: para o] *u(r)so come(r)*. (1;11.12)
- b. TOM: ponh(o) aí pa(ra) *faze(re)s* [?] # (es)tá? (2;8.9)
- (80) INI: ce(re)ais # pa(ra) *eu comer*. (2;5.24)

Como os dados em (79) e (80) permitem observar, estes primeiros infinitivos flexionados são marcados pela presença de um sujeito (em (79a) e (80)), pela presença de morfologia flexional (em (79b)) ou por ambas. A presença de um sujeito lexicalmente realizado é, como se viu na secção 1, a única forma de identificar um infinitivo flexionado na 1.^a ou na 3.^a pessoa do singular, já que, nestes casos, não existe marcação morfológica explícita na forma verbal. Sendo a 1.^a e a 3.^a pessoas do singular as pessoas mais frequentes nos dados dos primeiros estádios de produção de infinitivos flexionados, isso significa que grande parte dos infinitivos flexionados identificados nestes estádios é assinalada pela presença de um sujeito e não por flexão (vejam-se as tabelas em Santos et al. 2013). Assim, importa perceber se há de facto razões para tomar estes infinitivos como verdadeiros infinitivos flexionados e não como infinitivos pessoais. Em primeiro lugar, sabemos que as formas de 1.^a e 3.^a pessoas do singular emergem mais cedo do que outras formas de pessoa/número, mesmo em outros tempos e modos verbais; por isso, será natural que também no caso dos infinitivos flexionados essas sejam as primeiras formas observadas. Em segundo lugar, ocorrem também nestes dados casos de morfologia observável (de que é exemplo (79b)), sendo que, em todo o *corpus*, apenas se encontrou um caso em que a morfologia é esperada mas não é produzida (uma 3.^a pessoa do plural, assinalada apenas pela presença de um sujeito). Em terceiro lugar, os sujeitos que ocorrem com estes infinitivos têm, quando pronomes, a forma esperada de nominativo, como em (80). Santos et al. (2013) discutem a realização de sujeitos pronominais de 1.^a pessoa do singular, em que é visível o contraste de Caso (*eu* – nominativo; *me* – acusativo/dativo; *mim* – oblíquo), numa subparte inicial deste *corpus*: mostram que ocorrem 18 casos de sujeitos nominativos, que contrastam apenas com dois casos desviantes marcados com oblíquo (não sendo, obviamente, estes últimos tratados como casos de infinitivo flexionado). As crianças tratam, assim, os infinitivos flexionados como capazes de legitimar um sujeito nominativo, como se espera de acordo com a gramática adulta. Há, portanto, razões para dizer que estes primeiros infinitivos com sujeitos nominativos são verdadeiros infinitivos flexionados.

Outra observação relevante sobre este estágio inicial de produção de infinitivos flexionados diz respeito ao momento em que se observa a sua emergência, definido com referência a idade: nesse caso, poderemos dizer que, em duas das três crianças observadas, os infinitivos flexionados emergem perto dos 2 anos (1;11 no caso da criança TOM,

2;1 no caso da criança INI), observando-se na terceira criança (INM) uma emergência um pouco mais tardia da estrutura: 2;9. No entanto, o mais interessante é que os dados longitudinais destas três crianças permitem sugerir uma relação entre a emergência de infinitivos flexionados (em contextos de infinitivas introduzidas por *para*) e a emergência de complementadores (*que*) realizados: TOM e INI produzem o primeiro complementador *que* em completiva aos 2;1 e INM aos 2;6 (veja-se Santos 2006; Santos et al. 2013 para a relação aqui sugerida). Isto significa que (pelo menos alguns contextos de) infinitivos flexionados, que se constituem como domínios de subordinação suficientemente completos e especificados para permitirem a legitimação de sujeitos, surgem, em português europeu, a par de completivas finitas. Emergindo estes primeiros infinitivos flexionados em subordinadas finais (adjuntas ou relativas finais) encabeçadas pelo complementador *para*, a emergência paralela de *para* e do complementador *que*, a ser confirmada por mais dados, estará de acordo com o que se espera se as crianças tratarem efetivamente ambas as formas como complementadores.

Vejamos, agora, o que podemos considerar um segundo estágio de emergência de infinitivos flexionados e que corresponde à ocorrência de infinitivos flexionados em outros contextos. A observação do *corpus* de produção espontânea na base desta análise permitiu identificar numa das crianças, a partir dos 2;9, a emergência de infinitivos flexionados num segundo conjunto de contextos: infinitivos preposicionados independentes ou em complementos de verbos percetivos (em (81a) e (81b), respetivamente) e complementos do verbo *deixar* (em 82a)) ou do verbo percetivo *ver* (em 82b)). Estes são, a par das finais de *para*, que se mantêm como os contextos mais frequentes, os contextos de infinitivo flexionado que se observam nos dados até aos 3;10.

- | | | |
|---------|--|----------|
| (81) a. | TOM: os bonecos todos a fugi(r)em . | (2;9.7) |
| b. | TOM: ainda vi # os [/] os [/] os [/] os senhores a [/] a jogarem # ténis . | (3;0.22) |
| (82) a. | TOM: deixa eu # vi(rar) [/] virar . | (3;0.22) |
| b. | TOM: viste # ela # sa(ir) [/] sair? | (3;1.25) |

A emergência de infinitivos flexionados em subordinadas complemento de verbos percetivos e causativos foi confirmada num estudo experimental, de produção provocada. Santos/Gonçalves/Hyams (2015) aplicaram um teste de completamento, visando a produção de estruturas com verbos causativos, percetivos e de controlo de objeto, a crianças entre os 3 e os 5 anos, e a um grupo de controlo adulto. Os resultados obtidos mostram uma frequência elevada de orações de infinitivo flexionado como complemento de verbos causativos e de permissão (*mandar, deixar*), mesmo no grupo mais jovem, sendo esta estratégia preferida em detrimento de uma estrutura de elevação de sujeito para objeto, ou seja, com infinitivo não flexionado. A par disso, o mesmo estudo mostra uma elevada percentagem de complementos de verbos percetivos (cerca de 50%) que correspondem a construções de infinitivo preposicionado, com infinitivo flexionado, como em (83a), ou não flexionado, como em (83b).

- (83) a. ... viu estes dois cavalos a dançarem. (3;08,23)
 b. O pai tigre viu os filhotes a cair. (3;09,12)

Os resultados deste trabalho experimental são relevantes ainda no que diz respeito à ocorrência de infinitivos flexionados com verbos de controlo de objeto. Vimos na secção 2.3 que um verbo de controlo de objeto pode ter um complemento oracional de infinitivo flexionado, mantendo, ainda assim, uma interpretação controlada do sujeito encaixado (pelo menos, em casos de 3.^a pessoa). Este facto não tem sido suficientemente discutido na literatura. No trabalho experimental que agora referimos, os contextos testados implicavam complementos oracionais cujos sujeitos tinham uma referência de 3.^a pessoa do plural e eram forçosamente interpretados como controlados pelo objeto direto. As respostas dos adultos confirmam a naturalidade do infinitivo flexionado a par do não flexionado nesses contextos, correspondendo cada um dos casos a cerca de 50% das respostas. No caso das crianças, a produção de estruturas adultas com verbos de controlo de objeto é, em geral, mais baixa. No entanto, no grupo dos 5 anos, a produção de complementos oracionais de infinitivo flexionado com verbos de controlo de objeto atinge níveis próximos dos atingidos pelos adultos (veja-se (84)).

- (84) ...ensinou os gatos a nadarem. (5;7,10)

A par destas estruturas-alvo, as crianças, ao contrário dos adultos, produzem ainda infinitivos flexionados em estruturas não esperadas, sob verbos de controlo de objeto. Estas estruturas, de que (85) é um exemplo, são contextos em que a criança parece reinterpretar o verbo de controlo de objeto como um verbo com um único argumento interno, tomando o sintagma nominal que esperávamos que ocorresse como complemento direto do verbo (*os patinhos*) como o sujeito da oração de infinitivo flexionado. Note-se que, no caso do verbo *proibir*, em (85), quando o mesmo seleciona uma completiva finita, esta corresponde ao único argumento interno (ver secção 2.3), como se mostra em (86).

- (85) ... proibiu de os patinhos irem ao pé do crocodilo. (5;01,00)

- (86) Ele proibiu que os patinhos fossem ao pé do crocodilo.

Os dados até agora apresentados sugerem que as crianças são capazes de produzir infinitivos flexionados muito precocemente e que os produzem num conjunto determinado de contextos lícitos, começando por subordinadas introduzidas por *para* e com valor final e, mais tarde, em infinitivos preposicionados, independentes ou selecionados por verbos percetivos, e ainda como complementos não preposicionados de verbos causativos e percetivos. Este conjunto de contextos corresponde a um subconjunto dos casos em que o infinitivo flexionado tem um comportamento padrão, legitimando sujeitos lexicalmente realizados ou sujeitos nulos não controlados.

Contudo, os dados que acabámos de referir revelam que as crianças em idade pré-escolar produzem ainda infinitivos flexionados cujos sujeitos são controlados, nomeadamente como complementos de verbos de controlo de objeto. Este tipo de infinitivos flexionados, que chamaremos de não padrão por oposição aos restantes contextos, é uma possibilidade na gramática adulta (como descrito em 2.3. e em 4). A sua existência na gramática adulta aumenta consideravelmente a ambiguidade do *input*. Terão as crianças em idade pré-escolar adquirido plenamente as propriedades sintáticas e semânticas do infinitivo flexionado nestes dois contextos? Não existem ainda dados que permitam construir uma resposta a esta pergunta. No entanto, Pires/Rothman/Santos (2011) sugerem que, embora as crianças mostrem, em geral, conhecimento adulto das propriedades morfológicas e de distribuição do infinitivo flexionado desde as primeiras idades testadas (6–7 anos), antes dos 9–10 anos não mostram absoluta convergência com a gramática adulta no que diz respeito à interpretação dos sujeitos de infinitivos flexionados. Em particular, o estudo mostra que, quando interpretam enunciados como (87), em que se espera uma interpretação não controlada do sujeito nulo do infinitivo flexionado, as crianças mais jovens podem preferir a interpretação em que o sujeito da superior controla o sujeito da oração encaixada (num teste de escolha de imagem, escolhem a imagem que mostra o Mickey a lavar o carro, em detrimento da imagem que mostra o Donald e a Margarida a lavarem o carro). O tipo de contextos testados neste estudo é restrito, pelo que não sabemos ainda até que ponto e em que contextos as crianças interpretam os sujeitos de infinitivos flexionados em contextos padrão como pronomes, não controlados.

(87) O Mickey ficou satisfeito por lavarem o carro.

O estudo de Pires/Rothman/Santos (2011) tinha, no entanto, um outro objetivo: replicando o estudo de Pires/Rothman (2009) para o português brasileiro, pretendia-se mostrar que o infinitivo flexionado não é necessariamente uma aquisição tardia, pelo menos em português europeu. Pires/Rothman (2009) tinham mostrado que as crianças que adquirem o português brasileiro só mostram conhecimento das propriedades morfológicas, sintáticas e semânticas do infinitivo flexionado a partir dos 10–12 anos. Sabendo que o infinitivo flexionado não é parte das variedades coloquiais do português brasileiro, mas apenas da norma culta, os autores sugeriram que esta aquisição tardia é, na verdade, resultado de exposição à norma em contexto de escolarização. Dado que, em português europeu, o infinitivo flexionado é parte das variedades coloquiais e, portanto, faz parte do *input* a que as crianças estão expostas nos primeiros anos de vida, predizia-se que a aquisição, no português europeu, acontecesse muito mais precocemente. Os dados confirmaram-no.

As variedades coloquiais do português brasileiro incluem não o infinitivo flexionado, mas o que chamámos infinitivo pessoal (secção 2.1), uma estrutura infinitiva sem flexão de pessoa e número e com sujeito lexical marcado com nominativo. Alguns trabalhos, nomeadamente Pires (2002), têm tratado estas formas como sintaticamente

equivalentes a infinitivos flexionados, embora não exibindo contrastes morfológicos com realização fonética. A ausência de flexão nestas formas resultaria da perda generalizada de flexão verbal no português brasileiro. Se o infinitivo pessoal partilha propriedades sintáticas com o infinitivo flexionado e se o infinitivo pessoal está disponível nas variedades coloquiais do português brasileiro, logo, no *input* disponível para as crianças, a questão que se coloca é a de saber se esse infinitivo pessoal é adquirido tão precocemente como o infinitivo flexionado no português europeu. Pretendendo responder a esta questão, Rothman et al. (2013) observaram dados de produção espontânea de três crianças que adquirem o português brasileiro: Ana (2;4.11–2;10.29), A. C. (1;8–3;7) e Gab. (1;10–3;6).²¹ Os resultados obtidos mostram que não só os infinitivos pessoais em português brasileiro emergem em idades comparáveis às da emergência de infinitivo flexionado no português europeu (o infinitivo pessoal mais precoce ocorre aos 2;3), como emergem nos mesmos contextos: subordinadas de valor final introduzidas por *para* (veja-se (88)).

- (88) a. p(r)a [: para] eu co(r)ta(r) a massinha. (Ana - 2;10.29)
 b. pega ela p(r)a [*] eu fazer. (A.C. - 3;0)

6 Síntese

Neste artigo, ocupámo-nos das propriedades do infinitivo flexionado na gramática adulta do português padrão e na aquisição do português como L1.

Relativamente à gramática do adulto, focámos essencialmente a distribuição do infinitivo flexionado em contextos de frases subordinadas, por serem os contextos em que aquela forma verbal é mais frequente, e a interpretação do sujeito da frase infinitiva, em particular nos casos em que o mesmo é foneticamente nulo.

Mostrámos que o infinitivo apresenta a seguinte distribuição:

- (i) Contextos em que o infinitivo flexionado tem uma distribuição livre
 - a) contextos que não envolvem controlo obrigatório do sujeito (i.e., completivas de sujeito, frases adjuntas, relativas finais e completivas de nomes e adjetivos);
 - b) com verbos de elevação de sujeito para objeto, com verbos causativos, percetivos e com o verbos de permissão *deixar*;
 - c) na construção de infinitivo preposicionado.
- (ii) Contextos em que o infinitivo flexionado tem uma distribuição restrita
 - a) contextos de complementação selecionados por verbos de controlo de sujeito: o infinitivo flexionado ocorre apenas em completivas de tempo independente selecionadas por verbos sem orientação temporal especificada;

²¹ Os dados de Ana são do *corpus* de Magalhães (2006), os de A.C. e Gab. foram cedidos por Ruth Lopes.

- b) contextos de controlo de objeto (obrigatório ou preferencial): o infinitivo flexionado tem uma distribuição mais livre do que nos contextos de controlo de sujeito, embora seja preferencialmente usado em orações introduzidas por *para*.
- c) com verbos de elevação de sujeito para sujeito, o infinitivo flexionado pode ocorrer, não se registando elevação.

Relativamente à interpretação do sujeito nulo em orações de infinitivo flexionado, os dados apresentados permitem-nos chegar às seguintes conclusões:

- (i) Contextos em que o sujeito nulo não necessita de um antecedente na oração superior que fixe a sua referência
 - a) completivas com a função de sujeito, frases adjuntas (adverbiais) e frases relativas finais; no entanto, quando, na frase superior, existe um constituinte semanticamente adequado e com traços de pessoa e número idênticos aos da forma infinitiva, a interpretação preferencial é a de correferência com esse constituinte;
 - b) complementos de verbos de elevação;
 - c) complementos de verbos que, quando selecionam uma completiva de infinitivo não flexionado, obrigam a controlo do sujeito.
- (ii) Contextos em que o sujeito nulo tem a sua referência fixada por um constituinte que ocorre na frase
 - a) contextos de controlo de objeto;
 - b) na construção de infinitivo preposicionado.

Relativamente à aquisição do infinitivo flexionado por falantes que têm o português europeu como L1, mostrámos, com base em dados longitudinais de fala espontânea, que o infinitivo flexionado emerge precocemente e nos contextos adequados. Mostrámos, ainda, que os diferentes contextos emergem em estádios diferentes: as crianças começam por produzir infinitivo flexionado em orações subordinadas finais (adjuntas adverbiais e relativas finais), para, num segundo momento, o produzirem na construção de infinitivo preposicionado (independente ou como complemento de verbos percetivos) e em orações de infinitivo não preposicionado dependentes de verbos percetivos ou do verbo *deixar*. Dados de trabalho experimental confirmam os complementos de percetivos e os complementos dos verbos *mandar* e *deixar* como contextos de produção de infinitivos flexionados. Os mesmos dados mostram ainda produção de infinitivos flexionados em complementos de verbos de controlo de objeto, por vezes não conformes à gramática adulta.

Finalmente, a comparação de dados de crianças falantes de português europeu com dados de crianças falantes de português brasileiro mostrou que o infinitivo flexionado emerge em português europeu em idades e contextos comparáveis às idades e aos contextos de emergência do infinitivo pessoal em português brasileiro.

7 Referências

- Ambar, Manuela (1992), *Para uma Sintaxe da Inversão Sujeito-Verbo em Português*, Lisboa, Edições Colibri.
- Ambar, Manuela (2007), *Verb movement and Tense – EPP and T-completeness*, in: M. Cecilia Picchi/Alan Pona (edd.), *Proceedings of the XXXII Incontro di Grammatica Generativa*, Alessandria, Edizioni dell’Orso, 1–20.
- Barbosa, Pilar (2009), *Two kinds of subject «pro»*, *Studia Linguistica* 63, 2–58, DOI: 10.1111/j.1467–9582.2008.01153.x.
- Barbosa, Pilar/Cochofel, Fátima (2005), *O infinitivo preposicionado em português*, in: Inês Duarte/Isabel Leiria (edd.), *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 387–400.
- Barbosa, Pilar/Raposo, Eduardo B. Paiva (2013), *Subordinação argumental infinitiva*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 2, cap. 37, Lisboa, Fundação Calouste Gulbenkian, 1901–1977.
- Bechara, Evanildo (1991), *As fases da língua portuguesa escrita*, in: Dieter Kremer (ed.), *Actes du XVIII Congrès International de Linguistique et de Philologie Romanes*, vol. 3, Tübingen, Niemeyer, 68–76.
- Camões, Luís de (1572), *Os Lusíadas*, Lisboa, disponível em: Biblioteca Nacional Digital, <http://purl.pt/1/3/#/14> (30.09.2015).
- Cunha, Celso/Cintra, Luís Filipe Lindley (1990, 1984), *Nova Gramática do Português Contemporâneo*, Lisboa, João Sá da Costa.
- Declerck, Renaat (1991), *Tense in English: Its structure and use in discourse*, London/New York, Routledge.
- Duarte, Inês/Gonçalves, Anabela/Santos, Ana Lúcia (2012), *Infinitivo flexionado, independência temporal e controlo*, in: Armada Costa/Cristina Flores/Nélia Alexandre (edd.), *Textos Seleccionados do XXVII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 217–244.
- Duarte, Inês/Santos, Ana Lúcia/Alexandre, Nélia (2014), *How relative are purpose relative clauses?*, *Probus* 27:2, 237–269, DOI: 10.15.15/probus-2014–0002.
- Gonçalves, Anabela/Cunha, Luís Filipe/Silvano, Purificação (2010), *Interpretação temporal dos domínios infinitivos na construção de reestruturação do Português europeu*, in: Ana Maria Brito et al. (edd.), *Textos seleccionados do XXV Encontro Nacional da Associação Portuguesa de Linguística*, Porto, APL, 435–447.
- Gonçalves, Anabela/Santos, Ana Lúcia/Duarte, Inês (2014), *(Pseudo-)Inflected infinitives and control as Agree*, in: Karen Lahousse/Stefania Marzo (edd.), *Romance Languages and Linguistic Theory 2012*, Amsterdam/Philadelphia, Benjamins, 181–180.
- Herculano, Alexandre (1844), *Eurico o Presbítero*, Biblioteca Digital, Coleção Clássicos da Literatura Portuguesa, Porto Editora, <http://www.portoeditora.pt/assets/classicos/eurico/index.html> (30.09.2015).
- Herculano, Alexandre (2010, 1851), *A Abóbada*, in: *Lendas e Narrativas*, Tomo I, Lisboa, Livraria Bertrand.
- Hornstein, Norbert/Martins, Ana Maria/Nunes, Jairo (2008), *Perception and causative structures in English and European Portuguese: ϕ -feature agreement and the distribution of bare and prepositional infinitives*, *Syntax* 11, 205–229, DOI: 10.1111/j.1467–9612.2008.00105.x.
- Jones, Charles (1991), *Purpose clauses. Syntax, thematics and semantics of English purpose constructions*, Dordrecht, Kluwer.
- Kamp, Hans/Reyle, Uwe (1993), *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and Discourse Representation Theory*, Dordrecht, Kluwer.
- Landau, Idan (2010), *The explicit syntax of implicit arguments*, *Linguistic Inquiry* 41, 357–388.

- Landau, Idan (2013), *Control in Generative Grammar*, New York, Cambridge University Press.
- Magalhães, Telma V. (2006), *O sistema pronominal sujeito e objeto na aquisição do português europeu e do português brasileiro*, tese de Doutoramento, Universidade Estadual de Campinas.
- Mensching, Guido (2000), *Infinitive constructions with specified subjects. A syntactic analysis of the Romance languages*, Oxford, Oxford University Press.
- Modesto, Marcello (2010), *What Brazilian Portuguese says about control: Remarks on Boeckx & Hornstein*, *Syntax* 13, 78–96, DOI: 10.1111/j.1467-9612.2009.00136.x.
- Pires, Acrísio (2002), *Cue-based change: Inflection and subjects in the history of Portuguese infinitives*, in: David Lightfoot (ed.), *Syntactic effects of morphological change*, Oxford, Oxford University Press, 143–159.
- Pires, Acrísio/Rothman, Jason (2009), *Acquisition of Brazilian Portuguese in late childhood: implications for syntactic theory and language change*, in: Acrísio Pires/Jason Rothman (edd.), *Minimalist Inquiries into Child and Adult Language Acquisition: Case Studies Across Portuguese*, Berlin/Boston, De Gruyter, 129–154.
- Pires, Acrísio/Rothman, Jason/Santos, Ana Lúcia (2011), *L1 acquisition across Portuguese dialects: Modular and interdisciplinary interfaces as sources of explanation*, *Lingua* 121:4, 605–622, DOI: 10.1016/j.lingua.2010.06.002.
- Pountain, Chris (1995), *Romance infinitives with overt subjects: a pragmatic approach*, in: Tom F. Earle/Nigel Griffin (edd.), *Portuguese, Brazilian and African studies: Studies presented to Clive Williams on his retirement*, Warminster, Aris & Philips, 11–25.
- Raposo, Eduardo (1987), *Case Theory and Infl-to-Comp: The inflected infinitive in European Portuguese*, *Linguistic Inquiry* 18, 85–109.
- Raposo, Eduardo (1989), *Prepositional infinitival constructions in European Portuguese*, in: Osvaldo Jaeggli/Kenneth Safir (edd.), *The Null Subject Parameter*, Dordrecht, Kluwer, 277–305.
- Rothman, Jason, et al. (2013), *How early after all?: Inflected infinitives in European and Brazilian Portuguese L1 production*, in: Stavroula Stavrakaki/Marina Lalioti/Polyxeni Konstantinopoulou (edd.), *Advances in Language Acquisition*, Newcastle upon Tyne, Cambridge Scholars Publishing, 164–172.
- Santos, Ana Lúcia (2006), *Minimal Answers. Ellipsis, Syntax and Discourse in the acquisition of European Portuguese*, tese de Doutoramento, Universidade de Lisboa [publicado em 2009, Amsterdam/Philadelphia, Benjamins].
- Santos, Ana Lúcia/Gonçalves, Anabela/Hyams, Nina (2015), *Aspects of the acquisition of control and ECM-type verbs in European Portuguese*, *Language Acquisition* 22, 1–35, DOI:10.1080/10489223.2015.1067320.
- Santos, Ana Lúcia, et al. (2013), *Early or late acquisition of inflected infinitives in European Portuguese?: Evidence from spontaneous production data*, in: Misha Becker/John Grinstead/Jason Rothman (edd.), *Generative Linguistics and Acquisition: Studies in Honor of Nina M. Hyams*, Amsterdam/Philadelphia, Benjamins, 65–88.
- Santos, Ana Lúcia, et al. (2014), *A corpus of European Portuguese child and child-directed speech*, in: *Proceedings of the 9th Conference on Language Resources and Evaluation – LREC 2014*, Reykjavik, European Language Resources Association (ELRA), 1488–1491.
- Sten, Holger (1952), *L'«infinitivo impessoal» et l'«infinitivo pessoal» en portugais moderne*, *Boletim de Filologia* 13, 83–142.
- Szabolcsi, Anna (2009), *Overt nominative subjects in infinitival complements in Hungarian*, in: Marcel den Dikken/Robert M. Vago (edd.), *Approaches to Hungarian: Papers from the 2007 New York Conference*, Amsterdam/Philadelphia, Benjamins, 251–276.
- Tóth, Ildikó (2000), *Inflected Infinitives in Hungarian*, tese de Doutoramento, Tilburg University.

Maria Lobo

18 O gerúndio flexionado no português dialetal

Abstract: Neste capítulo, descreve-se o funcionamento do gerúndio flexionado, isto é, de uma forma de gerúndio com morfologia de concordância de pessoa, que se encontra em algumas variedades dialetais do português europeu. Depois de se fazer a caracterização desta forma, descreve-se a sua distribuição geográfica – limitada aos dialetos meridionais e insulares e coincidindo com a área em que o gerúndio é mais produtivo. Descreve-se ainda os contextos sintáticos em que o gerúndio flexionado pode surgir (predominantemente o de orações adverbiais), e as suas propriedades sintáticas, levantando hipóteses sobre o papel que desempenha a morfologia de pessoa. Mostra-se que as gerundivas flexionadas se aproximam em determinados aspetos das orações finitas (podem ser introduzidas por conectores típicos de domínios finitos), que a flexão permite identificar independentemente o sujeito nulo e facilita a ordem sujeito-verbo, ainda que não seja condição necessária para que esta ordem de palavras se manifeste.

Keywords: gerúndio flexionado, concordância, finitude, português dialetal

1 Introdução

Entre as formas verbais tradicionalmente classificadas como não finitas, por não apresentarem flexão de tempo, consideram-se habitualmente o infinitivo, o particípio e o gerúndio. Em português, o infinitivo pode ocorrer sem qualquer marca de flexão para além do morfema *-r* de infinitivo que se junta ao tema verbal (1), ou pode, em determinados contextos sintáticos, ter morfologia de pessoa, concordando com o sujeito gramatical (2).¹ A primeira e a terceira pessoas do singular, como acontece com muitos tempos verbais, não têm morfologia visível de pessoa.

- (1) cant-a-r
beb-e-r
dorm-i-r

- (2) (Isto é para...)
eu lav-a-r
tu lav-a-r-es

¹ As formas flexionadas de infinitivo são raras nas línguas naturais, estando registadas, para além do português, no galego e no sardo (Raposo 1987; Longa 1994; Vincent 1998; Scida 2004). Cf. também ↗ 17 O infinitivo flexionado na gramática do adulto e na aquisição de L1.

ele lav-a-r
 nós lav-a-r-mos
 vós lav-a-r-des (vocês lav-a-r-em)
 eles lav-a-r-em

O gerúndio, em contrapartida, na variedade padrão do português, não flexiona em pessoa, mesmo quando tem um sujeito expresso (3). Trata-se de uma forma não finita que se caracteriza pela junção do sufixo *-ndo* ao tema verbal (4).

- (3) Estando eu cheio de febre, a mãe chamou o médico
 Estando tu cheio de febre,...
 Estando ele cheio de febre,...
 Estando nós cheios de febre,...
 Estando vós/vocês cheios de febre ,...
 Estando eles cheios de febre,...

- (4) cant-a-ndo
 beb-e-ndo
 dorm-i-ndo

Em algumas variedades dialetais do português europeu, é possível encontrar formas de gerúndio que flexionam em pessoa, à semelhança do que acontece com o infinitivo (5):

- (5) Eu cantando Nós cantândomos
 Tu cantandos (Vós cantandeis)
 Ele cantando Eles cantandem

É deste fenómeno que trata o presente capítulo. Partindo sobretudo dos dados de um corpus de transcrições de fala espontânea de falantes de diferentes regiões de Portugal continental e insular (CORDIAL-SIN),² mas também de registos dispersos de gerúndio flexionado encontrados em diferentes trabalhos, descreveremos a distribuição geográfica do gerúndio flexionado no português (secção 2), a sua distribuição sintática comparativamente à do gerúndio simples (secção 3), as propriedades que têm as gerundivas flexionadas e o papel que desempenha a morfologia de pessoa (secção 4).

² CORDIAL-SIN – Corpus dialetal para o estudo da sintaxe, desenvolvido no Centro de Linguística da Universidade de Lisboa. URL: <http://www.clul.ul.pt/en/resources/411-cordial-corpus>.

2 História e geografia do gerúndio flexionado

As formas flexionadas de gerúndio, identificadas desde Nunes (1902) e referidas em trabalhos variados de dialetologia portuguesa,³ estão também registadas em variedades do galego (Carballo Calero 1966; 1975; Carré Alvarellos 1967; Longa 1993; 1994; ALGa⁴) e no napolitano antigo (Loporcaro 1986; Vincent 1996; 1998), como referido em Lobo (2008). Trata-se de uma forma de gerúndio a que são adicionados morfemas de pessoa, à semelhança do que acontece com o infinitivo flexionado.

O gerúndio flexionado não está registado em textos do português antigo, sendo provavelmente uma forma inovadora. A favor do seu surgimento tardio na história do português, para além da não atestação em textos antigos, estão vários fatores:

- i) nos poucos casos em que foi registada uma forma de segunda pessoa de plural,⁵ ela assume uma forma analógica – *víndeis* (Martins 1954) – e não a forma que se esperaria encontrar, caso o gerúndio flexionado tivesse surgido em época mais antiga;⁶
- ii) a sua distribuição geográfica corresponde à área das variedades meridionais e insulares do português, conhecida por ser mais inovadora, relativamente a variedades mais setentrionais.

No Mapa 1, estão sinalizadas com um círculo preto as localidades onde se registou gerúndio flexionado no CORDIAL-SIN e com um círculo cinzento as localidades onde há registo de gerúndio flexionado em monografias dialetais.⁷

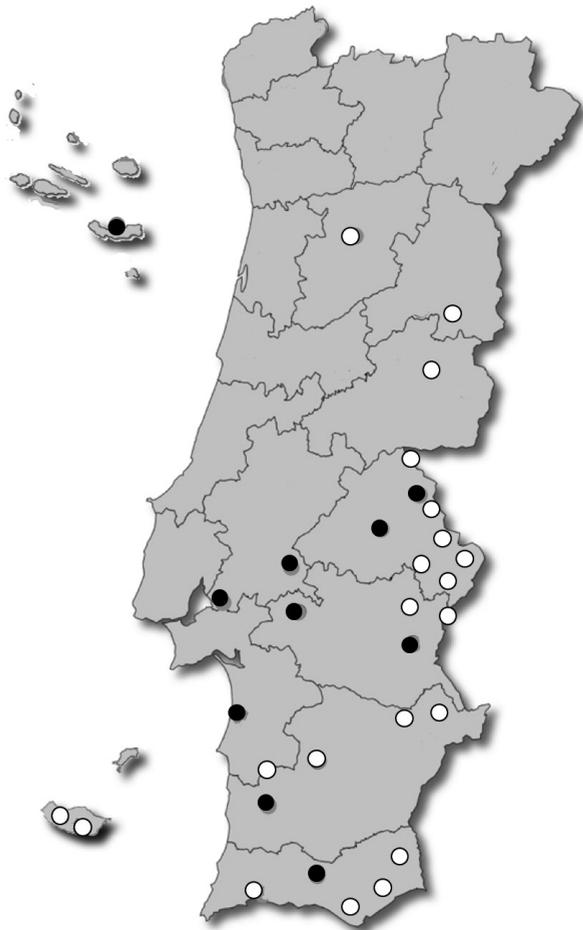
3 Para além das ocorrências registadas em monografias, há dois trabalhos mais recentes que se ocupam do gerúndio flexionado: Mota (1997) e Ribeiro (2002).

4 No ALGa (Atlas Lingüístico Galego), o gerúndio flexionado está registado nos Mapas 27–32 do vol. I.1.

5 Na área onde o gerúndio flexionado é mais produtivo, a segunda pessoa do plural (*vós*) não é usada, sendo substituída pela forma *vocês*, que desencadeia concordância de terceira pessoa do plural.

6 A queda do *-d-* nas terminações verbais de segunda pessoa do plural está datada do século XV (e.g. Cardeira 2005; Brocardo 2014). A terminação *-eis* não é a terminação esperada caso tivesse ocorrido a queda do *-d-* no gerúndio. No galego, em que a queda do *-d-* não ocorreu (confronte-se a forma portuguesa *comeis* com a forma galega *comedes*), está registada a forma *víndodes* (Carballo Calero 1975).

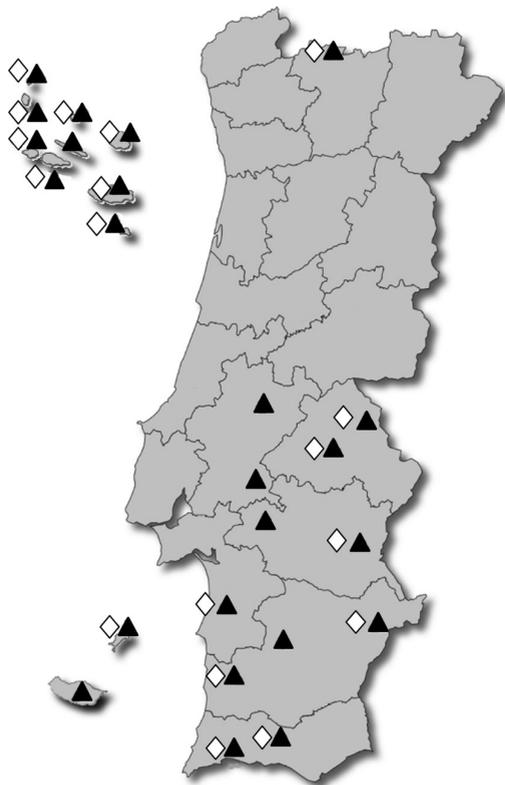
7 O gerúndio flexionado foi registado nas seguintes localidades referidas em monografias dialetais: Ervedosa do Douro (Azevedo 1928); Quadrazais (Braga 1971); Póvoa de Atalaia (Martins 1954); Nisa (Carreiro 1948); Escusa (Baptista 1967); Arronches (Paulino 1959); Campo Maior, Elvas e Alandroal (Matias 1974); Beja (Delgado 1951); Baleizão (Delgado 1970); Colos/Formalhas (Guerreiro 1968); Lagos (Nunes 1902); Olhão (Palma 1967); Monte Gordo (Ratinho 1959); Odeleite (Segura da Cruz 1969); Cedillo (Vilhena 1965); Olivença (Matias 1974); Canhas e Câmara de Lobos (Rezende 1961).



Mapa 1: Distribuição geográfica do gerúndio flexionado (adaptado de Lobo 2008).

Por que razão se desenvolve o gerúndio flexionado apenas em determinadas zonas?

Se analisarmos a localização geográfica das variedades em que surge o gerúndio flexionado, verificamos que corresponde *grosso modo* à das variedades em que o gerúndio se mantém mais produtivo na língua. A área do gerúndio flexionado coincide aproximadamente com a área em que se mantém o gerúndio na construção progressiva com *estar* (cf. Carrilho/Pereira 2011) e em estruturas predicativas (cf. Pereira 2014). Assim, o facto de o gerúndio flexionado estar registado no galego, mas não nas variedades portuguesas setentrionais, não é de estranhar, uma vez que o galego mantém a construção progressiva com gerúndio, ao passo que o português setentrional usa a construção com *estar a + infinitivo*. Assim, a manifestação de flexão no gerúndio parece coincidir com a existência de infinitivo flexionado, mas também com a produtividade do uso do gerúndio.



Mapa 2: Distribuição geográfica de *estar* + gerúndio (▲) e de gerúndio predicativo (◇) no CORDIAL-SIN (adaptado de Carrilho/Pereira 2011 e de Pereira 2014).

Verificamos, assim, que as regiões onde surge o gerúndio flexionado coincidem com aquelas em que o gerúndio é mais produtivo e que correspondem à área dos dialetos meridionais e insulares.

3 Distribuição sintática de gerúndio flexionado: contextos de gerúndio flexionado e não flexionado

Nesta secção, veremos quais os contextos sintáticos em que surge o gerúndio flexionado por oposição ao gerúndio simples. Antes de considerarmos os contextos sintáticos, contudo, vejamos quais as pessoas verbais em que se regista concordância no gerúndio no CORDIAL-SIN.

Como vimos acima, as formas de primeira e terceira pessoa do singular do gerúndio não têm morfologia de concordância visível. Estes contextos serão sempre contextos ambíguos entre gerúndio flexionado e não flexionado. Assim, só pode

surgir flexão de pessoa, morfologicamente visível, nas restantes pessoas gramaticais. Contudo, como se pode observar na Tabela 1, nem todas as pessoas gramaticais são igualmente produtivas no corpus. No CORDIAL-SIN, estão atestadas 42 ocorrências de gerúndio flexionado, distribuídas por dez localidades. As ocorrências de segunda pessoa do singular são escassas, o que não é de estranhar, uma vez que as situações de diálogo contempladas não correspondem ao grau de informalidade associado ao «tu». As formas de primeira pessoa do plural também são escassas. A maioria das ocorrências corresponde a formas de terceira pessoa do plural.⁸

Tabela 1: Ocorrências de gerúndio flexionado por pessoa verbal no CORDIAL-SIN.⁹

Localidade	2S	1P	3P	Total
AAL 4	0	0	6	6
ALC 18	0	0	1	1
CBV 12	1	1	4	6
CPT 39	0	1	0	1
LUZ 26	0	0	3	3
LVR 17	0	0	9	9
MLD 35	0	0	2	2
PAL 5	0	0	7	7
PIC 20	2	0	0	2
STJ 29	0	0	5	5
Total	3	2	37	42
			88%	

Nestas localidades, para além das formas de gerúndio flexionado, ocorrem formas sem morfologia de concordância. Das formas sem morfologia visível, 548 são formas de primeira ou terceira pessoa do singular e 132 são formas em que a pessoa gramatical é 2S ou 1P ou 3P (i.e. pessoas nas quais poderia ter surgido flexão). Assim, como se vê, sendo a 1S e a 3S mais frequentes no corpus do que 2S, 1P e 3P, os contextos em que poderia ter surgido flexão tornam-se reduzidos.

⁸ Usamos as seguintes abreviaturas para as diferentes pessoas verbais: 1S – primeira pessoa do singular; 2S – segunda pessoa do singular; 3S – terceira pessoa do singular; 1P – primeira pessoa do plural; 2P – segunda pessoa do plural; 3P – terceira pessoa do plural.

⁹ Na identificação das localidades, usamos as seguintes abreviaturas: AAL – Alto Alentejo; ALC – Alcochete; CBV – Cabeço de Vide; CPT – Carrapatelo; LUZ – Luzianes; LVR – Lavre; MLD – Melides; PAL – Porches e Alte; PIC – Pico; STJ – Santa Justa.

Tabela 2: Ocorrências de gerúndio não flexionado e flexionado nas localidades com gerúndio flexionado do CORDIAL-SIN.

Localidade	Gerúndio simples		Gerúndio Flexionado	Total
	1S, 3S	2S, 1P, 3P	2S, 1P, 3P	
AAL 4	37	4	6	47
ALC 18	13	2	1	16
CBV 12	6	2	6	44
CPT 39	50	11	1	62
LUZ 26	115	44	3	162
LVR 17	35	5	9	49
MLD 35	73	11	2	86
PAL 5	43	10	7	60
PIC 20	120	34	2	156
STJ 29	26	9	5	40
Total	548	132	42	722

Apesar de haver variação entre formas flexionadas e não flexionadas de gerúndio na segunda pessoa do singular e nas pessoas do plural, na realidade, a ocorrência de gerúndio flexionado ou não flexionado não é aleatória, se considerarmos o contexto sintático.

Em que contextos sintáticos pode ocorrer o gerúndio e de que forma é que esses contextos se correlacionam com a presença de gerúndio flexionado? Na variedade padrão do português europeu, o gerúndio pode ocorrer em diferentes contextos sintáticos e pode assumir diferentes valores aspetuais. Os principais contextos em que ocorre o gerúndio (cf. Lobo 2006) são os seguintes:

- i) alguns complexos verbais (com *ir, vir*) (6);
- ii) orações adverbiais (7);
- iii) orações periféricas pospostas (8);
- iv) algumas estruturas predicativas (9);
- v) algumas estruturas imperativas (10).

Contudo, não ocorre habitualmente no progressivo com *estar* + gerúndio (11), sendo antes usada a perífrase com *estar a* + infinitivo. A perífrase com *estar* + gerúndio é característica das variedades meridionais e insulares do português (Carrilho/Pereira 2011), bem como da variedade brasileira.

(6) Eles vão fazendo o que podem.

(7) Estudando um pouco mais, terás melhores resultados.

- (8) Os ladrões fugiram, tendo sido capturados uma semana depois.
- (9) Chegaram a casa cantando alegremente / a cantar alegremente.
- (10) Andando!
- (11) Todos estavam cantando / a cantar.

Também no CORDIAL-SIN, as formas de gerúndio podem ocorrer em diferentes contextos sintáticos. Contudo, como veremos, as formas flexionadas não surgem em todos eles. Encontramos formas de gerúndio em:

- i) complexos verbais com *estar*, *andar*, *ficar* e *começar* (em que a variedade standard usa *a* seguido de infinitivo) (12);
- ii) complexos verbais com *ir* e *vir* (13);
- iii) estruturas predicativas (14);
- iv) orações adverbiais com valores temporais, condicionais, concessivos ou causais (15);
- v) orações parentéticas (16);
- vi) orações comparativas (17);
- vii) orações independentes (18):
- (12) a. Estava limpando o trigo das palhas (LUZ 8)
 b. Eu fiquei gostando muito dela. (CPT 36)
 c. Andava guardando um rebanhito de porcos e uma vez fui ao Vale da Serva. (LUZ 49)
 d. Até levava pilheiras e a gente começava dizendo uns aos outros (...) para fazer um relheiro alto, para fazer (...) uma coisa grande (...) para as mudanças da máquina. (CBV 21)
- (13) a. Iam terrando, por ali acima, cavando. (LUZ 2)
 b. Vínhamos vindo atrás delas. (LVR 11)
- (14) a. O tempo que os homens estavam ali à espera (...) contando aquilo... (LUZ 16)
 b. Chegou um carro branco. Aí pela aldeia acima procurando onde é que morava o Cristiano Cruz Damásio. (STJ 6)
- (15) a. Estandem juntos os dois, lá pensaram eles a fazer o seguinte (AAL 54)
 b. Sendo carro, fica com dois metros e vinte. (AAL 40)
 c. Mesmo pagando o dinheiro, não há quem queira ir fazer. (PAL 11)
 d. E eu, como sendo o mais velho, (...) é que fui sempre o mais escravo. (AAL 35)
- (16) a. Ó Antolino, eu gostava que tu me fizesses a mim – estândomos a gente a falar – o que vocês costumam a fazer aos gaiatos, bater-lhe com as maniotas na mão. (CBV 41)
 b. E então, diziam: «Ah, mas...» dizendo eu: E quem trouxe essas sementes para aqui? (PAL 2)
- (17) a. Mas isto espeta no esterco e devolve aquilo muito mais rápido (...) do que sendo outra coisa qualquer. (CBV 10)

- b. Isto os arrendamentos, já se sabe, que não há como as coisas estandem nas mãos dos donos, está a compreender? (AAL 29)

(18) Isso sendo com qualquer terra, mas assim próprio, próprio propriamente não é com qualquer terra. (LUZ 18)

Como se pode ver na Tabela 3, os contextos mais produtivos são os de complexos verbais (58,6%) e orações adverbiais (32,6%):

Tabela 3: Distribuição da totalidade das formas de gerúndio (flexionado e não flexionado) pelos diferentes contextos sintáticos.

Localidade	Complexo verbal com <i>estar, ficar, começar, andar</i>	Complexo verbal (outros V: <i>ir, vir</i>)	Oração Independente	Estruturas Predicativas	Orações Adverbiais	Parentéticas	Comparativas	Total
AAL	0	17	0	2	26	0	2	47
ALC	0	8	0	0	8	0	0	16
CBV	2	22	1	2	15	1	1	44
CPT	9	19	0	14	20	0	0	62
LUZ	15	77	2	20	48	0	0	162
LVR	1	17	0	0	31	0	0	49
MLD	20	35	0	4	27	0	0	86
PAL	9	32	0	2	15	2	0	60
PIC	32	89	0	9	26	0	0	156
STJ	1	18	0	2	19	0	0	40
Total	89	334	3	55	235	3	3	722
	12,3%	46,3%	0,4%	7,6%	32,6%	0,4%	0,4%	

Se considerarmos os diferentes contextos sintáticos, verificamos que a maioria das formas de gerúndio flexionado ocorre em orações adverbiais, com valores temporais, causais ou condicionais, como se pode observar na Tabela 4. Há uma única ocorrência de gerúndio flexionado em complexos verbais.

Tabela 4: Distribuição de gerúndio flexionado por contexto sintático no CORDIAL-SIN.

Localidade	Complexos verbais	Orações adverbiais	Parentéticas	Estruturas predicativas	Comparativas	Total
AAL	0	5	0	0	1	6
ALC	0	1	0	0	0	1
CBV	1	4	1	0	0	6
CPT	0	1	0	0	0	1
LUZ	0	3	0	0	0	3
LVR	0	9	0	0	0	9
MLD	0	2	0	0	0	2
PAL	0	4	1	2	0	7
PIC	0	2	0	0	0	2
STJ	0	5	0	0	0	5
Total	1	36 85,7%	2	2	1	42

Vejam-se alguns exemplos de gerúndios flexionados retirados do corpus, ilustrativos dos diferentes contextos sintáticos – complexos verbais (19); orações adverbiais (20); parentéticas (21); estruturas predicativas (22); comparativas (23):

(19) Estão me chamandem. (CBV 39)

- (20) a. Estandem juntos os dois, lá pensaram eles a fazer o seguinte (AAL 54)
b. Então estandes em pé e a pequena nascendo, ela morre-te! (PIC 27)

(21) Enquanto os outros ouvem as coisas, ou vêem, (...) e de si mesmo não têm inteligência, muitas vezes, para descobrir qualquer coisa – ulgandem-se eles inteligentes! (PAL 5)

(22) E, às vezes, anda aquele rebanho junto, (...) e, depois, às vezes, vão dormir pensandem (...) que são todas fêmeas e há (...) algum macho ali no meio! (PAL 13)

(23) Isto os arrendamentos, já se sabe, que não há como as coisas estandem nas mãos dos donos, está a compreender? (AAL 29)

Comparemos agora os contextos em que ocorre o gerúndio simples com formas verbais em que poderia surgir flexão (segunda do singular, primeira do plural e terceira do plural) com os contextos de gerúndio flexionado:

Tabela 5: Distribuição de gerúndio flexionado (GF) e não flexionado (GS) para as pessoas 2S, 1P e 3P por contexto sintático no CORDIAL-SIN.

Localidade	Complexos verbais		Orações adverbiais		Paren-téticas	Estruturas predicativas		Compa-rativas	Total	
	GF	GS	GF	GS	GF	GF	GS	GF	GF	GS
AAL	0	1	5	3	0	0	0	1	6	4
ALC	0	2	1	0	0	0	0	0	1	2
CBV	1	1	4	1	1	0	0	0	6	2
CPT	0	4	1	5	0	0	2	0	1	11
LUZ	0	30	3	6	0	0	8	0	3	44
LVR	0	5	9	0	0	0	0	0	9	5
MLD	0	9	2	1	0	0	1	0	2	11
PAL	0	9	4	1	1	2	0	0	7	10
PIC	0	29	2	2	0	0	3	0	2	34
STJ	0	8	5	0	0	0	1	0	5	9
Total	1	98	36	19	2	2	15	1	42	132
	99		55		2		17		1	174

Como se pode observar na Tabela 5, em complexos verbais, a maioria das formas não apresenta flexão (98/99, cerca de 99%), havendo uma única ocorrência de gerúndio flexionado. Pelo contrário, em orações adverbiais, a maioria das formas tem flexão (36/55, 65,5%). Todas as parentéticas e comparativas têm o gerúndio flexionado; e, nas estruturas predicativas, a maioria das formas não tem flexão. Verifica-se, assim, que o gerúndio flexionado surge preferencialmente em contextos que correspondem a domínios oracionais autônomos relativamente à predicação principal. Isto não é de estranhar, já que o mesmo tipo de condicionamento se encontra relativamente às formas não finitas flexionadas do napolitano antigo (Vincent 1998) e relativamente ao infinitivo flexionado do português padrão.

Olhando com mais cuidado para as orações adverbiais gerundivas, verificamos que os contextos em que não surge flexão (excluindo as formas que são inequivocamente de 1S e 3S) são muitos deles casos em que também em domínios finitos a forma verbal pode aparecer na terceira pessoa do singular e não concordar com o que seria o seu sujeito. Dos 19 casos indicados na tabela acima, 9 têm verbos predicativos (*ser* ou *estar*) e sujeitos nulos ou pós-verbais (24), e 6 têm verbos inacusativos (*chegar*, *ir*, *vir*, *entrar* e *sair*), quase todos com sujeitos nulos ou pós-verbais (25). Ora, estes são contextos em que, como descrito em Carrilho (2003) e Cardoso/Carrilho/Pereira (2011), nestas variedades, o verbo surge frequentemente na terceira pessoa do singular, podendo considerar-se que tem um sujeito nulo expletivo.

- (24) a. ali quando estando em sacas, se não tivessem carregador, carregava-se para carregador. (MLD 47)
 b. E depois, em estando as grainhas tudo no cimo e que esteja completamente cozido, como a gente lhe chama, a gente prova, não é? (AAL 5)
 c. E quando sendo grandes, são cadelas. (LUZ 44)
- (25) a. Em saíndo de lá dos buracos, saem e trazem asas, e depois caem-lhe. (CPT 34)
 b. e depois quando vindo os nove dias de água, está assim podre. (LUZ 21)
 c. Aquilo tem umas mossazinhas; entrando àquelas mossas para dentro, já não saem (CBV 57)

Restam, pois, 5 casos em que se esperaria encontrar concordância, caso a flexão fosse categórica nas gerundivas adverbiais. Sendo este um fenómeno de sintaxe dialetal não *standard*, é expectável, contudo, que o sistema gramatical apresente alguma instabilidade, havendo variação entre formas flexionadas e não flexionadas, mesmo em contextos em que esperaríamos encontrar o gerúndio flexionado. O que é relevante reter é que há um padrão claro de distribuição sintática do gerúndio flexionado. A flexão não surge aleatoriamente em qualquer contexto. Está antes claramente relacionada com a presença de um domínio oracional autónomo.

4 Propriedades das orações gerundivas flexionadas: que papel desempenha a flexão no gerúndio?

Que especificidade têm as formas de gerúndio flexionado? Trata-se de uma mera marca de concordância ou tem consequências a outros níveis na gramática? Já vimos acima que a morfologia de pessoa no gerúndio está sobretudo associada a domínios oracionais autónomos. Vejamos agora outras propriedades que encontramos nas orações gerundivas das variedades consideradas.

Uma das possíveis consequências da presença de morfologia de pessoa no gerúndio está na caracterização da oração gerundiva enquanto domínio marcado como mais ou menos finito. De facto, a presença de flexão no gerúndio parece aproximar as orações gerundivas de domínios finitos. Por um lado, verificamos que, nas localidades consideradas, surgem orações gerundivas não dependentes, como vimos acima (24). Por outro lado, verificamos que, nestas localidades, para além de conectores adverbiais, como *bem* e *mesmo* (27), de origem preposicional, como *em* (28),¹⁰ ou combinados, como *em bem* (29), podem surgir conectores típicos de domínios finitos, tais como

¹⁰ Há um caso registado com gerúndio numa oração introduzida por *até*. Trata-se, no entanto, de um caso de audição duvidosa, que deixa por determinar se a forma verbal ocorre no gerúndio ou no infinitivo. Por isso, não o consideramos aqui. (Uma vaca, até indo <alt> ir </alt> para o matadouro, não tem lá nome disso. (ALC 22))

quando, onde, como e *caso* (30), a introduzir orações gerundivas adverbiais. A seguir ao conector *em* (66/234), o conector mais frequente é o conector *quando* (47/234). No entanto, enquanto as gerundivas introduzidas por *em* estão atestadas em praticamente todas as localidades, o mesmo não acontece com as gerundivas com *quando*, que se registam apenas em quatro das localidades com gerúndio flexionado, como se pode observar na Tabela 6. As gerundivas adverbiais sem conector (31) são maioritárias e estão atestadas em todas as localidades consideradas.

- (26) a. Isso sendo com qualquer terra, mas assim próprio, próprio propriamente não é com qualquer terra. (LUZ 18)
 b. Eu não sabendo, mas daqueles novos! (CBV 32)
- (27) a. Dantes punha-se a cruz até ao fundo, bem a cruz estando rasa, estava finto. (STJ 43)
 b. E mesmo sendo até um pai com filho, pois quase sempre fazem uma festa. (PIC 16)
- (28) a. Digo eu assim: Em o querendem levar, podem-no levar. (CBV 39)
 b. As vacas, em chegando pouco depois de o sol posto, catrapus, lá para dentro também das cercas. (LVR 11)
- (29) a. A gente em bem apanhando um sapato, enfiávamos lá os pés dentro. (STJ 51)
 b. Mas não sei que raio de ferrão tem o bicho que em bem elas aparecem, os animais, aquilo é o cabo dos trabalhos. (CBV 71)
- (30) a. Quando estando constipadinho, a minha patroa arreda logo comigo. (LUZ 32)
 b. Borregos é pequenos e quando sendem aí mais dum ano (...) já é carneiros. (MLD 52)
 c. Que elas quando começadem a aparecer... Só aparecem naquela coisa, depois so-mem-se; (LVR 33)
 d. Onde a gente morrendo, a gente que cá estamos, os velhos, não deve de aparecer já cá mais nenhuns! (MLD 21)
 e. E eu, como sendo o mais velho, (...) é que fui sempre o mais escravo. (AAL 35)
 f. Ele caso estando o bote também pegado, pois se vai também bota. (PIC 11)
- (31) Dandem trigo ou dandem cevada, (...) já é umas relvas. (LVR 10)

Tabela 6: Conectores que introduzem orações gerundivas adverbiais (flexionadas e não flexionadas) nas localidades com gerúndio flexionado.¹¹

Localidade	0	Em	Em bem	Bem	Quando	Mesmo	Onde	Como	Caso	Até	Total
AAL	11	14	0	0	0	0	0	1	0	0	26
ALC	5	2	0	0	0	0	0	0	0	1	8
CBV	5	5	4	0	0	1	0	0	0	0	15
CPT	1	19	0	0	0	0	0	0	0	0	20
LUZ	21	3	0	0	24	0	0	0	0	0	48
LVR	12	15	1	1	2	0	0	0	0	0	31
MLD	9	0	0	0	17	0	1	0	0	0	27
PAL	4	5	0	0	4	2	0	0	0	0	15
PIC	21	2	0	0	0	1	0	0	1	0	25
STJ	13	1	2	3	0	0	0	0	0	0	19
Total	102	66	7	4	47	4	1	1	1	1	234

O facto de, em muitas destas localidades, encontrarmos gerundivas introduzidas por conectores típicos de domínios finitos, bem como gerundivas independentes, mostra que o gerúndio, nestas variedades, se aproxima das formas finitas. A presença da flexão pode, assim, surgir associada a formas de gerúndio com um funcionamento diferente das da variedade *standard*.

Se pensarmos na finitude não como uma propriedade dicotómica, mas como um contínuo associado a diferentes propriedades (cf. Vincent 1998), as orações com o modo indicativo estarão num extremo da escala, as orações infinitivas não flexionadas no outro extremo e as gerundivas flexionadas estarão num grau intermédio da escala. A consideração da finitude não como uma propriedade binária, mas sim como um contínuo tem sido uma hipótese explorada por vários autores, que geralmente consideram o conjuntivo num grau intermédio da escala:

Indicativo > conjuntivo > infinitivo flexionado > infinitivo não flexionado

O lugar intermédio do conjuntivo é atribuível ao facto de este modo estar mais frequentemente associado a domínios dependentes. Há, contudo, muitos contextos em orações não dependentes do português que exigem o conjuntivo (orações imperativas, orações com determinados advérbios, como *talvez* ou *oxalá*), o que faz com que a questão seja mais complicada do que parece (↗23 O modo conjuntivo). No que diz

¹¹ Destes, há 17 ocorrências de gerundivas flexionadas sem conector, 13 com o conector *em*, 3 com *em bem*, 2 com *quando* e 1 com *mesmo*.

respeito ao gerúndio flexionado, o que parece claro é que esta forma pode estar associada a conectores típicos de domínios finitos, o que não acontece com o gerúndio não flexionado.

4.1 Gerúndio flexionado e identificação de sujeito nulo

Para além da aproximação de domínios finitos, a morfologia de pessoa no gerúndio pode desempenhar um papel na identificação do sujeito nulo. Tal como na variedade padrão do português, as variedades dialetais são sistemas linguísticos que permitem sujeitos nulos (↗21 Sujeitos nulos: gramática do adulto, aquisição de L1 e variação dialetal). Também as orações gerundivas podem ter sujeitos nulos. Estes sujeitos podem ser de diferentes tipos – expletivos (32a), arbitrários (32b) ou referenciais (32c), o que os aproxima dos sujeitos nulos das orações finitas.

- (32) a. Porque (...) a azeitona é uma das coisas que: chove agora, e, em estando bom, (...) já se trabalha. (AAL 30)
 b. Mesmo pagando o dinheiro, não há quem queira ir fazer. (PAL 11)
 c. Quando estando constipadinho, a minha patroa arreda logo comigo. (LUZ 32)

Contudo, enquanto na variedade padrão um sujeito nulo referencial de uma oração gerundiva é preferencialmente interpretado como correferente com o sujeito da oração principal, nas variedades dialetais com gerúndio flexionado, há uma menor restrição quanto à interpretação de sujeitos nulos nestas orações. Concretamente, nas variedades consideradas, a flexão no gerúndio permite identificar de forma independente o sujeito nulo da oração gerundiva. Como se pode verificar, nos exemplos seguintes, a flexão de pessoa no gerúndio facilita a identificação de sujeitos referencialmente disjuntos do sujeito da oração principal:

- (33) a. Então estandes em pé e a pequena nascendo, ela morre-te! (PIC 27)
 b. Em falandem no Albino carpinteiro das carroças, pronto, deixa que toda a gente (...) vai ter com ele. (AAL 51)
 c. Agora, em fazendem o resto das outras é que fica tudo ali junto, já. (AAL 24)
 d. Mas em se separândomos, o senhor pensa numa coisa e eu penso noutra. (CPT 27)

4.2 Gerúndio flexionado e ordem de palavras

Consideremos, finalmente, qual o papel que tem a flexão de pessoa na ordem de palavras em orações gerundivas, em particular na posição do sujeito relativamente ao verbo. Como descrito em vários trabalhos (Brito 1984; Ambar 1988; Barbosa 1995), na variedade padrão do português contemporâneo, o sujeito realizado de uma oração gerundiva tem de ocorrer em posição pós-verbal ou pós-auxiliar:

- (34) a. Estando nós a jantar, tocaram à campainha.
 b. *Nós estando a jantar, tocaram à campainha.

Nas variedades dialetais do português, contudo, é frequente encontrarmos sujeitos pré-verbais em orações gerundivas:

- (35) a. O gado estando habituado àquilo, conforme ouviam o toque do chocalho, do coiso, (...) ia tudo atrás. (CBV 50)
 b. E então hoje a coisa, estarem a pagar, eles tendem as coisas em casa, fazem a toda a hora, quando querem. (AAL 36)

Poder-se-ia pensar que esta diferença de ordem é consequência da presença de flexão visível, como assumido, por exemplo, em Barbosa (2002). Contudo, verificamos que existem variedades sem gerúndio flexionado (fora da zona geográfica do gerúndio flexionado) em que encontramos também sujeitos pré-verbais em orações gerundivas, a par de estruturas com inversão (36) (Fiéis/Lobo 2011a; 2011b):

- (36) a. Agora, em abalando a senhora ainda quero ter uma conversa, que ele falta-me aqui umas peças do tear, quero saber onde elas estão. (MST16)
 b. Agora em eu vindo para baixo, querem ir ver além os molhos [...] como é que são? (MST34)
 c. ele indo, eu acho que não volta. (PVC27)

Também em textos de português antigo, podemos encontrar variação entre ordens sujeito-verbo e verbo-sujeito em orações gerundivas (37)–(38) (Fiéis/Lobo 2011a; 2011b):¹²

- (37) a. Joham Rrodriguez estando no logar, veo sobrelle o concelho (...) (s. 15 CDF)
 b. Achando os abades ou os priores que sseus mōges auĩã feyto algũus erros pero seiã pequenos podemos castigar (1350? PP)
- (38) a. E, estando Elle em nós, recebemos seus dōes. (1504, Cat)
 b. E, elle estando em sua oraçon (...), veeo subitamente sobre elle hũũ grande lume do ceo (...) (s. 14, CGE)

É, assim, difícil estabelecer uma relação clara entre ordem de palavras em gerundivas e presença de flexão de pessoa.

Vejamos então quais os fatores que condicionam as diferentes ordens de palavras nas orações gerundivas adverbiais das variedades dialetais consideradas, incluindo

¹² Há, contudo, diferenças entre as variedades dialetais e o português antigo (Lobo/Fiéis 2015): nas variedades dialetais, independentemente de se tratar de variedades com ou sem gerúndio flexionado, a ordem sujeito-verbo é predominante (cerca de 60%), ao passo que no português antigo a inversão (ordem verbo-sujeito) é largamente predominante (cerca de 90%).

fatores como o tipo de verbo, o tipo de sujeito (pleno ou pronominal) e a presença ou ausência de conector.

Num total de 235 orações gerundivas adverbiais, 156 têm sujeitos nulos e apenas 79 têm sujeito realizado. Destas, como se pode ver na tabela abaixo, há uma preferência pela ocorrência do sujeito em posição pré-verbal (55 têm sujeitos pré-verbais ou pré-conector e 24 têm sujeitos pós-verbais). A presença ou ausência de conector não parece condicionar a posição do sujeito.¹³

Tabela 7: Ordem de palavras em orações gerundivas de acordo com presença/ausência de conector.

Localidade	Su-V		V-Su		Total	
	Sem conector (Su-V)	Com conector (Co-Su-V)	Sujeito pré-conector (Su-Co-V)	Sem conector (V-Su)		Com conector (Co-V-Su)
AAL	3	1	3	1	1	9
ALC	0	0	0	0	0	0
CBV	1	4	0	1	2	8
CPT	1	2	1	0	0	4
LUZ	3	8	1	1	1	14
LVR	1	3	2	4	4	14
MLD	4	0	1	2	0	7
PAL	0	3	4	1	0	8
PIC	5	0	1	3	1	10
STJ	2	1	0	1	1	5
Total	20	22	13	14	10	79

Como se pode ver na tabela abaixo, enquanto a posição pré-verbal pode estar associada a qualquer tipo de verbo, a posição pós-verbal está normalmente associada a sujeitos de verbos inacusativos e predicativos. Os verbos predicativos (como *ser* e *estar*) e os verbos inacusativos (como *chegar*, *ir*), como é sabido, têm um sujeito gramatical que funciona como argumento interno e não como argumento externo, o que explica que a inversão do sujeito seja mais frequente com este tipo de verbos.

¹³ Também nas localidades sem gerúndio flexionado a ordem sujeito-verbo é maioritária, estando mais dependente de fatores como o tipo de verbo do que da presença de conector (cf. Lobo/Fiéis 2015).

Tabela 8: Ordem de palavras em orações gerundivas de acordo com tipo de verbo.

	Su-V	V-Su	Total
Verbos predicativos e inacusativos	29	23	52
Outros verbos (transitivos e intransitivos)	26	1	27
Total	55	24	79

Se considerarmos agora a natureza do sujeito, verificamos que, em posição pré-verbal, ocorrem sujeitos de diferentes tipos – plenos, pronomes pessoais e pronomes demonstrativos –, ao passo que, em posição pós-verbal, ocorrem preferencialmente sujeitos plenos.

Tabela 9: Ordem de palavras em orações gerundivas de acordo com tipo de sujeito (*8 são com *a gente* = ‘nós’; **3 são com *a gente* = ‘nós’).

	Su-V	V-Su	Total
Pronome Pessoal	20	2	22
Pronome Demonstrativo	6	0	6
Sintagma Nominal	29*	22*	51
Total	55	24	79

Assim, de alguma forma, nestas variedades, a posição pós-verbal parece ser mais marcada, estando associada sobretudo a sujeitos que funcionam como argumentos internos e a sujeitos morfofonologicamente mais complexos.

5 Conclusões

Neste capítulo, tratámos de um fenómeno de variação morfossintática nas variedades dialetais do português europeu. Verificámos que:

- i) a presença de flexão de pessoa no gerúndio está circunscrita geograficamente às variedades meridionais e insulares;
- ii) há um padrão claro de distribuição sintática de gerúndio flexionado, que está associado sobretudo a orações adverbiais e que está praticamente ausente de complexos verbais;
- iii) a presença de flexão parece favorecer a ocorrência de conectores característicos de domínios finitos e permitir a identificação de sujeitos nulos de forma mais livre do que na variedade padrão.

6 Referências

- Ambar, Manuela (1988), *Para uma sintaxe da inversão sujeito-verbo em português*, tese de Doutorado, Universidade de Lisboa.
- Barbosa, Pilar (1995), *Null Subjects*, tese de Doutorado, MIT.
- Barbosa, Pilar (2002), *A propriedade do sujeito nulo e o princípio da projeção alargado*, in: Maria Helena Mira Mateus/Clara Nunes Correia (orgs.), *Saberes no Tempo. Homenagem a Maria Henriqueta Costa Campos*, Lisboa, Colibri, 51–71.
- Brito, Ana Maria (1984), *Sobre as noções de sujeito e argumento externo: semelhanças entre a estrutura de Fe e a estrutura de SN em português*, *Boletim de Filologia* 29, 421–478.
- Brocardo, M. Teresa (2014), *Tópicos de História da Língua Portuguesa*, Lisboa, Colibri.
- Carballo Calero, Ricardo (1966), *Gramática elemental del gallego común*, Vigo, Galáxia.
- Carballo Calero, Ricardo (1975), *No galego de Rianxo existe un xerúndio flexional*, *Grial* 50, 497–499 [republicado como *Um exemplo de Gerúndio Flexional*, in: Ricardo Carballo Calero (1981), *Problemas da Língua Galega*, Lisboa, Sá da Costa, 129–133].
- Cardeira, Esperança (2005), *Entre o Português Antigo e o Português Clássico*, Lisboa, Imprensa Nacional – Casa da Moeda.
- Cardoso, Adriana/Carrilho, Ernestina/Pereira, Sandra (2011), *On verbal agreement variation in European Portuguese: syntactic conditions for the 3SG/3PL alternation*, *Diacrítica* 25:1, 137–159.
- Carré Alvarellos, Leandro (1967), *Gramática Gallega*, La Coruña, Ed. Moret.
- Carrilho, Ernestina (2003), *Ainda «a unidade e a diversidade da língua portuguesa»: a sintaxe*, in: Ivo Castro/Inês Duarte (edd.), *Razões e Emoção: Miscelânea de estudos em homenagem a Maria Helena Mira Mateus*, vol. 2, Lisboa, Imprensa Nacional–Casa da Moeda, 19–41.
- Carrilho, Ernestina/Pereira, Sandra (2011), *Sobre a distribuição geográfica de construções sintáticas não-padrão em Português europeu*, in: *Textos Seleccionados do XXVI Encontro Nacional da Associação Portuguesa de Linguística*, CD-ROM, Lisboa, APL, 125–139.
- Cintra, Luís Filipe Lindley (1971), *Nova proposta de classificação dos dialectos galego-portugueses*, *Boletim de Filologia* 22, 81–116 [republicado em Cintra, L.F.L. (1983), *Estudos de Dialectologia Portuguesa*, Lisboa, Sá da Costa Ed.].
- Fiéis, Alexandra/Lobo, Maria (2011a), *Word order variation in Portuguese gerund clauses*, poster apresentado em *ICLaVe 6*, Freiburg.
- Fiéis, Alexandra/Lobo, Maria (2011b), *Word order variation in Portuguese gerund clauses: interaction between syntax and interface properties*, ms., Universidade Nova de Lisboa.
- Lobo, Maria (2006), *Dependências temporais: a sintaxe das orações subordinadas gerundivas do português*, *Veredas* 10:1–2, 59–75.
- Lobo, Maria (2008), *Varição morfo-sintáctica em dialectos do português europeu: o gerúndio flexionado*, *Diacrítica* 22(1), 25–55.
- Lobo, Maria/Fiéis, Alexandra (2015), *Ordem de palavras em orações gerundivas do português*, comunicação apresentada no *III Congresso Internacional de Linguística Histórica. Gallaecia (Simpósio Gramática, Variação e Mudança: Sujeito e Ordem de Palavras)*, Santiago de Compostela.
- Longa, Víctor (1993), *Gerundios Gallegos, Posición de Sujetos y Concordancia*, comunicação apresentada ao *XXIII Simposio de la S.E.L.*, Lérida.
- Longa, Víctor (1994), *The Galician Inflected infinitive and the Theory of UG*, *Catalan Working Papers in Linguistics* 4:1, 23–44.
- Loporcaro, Michele (1986), *L'infinito coniugato nell'Italia centro-meridionale: ipotesi genetica e ricostruzione storica*, *L'Italia dialettale* 49, 173–240.
- Mota, Maria Antonia (1997), *Les traits nombre et personne/nombre en portugais – l'oral dans ses variétés*, in: Mireille Bilger/Karel van den Eynde/Françoise Gadet (edd.), *Analyse linguistique et*

- approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste, Orbis / Supplementa*, tome 10, Leuven/Paris, Peeters, 339–345.
- Pereira, Sílvia A. (2014), *A sintaxe na classificação dos dialetos portugueses*, in: *Textos Seleccionados, XXIX Encontro Nacional da Associação Portuguesa de Linguística*, Porto, APL, 445–464.
- Raposo, Eduardo Paiva (1987), *Case Theory and Infl-to-Comp: The Inflected Infinitive in European Portuguese*, *Linguistic Inquiry* 18, 85–109.
- Ribeiro, Raquel (2002), *As ocorrências da forma de gerúndio na variedade padrão e numa variedade dialectal do Português Europeu*, dissertação de Mestrado, FCSH-Universidade Nova de Lisboa.
- Scida, Emily E. (2004), *The Inflected Infinitive in Romance Languages*, New York, NY, Routledge.
- Vincent, Nigel (1996), *Appunti sulla sintassi dell'infinito coniugato in un testo napoletano del '300*, in: Paola Benincà et al. (edd.), *Italiano e dialetti nel tempo. Saggi di grammatica per G. Lepschy*, Roma, Bulzoni, 387–496.
- Vincent, Nigel (1998), *On the grammar of inflected non-finite forms (with special reference to Old Neapolitan)*, in: Iorn Korzen/Michael Herslund (edd.), *Clause Combining and Text Structure*, Copenhagen, Samfundslitteratur, 135–158.

Fontes dos dados

- ALGa – Francisco Fernández Rei (coord.) (1990), *Atlas Lingüístico Galego*, vol. I.1: *Morfología Verbal*, La Coruña, Fundación «Pedro Barrié de la Maza, Conde de Fenosa».
- Azevedo, Celestino M. S. de (1928), *Linguagem popular de Ervedosa do Douro*, *Revista Lusitana* 27, 86–197.
- Baptista, Cândida da S. C. (1967), *O Falar da Escusa*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Braga, Franklin C. (1971), *Quadrazais. Etnografia e Linguagem*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Carreiro, Maria Eduarda V. (1948), *Monografia Linguística de Nisa*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- CORDIAL-SIN: Corpus Dialectal para o Estudo da Sintaxe/Syntax-oriented Corpus of Portuguese Dialects*, Martins, Ana Maria (coord.) ([2000–] 2010), Lisboa, Centro de Linguística da Universidade de Lisboa, URL: <http://www.clul.ul.pt/en/resources/411-cordial-corpus>.
- Delgado, Manuel Joaquim (¹1951, ²1983), *A Linguagem Popular do Baixo Alentejo e o Dialecto Barranquenho*, Beja, Ed. Assembleia Distrital de Beja.
- Delgado, Maria Carolina Saramaga (1970), *O Falar de Baleizão*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Guerreiro, António Machado (1968), *Colos (Alentejo) – Elementos Monográficos*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Martins, Maria José Dias (1954), *Etnografia, Linguagem e Folclore de uma pequena região da Beira Baixa (Póvoa de Atalaia, Alcongosta, Tinalhas e Sobral do Campo)*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Matias, Maria de Fátima de Resende F. (1974), *Bilinguismo e níveis sociolinguísticos numa região luso-espanhola (Concelhos do Alandroal, Campo Maior, Elvas e Olivença)*, dissertação de Licenciatura, Coimbra, Universidade de Coimbra [publicado em 1984 como Separata da Revista Portuguesa de Filologia 18 e 19].
- Nunes, José Joaquim (1902), *Dialectos Algarvios*, *Revista Lusitana* 7, 33–55.
- Palma, Branca Marília Seixal (1967), *O Falar dos Pescadores de Olhão*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.

- Paulino, Maria de Lurdes Semedo (1959), *Arronches – estudo de linguagem e etnografia*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Ratinho, Maria Filipe Mariano (1959), *Monte Gordo. Estudo Etnográfico e Linguístico*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Rezende, Maria Ângela Leotte (1961), *Canhas e Câmara de Lobos. Estudo monográfico e linguístico*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Segura da Cruz, Luísa (1969), *O Falar de Odeleite*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.
- Vilhena, Maria da Conceição (1965), *Falares de Herrera e Cedillo*, dissertação de Licenciatura, Faculdade de Letras da Universidade de Lisboa.

Madalena Colaço

19 Especificidades das estruturas de coordenação: padrões de concordância

Abstract: Este capítulo centra-se na descrição dos padrões de concordância em construções que envolvem coordenação em português europeu. Mostrar-se-á que, mediante certas condições estruturais, a língua admite dois padrões de concordância: concordância plena e concordância parcial. Relativamente à concordância sujeito-verbo, os dados mostrarão que, em construções com sujeitos pós-verbais, para além de o verbo poder concordar com valores resultantes de regras de resolução de traços, pode também, sob certas condições estruturais, concordar apenas com o primeiro constituinte nominal. Serão apresentados argumentos que apontam no sentido da biorracionalidade das estruturas que originam esta forma de concordância. No que diz respeito à concordância que se estabelece no interior de constituintes nominais, os dados mostrarão que, enquanto os elementos pós-nominais concordam com valores resultantes de regras de resolução de traços, os elementos pré-nominais manifestam obrigatoriamente concordância parcial, o que decorre de aspetos de natureza linear e estrutural.

Keywords: concordância plena, concordância parcial, ordem de palavras, sujeito pós-verbal, constituinte pré-nominal

1 Os valores dos traços de concordância de expressões nominais coordenadas

Em português europeu, quando um constituinte de natureza nominal possui uma estrutura de coordenação tendo como núcleo a conjunção aditiva *e*, os valores dos traços de concordância ou traços- ϕ – pessoa, número, género – que lhe estão associados decorrem, tipicamente, de uma combinação dos valores que os termos coordenados possuem relativamente a esses mesmos traços. Os resultados desta combinação seguem os seguintes padrões, facto que se torna visível quando o constituinte coordenado entra numa relação de concordância com outra categoria:

- a) Traço de pessoa: a 1.^a pessoa prevalece sobre as restantes; a 2.^a pessoa prevalece sobre a 3.^a.¹

¹ A forma verbal usada em (3), *nasceram* (equivalente, neste contexto, a *nascestes*), justifica-se pelo facto de, no português europeu atual, o pronome pessoal de segunda pessoa do plural *vós* ter um uso circunscrito a contextos de muita formalidade ou, alternativamente, associado a variedades dialetais. Nos restantes contextos, prevalece o uso do pronome *vocês*.

- (1) a. [[Eu]_{1ªPess} e [tu]_{2ªPess}]_{1ªPess} nascemos na mesma cidade.
 b. [[Tu]_{2ªPess} e [eu]_{1ªPess}]_{1ªPess} nascemos na mesma cidade.
- (2) a. [[Eu]_{1ªPess} e [ele]_{3ªPess}]_{1ªPess} nascemos na mesma cidade.
 b. [[Ele]_{3ªPess} e [eu]_{1ªPess}]_{1ªPess} nascemos na mesma cidade.
- (3) a. [[Tu]_{2ªPess} e [ele]_{3ªPess}]_{2ªPess} nasceram (/nascestes) na mesma cidade.
 b. [[Ele]_{3ªPess} e [tu]_{2ªPess}]_{2ªPess} nasceram (/nascestes) na mesma cidade.
- b) Traço de número: a coordenação de dois valores singulares tem como resultado um valor plural; o valor plural prevalece sobre o singular.
- (4) a. [[Eu]_{SG} e [tu]_{SG}]_{PL} nascemos na mesma cidade.
 b. [[Eu]_{SG} e [eles]_{PL}]_{PL} nascemos na mesma cidade.
 c. [[Eles]_{PL} e [eu]_{SG}]_{PL} nascemos na mesma cidade.
- c) Traço de género: o valor masculino prevalece sobre o feminino.
- (5) a. A estante do João tem [[livros]_{Masc} e [revistas]_{Fem}]_{Masc} antigos.
 b. A estante do João tem [[revistas]_{Fem} e [livros]_{Masc}]_{Masc} antigos.

Segundo vários autores (Corbett 1983; 2006; Sag et al. 1985; Aoun/Benmamoun/Sportiche 1994; Munn 1999; Johannessen 1998; entre outros), esta combinação dos valores dos referidos traços resulta da aplicação das chamadas regras de resolução. A intervenção de regras de resolução poderá ter como resultado uma de duas situações: ou existe a prevalência dos valores dos traços de um dos constituintes nominais coordenados (como acontece, em português europeu, por exemplo, relativamente ao traço de pessoa ou de género, veja-se (1)–(3) e (5), respetivamente), ou então verifica-se a «criação» de um novo valor (como acontece com o traço de número, quando, como em (4a), a coordenação de dois constituintes nominais singulares dá origem a um constituinte nominal plural).

2 Concordância plena *versus* concordância parcial

Em línguas como o português europeu, os fenómenos de concordância não estão, normalmente, dependentes de questões de ordem linear. No entanto, em construções que envolvem a coordenação aditiva de constituintes nominais, a posição relativa dos elementos entre os quais se estabelece a relação de concordância é determinante.²

² Este facto é apontado em diversas gramáticas do português (cf., e.g., Epiphanyo da Silva Dias ¹1918; Said Ali ¹1931; Bechara ¹1961; e, mais recentemente, Mateus et al. 2003 e Raposo et al. 2013), embora em nenhuma delas seja tratado com profundidade.

Assim, a ocorrência de um sujeito nominal com uma estrutura coordenada numa posição pré-verbal obriga à realização de concordância plena (entendida como resultante da intervenção das referidas regras de resolução). No entanto, quando o sujeito ocorre em posição pós-verbal, o verbo pode concordar apenas com a primeira expressão nominal, o que conduz à ideia de que o português europeu é uma língua com concordância parcial. Também no interior de constituintes nominais em que ocorre uma coordenação de nomes ou expressões nominais, a ocorrência de concordância plena ou de concordância parcial depende igualmente de questões de ordem linear. Estes aspetos serão desenvolvidos nas secções seguintes.

2.1 Concordância sujeito-verbo

Como foi referido atrás, o português europeu dispõe da possibilidade de, na presença de uma estrutura de coordenação aditiva, o verbo entrar numa relação de concordância apenas com a primeira expressão nominal coordenada. No entanto, como foi também já notado, esta possibilidade restringe-se às construções com sujeitos pós-verbais, como se verifica pelo contraste seguinte:

- (6) a. Durante a cerimónia, o presidente e o representante dos alunos discursaram.
- b. *Durante a cerimónia, o presidente e o representante dos alunos discursou.

- (7) a. Durante a cerimónia, discursaram o presidente e o representante dos alunos.
- b. Durante a cerimónia, discursou o presidente e o representante dos alunos.

Estes dados mostram, então, que, enquanto os sujeitos pré-verbais obrigam à realização de concordância plena, os sujeitos pós-verbais admitem, alternativamente, a concordância parcial.

Constata-se, por outro lado, pelo contraste entre (7b) e (8), que a concordância parcial corresponde, em português europeu, a concordância com o primeiro termo coordenado e não a concordância com o constituinte mais próximo.

- (8) *Durante a cerimónia, os representantes dos alunos e o presidente discursou.

A opção por uma destas possibilidades permite distinguir grupos de línguas, com base em duas formas distintas de concordância parcial. Assim, por um lado, existem línguas com concordância com o primeiro termo coordenado (*first conjunct agreement*), em que o controlador da concordância é, sistematicamente, o primeiro termo coordenado, e, por outro lado, línguas com concordância com o termo coordenado mais próximo (*closest conjunct agreement*), em que a concordância é controlada pelo termo coordenado linearmente mais próximo da categoria controlada.

A coexistência de dois padrões de concordância alternativos – concordância plena e concordância parcial – estende-se também à relação que se estabelece entre o

sujeito e um predicador não verbal, em construções que envolvem domínios oracionais não frásicos. Neste caso, poderá entrar em jogo, para além dos traços de pessoa e número, o traço de género, que apresenta o mesmo tipo de variação:

- (9) a. Duas meninas e três meninos ficaram adoentados.
 b. Ficaram adoentadas duas meninas e três meninos.

O fenómeno da concordância parcial envolvendo estruturas de coordenação verifica-se em várias outras línguas, românicas e não só, e tem sido objeto de alguma atenção por parte de vários autores, em trabalhos mais ou menos recentes. Nestes trabalhos, as propostas de análise sugeridas divergem, fundamentalmente, em relação a dois aspetos: (i) o nível estrutural em que ocorre a coordenação, ou seja, as categorias coordenadas (nomeadamente, as opiniões dividem-se entre coordenação de constituintes nominais e coordenação de constituintes oracionais), (ii) os mecanismos que permitem a realização das duas formas de concordância. Estes serão, pois, os dois aspetos cruciais em que se centrarão as considerações que se seguem.

2.1.1 Coordenação de constituintes nominais *versus* coordenação de constituintes oracionais

As descrições efetuadas pelos diferentes autores conduzem à conclusão de que as construções em que se verifica a concordância parcial entre o sujeito e o verbo divergem, nas diferentes línguas, numa questão fundamental que se prende com a configuração estrutural que permite a sua realização.

Assim, ou se considera – como em Munn (1996; 1999), van Koppen (2005; 2008), entre outros – que ambas as formas de concordância, plena e parcial, são possíveis em construções em que está presente uma coordenação de constituintes nominais, ou então, sugere-se – como em Aoun/Benmamoun/Sportiche (1994), por exemplo – que a concordância se realiza de forma plena quando existe uma coordenação de constituintes nominais, mas a concordância parcial só se torna possível quando os constituintes coordenados são de natureza oracional. Na verdade, assumir esta segunda perspetiva significa rejeitar a hipótese de a concordância se estabelecer efetivamente de forma parcial, ou seja, apenas com um dos termos coordenados, já que, neste caso, a concordância se estabelece, de forma regular, no interior de cada oração coordenada. Numa hipótese mais conciliadora, autores como Johannessen (1996; 1998) consideram, dependendo das línguas, que a concordância parcial pode existir verdadeiramente, em estruturas de coordenação nominal, ou ser apenas aparente, em estruturas de coordenação oracional.

A discussão em torno da identificação da natureza categorial dos termos coordenados e da relação deste aspeto com o nível em que ocorre a coordenação tem sido relacionada com o modo como as expressões nominais envolvidas nestas duas formas de concordância são interpretadas. Concretamente, a distinção estabelece-se entre

uma interpretação plural e uma interpretação não plural. Na verdade, nas línguas em que existe verdadeiramente concordância parcial entre o sujeito e o verbo, esta coexiste com uma interpretação plural do sujeito (veja-se, por exemplo, os dados apresentados em Johannessen 1998 e Sadler 1999; 2001). Este não parece, no entanto, ser o caso do português europeu. Com efeito, o que verificaremos em seguida é que, nos casos em que é possível a concordância entre o verbo e apenas a primeira de uma sequência de expressões nominais, o sujeito não denota pluralidade semântica. Tendo em conta que a conjunção copulativa com valor aditivo atua como um operador de formação de conjuntos, a entidade denotada por uma expressão nominal coordenada será uma entidade plural, o que fará com que essa expressão possa ser alvo de uma interpretação grupal, para além de uma interpretação distributiva. Logo, a inexistência de pluralidade semântica – e, conseqüentemente, a impossibilidade de uma interpretação grupal – será um indício da inexistência de coordenação de expressões nominais. Este facto permitir-nos-á argumentar em favor da ideia de que, em português europeu – tal como acontece em algumas outras línguas –, não existe verdadeiramente concordância parcial entre o sujeito e o verbo.

Os dados que apresentaremos em seguida mostrarão, então, que a sequência de expressões nominais que ocorre nas construções em análise só conduz à formação de uma pluralidade semântica quando a concordância com o verbo se realiza de uma forma plena e não quando o verbo concorda apenas com a primeira dessas expressões. Para sustentarmos esta ideia, usaremos alguns testes que foram utilizados em trabalhos de outros autores para outras línguas (como, por exemplo, Aoun/Benmamoun/Sportiche 1994 e Munn 1999). Por outro lado, apresentaremos evidências que nos permitirão fundamentar a ideia de que, em português europeu, as construções que aparentam manifestar concordância parcial possuem, na verdade, uma estrutura bioracional.

Para começar, mostraremos que a sequência formada pelas expressões nominais não conduz à formação de pluralidade, o que contraria a ideia de estarmos perante uma estrutura de coordenação de constituintes nominais. Assim, os testes que apresentamos em (a) a (c) podem ser encarados como formas de diagnosticar a incompatibilidade, na língua em análise, entre a concordância parcial e a pluralidade semântica do sujeito pós-verbal.

(a) Construções com sujeitos obrigatoriamente plurais

Os verbos recíprocos caracterizam-se por poderem ocorrer numa construção em que existem dois argumentos cuja posição pode ser invertida sem que essa inversão afete a interpretação. Paralelamente, admitem uma outra construção com um único argumento – o sujeito –, que deve ser plural. Essa pluralidade pode, eventualmente, resultar da presença de uma estrutura de coordenação copulativa. Veja-se (10a) a (10c):

- (10) a. O Pedro partilha com a Maria o gosto por desportos radicais.
- b. O Pedro e a Maria partilham o gosto por desportos radicais.
- c. Eles partilham o gosto por desportos radicais.

Nesta segunda construção, o sujeito é obrigatoriamente plural, como se verifica pelo confronto entre (10b e c) e (11). No entanto, esta pluralidade pode ser apenas semântica, pelo que o sujeito, nesta construção, poderá não estar associado a traços morfológicos de plural, desde que denote uma pluralidade de entidades, como se verifica em (12).

(11) *O Pedro partilha o gosto por desportos radicais.

(12) O grupo partilha o gosto por desportos radicais.

A observação de frases em que ocorrem os verbos em questão, nomeadamente no que diz respeito a fenómenos relacionados com concordância, permite-nos constatar a impossibilidade de realização de concordância parcial. Com efeito, verificamos que estes verbos apenas podem ocorrer em frases com sujeitos coordenados em que o verbo exhibe marcas morfológicas de plural, ou seja, quando a concordância se realiza de forma plena:

- (13) a. Partilham o gosto pela aventura o Pedro e a Maria.
- b. *Partilha o gosto pela aventura o Pedro e a Maria.

Se assumirmos que a agramaticalidade de (13b) é atribuível ao facto de a entidade com a qual se relaciona o predicador verbal não denotar pluralidade semântica, então o contraste em (13) conduz-nos à conclusão de que o sujeito é semanticamente plural apenas na construção em que a concordância se realiza de forma plena.

(b) Ocorrência de expressões que implicam uma interpretação grupal do sujeito

A possibilidade de ocorrência de expressões como *em conjunto* tomando como antecedente o constituinte com a relação gramatical de sujeito pode ser vista como um indicador da presença de pluralidade semântica do referido constituinte, uma vez que tais expressões têm obrigatoriamente como antecedente uma entidade semanticamente plural, embora a expressão nominal que denota essa entidade possa não estar associada ao valor morfológico plural. Esta ideia é suportada pelos dados que apresentamos em (14) a (16):

- (14) a. O Pedro e a Maria trabalharam em conjunto durante a aula.
- b. Os alunos trabalharam em conjunto durante a aula.

(15) *O Pedro trabalhou em conjunto durante a aula.

(16) A turma trabalhou em conjunto durante a aula.

Expressões como *em conjunto* não podem ocorrer em construções com concordância parcial, mas apenas naquelas em que a concordância é plena. É o que constatamos através dos contrastes presentes em (17):

- (17) a. Trabalharam em conjunto o Pedro e a Maria.
 b. *Trabalhou em conjunto o Pedro e a Maria.

Tendo em conta, como já foi referido, que expressões como *a* que ocorre nos exemplos acima obrigam à presença de um antecedente semanticamente plural, os contrastes que se estabelecem em (17) mostram que, nas construções em que o verbo ocorre no singular, os constituintes nominais envolvidos não se associam na formação de uma expressão complexa que denote uma entidade plural. Confirma-se, pois, uma distinção relevante entre as construções em que a concordância se realiza com o primeiro constituinte e aquelas em que a concordância é plena.

(c) Ocorrência de orações relativas contendo pronomes recíprocos

A impossibilidade da ocorrência, em construções que exibem concordância parcial, de uma oração relativa contendo um pronome recíproco que tome como antecedente uma entidade plural denotada conjuntamente pelas expressões nominais que ocorrem, na frase raiz, em posição pós-verbal tem sido encarada – por exemplo, por Aoun/Benmamoun/Sportiche (1994) – como um argumento para defender que, nas construções em que se verifica este tipo de concordância, a coordenação se verifica, não a nível do constituinte nominal sujeito, mas sim a nível oracional. Em português europeu, contrastes como os presentes em (18) mostram que, tal como acontece noutras línguas, a ocorrência da oração relativa apenas é possível quando a concordância entre o sujeito e o verbo se realiza de forma plena, ou seja, quando o verbo ocorre no plural:

- (18) a. *Chegou o rapaz e a rapariga que se conheceram na internet.
 b. Chegaram o rapaz e a rapariga que se conheceram na internet.

O contraste registado em (18) não poderá ser imputado apenas a uma eventual ausência de pluralidade sintática do antecedente da relativa, dado que o pronome recíproco que nela ocorre apenas parece exigir uma pluralidade em termos semânticos. Veja-se a gramaticalidade de sequências como (19), em que a expressão nominal com a relação gramatical de sujeito, embora denotando uma entidade semanticamente plural, é singular do ponto de vista morfológico:

- (19) Está na festa um casal que se conheceu há precisamente dez anos.

Assim, o conjunto de propriedades assinaladas permite-nos afirmar que, em português europeu, em construções em que, em virtude das propriedades do verbo ou de

outro elemento presente na frase, o sujeito é obrigatoriamente semanticamente plural, a concordância parcial não é possível. Desta forma se demonstra a incompatibilidade que se verifica, nesta língua, entre a concordância parcial e a pluralidade semântica do sujeito. Esta incompatibilidade manifesta-se, da mesma forma, em várias outras línguas. Veja-se, por exemplo, o caso do árabe libanês e do árabe marroquino, em que, segundo a descrição de Aoun/Benmamoun/Sportiche (1999), se verifica o mesmo tipo de restrição. Os dados (20) e (21) mostram que, em construções com verbos que requerem um sujeito semanticamente plural, a concordância parcial não é, também nestas línguas, possível:

- (20) a. *Itaʔa kariim w marwaan.
 encontrar.3SG.MASC Karim e Marwan
 ‘Karim e Marwan encontraram-se.’
- b. Itaʔo kariim w marwaan.
 encontrar.3PL.MASC Karim e Marwan
 ‘Karim e Marwan encontraram-se.’
 (árabe libanês, Aoun/Benmamoun/Sportiche 1999, 670)
- (21) a. *tlaqa ʕomar w karim.
 encontrar.3SG.MASC Omar e Karim
 ‘Omar e Karim encontraram-se.’
- b. tlaqaw ʕomar w karim.
 encontrar.3PL.MASC Omar e Karim
 ‘Omar e Karim encontraram-se.’
 (árabe marroquino, Aoun/Benmamoun/Sportiche 1999, 671)

Em seguida, apresentaremos alguns argumentos em favor da ideia de que, na realidade, as construções com concordância parcial têm, em português europeu, uma estrutura bioracional e não envolvem um sujeito composto por coordenação.

(d) Possibilidade de ocorrência de mais do que um advérbio temporal com escopo oracional

Nas construções correspondentes a frases simples em que a concordância entre o verbo e o sujeito coordenado é plena, não é possível a ocorrência de mais do que um advérbio temporal com escopo oracional, se o valor temporal que exprimem não for compatível. Este facto pode ser ilustrado por exemplos como em (22):

- (22) a. Telefonaram hoje o Pedro e a Maria.
 b. Hoje telefonaram o Pedro e a Maria.
 c. *Telefonaram hoje o Pedro e ontem a Maria.
 d. *Hoje telefonaram o Pedro e ontem a Maria.

No entanto, se a concordância verbal se estabelecer apenas com o primeiro constituinte nominal, torna-se possível a ocorrência de tantos advérbios (ou expressões adverbiais) temporais quantos os constituintes presentes. É o que se verifica em (23) e (24):

- (23) a. Telefonou hoje o Pedro e ontem a Maria.
 b. Hoje telefonou o Pedro e ontem a Maria.
- (24) a. Telefonou hoje o Pedro, ontem a Maria e no domingo o Jorge.
 b. Hoje telefonou o Pedro, ontem a Maria e no domingo o Jorge.

Assumindo que os advérbios ou expressões adverbiais temporais como os utilizados nas frases de cima, que exprimem tempos incompatíveis, têm escopo sobre domínios oracionais, os dados apresentados apontam para a existência de uma distinção estrutural entre as construções em que a concordância é plena e aquelas em que o verbo concorda apenas com o primeiro constituinte nominal, uma vez que apenas as primeiras aparentam conter uma estrutura de coordenação de constituintes nominais. Com efeito, a possibilidade de ocorrência de mais do que um advérbio de escopo oracional exprimindo tempos incompatíveis é um indicador da presença de mais do que um domínio oracional.

(e) Ocorrência do advérbio de denotação predicativa *também*

A ocorrência do advérbio *também* enquanto advérbio de denotação predicativa afetando apenas uma parte da sequência pós-verbal apenas se torna possível quando a concordância é parcial, como podemos observar pelo contraste de gramaticalidade que se verifica em (25):³

- (25) a. Telefonou o Pedro e [a Maria também].
 b. *Telefonaram o Pedro e [a Maria também].

A frase que apresentamos em (26) comprova, inclusivamente, a possibilidade de ocorrerem tantos advérbios quantos constituintes nominais, quando a concordância verbal é realizada de forma parcial:

- (26) Telefonou o Pedro, a Maria também e o Jorge também.

Estes factos reforçam a constatação que temos vindo a fazer de que, quando a concordância é plena, o constituinte sujeito possui, efetivamente, uma estrutura

³ Note-se que a agramaticalidade de (25b) só existe se tomarmos o advérbio como estando relacionado apenas com o segundo constituinte nominal que ocorre em posição pós-verbal.

coordenada, ou seja, existe uma coordenação de constituintes nominais. Esta assunção permite justificar a agramaticalidade de (25b), comparável à de uma sequência em que o sujeito ocorre em posição pré-verbal, como acontece em (27):

(27) *O Pedro e [a Maria também] telefonaram.

Pelo contrário, nos contextos em que a concordância verbal se estabelece apenas com o primeiro constituinte nominal, o facto de ser possível a ocorrência do advérbio *também* a afetar apenas essa expressão pode ser visto como um indício da presença de dois domínios oracionais. Note-se que também para outras línguas que, relativamente a este aspeto, apresentam um comportamento semelhante ao do PE, têm sido apresentados argumentos que fundamentam a hipótese bioracional. Assim, por exemplo, Aoun/Benmamoun/Sportiche (1999), com base em dados do árabe libanês e do árabe marroquino, mostram que, também nestas línguas, a verdadeira concordância parcial não existe, uma vez que as construções em que o verbo concorda com o primeiro de uma sequência de constituintes nominais unidos por uma conjunção aditiva apresentam domínios oracionais distintos.

Em síntese, o facto de, por um lado, se verificar uma incompatibilidade entre a realização da concordância verbal com o primeiro constituinte nominal e uma interpretação plural do sujeito e de, por outro lado, existirem evidências da presença de uma estrutura bioracional conduz-nos à ideia de que as construções em que o verbo concorda apenas com o primeiro constituinte nominal e aquelas em que o verbo concorda com os valores resultantes de regras de resolução de traços são, efetivamente, construções com propriedades distintas, nomeadamente no que concerne ao nível estrutural em que ocorre a coordenação.

2.1.2 Concordância parcial e movimento simultâneo

A diversidade do comportamento das diferentes línguas face ao fenómeno da concordância parcial encontra correspondência na multiplicidade de propostas que tentam explicar este padrão de concordância. A questão levanta-se de forma mais veemente relativamente às línguas em que a concordância parcial coexiste com a concordância plena.

Embora a questão assumia contornos mais amplos quando consideramos as línguas em que a concordância parcial pode coexistir não só com uma interpretação plural do sujeito, mas também com a ocorrência do sujeito em posição pré-verbal, referimo-nos, neste caso, concretamente à situação que se verifica em português europeu, língua em que a concordância do verbo com o primeiro constituinte coordenado se restringe, como vimos, às construções com sujeitos pós-verbais. A questão de que partimos é, então, a seguinte: por que razão apenas a posição pós-verbal do sujeito permite esta forma de concordância? Ou, por outras palavras: por que razão

apenas as construções com sujeitos pós-verbais permitem obter uma estrutura que, sendo bioracional, aparenta corresponder a uma situação de concordância com o primeiro termo coordenado?

Nas propostas dos vários autores, os mecanismos propostos para explicar a concordância parcial são variáveis. Aoun/Benmamoun/Sportiche (1999), por exemplo, partindo de dados de línguas que se comportam de uma forma semelhante ao português europeu, defendem a existência de uma estrutura bioracional em que o efeito de concordância parcial é obtido por uma conjugação de duas operações da gramática: elevação do nó direito, que se verifica quando existe a partilha de um argumento interno por dois núcleos verbais, e movimento simultâneo, de que resulta a colocação de um único verbo a preceder os elementos coordenados. Numa perspetiva diferente, Munn (1999) defende a ideia de que a concordância parcial é compatível com uma estrutura de coordenação de constituintes nominais, sugerindo que aquilo que distingue esta forma de concordância da concordância plena é o mecanismo de que resulta a partilha de valores dos traços relevantes, em cada um dos casos. Assim, o autor estabelece a distinção entre dois tipos de relações estruturais – no quadro teórico adotado, a relação especificador–núcleo e a relação de regência – para dar conta desta diferença. Mais recentemente, van Koppen (2005; 2008), com base em dados de línguas em que a concordância parcial coexiste com a pluralidade de sujeitos pós-verbais (nomeadamente, o norueguês), sugere que aquilo que torna simultaneamente possível a concordância plena e concordância parcial é uma configuração estrutural em que existem dois alvos equidistantes de uma sonda, o que se verifica quando o verbo está em T e o sujeito não se moveu para fora de VP, ou seja, quando o sujeito é pós-verbal. Assim, quando o DP complexo permanece no interior de VP, obtendo-se a ordem verbo–sujeito, tanto o constituinte composto como o primeiro termo coordenado estão igualmente acessíveis, pelo que a concordância poderá ser, respetivamente, plena ou parcial. Outros autores, como Walkow (2013) e Bhatt/Walkow (2013), explicam a concordância parcial com sujeitos pós-verbais a partir da ideia de que, nas línguas que admitem este tipo de concordância, o traço EPP e os traços- ϕ podem ser legitimados de forma independente, podendo estes últimos ser validados através da relação *agree* entre T e o primeiro termo coordenado, estando este numa posição interna a VP.

Em Colaço (2005), é sugerida uma hipótese de análise com base nos dados do português europeu que parte da ideia de que a concordância do verbo com o primeiro constituinte nominal se verifica numa construção que aparenta ser de concordância parcial, mas que, na realidade, apresenta uma estrutura bioracional com apagamentos na componente fonológica. Nesta proposta, assume-se que o contraste de gramaticalidade que se verifica em (28) decorre de aspetos de natureza estrutural, nomeadamente da relação de c-comando:

- (28) a. Discursou o presidente e o representante dos alunos.
 b. *O presidente discursou e o representante dos alunos.

Concretamente, o verbo deve c-comandar as posições vazias nos termos coordenados, o que pode ser conseguido através do movimento simultâneo do verbo. Para além do verbo, também um eventual argumento interno poderá ser alvo de movimento simultâneo, o que acontece em construções como a ilustrada em (29), que mostra que também o argumento do verbo deve ocorrer numa posição que lhe permita o c-comando das posições vazias nos termos coordenados. Assume-se, assim, que, em (29), o verbo e o argumento, para além de precederem o sujeito, ocorrem numa posição estruturalmente mais alta que lhes permite estabelecer a relação relevante.

- (29) a. Entrou na Faculdade o Pedro e o João.
 b. *Entrou o Pedro e o João na Faculdade.

Da ocorrência do movimento destes constituintes resulta uma estrutura em que, no interior dos termos coordenados, apenas o sujeito é realizado foneticamente, o que conduz a uma sequência linear de dois constituintes nominais – na realidade, os sujeitos de dois domínios oracionais distintos – entre os quais ocorre uma junção coordenativa, aparentando tratar-se de uma construção com um sujeito composto com concordância do verbo com o primeiro termo coordenado.

De forma a explicar de que modo se verifica a partilha de traços- ϕ , propõe-se, em Colaço (2005), que ConjP ocorre num nível inferior a TP, nomeadamente ao nível de vP, e que o sujeito pós-verbal ocorre na posição de especificador de vP. O verbo, em T, c-comanda os dois termos coordenados, como se vê em (30):

- (30) [_{TP} discursou [_{ConjP(vP)} [_{vP} o presidente [-]] [_{Conj'} e [_{vP} o representante dos alunos... [-]]]]]

2.2 Concordância no interior de constituintes nominais

Tal como acontece em várias outras línguas, também em português europeu as relações de concordância que se estabelecem no interior de constituintes nominais que incluem uma estrutura de coordenação apresentam algumas particularidades.

Em Colaço (2005; 2006; 2012), mostrou-se que, em português europeu, a concordância que se estabelece no interior de constituintes nominais coordenados é sensível a questões de linearidade, uma vez que apenas elementos que precedem as expressões nominais coordenadas podem concordar apenas com um dos termos coordenados, o primeiro.

Nos dados que apresentamos em (31) e (32), constata-se que os quantificadores e os adjetivos em posição pré-nominal (na interpretação em que têm escopo sobre todo o constituinte coordenado) concordam obrigatoriamente com o primeiro termo, não sendo possível, neste caso, que a concordância se estabeleça de forma plena. Embora seja de notar o contraste entre as alíneas a. e b. – que mostra uma menor aceitação da construção nos casos em que, estando os constituintes nominais associados a valores de género distintos, o primeiro termo, com o qual se estabelece a concordância, é do

género feminino –, o contraste ente as alíneas b. e c. é evidente, mostrando a impossibilidade de concordância plena.

- (31) a. Todos os livros e as revistas que comprei estão em cima da mesa.
 b. ?Todas as revistas e os livros que comprei estão em cima da mesa.
 c. *Todos as revistas e os livros que comprei estão em cima da mesa.
- (32) a. O jornal tem variados títulos e ilustrações.
 b. ?O jornal tem variadas ilustrações e títulos.
 c. *O jornal tem variados ilustrações e títulos.

No entanto, a concordância parcial não é possível com elementos pós-nominais, que concordam necessariamente de forma plena. Este facto reflete-se nos contrastes presentes em (33) e (34), em que é visível a obrigatoriedade da concordância plena (note-se que a interpretação relevante de (34) é aquela em que o adjetivo modifica os dois nomes):

- (33) a. Os livros e as revistas que comprei estão todos em cima da mesa.
 b. *Os livros e as revistas que comprei estão todas em cima da mesa.
- (34) a. O jornal tem títulos e ilustrações variados.
 b. *O jornal tem títulos e ilustrações variadas.

Trabalhos sobre outras línguas têm mostrado que o fenómeno da concordância parcial no interior de constituintes nominais nem sempre depende da ordem de palavras. Villavicencio/Sadler (2005), por exemplo, contrariando Munn (1996; 1999), atestam a possibilidade de, em português brasileiro, também os adjetivos pós-nominais concordarem parcialmente com o segundo nome coordenado, contrariamente ao que se verifica na variedade europeia da língua. Veja-se (35), em que o adjetivo ocorre numa forma feminina, embora modifique ambos os nomes:

- (35) As maldições se cumpriam no povo e gente hebreia. (Villavicencio/Sadler 2005, 3)

Também Camacho (2003) e Demonte/Pérez-Jiménez (2012) referem esta possibilidade em espanhol, ilustrada em (36) e (37), dados apresentados pelas autoras:⁴

- (36) el modelo y sistema actual
- (37) su escasa flora y fauna acuática

⁴ Nestes dados, não é claro que o adjetivo pós-nominal modifique toda a estrutura coordenada. Os dados seriam mais claros se os nomes envolvidos na coordenação tivessem valores diferentes para o traço de género ou de número. No entanto, as autoras referem-nos como instâncias de concordância parcial.

Assim, tal como sucede relativamente à concordância sujeito–verbo descrita na secção anterior, existem, por um lado, línguas que apenas admitem concordância parcial entre elementos que precedem a estrutura coordenada e o primeiro termo coordenado (*first conjunct agreement*) e, por outro lado, línguas que admitem a concordância, quer de elementos pré-nominais, quer de elementos pós-nominais, com o termo coordenado mais próximo (*closest conjunct agreement*).

A observação dos dados do português europeu conduziu-nos, nos trabalhos anteriores já referidos, à proposta de que, nesta língua, a concordância parcial de elementos pré-nominais com o primeiro termo coordenado decorre da relação *agree* (no sentido de Chomsky 2001) sob uma condição de c-comando assimétrico local. Nesta proposta, a necessidade de c-comando assimétrico foi usada para explicar a impossibilidade de os elementos pré-verbais concordarem com a categoria resultante da coordenação, ConjP, enquanto o requisito de localidade foi usado para explicar a impossibilidade de os referidos elementos concordarem parcialmente com o segundo termo coordenado, visível na agramaticalidade de (38) e (39):

(38) *Todas os livros e as revistas que comprei estão em cima da mesa.

(39) *O jornal tem variadas títulos e ilustrações.

Nesta perspetiva, o contraste observado em português europeu no que diz respeito à forma como se realiza a concordância com elementos que precedem ou seguem a estrutura coordenada está diretamente relacionado com a assimetria da estrutura interna da categoria que resulta da coordenação, uma vez que apenas o termo coordenado que, nesta configuração, ocupa a posição de especificador (ou seja, o primeiro termo) está acessível à operação sintática de que resulta a concordância parcial (este é um aspeto que retomaremos adiante):

(40) [_{ConjP} [XP] [_{Conj} Conj [YP]]]

Em Colaço (2005; 2006), não foram, no entanto, explicados alguns aspetos determinantes para uma compreensão global do fenómeno da concordância parcial em português europeu, nomeadamente:

(i) A interferência do valor do traço de género na aceitabilidade da concordância parcial, que se verifica em (41) (tal como anteriormente, as expressões nominais seguintes devem ser consideradas apenas na interpretação relevante, em que o adjetivo modifica toda a estrutura coordenada):

- (41) a. O jornal tem variados títulos e ilustrações.
b. ?O jornal tem variadas ilustrações e títulos.

Com efeito, existe um grupo de falantes nativos do português europeu que rejeita sistematicamente a interpretação de frases como (41b) como instâncias de concordância parcial, admitindo unicamente a interpretação alternativa em que o adje-

tivo modifica apenas o primeiro termo coordenado. Por outro lado, os falantes que aceitam a interpretação em que o adjetivo modifica os dois termos coordenados reconhecem-lhe, frequentemente, algum grau de estranheza.

- (ii) Dados como os apresentados em (42) mostram que, quando a coordenação envolve constituintes com núcleos nominais singulares, o adjetivo pré-nominal em concordância com o primeiro termo coordenado modifica apenas esse termo, não podendo, neste caso, modificar os dois termos coordenados, o que inviabiliza a concordância parcial. Assim, a interpretação de (42a) e (42b) contrasta com a de (43), em que o adjetivo ocorre em posição pós-nominal, concordando, neste caso, de forma plena:

- (42) a. Vende-se casa com amplo jardim e terraço. (só o jardim é amplo)
 b. Vende-se casa com amplo jardim e terraços. (só o jardim é amplo)
 c. Vende-se casa com amplos jardins e terraço. (só os jardins são amplos)

- (43) Vende-se casa com jardim/jardins e terraço/terraços amplos.
 (só os terraços são amplos ou os jardins e os terraços são amplos)

Acresce ainda que não é possível a ocorrência do adjetivo pré-nominal numa forma plural quando os constituintes coordenados são ambos singulares, o que mostra a impossibilidade de concordância plena:

- (44) *Vende-se casa com amplos jardim e terraço.

A observação destes dados parece sugerir que as condições que legitimam a concordância parcial em género não são necessariamente as mesmas que legitimam a concordância parcial em número. Assim, pelo que foi dito anteriormente, chegamos às seguintes conclusões acerca do português europeu:

- a) Relativamente ao valor de género, o adjetivo pré-nominal concorda obrigatoriamente com o primeiro termo coordenado, embora a construção seja menos aceitável (ou aceitável para um menor número de falantes) quando o primeiro termo tem o género feminino.
- b) No que diz respeito ao valor de número, o adjetivo apenas pode modificar ambos os termos coordenados se ocorrer numa forma plural em posição pré-nominal, sendo ambos os termos coordenados plurais, ou se ocorrer numa forma plural em posição pós-verbal, independentemente do valor de número dos termos coordenados.

Os dados que apresentamos em seguida mostram, no entanto, que certos casos contrariam parcialmente a conclusão enunciada em (b). Com efeito, quando os constituintes coordenados integram nomes abstratos, é possível a concordância parcial com adjetivos pré-nominais sendo os constituintes singulares, continuando a

concordância plena a verificar-se com adjetivos pós-nominais, como se vê, respetivamente, em (45) e (46):

(45) A Ana reagiu com inesperada coragem e determinação.

(46) A Ana reagiu com coragem e determinação inesperadas.

Note-se que o facto de, em (46), o adjetivo pós-verbal ter uma forma plural mostra que ConjP tem o valor plural. A inexistência desta pluralidade em (45) não inviabiliza, no entanto, uma interpretação em que o adjetivo modifica ambos os termos coordenados.

Em Colaço (2012), foi sugerida a ideia de que o contraste que se verifica entre estes dados e os apresentados na secção anterior decorre de aspetos gramaticais mais abrangentes, relacionados com questões de determinação e referência. Com efeito, em português europeu, verifica-se um contraste no que concerne à possibilidade de ocorrência de apenas um determinante no interior de constituintes que integram expressões nominais coordenadas, pois essa possibilidade apenas parece existir quando essas expressões nominais têm como núcleo nomes abstratos. Este contraste decorre, assim, de propriedades semânticas associadas aos nomes envolvidos e está ilustrado em (47) e (48):

- (47) a. Admiro muito a coragem e determinação da Ana.
b. Admiro muito a coragem e a determinação da Ana.

- (48) a. *Admiro muito o jardim e terraço da casa da Ana.
b. Admiro muito o jardim e o terraço da casa da Ana.

Com efeito, em construções que envolvem nomes concretos, a ocorrência de apenas um determinante está limitada aos contextos em que ambos os nomes são plurais, como acontece em (49):

- (49) a. Admiro muito os jardins e terraços da casa da Ana.
b. Admiro muito os jardins e os terraços da casa da Ana.

Na verdade, existem motivos para pensar que, em frases como (49a) e (49b), a coordenação não ocorre no mesmo nível estrutural. Veja-se, por exemplo, que apenas quando ambos os termos coordenados integram determinantes, como em (49b), se torna possível a ocorrência de modificadores:

- (50) a. ??Admiro muito os jardins fascinantes e terraços da casa da Ana.
b. Admiro muito os jardins fascinantes e os terraços da casa da Ana.

- (51) a. ??Admiro muito os jardins e terraços de pedra da casa da Ana.
b. Admiro muito os jardins e os terraços de pedra da casa da Ana.

Corroborar esta ideia o facto de (52) ser gramatical apenas na interpretação em que o modificador adjetival modifica toda a estrutura coordenada. Em (53), representamos, muito simplificadamente, a diferença estrutural em questão:

(52) Admiro muito os jardins e terraços amplos da casa da Ana.

- (53) a. ??Admiro muito os [[jardins] e [terraços amplos]] da casa da Ana.
 b. Admiro muito os [[jardins] e [terraços] amplos] da casa da Ana.

Note-se que é possível estabelecer uma correlação entre o comportamento dos determinantes e o comportamento dos adjetivos pré-nominais que participam nas construções em análise. Com efeito, quando a expressão nominal que integra a coordenação inclui determinantes, a ocorrência de um adjetivo pré-nominal plural a modificar ambos os termos coordenados plurais apenas se torna possível nas construções em que existe apenas um determinante. Assim, enquanto (54a) permite que o adjetivo seja interpretado como modificador dos dois nomes coordenados, (54b) apenas disponibiliza a interpretação em que o adjetivo modifica unicamente o primeiro nome:

- (54) a. Admiro os amplos terraços e varandas da casa da Ana.
 b. Admiro os amplos terraços e as varandas da casa da Ana.

Pelo que foi mostrado anteriormente, assumimos que, em (54a), o determinante e o adjetivo concordam com o primeiro termo coordenado. Neste sentido, uma forma de dar conta dos contrastes observados consiste em pensar que a coordenação ocorre, em cada caso, em níveis estruturais diferentes. Assumimos, então, que, enquanto a ocorrência de dois determinantes implica a presença de uma estrutura de coordenação de DPs, a ocorrência de apenas um determinante se verifica na presença de uma estrutura de coordenação de uma categoria hierarquicamente inferior, nomeadamente NP ou N. De aqui decorre, sendo este um aspeto determinante, que, nas construções com concordância parcial, o constituinte pré-nominal ocupa uma posição externa e estruturalmente superior à categoria que resulta da coordenação.

Seguindo esta linha de pensamento, quando a coordenação ocorre num nível inferior a DP, o determinante ocupa uma posição estrutural externa e superior a ConjP. O mesmo assumiremos que acontece com o adjetivo pré-nominal, se tivermos em conta que, como vimos atrás, os contextos que permitem a ocorrência de um adjetivo pré-nominal a modificar os dois termos coordenados coincidem com os contextos que permitem a ocorrência de um único determinante associado a toda a coordenação.

Note-se que esta ideia permite compreender por que razão um quantificador como *toda(s)/toda(s)* concorda parcialmente com o primeiro termo coordenado, mesmo em construções em que ambas as expressões nominais envolvidas integram determinantes, como se mostrou nos exemplos que retomamos em (55), assumindo que o quantificador ocupa uma posição estrutural superior a DP. A ocorrência do quantifi-

gador numa posição exterior à coordenação conduz à interpretação em que ambos os termos coordenados são por ele afetados (uma interpretação alternativa de (55a) e (55b) em que o quantificador afeta apenas o primeiro termo coordenado resulta da sua ocorrência no interior desse termo).

- (55) a. Todos os livros e as revistas que comprei estão em cima da mesa.
 b. ?Todas as revistas e os livros que comprei estão em cima da mesa.
 c. *Todos as revistas e os livros que comprei estão em cima da mesa.

Em Colaço (2012), foi sugerido que a impossibilidade de os determinantes e os adjetivos pré-nominais concordarem de forma plena com o constituinte coordenado e, ao mesmo tempo, a obrigatoriedade de concordarem com o primeiro termo podem ser explicadas, como já havia sido sugerido em trabalhos anteriores, a partir da existência de uma condição estrutural de c-comando assimétrico. Com efeito, como foi referido atrás, apesar de as categorias pré-nominais estabelecerem a relação *agree* tanto com ConjP como com o primeiro termo coordenado, apenas com o primeiro termo coordenado se verifica uma relação de c-comando assimétrico, sendo esta uma condição para a verificação dos traços em questão.

2.2.1 A reinterpretação da concordância parcial em género como concordância plena

Resta-nos agora retornar a uma das questões que enunciámos atrás, relacionada com o contraste notado no que diz respeito ao traço de género, ilustrada em (55) e também nos dados que retomamos em (56):

- (56) a. O jornal tem variados títulos e ilustrações.
 b. ?O jornal tem variadas ilustrações e títulos.
 c. *O jornal tem variados ilustrações e títulos.

Como referimos anteriormente, alguns falantes nativos do português europeu rejeitam, ou consideram muito estranha, a concordância de um elemento pré-nominal com o primeiro termo coordenado quando este tem como núcleo um nome do género feminino. Para este grupo de falantes, quando o primeiro termo é feminino, a única solução consiste em optar por uma ordem de palavras diferente, de que resulte ou a colocação do termo coordenado masculino em primeiro lugar ou a ocorrência do adjetivo em posição pós-nominal. Note-se que não se trata de uma não aceitação da concordância parcial, uma vez que a concordância plena é unanimemente considerada também inaceitável, como se verifica em (55c) e (56c), que mostram a impossibilidade de um elemento pré-nominal concordar com os traços do constituinte coordenado resultantes da aplicação das regras de resolução quando o primeiro termo é do género feminino.

Este fenómeno poderá ser parcialmente explicado por uma reinterpretação, por este grupo de falantes, da concordância que se verifica nas frases em que o primeiro

termo é masculino – como acontece em (55a) e (56a) – como uma instância de concordância plena, assumindo-se que a forma masculina do elemento pré-nominal corresponde a uma forma resultante de resolução de valores de traços. Essa assunção poderá levar esses falantes a rejeitar frases como (55b) ou (56b), embora não explique, no entanto, a rejeição ainda mais clara de (55c) e (56c), que mostra que o facto de o primeiro termo coordenado e o elemento pré-nominal não partilharem o valor do traço de género inviabiliza determinantemente a sua interpretação como instâncias de concordância plena.

3 Questões para investigação futura

Futuros desenvolvimentos do estudo da concordância no interior de constituintes que envolvem coordenação poderão permitir uma maior compreensão deste fenómeno em português europeu. Tendo em conta que se trata de um aspeto relativamente ao qual os juízos dos falantes nativos divergem, a observação de um conjunto de dados mais alargado poderá revelar novos contornos que, eventualmente, permitam uma análise mais abrangente. Por outro lado, um estudo de *corpora* permitirá, de uma forma mais rigorosa e objetiva, estabelecer os contrastes relevantes, com base em produções reais dos falantes.

Uma questão que julgamos merecer alguma atenção futura diz respeito à relação entre a ordem de palavras e a concordância parcial no interior de constituintes nominais. Temos estado, desde o início, com base nos dados observados, a assumir que o português europeu não permite a concordância parcial de um adjetivo em posição pós-nominal, que, neste caso, concordaria com o segundo termo coordenado. No entanto, apenas uma observação mais atenta dos dados permitirá confirmar este facto. Com efeito, como já referimos, existem várias línguas em que a concordância parcial com adjetivos pós-nominais é possível. O dado que apresentamos em (57) mostra que, em português europeu, este aspeto deve ser descrito com maior rigor. Com efeito, dados como este atestam a produção de expressões em que um adjetivo pós-nominal, afetando uma coordenação de elementos nominais, não concorda com valores resultantes de regras de resolução de traços do constituinte coordenado, mas apenas com os valores do termo mais próximo, o segundo:

(57) Possuem vozes melodiosas e normalmente com um timbre e ritmo agradável.

(http://www.anje.pt/feira/media/super_vendedor.pdf)

Ao mesmo tempo, o dado que apresentamos em (58) ilustra um fenómeno que já referimos atrás, que está atestado, por exemplo, para o espanhol por Camacho (2003) e por Demonte/Pérez-Jiménez (2012), e também para o português brasileiro por Villavicencio/Sadler (2005), que consiste na possibilidade de o adjetivo pós-nominal concordar em género com o segundo termo coordenado:

- (58) Ali surgem povos e civilizações tão antigas como a do Egipto
(<http://www.prof2000.pt/users/filipe/pessoa/religoes/antigas.htm>)

4 Conclusão

O objetivo deste capítulo foi descrever o funcionamento do português europeu no que diz respeito aos padrões de concordância que se estabelecem na presença de estruturas coordenadas. Os dados apresentados permitiram-nos concluir que, nesta língua, a concordância sujeito-verbos é sempre plena, uma vez que as construções com sujeitos pós-verbais que aparentam exibir concordância parcial possuem, na verdade, uma estrutura bioracional e não uma estrutura mono-oracional com um sujeito composto. Pelo contrário, no interior de constituintes nominais, a concordância que se estabelece com elementos pré-nominais é obrigatoriamente parcial. Concluímos com a constatação de que este é um tópico pouco estudado nesta língua e de que uma observação mais alargada de dados empíricos poderá conduzir a uma compreensão mais abrangente deste fenómeno.

5 Referências

- Aoun, Joseph/Benmamoun, Elabbas/Sportiche, Dominique (1994), *Agreement, word order and conjunction in some varieties of Arabic*, *Linguistic Inquiry* 25:2, 195–220.
- Aoun, Joseph/Benmamoun, Elabbas/Sportiche, Dominique (1999), *Further Remarks on First Conjunct Agreement*, *Linguistic Inquiry* 30:4, 669–681.
- Bechara, Evanildo (1999, ¹1961), *Moderna Gramática Portuguesa*, Rio de Janeiro, Editora Lucerna.
- Bhatt, Rajesh/Walkow, Martin (2013), *Locating agreement in grammar: an argument from agreement in conjunctions*, *Natural Language and Linguistic Theory* 3:4, 951–1013.
- Camacho, José (2003), *The Structure of Coordination: Conjunction and Agreement Phenomena in Spanish and other Languages*, Dordrecht, Kluwer.
- Chomsky, Noam (2001), *Derivation by Phase*, in: Ken Hale (ed.), *A Life in Language*, Cambridge MA, The MIT Press, 1–54.
- Colaço, Madalena (2005), *Configurações de Coordenação Aditiva: Tipologia, Concordância e Extração*, tese de Doutoramento, Universidade de Lisboa.
- Colaço, Madalena (2006), *Concordância parcial no interior de constituintes nominais em Português Europeu*, in: Fátima Oliveira/Joaquim Barbosa (edd.), *Textos Seleccionados do XXI Encontro da APL*, Lisboa, APL, 273–284.
- Colaço, Madalena (2012), *Retomando a questão da concordância parcial no interior de constituintes nominais coordenados*, in: Armada Costa/Isabel Falé/Pilar Barbosa (edd.), *Textos Seleccionados do XXVII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Edições Colibri/APL, 153–170.
- Corbett, Greville (1983), *Hierarchies, Targets and Controllers: Agreement patterns in Slavic*, London, Croom Helm.
- Corbett, Greville (2006), *Agreement*, Cambridge, Cambridge University Press.

- Demonte, Violeta/Pérez-Jiménez, Isabel (2012), *Closest conjunct agreement in Spanish DPs. Syntax and beyond*, *Folia Linguistica* 46:1, 21–74.
- Epiphânio da Silva Dias, Augusto (1970, ¹1918), *Syntaxe Historica Portuguesa*, Lisboa, Livraria Clássica Editora.
- Johannessen, Janne B. (1996), *Partial agreement and coordination*, *Linguistic Inquiry* 27:4, 661–676.
- Johannessen, Janne B. (1998), *Coordination*, New York, Oxford University Press.
- van Koppen, Marjo (2005), *One Probe – Two Goals: Aspects on Agreement of Dutch Dialects*, tese de Doutoramento, Leiden University, LOT-dissertations 105.
- van Koppen, Marjo (2008), *Agreement with coordinated subjects. A comparative perspective*, *Linguistic Variation Yearbook* 7, 121–161.
- Mateus, Maria Helena Mira, et al. (orgs.) (2003), *Gramática da Língua Portuguesa*, 5.ª ed. revista e aumentada, Lisboa, Caminho.
- Munn, Alan B. (1993), *Topics in the syntax and semantics of coordinate structures*, tese de Doutoramento, Universidade de Maryland.
- Munn, Alan B. (1996), *First conjunct agreement without Government: a reply to Aoun, Benmamoun and Sportiche*, ms.
- Munn, Alan B. (1999), *First conjunct agreement: Against a clausal analysis*, *Linguistic Inquiry* 30, 643–668.
- Raposo, Eduardo B. Paiva, et al. (orgs.) (2013), *Gramática do Português*, Lisboa, Fundação Calouste Gulbenkian.
- Sadler, Louisa (1999), *Non-distributive Features in Welsh Coordination*, in: *Proceedings of the LFG99 Conference, University of Manchester*, Stanford CA, CSLI, 1–19, <http://web.stanford.edu/group/cslipublications/cslipublications/LFG/4/lfg99sadler.pdf> (06.01.2016).
- Sadler, Louisa (2001), *Agreement Asymmetries and the Structure of Coordination*, handout apresentado em *The 4th NWCL Conference: Coordination: Syntax, Semantics and Pragmatics*, Salford University.
- Sag, Ivan, et al. (1985), *Coordination and How to Distinguish Categories*, *Natural Language and Linguistic Theory* 3, 117–171.
- Said Ali, Manuel (1971, ¹1931), *Gramática Histórica da Língua Portuguesa*, Rio de Janeiro, Livraria Acadêmica.
- Villavicencio, Aline/Sadler, Louisa (2005), *Agreement Patterns in Corpora*, <http://citeseerx.ist.psu.edu/.../download?doi=10> (30.09.2015).
- Walkow, Martin (2013), *When can you agree with a closest conjunct?*, in: *Proceedings of the 31st West Coast Conference on Formal Linguistics*, Cambridge, Cascadilla Press, 474–483.

Anabela Gonçalves, Ernestina Carrilho e Sandra Pereira

20 Predicados complexos numa perspetiva comparativa

Abstract: Neste capítulo apresentam-se aspetos sintático-semânticos caracterizadores dos predicados complexos, entendidos como sequências de dois (ou mais) verbos que participam numa relação de complementação infinitiva e que, em determinados contextos, formam uma unidade sintático-semântica. Numa perspetiva comparativa, com dados de outras línguas e de diferentes variedades de português, são identificadas as características gerais definidoras de predicados complexos (verbais) em português europeu e distinguem-se duas construções, designadas por *reestruturação* e *fazer-Infinitivo*. Para cada uma destas construções, são apresentadas propriedades diferenciadoras e são identificados elementos dos diferentes grupos de verbos que podem ocorrer em predicados complexos.

Keywords: construções infinitivas, predicado complexo sintático, construção de reestruturação, construção *fazer-Inf*

1 Introdução

Em português europeu, como noutras línguas românicas, existem verbos que selecionam como complemento orações (completivas) de infinitivo não flexionado e que manifestam um comportamento sintático particular, podendo ocorrer (i) em construções de complementação típicas, mantendo as suas propriedades de verbos principais, como nos exemplos (1a) e (1b), e (ii) em construções em que formam uma unidade sintática com o verbo do complemento infinitivo, exibindo um comportamento próximo do dos verbos auxiliares, no sentido em que o seu complemento é transparente para determinados processos sintáticos que deveriam aplicar-se internamente a esse complemento, como no exemplo (1c). Em particular, neste exemplo, o clítico *a* realiza-se em adjacência ao verbo que seleciona a oração infinitiva (*querer*), e não ao verbo que o seleciona (*interromper*), construção que se designa como *subida de clítico*, uma vez que o pronome *se* realiza numa posição mais alta na frase do que a esperada, que seria a de adjacência ao verbo de que depende semântica e/ou sintaticamente – o verbo encaixado.

- (1) a. O professor quer [que os alunos não interrompam a exposição].
- b. O professor não quer [interrompê-la].
- c. O professor não a [quer interromper].

Na literatura sobre o assunto, a sequência formada pelos dois (ou mais) verbos nesta última construção tem sido designada como *predicado complexo*, termo introduzido por Burzio (1986).¹

Um dos marcos fundamentais no estudo dos predicados complexos é o trabalho de Kayne (1975), sobre os verbos causativos do francês.² Nesse trabalho, o autor mostra que verbos pertencentes a domínios frásicos distintos (a frase matriz e a frase encaixada) podem comportar-se como uma unidade sintática face a determinados fenómenos sintáticos, como a colocação dos pronomes pessoais átonos (clíticos) dependentes do verbo do complemento infinitivo: em certas construções causativas do francês, tais clíticos realizam-se obrigatoriamente em adjacência ao verbo superior, o verbo causativo. Esta subida de clítico encontra-se ilustrada em (2b–c), em contraste com (2d–e). (Sobre colocação de clíticos em português, 715 A colocação dos pronomes clíticos em sincronia e diacronia).

- (2) a. Elle fera manger ce gâteau à Jean.
 b. Elle le fera manger à Jean.
 c. Elle lui fera manger ce gâteau.
 d. *Elle fera le manger à Jean.
 e. *Elle fera lui manger ce gâteau. (Kayne 1975, 255–256)

Tendo em consideração dados do italiano, Rizzi (1976; 1978; 1982) considera um outro conjunto de verbos (como os equivalentes de *querer*, *poder*, *ir* e *vir*, entre outros), cujos membros também formam uma unidade sintática com o verbo do seu complemento infinitivo. Tal como acontece com os verbos causativos do francês, estudados por Kayne, com alguns verbos do italiano que pertencem às classes descritas por Rizzi, também se verifica subida de clítico (cf. (3a)), embora opcionalmente (cf. (3b)):

- (3) a. Piero ti verrà a parlare di parapsicologia.
 b. Piero verrà a parlarti di parapsicologia. (Rizzi 1982, 1)

Embora o termo *predicado complexo* seja usado para referir globalmente construções como as exemplificadas em (2a–c) e em (3a), existem algumas diferenças entre elas que motivam diferentes designações: no caso das construções com verbos causativos

¹ Num sentido mais lato, o termo *predicado complexo* aparece também por vezes associado a uma maior diversidade de outras sequências, podendo envolver um verbo e um elemento predicador de outra categoria, como um nome, um adjetivo, uma partícula, não necessariamente numa relação de complementação (cf. e.g. Duarte 2003; Svenonius 2008). Usamos aqui, na esteira de Burzio 1986, uma noção mais restritiva de *predicado complexo*.

² A ideia de que existem dois tipos de complementos infinitivos, um que constitui um domínio frásico independente e outro em que se observam efeitos de transparência que anulam a autonomia do mesmo, pode encontrar-se já num texto de Gunnar Bech, dos anos 50, para o alemão (Bech 1955–1957).

(e percetivos, como se verá adiante), usaremos, na linha de Kayne (1975), o termo *fazer-Inf* e, para as construções idênticas às do italiano em (3), o termo *reestruturação*, como em Rizzi (1976; 1978; 1982). Tal como nas línguas acima referidas, as duas construções estão disponíveis em português (cf. (4) para a construção *fazer-Inf* e (5c), ou (1c), para a construção de reestruturação):

- (4) a. Ela mandou comer o bolo ao João.
 b. Ela mandou-o comer ao João. [*o* = *o bolo*]
 c. Ela mandou-lhe comer o bolo. [*lhe* = *ao João*]
- (5) a. O João não quis conhecer a Marta.
 b. O João não quis conhecê-la. [*la* = *a Marta*]
 c. O João não a quis conhecer. [*a* = *a Marta*]

A possibilidade de os clíticos ocorrerem em adjacência ao verbo superior, ilustrada em (4b–c) e em (5c), bem como outras propriedades que revelam que o complemento infinitivo selecionado pelos verbos de reestruturação (e também pelos causativos e percetivos) é transparente para determinados fenómenos sintáticos, levou muitos gramáticos luso-brasileiros a considerar que alguns destes verbos são auxiliares (veja-se, em particular, Soares Barbosa 1822; Said Ali 1908; 1931; Epiphany da Silva Dias 1918; Bechara 1961, 1999; Cunha/Cintra 1984). Contudo, como vários autores têm assinalado (para o português europeu, ver Gonçalves 2001) tais verbos não devem ser classificados como auxiliares. O facto de, em alguns casos, o comportamento dos verbos de reestruturação se aproximar do dos auxiliares típicos tem sido explicado por mecanismos que criam efeitos de transparência sintática semelhantes aos que se observam em sequências de verbo auxiliar–verbo principal, ou seja, que dão origem a um elevado grau de sintatização entre os dois verbos do predicado complexo. É este segundo tipo de abordagem que se apresentará no presente capítulo.³

2 A identificação de predicados complexos

2.1 Propriedades dos predicados complexos sintáticos: os estudos seminais

Para dar conta do comportamento dos verbos que permitem a formação de predicados complexos, têm sido propostas análises nem sempre coincidentes. Por exemplo, uma das questões mais debatidas na literatura é a de saber se as construções desta natureza são bioracionais (isto é, se o complemento infinitivo mantém a sua autonomia sintática) ou monoracionais (caso em que a fronteira frásica que corresponde ao

³ Sobre a distinção entre verbos de reestruturação e os auxiliares típicos, veja-se a secção 3.2.2.

complemento infinitivo é, de certa forma, inexistente). No entanto, é consensualmente aceite que um predicado complexo consiste numa sequência de dois ou mais elementos (verbos, no caso que nos ocupa neste capítulo) que apresentam um forte grau de coesão sintático-semântica.

No seu trabalho de 1975, Kayne analisa os verbos causativos do francês (e.g. *faire* seguido de complemento infinitivo) numa construção que designa como «*faire-Infinitif*» e na qual o constituinte que, aparentemente, corresponde ao sujeito do verbo no infinitivo apresenta as seguintes propriedades:

- (i) ocorre após a sequência <verbo causativo-verbo infinitivo> (cf. (6a) e (7a));
- (ii) é substituível por um pronome clítico acusativo (como *le* ou, em português, *o*) ou dativo (como *lui*, ou *lhe* em português), dependendo das propriedades sintáticas do verbo infinitivo; assim, contrariamente ao que acontece no caso dos sujeitos típicos, que apenas podem ser substituídos pelo pronome pessoal na sua forma nominativa (como *ele*), na construção em apreço o pronome assume a forma acusativa, típica do objeto direto, no contexto de verbos intransitivos (cf. (6b)), e a forma dativa, típica do objeto indireto, no contexto de verbos transitivos (cf. (7b));
- (iii) quando assume a forma clítica, ocorre obrigatoriamente em adjacência ao verbo causativo e não ao verbo de que depende semanticamente (o infinitivo) – i.e. existe subida de clítico –, como se conclui da agramaticalidade de (6c) e (7c).

- (6) a. Elle fera partir ses amis. (Kayne ¹1975, 1977, 255, ex. (1))
 b. Elle les fera partir.
 c. *Elle fera les partir.
- (7) a. Elle fera manger ce gâteau à Jean. (Kayne ¹1975, 1977, 255, ex. (1))
 b. Elle lui fera manger ce gâteau.
 c. *Elle fera lui manger ce gâteau.

O facto de o alegado sujeito do verbo infinitivo corresponder a formas pronominais que não são características dos verdadeiros sujeitos e de os clíticos se realizarem em adjacência ao verbo causativo e não ao verbo que os seleciona constituem argumentos empíricos para a classificação da sequência de verbos como predicado complexo. Na realidade, o complemento infinitivo não se apresenta como um domínio autónomo, visto que constituintes desse domínio podem ser realizados em posições relacionadas com o primeiro verbo.

Largamente inspirado no estudo de Kayne (1975), ainda que apresente uma análise distinta, Burzio (1986) mostra que a mesma construção se encontra disponível com os verbos causativos do italiano (*fare* ‘fazer’/‘mandar’ e *lasciare* ‘deixar’) e que, nesta língua, também os verbos percetivos (*vedere* ‘ver’, *sentire* ‘ouvir’) podem desencadear a formação de um predicado complexo, embora em alguns aspetos distinto daquele em que ocorre um verbo causativo.

No estudo de Rizzi (1982), que constitui uma versão revista de um trabalho inaugural do mesmo autor, datado de 1978, é considerado um outro tipo de predica-

dos complexos em italiano, que resultam de um processo sintático a que o autor chama *reestruturação*. Em italiano, a formação deste tipo de predicados complexos é desencadeada por alguns verbos que o autor classifica como modais (e.g. *potere* ‘poder’, *volere* ‘querer’), aspetuais (e.g. *continuare* ‘continuar’, *cominciare* ‘começar’) e de movimento (e.g. *andare* ‘ir’, *venire* ‘vir’) e que apresentam comportamentos sintáticos distintos dos de outros verbos que pertencem às mesmas classes, selecionando igualmente complementos infinitivos.

Em primeiro lugar, com alguns verbos das referidas classes, os clíticos dependentes do verbo do complemento infinitivo (que designaremos como *verbo encaixado*) podem ocupar uma de duas posições: em adjacência a este verbo ou em adjacência ao verbo que seleciona a completiva (o *verbo matriz*). Esta última posição ilustra a já referida construção de *subida de clítico*. Os exemplos em (8) e (9) apresentam o contraste entre os verbos do italiano relativamente à subida de clítico (dados retirados de Rizzi 1982):

- (8) a. Piero verrà a parlarti di parapsicologia.
 b. Piero ti verrà a parlare di parapsicologia.
- (9) a. Piero deciderà di parlarti di parapsicologia.
 b. *Piero ti deciderà di parlare di parapsicologia.

De acordo com Rizzi (1982), a subida de clítico em (8b) é uma propriedade caracterizadora dos predicados complexos de reestruturação. A ideia é que, se o complemento infinitivo constituir um domínio totalmente autónomo, qualquer processo que envolva os elementos desse domínio, como a posição relativa entre o verbo e os clíticos que dele dependem, só pode operar internamente ao mesmo, como acontece em (9). Neste caso, não se verifica reestruturação. Assim, tendo em conta os dados em (8)–(9), Rizzi conclui que *venire* desencadeia reestruturação, mas *decidere* não.

A segunda propriedade distintiva dos verbos que admitem reestruturação em italiano é, segundo Rizzi (1982), a possibilidade de o objeto direto do verbo infinitivo em frases ativas ser realizado como sujeito em frases passivas pronominais.⁴ Assim, como acontece no caso da subida de clítico, os verbos que selecionam complementos

4 Nas orações passivas pronominais ocorre sempre o clítico *se*, sem que se realize qualquer verbo auxiliar. Considerem-se os exemplos em (i), retirados de Duarte (2013, 444 e 445):

- (i) a. As obras mais recentes publicaram-se numa editora espanhola para garantir maiores tiragens.
 b. Já se publicaram em português todos os romances de Salman Rushdie.

Nestas orações, o verbo ocorre sempre na 3.^a pessoa e concorda em número e pessoa com um constituinte que, na oração ativa, corresponde ao seu objeto (no caso dos exemplos em (i), os constituintes *as obras mais recentes* e *todos os romances de Salman Rushdie*). Assim, o objeto do verbo na oração ativa realiza-se como sujeito na oração passiva pronominal, como acontece nas passivas em que ocorre o auxiliar *ser*:

infinitivos distribuem-se por duas classes, como mostra o contraste entre (10b) e (11b) (dados de Rizzi 1982):

- (10) a. Si vuole vendere queste case a caro prezzo.
 b. Queste case si vogliono vendere a caro prezzo.
- (11) a. Si è promesso di costruire le nuove case popolari entro un anno.
 b. *Le nuove case popolari si sono promesse di costruire entro un anno.

O constituinte *queste case*, correspondente ao objeto direto do verbo no infinitivo na frase ativa em (10a), realiza-se, na passiva pronominal em (10b), como sujeito do verbo matriz, *volere*. Podemos dizer que este constituinte, selecionado pelo verbo da frase encaixada, como objeto, se move para uma posição mais alta na frase – a de sujeito da frase matriz, designando-se esta construção como *movimento longo de objeto* (Rizzi 1978; 1982). Como se verifica pela agramaticalidade de (11b), nem todos os verbos que selecionam complementos infinitivos admitem esta construção. A conclusão de Rizzi (1982) é a de que o movimento longo de objeto só se verifica em contextos de reestruturação, visto que só neste caso a oração infinitiva é, de uma forma que caracterizaremos adiante, transparente, permitindo que elementos que pertencem ao domínio do verbo encaixado se realizem no domínio do verbo matriz.

Finalmente, Rizzi (1982) nota que a escolha do auxiliar que ocorre nos tempos compostos é uma outra propriedade que permite caracterizar distintivamente os verbos de reestruturação do italiano. Esta língua, tal como o francês, dispõe de dois auxiliares de tempos compostos, *avere* (em português, *ter*) e *essere* (em português, *ser*), cuja distribuição é prevista a partir das propriedades do verbo principal. Assim, em frases simples, um verbo como *volere* seleciona o auxiliar *avere*, ao passo que um verbo como *tornare* seleciona o auxiliar *essere*:

- (12) a. Gianni ha voluto i libri.
 b. *Gianni è voluto i libri.
- (13) a. Gianni è tornato a casa.
 b. *Gianni ha tornato a casa.

Em frases complexas, alguns verbos que selecionam *avere* podem também coocorrer com *essere*, desde que seja este o auxiliar selecionado pelo verbo encaixado. É o que acontece em frases em que a oração infinitiva é complemento de *volere*, como se ilustra em (14); porém, no contexto do verbo *promettere*, o único auxiliar possível é

(ii) a. As obras mais recentes foram publicadas numa editora espanhola para garantir maiores tiragens.
 b. Já foram publicados em português todos os romances de Salman Rushdie.

avere, mesmo que o verbo encaixado seleccione *essere*, como se conclui da agramaticalidade de (15b):

- (14) a. Gianni ha voluto tornare a casa.
 b. Gianni è voluto tornare a casa.
- (15) a. Mario ha promesso di tornare a casa.
 b. *Mario è promesso di tornare a casa.

O facto de em (14) ser possível a ocorrência do auxiliar selecionado pelo verbo encaixado constitui, segundo Rizzi (1982), uma característica de um predicado complexo, uma vez que as propriedades desse verbo podem determinar alterações na frase matriz, como se não existisse uma fronteira frásica. Por essa razão, nos casos em que há mudança do auxiliar, é obrigatória a subida de clítico, como se observa em (16c–d):

- (16) a. Gianni ha voluto andarci.
 b. *Gianni ci ha voluto andare.
 c. Gianni ci è voluto andare.
 d. *Gianni è voluto andarci.

Nos exemplos que nos têm vindo a ocupar, a formação do predicado complexo tem apenas efeitos sintáticos. No entanto, em algumas línguas, nomeadamente as polisintéticas,⁵ de raiz não indo-europeia, como o mowhak, ou o chichewa, a complexidade deste tipo de predicados manifesta-se não só na sintaxe, mas também na morfologia, já que os elementos que constituem o predicado complexo formam uma unidade morfológica. Nestes casos, um dos elementos do predicado complexo é um afixo, o que obriga à formação da unidade morfológica, impede a ocorrência de material lexical entre os dois membros do predicado e resulta num único domínio frásico, propriedades visíveis no exemplo (17), do chichewa:

- (17) Mtsikana anau-gw-ets-a mtsuko.
 rapariga FLEX-cair-fez-ASP cântaro
 'A rapariga fez cair o cântaro.' (Baker 1988, 148)

No entanto, como Baker (1996) salienta, esta característica não deve fazer parte da definição dos predicados complexos, visto que, em línguas analíticas, como o português, o francês ou o italiano, os membros do predicado complexo preservam a sua

5 Sapir (1921, 128) caracteriza as línguas polisintéticas da seguinte forma: «A polysynthetic language, as its name implies, is more than ordinarily synthetic. The elaboration of the word is extreme. Concepts which we should never dream of treating in a subordinate fashion are symbolized by derivational affixes or «symbolic» changes in the radical element, while the more abstract notions, including the syntactic relations, may also be conveyed by the word».

autonomia morfológica, sendo mesmo possível inserir material lexical entre eles, como se ilustra em (18) para o português:

- (18) a. Ela mandou-o logo comer ao João.
 b. Ele não a quis mesmo conhecer.

Em síntese, a partir dos estudos seminais de Kayne (1975), Rizzi (1978; 1982) e Burzio (1986), são normalmente tomadas como indicadoras da formação de predicados complexos, em sequências verbais introduzidas por verbos causativos (e.g. *mandar*), percetivos (e.g. *ouvir*) ou verbos de reestruturação, como *querer* e verbos modais (e.g. *poder*) e aspetuais (e.g. *continuar*), as seguintes propriedades sintáticas: (i) subida de clítico; (ii) movimento longo de objeto; e, (iii) em línguas que dispõem de mais do que um auxiliar para formação de tempos compostos, manifestação do auxiliar do verbo infinitivo em formas de tempo composto do predicado complexo. Em português europeu, os predicados complexos são, globalmente, caracterizados pelas propriedades (i) e (ii).

2.2 Dois tipos de predicados complexos em português europeu

Como afirmámos anteriormente, desde os primeiros estudos para o francês (Kayne 1975) e para o italiano (Rizzi 1978; 1982; Burzio 1986), tem sido demonstrado que, nas frases que integram um predicado complexo nas línguas românicas, um verbo que seleciona um complemento infinitivo constitui um núcleo sintática e semanticamente coeso com o verbo deste complemento. Dito de outra forma, os dois verbos manifestam um elevado grau de sintatização, o que determina que, em certos contextos, se comportem como se formassem uma unidade do ponto de vista sintático.

O termo *complemento* permite excluir, desde logo, os casos em que a oração infinitiva tem a função sintática de sujeito ou de adjunto (cf. (19) e (20)). Por sua vez, o termo *infinitivo* exclui os casos em que a oração selecionada pelo primeiro verbo, ainda que desempenhe a função sintática de complemento, é finita (ou seja, apresenta o verbo nos modos indicativo ou conjuntivo; cf. (21)); de facto, tomando, por agora, a subida de clítico como critério para identificação de um predicado complexo também para o português europeu, verifica-se que a mesma não opera nestes contextos:

- (19) a. Não surpreendeu a Ana ter encontrado o Pedro.
 b. Não surpreendeu a Ana tê-lo encontrado. [(l)o= o Pedro]
 c. *Não o surpreendeu a Ana ter encontrado.
- (20) a. O João não quer sair sem ver a Marta.
 b. O João não quer sair sem a ver. [a = a Marta]
 c. *O João não a quer sair sem ver.

- (21) a. A Maria não quer que a Ana leia o livro.
 b. A Maria não quer que a Ana o leia. [*o = o livro*]
 c. *A Maria não o quer que a Ana leia.

A observação destes dados conduz à primeira generalização sobre os contextos em que um predicado complexo pode ocorrer, a saber, completivas de infinitivo. Numa língua que, como o português, dispõe de infinitivo flexionado, coloca-se, agora, uma segunda questão – a de saber se este é um contexto que legitima a formação de um predicado complexo. A resposta é negativa, como se conclui da agramaticalidade da subida do clítico em (22c):

- (22) a. O presidente lamentou terem os polícias agredido os manifestantes.
 b. O presidente lamentou terem nos os polícias agredido. [*(n)os = os manifestantes*]
 c. *O presidente lamentou os terem os polícias agredido.

Assim, a generalização acima deverá ser reformulada para o português europeu: os predicados complexos ocorrem exclusivamente em contextos de infinitivo não flexionado. São, assim, potenciais desencadeadores da formação de predicados complexos os verbos causativos e perceptivos, os verbos de controlo e os verbos de elevação.⁶ Contudo, como acontece em italiano e em francês, o facto de um verbo pertencer a uma destas classes não garante a formação de um predicado complexo. Veja-se, neste sentido, o contraste entre (23) e (24), em que os verbos matriz (*querer* e *prometer*) pertencem ambos à classe dos verbos de controlo:⁷

6 Neste capítulo, assumimos as análises clássicas de controlo e de elevação (Chomsky 1981, entre outros). Assim, o termo *controlo* designa estruturas em que, quando o infinitivo não flexionado é usado na frase encaixada, o sujeito desta frase é nulo e tem um antecedente fixo na frase superior: o sujeito (em (i)), o objeto direto (em (ii)) ou o objeto indireto (em (iii)). Esta relação entre antecedente (na frase matriz) e sujeito nulo da frase encaixada é assinalada através de índices nos exemplos que se seguem.

- (i) Os meninos; querem [-]_i comprar uma bicicleta.
 (ii) Os pais obrigaram os meninos; a [-]_i comer a sopa.
 (iii) Os pais permitiram aos meninos; [-]_i chegar depois da meia-noite.

Por sua vez, o termo *elevação* designa estruturas em que um constituinte com uma função sintática na frase superior (nos casos que nos interessam neste capítulo, o sujeito) corresponde a um argumento do verbo da encaixada, posição a partir da qual foi movido, como em (iv):

- (iv) a. Os meninos parecem [~~os meninos~~] dormir.
 b. Os meninos devem [~~os meninos~~] dormir.

Os verbos matriz que ocorrem na construção de controlo designam-se como *verbos de controlo*, os que ocorrem na construção de elevação, como *verbos de elevação*.

7 Cf. também variação nas manifestações típicas de alguns destes verbos, mencionada na secção 2.3, adiante.

- (23) a. O treinador não quis contratar os jogadores desempregados.
 b. O treinador não quis contratá-los.
 c. O treinador não os quis contratar.
- (24) a. O treinador não prometeu contratar os jogadores desempregados.
 b. O treinador não prometeu contratá-los.
 c. *O treinador não os prometeu contratar.

A distribuição dos verbos que selecionam um complemento infinitivo por dois grupos (os verbos de controlo e os de elevação, que ocorrem na construção de reestruturação, e os verbos causativos e percetivos, que ocorrem na construção *fazer-Inf*) não é um traço inovador do português. Com efeito, os dados do italiano apresentados na secção anterior revelam um comportamento idêntico nesta língua. Interessa, portanto, verificar se as propriedades que caracterizam os predicados complexos em italiano se aplicam ao português.

O contraste que os exemplos (23) e (24) ilustram permite afirmar que a subida de clítico também opera em português europeu em contextos de reestruturação. Assim, tendo em conta esta propriedade, diríamos que o verbo *querer*, mas não o verbo *prometer*, é um verbo de reestruturação.⁸ Como se referiu a propósito dos dados do italiano, sendo a colocação do clítico um processo local (isto é, por defeito, o clítico associa-se ao verbo de que depende sintática e/ou semanticamente), espera-se que, em frases complexas, o clítico se realize junto do verbo encaixado, se dele depender.⁹

8 De acordo com Fiéis/Madeira (2012), existe uma gradação de aceitabilidade destas construções, que depende da presença de elementos proclisadores (ou seja, de elementos que desencadeiam a próclise e que, segundo as autoras, facilitam a subida de clítico) e do próprio tipo de clítico. Embora se verifique, efetivamente, alguma variação entre os falantes de português europeu, é possível encontrar uma escala de aceitabilidade que uniformiza alguns juízos (Gonçalves 1999; Fiéis/Madeira 2012). Assim, geralmente, como *querer* comportam-se *conseguir*, *desejar*, *tentar*, *poder*, *dever*, *estar (a)*, *ir*, *vir*, entre outros, enquanto *ousar*, *jurar*, *evitar* se comportam como *prometer*. Já o verbo *decidir* é um dos que revela menor consenso entre os falantes, que ora o aproximam de *querer* ora o aproximam de *prometer*.

9 Quando se associa ao verbo matriz, o clítico pode ocorrer em próclise (antes do verbo), em ênclise (depois do verbo) ou em mesóclise (no interior do verbo), de acordo com as regras gerais de colocação dos clíticos (≥15 A colocação dos pronomes clíticos em sincronia e diacronia). Vejam-se os exemplos (i), nos quais se ilustram as três posições, considerando o contexto de subida de clítico:

- (i) a. Estás-me a ouvir? (ênclise)
 b. Não me estás a ouvir? (próclise)
 c. Estar-me-ás a ouvir? (mesóclise)

Ainda que os três padrões sejam admitidos em contexto de subida de clítico, a generalidade dos falantes considera que a próclise é o que produz melhores resultados. A ênclise é, muitas vezes, considerada estranha, sobretudo quando o clítico corresponde às formas acusativas *o*, *a*, *os*, *as*, caso em que a forma verbal e o clítico sofrem processos fonológicos que são responsáveis pela estranheza das frases. Veja-se o contraste entre (ia) e (ii):

- (ii) ^{*/}*Está-lo a ouvir?

É isso que se observa em (24), no contexto do verbo *prometer*, mas não em (23), no contexto do verbo *querer*. Este contraste foi notado, para o português europeu, em Martins (1995) e em Gonçalves (1999).

É de notar que o contraste entre (23c) e (24c) sugere que a hipótese de explicação do processo de formação de um predicado complexo de reestruturação não pode residir apenas na classe a que o verbo matriz pertence. Com efeito, os dois verbos matriz são verbos de controlo, mas apenas *querer* potencia a constituição de um predicado dessa natureza. Para além disso, o facto de o mesmo verbo apresentar um comportamento distinto em línguas diferentes constitui evidência empírica adicional a favor desta ideia: vejam-se, a este propósito, os exemplos do francês em (25), em que o verbo matriz, *vouloir*, não admite a formação de um predicado complexo (ver (25c)), ao contrário dos seus correspondentes em português e em italiano, respetivamente *querer* e *volere*:

- (25) a. Je veux donner le livre à Jean.
 b. Je veux lui donner le livre. [*lui* = à Jean]
 c. *Je lui veux donner le livre.

A segunda propriedade estabelecida por Rizzi (1982) para identificar predicados complexos de reestruturação é, como já referimos, o movimento longo de objeto. Tal como em italiano, existe um contraste entre os verbos do português europeu que selecionam complementos infinitivos. Considerem-se os exemplos seguintes:

- (26) Os inquilinos salvaram os móveis quando a casa já estava a arder.
 (27) Quiseram-se salvar os móveis quando a casa já estava a arder.
 (28) *Prometeram-se salvar os móveis quando a casa já estava a arder.

Em (26), uma frase ativa, o constituinte *os móveis* é selecionado pelo verbo *salvar*, tendo a função sintática de objeto direto. Na frase passiva pronominal em (27), este constituinte, sendo igualmente selecionado pelo verbo *salvar*, que ocorre no complemento infinitivo, é realizado como sujeito do verbo matriz. Assim, como acontece em italiano, também alguns verbos de controlo do português europeu admitem a construção de movimento longo de objeto, caracterizando-se, por isso, como verbos de reestruturação. Também como em italiano, verifica-se um contraste entre verbos de controlo, como se conclui do contraste entre (27) e (28). Note-se que os mesmos verbos que admitem subida de clítico admitem movimento longo de objeto e os que não admitem a primeira construção não admitem a segunda, o que revela, por parte destes

Para evitar que fatores que não são de natureza sintática interfiram na gramaticalidade das frases, a generalidade dos exemplos do português europeu que apresentamos neste capítulo constituem casos de próclise.

verbos, um comportamento consistente quanto à possibilidade (ou impossibilidade) de formação de um predicado complexo.

A terceira propriedade apontada por Rizzi (1982) para identificar um predicado complexo de reestruturação diz respeito à mudança de auxiliar nos tempos compostos. Tal propriedade não se verifica em português europeu, uma vez que, nesta língua, o único auxiliar dos tempos compostos é *ter*.

Para além de verbos de controlo como *querer*, *desejar*, *conseguir*, *tentar*, entre outros, o grupo dos verbos de reestruturação integra, ainda, os semiauxiliares aspetuais (e.g. *estar (a)*, *começar (a)*, *continuar (a)*), modais (e.g., *poder*, *dever*) e temporais (e.g. *ir*), que podem ser incluídos na classe dos verbos de elevação (Gonçalves 1992, entre outros). Vejam-se os exemplos em (29) a (34), para cada uma destas classes de verbos e para cada propriedade de predicado complexo anteriormente identificada para o português europeu:

(i) Possibilidade de subida de clítico

- (29) a. A Ana comprou uma nova gramática mas ainda não começou a consultá-la.
 b. A Ana comprou uma nova gramática mas ainda não a começou a consultar.

- (30) a. Os alunos acabaram o trabalho e disseram que já podem entregá-lo.
 b. Os alunos acabaram o trabalho e disseram que já o podem entregar.

- (31) a. A Maria não contou a verdade aos pais porque eles não vão compreendê-la.
 b. A Maria não contou a verdade aos pais porque eles não a vão compreender.

(ii) Movimento longo de objeto

- (32) Já se começaram a construir casas para os desalojados das cheias.
 (33) Os resultados do concurso já se podem consultar no *Diário da República*.
 (34) Todos os resultados do concurso se vão publicar no *Diário da República*.

Relativamente aos verbos causativos (*mandar*, *deixar*, *fazer*) e aos percetivos (*ver*, *ouvir*, *sentir*), o português europeu aproxima-se, de novo, do francês e do italiano, no sentido em que também esses verbos podem encabeçar um predicado complexo do tipo *fazer-Inf* quando selecionam complementos de infinitivo não flexionado. Vejam-se, a este propósito, os exemplos em (35) a (38):

(i) Subida de clítico com causativos (cf. (35)) e percetivos (cf. (36))

- (35) a. O professor mandou consultar a pauta aos alunos.
 b. O professor mandou-a consultar aos alunos. [*a* = *a pauta*]
 c. O professor mandou-lhes consultar a pauta. [*lhes* = *aos alunos*]

- (36) a. O João ouviu cantar a *Aida* a uma cantora lírica australiana.
 b. O João ouviu-a cantar a uma cantora lírica australiana.
 c. ?O João ouviu-lhe cantar a *Aida*.¹⁰
- (ii) Movimento longo de objeto com causativos (cf. (37)) e percetivos (cf. (38))¹¹
- (37) Mandaram-se construir novas casas para os desalojados das cheias.
- (38) Ouviram-se saltar rolhas de garrafas de champanhe.

Ainda que todos os verbos que entram na formação de predicados complexos manifestem um comportamento idêntico relativamente à subida de clítico e ao movimento longo de objeto, como se conclui da observação dos exemplos que foram apresentados nesta secção, existem, em português europeu, diferenças entre reestruturação e *fazer-Inf*.

São várias as propriedades que distinguem os dois tipos de predicados complexos em português europeu (cf. Gonçalves 1999).¹² Em primeiro lugar, como também assinalado em Raposo (1981), os predicados complexos do tipo *fazer-Inf*, que integram um verbo causativo ou um percetivo, implicam uma alteração à ordem dos constituintes do domínio infinitivo, associada a uma reatribuição de funções sintáticas. Com efeito, neste contexto, como já se descreveu anteriormente para o francês (cf. (6) e (7)), o constituinte *que*, do ponto de vista semântico, corresponde ao sujeito lógico do verbo encaixado, apresenta, do ponto de vista sintático, propriedades de objeto direto ou de objeto indireto, consoante a transitividade do verbo, e ocorre em posição pós-predicado complexo. Assim, quando o verbo encaixado é intransitivo, o constituinte em causa pode ser substituído pelo clítico acusativo *o*, que identifica o objeto direto (cf. (39)); pelo contrário, quando o verbo encaixado é transitivo, o referido constituinte, quando pronominalizado, assume a forma dativa do clítico, *lhe*, que identifica o objeto indireto (cf. (40)):

- (39) a. O treinador mandou correr os atletas.
 b. O treinador mandou-os correr.
- (40) a. O Luís mandou arranjar o carro da mulher ao mecânico mais antigo da cidade.
 b. O Luís mandou-lhe arranjar o carro da mulher.

10 A presença do clítico dativo em predicados complexos com verbos percetivos não é igualmente admitida pelos falantes de português europeu. Ver também adiante, nota 18.

11 Com base numa pesquisa no *corpus* CRPC, Vasconcelos (2013) mostra que o movimento longo de objeto é mais frequente em estruturas com causativos, percetivos e verbos de elevação como os semiauxiliares do que com os verbos de controlo (*querer, conseguir, tentar, pretender*, por exemplo).

12 Embora estejam relatadas várias diferenças entre os dois tipos de predicados complexos, seleccionámos apenas três, a título ilustrativo. Para mais informação sobre o português europeu, ver Gonçalves (1999).

Dado que os constituintes assinalados em (39a) e (40a) não manifestam as propriedades típicas de um sujeito (nomeadamente, não são substituíveis pela forma nominativa do pronome pessoal, *ele/ela/eles/elas*), passaremos a designá-los como «causado» (do inglês, *causee*). Voltaremos a esta questão na secção 4.

Em segundo lugar, os dois tipos de predicados complexos distinguem-se quanto à passiva: o complemento infinitivo dos verbos de reestruturação pode ocorrer na forma passiva (cf. (41)), ao contrário do que acontece no contexto de *fazer-Inf* (cf. (42)):

- (41) a. O Pedro já conseguiu [ser apresentado à Maria (pelo Mário)].
b. O Pedro já lhe conseguiu ser apresentado (pelo Mário).

(42) *O Pedro já mandou ser apresentado o Pedro à Maria (pelo Mário) nessa noite.

Finalmente, na construção *fazer-Inf* observa-se uma maior restrição quanto à natureza dos clíticos que sofrem o fenómeno que designámos como subida de clítico. Considerem-se os exemplos (43) e (44):

- (43) a. Todos os adolescentes se querem barbear com a máquina do pai.
b. O Luís nunca se quis queixar das dores que tinha antes de ser operado.

- (44) a. *O João mandou-se_i barbear (a)o Pedro_i.
b. *O João mandou-se_i queixar das dores (a)o Luís_i.

Enquanto em (43), um contexto de reestruturação, a ocorrência de clíticos reflexos (em (a)) e inerentes¹³ (em (b)) produz um resultado gramatical, em (44), exemplificativo de *fazer-Inf*, os mesmos clíticos não podem realizar-se em adjacência ao verbo causativo se tiverem como antecedente o causado. Esta diferença decorre de propriedades do complemento infinitivo, em particular no que diz respeito à natureza e à posição do sujeito/causado neste complemento (Gonçalves 1999). Nas secções 3 e 4, caracterizaremos com mais pormenor os dois tipos de predicados complexos em português europeu.

2.3 Aspetos de variação na formação de predicados complexos

A possibilidade de formação de predicados complexos é um dos aspetos diferenciadores entre português europeu e português brasileiro. A ausência de predicados complexos tem sido normalmente apontada como fenómeno característico do português

¹³ Os clíticos inerentes são os que fazem parte integrante do próprio verbo, como em *queixar-se*, *suicidar-se*, *portar-se* (*bem/mal*), entre outros.

brasileiro (Cyrino 2008; 2010; entre outros), em evidência na agramaticalidade dos seguintes exemplos (Cyrino 2010, 200s.):

- (45) a. *O João mandou comer a sopa à Ana. (português brasileiro)
 b. *O João mandou-lhe comer a sopa. (português brasileiro)
- (46) a. *O João viu sair a Maria. (português brasileiro)
 b. *O João viu-a sair. (português brasileiro)

Além da inexistência da construção *fazer-Inf* exemplificada em (45) e (46), em português brasileiro os clíticos são realizados no domínio do verbo encaixado, sem que ocorra subida do clítico, o que mostra também a ausência de predicados complexos de reestruturação nesta variedade:

- (47) João pode/quer te ver. (português brasileiro; adaptado de Cyrino 2010, 200, ex. (23))

Esta diferença entre português europeu e português brasileiro decorre de diferenças na estrutura associada ao domínio do verbo infinitivo, que não permite a formação de um predicado complexo, pelo que o clítico ocorre obrigatoriamente nesse domínio.

Os contextos aqui considerados, que potenciam reestruturação em português europeu padrão, apresentam ocasionalmente manifestações não padrão de infinitivo flexionado (descritas para variedades de português europeu, assim como para variedades de português em África; cf. e.g. Gonçalves/Santos/Duarte 2014; A. Gonçalves 2014; P. Gonçalves 2015), excluindo-se nesse caso a possibilidade de formação de um predicado complexo:

- (48) a. Vás seres apanhado ...
 b. Começas a teres o que queres ...
 (português europeu não padrão, in Carrilho/Lobo 1999)
 c. Se a gente andarmos a apanharmos cana...
 (português europeu não padrão, in Carrilho/Lobo 1999)
- (49) a. As pessoas preferem ganharem naquela hora mesmo.
 b. As mulheres devem fabricarem panelas de barro e depois vão venderem.
 (português de Moçambique, in Gonçalves 2015, 11 (exs. 1a–b))

Finalmente, notamos a existência de variação quanto aos verbos matriz que permitem subida do clítico (cf. também nota 8, acima). Em (50), por exemplo, ocorre subida de clítico no contexto de um verbo (*permitir*) que, para alguns falantes, exclui tal possibilidade.

- (50) Sou professor provisório sem preparação específica para a docência nem estágio, porque o Estado nunca me permitiu fazer. (CETEMPúblico, Ext=753989)

Com base em dados do corpus dialetal CORDIAL-SIN, Magro (2005) encontra também contextos adicionais de subida de clítico, concluindo que variedades do português europeu diferentes da variedade padrão são normalmente mais permissivas em relação a este fenómeno.¹⁴

Uma vez que a subida de clítico constitui um dos critérios considerados na definição de predicados complexos, esta variação evidencia uma potencial margem de alargamento da identificação de predicados complexos em português europeu.

3 A construção de reestruturação em português europeu

Os dados apresentados na secção 2 mostram que existem verbos de controlo e de elevação do português europeu que, selecionado um complemento infinitivo, podem formar um predicado complexo de reestruturação com o verbo deste complemento. No entanto, como se mostrou em (23c) e (24c), repetidos em (51), apenas um subconjunto de verbos das mesmas classes pode, de facto, legitimar a formação do predicado complexo:

- (51) a. O treinador não os quis contratar.
b. *O treinador não os prometeu contratar.

A primeira questão que se coloca é a de saber de que forma a subida de clítico e o movimento longo de objeto são indicadores de que os dois verbos formam um predicado complexo, isto é, de que os dois verbos constituem uma unidade sintática.

3.1 Da monoracionalidade...

Existem vários argumentos empíricos a favor da ideia, já herdada de Rizzi (1978; 1982), de que o verbo do complemento infinitivo e o próprio complemento perdem, em parte, a sua autonomia sintática, na construção de reestruturação. Esta observação

¹⁴ Entre os verbos assinalados por Magro (2005), combinados com as preposições introdutoras da oração completiva infinitiva, encontram-se *acabar (de)*, *custar (a)*, *usar (a)*, *ir (para)*, *ter (de)*.

tem levado diversos autores a propor que, nesta construção, existe um único domínio frásico (Rizzi 1978; 1982; Wurmbrand 2001; 2004; 2006; Cinque 2006). Assim, tal como descrito para o italiano por Rizzi, em português europeu são observáveis os seguintes factos:

- (i) Quando ocorre um verbo que pode desencadear a reestruturação, a coocorrência de modificadores com valores temporais que não se intersejam, afetando cada um deles um domínio distinto, produz um resultado agramatical (cf. (52)). O mesmo não acontece se o verbo não desencadear reestruturação (cf. (53)).

(52) *Os jornalistas, ontem, quiseram entrevistar esse candidato amanhã.

(53) Os jornalistas, ontem, prometeram entrevistar esse candidato amanhã.

Note-se que o verbo *querer*, como *prometer*, seleciona um complemento infinitivo com uma leitura de posterioridade (isto é, a situação descrita nesse complemento é necessariamente posterior à situação descrita na oração matriz), pelo que (52) deveria ser gramatical. Voltaremos a esta questão mais adiante.

- (ii) Se os dois verbos formarem um predicado complexo, sendo possível a subida de clítico, o advérbio de negação frásica não pode ocorrer no domínio encaixado, o que decorre do facto de os dois verbos formarem uma unidade, pelo que não é possível negar apenas a situação encaixada, como se conclui da agramaticalidade de (54b). Note-se que, se os dois verbos se mantiverem autónomos, não formando um predicado complexo, o complemento infinitivo pode ser negado de forma independente (cf. (54c)).

(54) a. Depois de terem faltado à aula, os alunos só queriam não enfrentar a professora.

b. *Depois de terem faltado à aula, os alunos só a queriam não enfrentar.

[a = a professora]

c. Depois de terem faltado à aula, os alunos só queriam não a enfrentar.

- (iii) Se a reestruturação se verificar, a completiva infinitiva não pode ser deslocada de forma independente, como se conclui do contraste entre (55a) e (55b):

(55) a. Enfrentá-la, os alunos não quiseram.

b. *Enfrentar, os alunos não a quiseram.

- (iv) A ocorrência de uma frase parentética entre os dois verbos do predicado complexo de reestruturação dá origem a uma sequência agramatical (cf. (56a)), facto que não se verifica se os dois verbos mantiverem a sua autonomia sintática (cf. (56b)).

(56) a. *Depois de terem faltado à aula, os alunos queriam-na – mas sabiam que não deviam – enfrentar.

b. Depois de terem faltado à aula, os alunos só queriam – mas sabiam que não deviam – enfrentá-la.

Os exemplos (52) a (56) mostram que, efetivamente, no contexto de reestruturação, os dois verbos revelam um forte grau de sintatização, ou seja, formam uma unidade

sintática, o que poderá ser um indicador de que tal construção tem uma estrutura monoracional.¹⁵

3.2 ... à bioracionalidade

A questão que nos ocupará na presente secção é a seguinte: o facto de os dois verbos formarem uma unidade sintática implica necessariamente a existência de um único domínio frásico? Dito de outra forma, nos casos em que se verifica reestruturação, existe apenas uma frase, como defendido, entre outros, por Rizzi (1978; 1982), Wurmbrand (2001; 2004; 2006) e Cinque (2006)? A resposta a esta questão é negativa, como proposto em Bok-Bennema/Kampers-Manhe (1994), Roberts (1997), Gonçalves (1999), Bok-Bennema (2006), entre outros.

3.2.1 Propriedades temporais dos domínios infinitivos na construção de reestruturação¹⁶

A impossibilidade de coocorrência de modificadores temporais com valores opostos exemplificada em (57) parece ser um argumento forte a favor da monoracionalidade da construção de reestruturação. Assim, a agramaticalidade dessa sequência seria naturalmente esperada: se existisse apenas um domínio frásico, toda a situação descrita se localizaria num único intervalo de tempo (sobre semântica temporal do português, ↗12 Semântica e pragmática). É isso que se verifica em frases simples, com ou sem o auxiliar de tempos compostos:

- (57) a. *O presidente, ontem, promulgou o decreto-lei hoje.
 b. *O presidente, ontem, já tinha promulgado o decreto-lei anteontem.

15 Note-se, porém, que esta unidade sintática não corresponde a uma unidade morfológica em português europeu, como em outras línguas românicas: como já referido, a adjacência entre os dois verbos pode ser interrompida:

(i) Depois de terem faltado à aula, os alunos não a queriam mesmo enfrentar.

16 As propriedades temporais dos domínios de infinitivo não flexionado que são descritas nesta secção caracterizam não só o português europeu como o português brasileiro. Porém, como referimos na secção 2.3, esta última variedade não dispõe de predicados complexos. Assim, considerar os mecanismos de dependência temporal é crucial para dar conta de diferenças entre verbos matriz em português europeu, distinguindo os que permitem reestruturação dos que não a permitem, mas tais mecanismos são irrelevantes em português brasileiro, variedade em que a reestruturação é bloqueada por razões de natureza sintática, mesmo em contextos em que se verifica dependência temporal.

Gonçalves/Cunha/Silvano (2010) mostram, no entanto, que o complemento infinitivo dos verbos de reestruturação pode possuir traços de temporalidade, localizando a situação que descreve num intervalo de tempo distinto daquele em que se localiza a situação matriz (isto é, aquela que é descrita através do verbo de reestruturação e seus argumentos), o que constitui um argumento empírico forte a favor da bioracionalidade das estruturas em que ocorre a reestruturação. Considere-se a frase (58):

(58) O presidente não quis promulgar o decreto-lei.

Nesta frase, em que ocorre um verbo de controlo (*querer*) que potencia a formação de um predicado complexo de reestruturação, o intervalo de tempo em que se localiza a situação descrita na encaixada (*promulgar o decreto-lei*) é posterior ao intervalo de tempo em que se localiza a oração matriz, ou seja, corresponde a um futuro não realizado (cf. Stowell 1982) no momento em que o presidente mostrou não querer promulgar o decreto-lei. Neste aspeto, a frase (58) não se distingue de (59), em que se verifica também uma relação de posterioridade entre o domínio infinitivo e o domínio matriz:

(59) O presidente não prometeu promulgar o decreto-lei.

Os dados em (58) e (59) mostram que verbos como *querer* e *prometer* influenciam de forma decisiva a localização temporal da eventualidade encaixada – esta é posterior à situação descrita pela matriz, o que explica a agramaticalidade de (60).

- (60) a. *O presidente quis ter promulgado o decreto-lei.
 b. *O presidente prometeu ter promulgado o decreto-lei.

Assim, estes dois verbos têm a mesma orientação temporal, entendida como uma propriedade inscrita na entrada lexical dos verbos que selecionam complementos oracionais e que determina a localização temporal da situação descrita na frase encaixada como anterior, posterior ou sobreposta à situação descrita na oração matriz (Duarte/Gonçalves/Santos 2012; Gonçalves/Santos/Duarte 2014).

No entanto, *querer* e *prometer* têm comportamentos distintos relativamente à reestruturação: apenas o primeiro admite a formação de um predicado complexo deste tipo, como se observa pelo contraste entre (61a) e (61b):

- (61) a. O presidente não o quis promulgar. [*o* = *o decreto-lei*]
 b. *O presidente não o prometeu promulgar.

Não sendo *querer* e *prometer* distintos relativamente à orientação temporal (posterioridade), as orações subordinadas a estes verbos manifestam comportamentos distintos no que diz respeito aos mecanismos de dependência temporal. Assim, quando o

verbo matriz é *querer*, o intervalo de tempo a partir do qual a situação descrita pelo complemento infinitivo é perspectivada (ou seja, o seu ponto de perspectiva temporal; Kamp/Reyle 1993) é o da oração matriz; neste caso, as duas situações partilham o mesmo domínio temporal, entendido como o intervalo de tempo ocupado por uma situação ou conjunto de situações temporalmente relacionadas (Declerck 1991). Por essa razão, a sequência (52), repetida em (62), é agramatical, mesmo que o requisito de posterioridade se verifique: a situação encaixada só pode tomar como ponto de perspectiva temporal o tempo da matriz e não o da enunciação, o que exclui advérbios de tempo dêiticos, como *amanhã*. Neste caso, constitui-se um único domínio temporal, e a oração encaixada é, portanto, temporalmente dependente da matriz.

(62) *Os jornalistas, ontem, quiseram entrevistar esse candidato amanhã.

Note-se que outros advérbios temporais com valor de futuro podem ocorrer, desde que integrem a situação encaixada no domínio temporal criado pela situação matriz:

(63) O presidente quis promulgar o decreto-lei no dia seguinte.

Pelo contrário, no contexto do verbo *prometer*, o ponto de perspectiva temporal do complemento infinitivo é tanto o intervalo de tempo em que se localiza a situação matriz como o momento da enunciação; neste sentido, as duas situações não partilham necessariamente o mesmo domínio temporal, pelo que (53), repetida em (64), é uma frase gramatical. Neste caso, a oração encaixada é, portanto, temporalmente independente da matriz.

(64) Os jornalistas, ontem, prometeram entrevistar esse candidato amanhã.

Os conceitos de orientação temporal e de (in)dependência temporal são cruciais para a reestruturação. Deste modo,

(i) Verbos que não têm uma orientação temporal especificada, não fixando, por isso, a localização temporal da situação encaixada (cf. (65)) não admitem reestruturação (cf. (66b)):

- (65) a. O arquiteto não disse transportar consigo a planta do edifício em recuperação.
(*sobreposição*)
b. O arquiteto não disse ter transportado consigo a planta do edifício em recuperação.
(*anterioridade*)
c. O arquiteto não disse ir transportar consigo a planta do edifício em recuperação.
(*posterioridade*)
- (66) a. O arquiteto não disse transportá-la consigo.
b. *O arquiteto não a disse transportar.

- (ii) Verbos que têm uma orientação temporal especificada, determinando a localização temporal da situação encaixada, desencadeiam reestruturação desde que tal situação seja obrigatória e exclusivamente integrada no domínio temporal criado pela situação matriz, sendo o seu ponto de perspetiva temporal apenas o tempo da matriz. Geralmente, o domínio infinitivo estabelece com o tempo da matriz uma relação de posterioridade ou de sobreposição.¹⁷ Desta forma, admitem reestruturação verbos de controlo como *querer*, *desejar*, *conseguir*, bem como verbos de elevação como os semiauxiliares aspetuais (*estar a*, *começar a*, *continuar a*), modais (*poder*, *dever*) e temporais (*ir*), mas não verbos de controlo como *prometer*, *decidir*, *jurar*, *lamentar*, entre outros.

Considerar que, na construção de reestruturação, o complemento infinitivo tem traços de temporalidade, ainda que se verifique uma relação de dependência temporal relativamente ao tempo da matriz, implica defender a bioracionalidade dessa construção. Tal análise tem algumas vantagens relativamente às abordagens que propõem a monoracionalidade da estrutura. Em primeiro lugar, explica por que razão se observam as mesmas relações temporais entre a oração encaixada e a matriz, exista ou não reestruturação; veja-se o contraste entre (67), sem reestruturação, e (68), com reestruturação:

- (67) a. *O presidente não quis promulgá-lo amanhã.
 b. O presidente não quis promulgá-lo no dia seguinte.
- (68) a. *O presidente não o quis promulgar amanhã.
 b. O presidente não o quis promulgar no dia seguinte.

Em segundo lugar, dá conta da diferença entre verbos de reestruturação e verbos auxiliares, como mostraremos na secção seguinte.

3.2.2 Verbos de reestruturação e verbos auxiliares

Como referido na Introdução, alguns verbos que admitem reestruturação foram classificados como auxiliares em diversas gramáticas do português, a par do verbo

¹⁷ Note-se que, em alguns casos, a subordinada parece ter uma interpretação de anterioridade mesmo que o verbo matriz tenha uma orientação temporal especificada para posterioridade:

- (i) O João queria ter ido ao cinema.

No entanto, as frases deste tipo só são gramaticais quando o verbo matriz ocorre no pretérito imperfeito, que assume, aqui, um valor modal (de contrafactualidade, neste caso). Assim, frases como (i) podem ser seguidas de *...mas não* [verbo encaixado]. Se esse valor não estiver presente, o resultado é agramatical. Compare-se (ii) com (iii):

- (ii) O João queria ter ido ao cinema (mas não foi).
 (iii) *O João quis ter ido ao cinema (mas não foi).

ter, auxiliar dos tempos compostos, e de *ser*, auxiliar da passiva (veja-se, em particular, Soares Barbosa 1822; Said Ali 1908; 1931; Epiphany da Silva Dias 1918; Bechara 1961, ³⁷1999; Cunha/Cintra 1984). Um dos argumentos para tal classificação consiste na possibilidade de subida de clítico, propriedade que identifica predicados complexos. Com efeito, no contexto dos verbos auxiliares acima referidos, os clíticos dependentes do verbo principal, que ocorre no particípio passado, associam-se ao verbo auxiliar, o que, em parte, pode ser considerado como subida de clítico:

- (69) a. A mãe do Zé só o tem visto aos fins de semana.
 b. Não me foram apresentadas provas concretas de falsificação de documentos.

Note-se, porém, que, contrariamente ao que acontece na construção de reestruturação, o clítico não pode associar-se aos verbos principais em (70):

- (70) a. *A mãe do Zé só tem visto-o aos fins-de-semana.
 b. *Foram apresentadas-me provas concretas de falsificação de documentos.

Um outro aspeto que pode aproximar os verbos de reestruturação dos auxiliares é a impossibilidade de negação do domínio do segundo verbo:

- (71) *A Teresa está tão aborrecida com o Miguel que só o quer não ver.

- (72) a. *A Teresa tem não saído com o Zé.
 b. *A documentação foi não entregue dentro do prazo legal de candidaturas.

No entanto, também neste aspeto há que distinguir entre verbos auxiliares e verbos de reestruturação. Com efeito, estes últimos podem ocorrer em contextos em que não se verifica reestruturação e, neste caso, a negação do domínio infinitivo produz um resultado gramatical (cf. (73)); pelo contrário, com os verbos auxiliares apenas é possível um operador de negação frásica, que precede obrigatoriamente o complexo <verbo auxiliar – verbo principal> (veja-se o contraste entre (72) e (74)).

- (73) A Teresa está tão aborrecida com o Miguel que só quer não o ver.

- (74) a. A Teresa não tem saído com o Zé.
 b. A documentação não foi entregue dentro do prazo legal de candidaturas.

As diferenças acima ilustradas constituem, assim, um argumento a favor da distinção entre verbos de reestruturação e verbos auxiliares. Em particular, mostram que os primeiros ocorrem em domínios bioracionais e os segundos, em domínios monoracionais. O facto de, quando se verifica reestruturação, existirem algumas propriedades que parecem apontar para a monoracionalidade da estrutura decorre da possibilidade de estes verbos selecionarem complementos defetivos, sintática e semanticamente, como defendido em Gonçalves (1999) e Gonçalves/Cunha/Silvano (2010), entre outros.

3.3 Síntese

No Quadro 1, sintetizam-se as propriedades apresentadas nas secções anteriores que caracterizam os predicados complexos de reestruturação em português europeu:

Quadro 1: Síntese das propriedades da construção de reestruturação em português europeu.

Evidência de formação de predicado complexo	Possibilidade de subida de clítico	<i>Os rapazes não <u>q</u> quiseram conhecer.</i>
	Movimento longo do objeto	<i>Quiseram-se salvar os móveis.</i>
Evidência / características de construção de reestruturação	Reduzida autonomia sintática do complemento infinitivo / forte grau de sintatização	- restrições sobre negação (em contexto de subida de clítico): * <i>Eles só a queriam não enfrentar.</i> - restrições sobre deslocação: * <i>Enfrentar, os alunos não a quiseram.</i> - impossibilidade de parentética entre os dois verbos: * <i>Eles queriam-na – mas sabiam que não deviam – enfrentar.</i>
	Bioracionalidade	- apesar de restrições sobre as dependências temporais entre a oração matriz e a oração encaixada, manutenção de temporalidade no complemento infinitivo: <i>O presidente não o quis promulgar no dia seguinte.</i>
	Verbos matriz	Subconjunto de verbos de controlo: <i>querer, desejar, conseguir, tentar.</i> Subconjunto de verbos de elevação: semiauxiliares modais (<i>poder, dever</i>), aspetuais (<i>estar (a), continuar (a), começar (a)</i>), temporais (<i>ir</i>).

4 A construção *fazer-Inf* em português europeu

O segundo tipo de construção de predicado complexo que aqui consideramos ocorre, como vimos, em contextos de complementação infinitiva de verbos causativos (como *mandar, deixar, fazer*) e percetivos (como *ver, ouvir, sentir*). Admitindo estes verbos, alternativamente, complementos oracionais de outro formato (cf. adiante, secção 4.2), interessa-nos, num primeiro momento, considerar exclusivamente os contextos que, com estes verbos, dão lugar à formação de um predicado complexo em português europeu, para caracterizarmos mais finamente as suas propriedades distintivas em relação à reestruturação. Restringimos pois o âmbito da primeira secção aos contextos em que o verbo encaixado é um infinitivo não flexionado e que permitem subida de clítico e movimento longo de objeto (cf., acima, 2.1 e 2.2). Em seguida, na secção 4.2, são apresentados alguns aspetos de comparação entre os diferentes tipos de comple-

mentos dos verbos causativos e perceptivos, integrando dados de variação intra e interlinguística.

4.1 Propriedades distintivas dos predicados complexos de tipo *fazer-Inf*

Das diferenças entre os dois tipos de predicados complexos, destacámos, na secção 2.2, as seguintes características identificadoras de *fazer-Inf*: além do requisito de subida do clítico, comum à reestruturação, também restrições sobre a natureza dos clíticos que ocorrem em *fazer-Inf*; a impossibilidade de o infinitivo ser passivizado; e a alteração à ordem de constituintes do domínio infinitivo, associada a uma reatribuição de funções sintáticas. Estas propriedades correlacionam-se com uma menor autonomia do complemento infinitivo nestas construções, que resulta numa «amálgama de duas orações numa só» (Barbosa/Raposo 2013, 1963) – caracterização na base da designação «união de orações» também atribuída a este tipo de predicado complexo (Raposo 1981). Neste caso, os dois verbos formam um núcleo verbal complexo, com consequências na ordenação e nas funções sintáticas dos constituintes da frase, assim como na limitação de outras manifestações sintáticas, sintetizadas de seguida.

A família das línguas românicas manifesta assimetrias relativamente à ocorrência de pronomes reflexos de interpretação dependente do causado no domínio infinitivo. Assim, na construção *fazer-Inf* o clítico ocorre obrigatoriamente no domínio infinitivo em francês, mas não em italiano nem em português europeu (os exemplos do francês e do italiano são retirados de Guasti 1993):

- (75) a. Il faisait se réveiller la fille.
 b. *Lui faceva svegliarsi la ragazza.
 c. *O João mandou barbear-se o Pedro.

No contexto de subida de clítico caracterizador de *fazer-Inf*, verifica-se a mesma restrição sobre pronomes reflexos (e inerentes, como vimos na secção 2.2) em português europeu.

- (76) *O João mandou-se_i barbear o Pedro_i.

A construção *fazer-Inf* distingue-se da reestruturação também pelo facto de o verbo infinitivo não poder ocorrer na forma passiva (cf. contraste em (41)–(42), aqui repetido como (77)–(78)):

- (77) a. O Pedro já conseguiu [ser apresentado à Maria (pelo Mário)].
 b. O Pedro já lhe conseguiu ser apresentado (pelo Mário).

- (78) a. *O Pedro já mandou [ser apresentado o João à Maria (pelo Mário)].
 b. *O Pedro já lhe mandou ser apresentado o João.

A passivização do verbo superior, envolvendo ou não *se* passivo (e movimento longo de objeto), pode no entanto ocorrer, realizando-se como sujeito o argumento interno do verbo infinitivo.

- (79) a. Estas casas foram mandadas construir a arquitetos famosos.
 b. Mandaram-se construir algumas casas a arquitetos famosos.

A propriedade mais específica dos predicados complexos *fazer-Inf* diz respeito à reorganização de funções sintáticas no complemento infinitivo, que afeta, particularmente, o constituinte classicamente designado como sujeito, a que temos vindo a chamar causado. Este constituinte, ao contrário dos sujeitos típicos, realiza-se como um sintagma preposicional em contextos de verbo transitivo (cf. (81)) e pode ser substituído por um clítico acusativo (em (80)) ou por um clítico dativo (em (81)), que correspondem, respetivamente, ao objeto direto e ao objeto indireto:

- (80) a. O treinador mandou correr os atletas. → O treinador mandou-os correr.
 b. Os pais do Luís ouviram cantar a Maria Callas. → Os pais do Luís ouviram-na cantar.
- (81) a. O Luís mandou arranjar o carro da mulher ao mecânico mais antigo da cidade.
 → O Luís mandou-lhe arranjar o carro da mulher.
 b. ?Os pais do Luís ouviram cantar La Traviata à Maria Callas.¹⁸
 → ?Os pais do Luís ouviram-lhe cantar La Traviata.

Relativamente a esta questão, encontram-se na literatura dois tipos de abordagens: aquelas em que o causado é caracterizado como o sujeito do verbo no infinitivo (cf. Kayne 1975; Burzio 1986; Villalba 1992; Guasti 1993; 1997; Folli/Harley 2007; entre outros) e aquelas em que o mesmo é reanalisado em termos da sua função sintática (cf. Raposo 1981; Alsina 1996; Gonçalves 1999; entre outros).

A ideia de que o causado, na construção *fazer-Inf*, corresponde ao sujeito, embora, por razões que não se exploram no presente capítulo, apresente propriedades que não são típicas desta função sintática, encontra fundamento em diferentes factos. Em primeiro lugar, é semanticamente dependente do verbo encaixado, o que explica a agramaticalidade de (82b):

18 Os falantes de português europeu manifestam alguma divergência em relação à gramaticalidade deste tipo de sequência. Como complemento de verbos percetivos, o infinitivo não flexionado com alteração de funções sintáticas é por vezes sentido como marginal para alguns falantes quando o verbo infinitivo é transitivo (cf. também Barbosa/Raposo 2013, 1962, n.79).

- (82) a. O professor fez ler o livro ao Pedro.
 b. *O professor fez ler o livro aos cães.

Note-se que, no contexto do mesmo verbo causativo, *fazer*, é possível a ocorrência do constituinte que gera a agramaticalidade em (82b), *aos cães*, se se alterar o verbo encaixado:

- (83) O professor fez comer a ração aos cães.

Adicionalmente, em algumas línguas, como o italiano, o causado pode qualificar-se como antecedente de expressões anafóricas que ocorrem na posição de objeto direto, como os pronomes reflexos (cf. (84a)), ou fixar a referência do sujeito nulo de frases encaixadas selecionadas pelo verbo do complemento infinitivo (cf. (84b)), propriedades que partilha com o sujeito típico:

- (84) a. Con le minacce, fecero accusare se stesso_i a Giovanni_i. (Burzio 1986)
 b. Ho fatto affermare di [-]_i averla visto a Giovanni_i. (Burzio 1986)

Os exemplos do português europeu correspondentes a (84) são, porém, agramaticais:

- (85) a. *O João {mandou lavar-se_i/mandou-se_i lavar} (a)os meninos_i.
 b. *O Miguel mandou dizer [-]_i ter visto a Ana ao Zé_i.

Adicionalmente, em português o causado apresenta outras propriedades que o distinguem do sujeito típico do verbo infinitivo, aproximando-o de um objeto (objeto direto ou objeto indireto):

- i. se o verbo encaixado for transitivo, o causado não pode estar associado a quantificadores flutuantes (como *todos* em (86)):

- (86) *Aos meninos, o professor mandou ler todos um livro.

Note-se que em contextos não causativos, os sujeitos podem estar associados a quantificadores flutuantes (cf. (87)), mas os objetos indiretos, não (cf. (88)):

- (87) Os meninos leram todos um livro.

- (88) *Aos meninos, a Marta deu as bolachas todos.

- ii. quando o objeto direto integra um quantificador distributivo, pode ser antecedente local de expressões de posse que ocorram no causado:

- (89) O Presidente mandou defender [cada proposta]_i ao seu_i autor.

Note-se que em contextos não causativos, o objeto direto não pode ser antecedente local de expressões de posse que ocorram no sujeito (cf. (90)), mas pode ser antecedente de expressões que ocorram no objeto indireto (cf. (91)):

- (90) *O seu_i autor defendeu [cada proposta]_i.
 (91) O professor deu [cada livro]_i ao seu_i autor.

4.2 *Fazer-Inf* e outros complementos de verbos causativos e percetivos em português europeu

Em português europeu os verbos causativos e percetivos apresentam um leque variado de complementos oracionais que constituem alternativas à formação de um predicado complexo de tipo *fazer-Inf* (cf. (92)): complementos finitos (ex. (93)), de infinitivo flexionado (ex. (94)), de infinitivo não flexionado com ordem sujeito-verbo (ex. (95)) – cf. e.g. Silva 1999; Gonçalves/Duarte 2001; Duarte/Gonçalves 2002; Martins 2004; Barbosa/Raposo 2013.¹⁹

- (92) a. A presidente da Assembleia mandou votar a lei aos deputados.
 b. [?]A presidente da Assembleia viu votar a lei aos deputados.²⁰
- (93) a. A presidente da Assembleia mandou que os deputados votassem a lei.
 b. A presidente da Assembleia viu que os deputados votaram a lei.
- (94) a. A presidente da Assembleia mandou os deputados votarem a lei.
 b. A presidente da Assembleia viu os deputados votarem a lei.
- (95) a. A presidente da Assembleia mandou todos os deputados votar a lei.
 b. A presidente da Assembleia viu os deputados votar a lei.

Os verbos percetivos podem ainda seleccionar um complemento infinitivo (flexionado ou não) introduzido pela preposição *a*, com a mesma distribuição do gerúndio, nos dialetos em que este ocorre (≈18 O gerúndio flexionado no português dialetal).²¹ Neste caso, o complemento infinitivo assume o valor aspetual de progressivo.

- (96) a. A presidente da Assembleia viu todos os deputados a votar(em) a lei.
 b. A presidente da Assembleia viu todos os deputados votando a lei.

¹⁹ Para outras variantes não padrão, ver Pereira (2012; 2015).

²⁰ Cf. nota 18.

²¹ Raposo (1989) designa esta construção como «Construção de Infinitivo Preposicionado» (*Prepositional Infinitive Construction*). Relativamente ao elemento que introduz o complemento infinitivo, *a*, existem divergências quanto à sua classificação: preposição (Raposo 1989) ou marcador aspetual (Duarte 1993; Gonçalves/Duarte 2001; Barbosa/Cochofel 2005).

De todos estes contextos, apenas (92) integra um predicado complexo do tipo *fazer-Inf*. Com efeito, considerando que a subida de clítico é um indicador da existência deste tipo de predicados, verifica-se que o clítico correspondente ao constituinte *a lei* (objeto direto do verbo *votar*) só pode associar-se ao verbo causativo ou ao verbo percetivo em (92), dando origem a (97), que contrasta com (98)–(100).

- (97) a. A presidente da Assembleia mandou-a votar aos deputados
 b. ?A presidente da Assembleia viu-a votar aos deputados.²²
- (98) a. *A presidente da Assembleia mandou-a que os deputados votassem.
 b. *A presidente da Assembleia viu-a que os deputados votaram.
- (99) a. *A presidente da Assembleia mandou-a os deputados votarem.
 b. *A presidente da Assembleia viu-a os deputados votarem.
- (100) a. *A presidente da Assembleia mandou-a os deputados votar.
 b. *A presidente da Assembleia viu-a os deputados votar.
- (101) *A presidente da Assembleia viu-a os deputados a votar(em).

Como se mostrou na secção 2.2, a existência de predicados complexos de tipo *fazer-Inf* não é um traço específico do português europeu. Pelo contrário, trata-se de uma construção que caracteriza transversalmente as línguas românicas, em particular na expressão de uma causativa (também referida «causativa românica», cf. Guasti 1997).

A este respeito, no seio das línguas românicas, a especificidade do português advém da diversidade de estruturas que apresenta. Em particular, o português europeu permite, a par de *fazer-Inf*, o tipo de complementos infinitivos flexionados exemplificados em (94), uma possibilidade que não se encontra igualmente disponível noutras línguas, nem mesmo em galego, que apresenta infinitivos flexionados noutros contextos (cf. Álvarez/Xove 2002, 308; Carrilho/Sousa 2010):

- (102) a. Ana fixo saír os alumnos. / Ana fixo os alumnos saír.
 b. *Ana fixo os alumnos saíren. (galego; in Carrilho/Sousa 2010, 4, ex. (7))

Além disso, o português (neste caso, a par do galego) dispõe também da construção com infinitivo não flexionado e com ordem sujeito-verbo no domínio deste infinitivo (cf. (95); em (102) exemplifica-se a mesma possibilidade em galego). Pelo contrário, em italiano, espanhol e francês, esta ordem é possível com os correspondentes do

22 Cf. nota 18.

verbo causativo *deixar* e dos verbos perçetivos (cf. (103) e (104)), mas não com os restantes causativos (cf. (105)).²³

- (103) a. Gianni ha lasciato Maria mangiare la mela.
 b. Deja el agua correr.
 c. Je laisserai Marie parler.

- (104) a. Gianni ha visto Carlo mangiare la mela.
 b. Veo (a) la televisión funcionar.
 c. J'ai vu Marie parler.

- (105) a. *Gianni ha fatto Maria parlare.
 b. *Juan hizo la televisión funcionar.²⁴
 c. *Jean a fait son amie partir.

Estes complementos infinitivos não flexionados em português apresentam ambiguidade com instâncias da construção *fazer-Inf* quando o causado é pronominalizado, uma vez que este é casualmente marcado como acusativo, numa estrutura designada de «Marcação Excepcional de Caso» (pelo verbo matriz). No caso de *fazer-Inf*, o causado/sujeito também pronominaliza em forma de acusativo quando o verbo infinitivo é intransitivo, o que potencia a ambiguidade:

- (106) a. A presidente da Assembleia mandou-os entrar.
 [=A presidente da Assembleia mandou os deputados entrar]
 [=A presidente da Assembleia mandou entrar os deputados] (*fazer-Inf*)
 b. A presidente da Assembleia viu-os entrar.
 [=A presidente da Assembleia viu os deputados entrar]
 [=A presidente da Assembleia viu entrar os deputados] (*fazer-Inf*)

Distinguem-se, no entanto, nas formas pronominais de terceira pessoa semanticamente dependentes de verbos infinitivos transitivos (neste contexto, um causado/ /sujeito realiza-se pronominalmente como dativo na construção de *fazer-Inf* (107b/d) e como acusativo na construção de infinitivo não flexionado (107a/c)):

- (107) a. A presidente da Assembleia mandou-os votar a lei.
 [=A presidente da Assembleia mandou os deputados votar a lei]

²³ Sobre a proximidade estrutural entre os complementos infinitivos de causativas e perçetivas em português e em inglês (*John saw/heard/made her hit Fred.*), veja-se Hornstein/Martins/Nunes (2008).

²⁴ Em espanhol, em alguns contextos, o causado pode ocorrer em posição pré-verbal, sendo introduzido pela preposição *a*:

- (i) Juan hizo a la radio funcionar. (Torrego 1998)
 (ii) La entrenadora hizo a la atleta repetir el ejercicio. (Torrego 2010)

Há, no entanto, algumas restrições fortes sobre esta possibilidade; ver Torrego (1998; 2010) e Hernanz (1999).

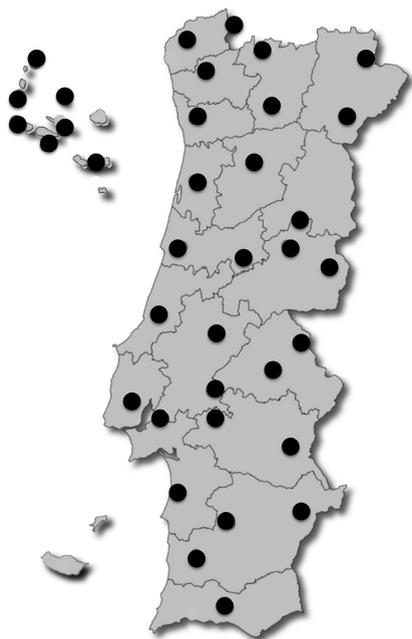
- b. A presidente da Assembleia mandou-lhes votar a lei.
[=A presidente da Assembleia mandou votar a lei aos deputados] (*fazer-Inf*)
- c. A presidente da Assembleia viu-os votar a lei.
[=A presidente da Assembleia viu os deputados votar a lei]
- d. ²A presidente da Assembleia viu-lhes votar a lei.
[=A presidente da Assembleia viu votar a lei aos deputados] (*fazer-Inf*)

As duas construções apresentam diferenças relacionadas com a diversidade estrutural de *fazer-Inf* e de construções de infinitivo não flexionado, evidenciadas, a título de exemplo, em (108), pelo comportamento divergente que as construções causativas exibem em relação à presença da negação no domínio infinitivo:²⁵

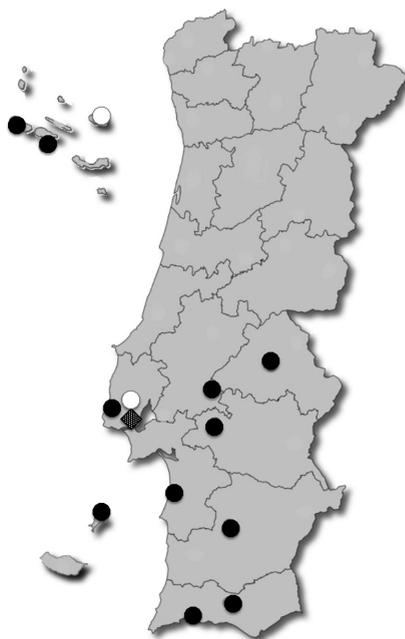
- (108) a. *A presidente da Assembleia mandou não votar a lei aos deputados. / *A presidente da Assembleia mandou-lhes não votar a lei. (*fazer-Inf*)
- b. A presidente da Assembleia mandou os deputados não votar a lei. / A presidente da Assembleia mandou-os não votar a lei.

Para concluir, consideram-se alguns dados empíricos da distribuição destas diferentes construções em português europeu. Contrastando com outras línguas românicas, destaca-se normalmente a prevalência, em português europeu, de construções diferentes de *fazer-Inf*, em especial a construção de «Marcação Excepcional de Caso», ou seja, aquela em que o infinitivo não flexionado ocorre com um causado preverbal cliticizável generalizadamente em acusativo (cf. e.g. Silva 1999; Gonçalves/Duarte 2001; Sousa 2004). Investigação recente (Pereira 2012), sobre dados de variedades regionais (e sociais) diferentes da variedade padrão do português europeu, reunidos no corpus dialetal CORDIAL-SIN, revelam no entanto que esta prevalência não deve ser homogeneamente entendida. Por um lado, Pereira (2012) encontra predomínio de *fazer-Inf* como complemento de verbos causativos. Em contextos de verbos perceptivos, apenas a construção de infinitivo preposicionado a ultrapassa. Por outro lado, a distribuição geográfica das variantes encontradas (cf. contraste entre Mapas 1 e 2) permite evidenciar uma área meridional e açoriana – uma área que contém o centro geográfico difusor da variedade padrão –, na qual se concentram as construções que diferem de *fazer-Inf*.

²⁵ No caso dos verbos perceptivos a presença de negação produz efeitos agramaticais mesmo em contexto em que não se forma um predicado complexo, por razões decorrentes das propriedades semânticas destes verbos.



Mapa 1: Distribuição geográfica de *fazer-Inf* com verbos causativos no CORDIAL-SIN (Pereira 2012).



Mapa 2: Distribuição geográfica de construções diferentes de *fazer-Inf* com verbos causativos no CORDIAL-SIN (Pereira 2012).

5 Conclusões gerais

Identificadas as principais características sintático-semânticas da formação de predicados complexos verbais em português europeu, destaca-se que:

- (i) o português europeu apresenta os dois tipos de predicados complexos sintáticos identificados nas línguas naturais, predicados complexos de reestruturação e de tipo *fazer-Inf*;
- (ii) como descrito na literatura especializada, estas duas construções apresentam diferenças sintáticas também em português europeu, sendo a mais relevante a menor autonomia sintática dos domínios infinitivos que ocorrem em predicados complexos de tipo *fazer-Inf*;
- (iii) o português europeu regista, a este respeito, importantes diferenças em relação ao português brasileiro, que não apresenta predicados complexos;
- (iv) no seio das línguas românicas, o português europeu se distingue por ter disponíveis os dois tipos de predicados complexos e pela diversidade de verbos que, de modo mais generalizado, entram em cada uma das construções;

- (v) no seio das línguas românicas, o português europeu se distingue pela diversidade de construções que concorrem com um predicado complexo (*fazer-Inf*) em construções com verbos causativos e percetivos;
- (vi) no domínio da variação interna ao português europeu, existe uma maior diversidade de verbos que admitem reestruturação;
- (vii) considerada a distribuição geográfica de construções alternativas a um predicado complexo *fazer-Inf*, existe uma oposição entre o predicado complexo, mais generalizadamente distribuído por todo o território, e outras construções, que se encontram geograficamente concentradas numa área de predomínio meridional.

6 Referências

- Alsina, Alex (1996), *The Role of Argument Structure in Grammar. Evidence from Romance*, CSLI Lecture Notes 62, Stanford, CSLI Publications.
- Alsina, Alex (1997), *Causatives in bantu and Romance*, in: Alex Alsina/Joan Bresnan/Peter Sells (edd.), *Complex predicates*, CSLI Lecture Notes 64, Stanford, CSLI Publications, 203–246.
- Álvarez, Rosario/Xove, Xosé (2002), *Gramática da Língua Galega*, Galáxia, Vigo.
- Baker, Mark (1988), *Incorporation*, Chicago, University of Chicago Press.
- Baker, Mark (1996), *The Polysynthesis Parameter*, Oxford, Oxford University Press.
- Barbosa, Pilar/Cochofel, Fátima (2005), *O Infinitivo Preposicionado em PE*, in: Inês Duarte/Isabel Leiria (edd.), *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 387–400.
- Barbosa, Pilar/Raposo, Eduardo B. Paiva (2013), *Subordinação verbal infinitiva*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 2, cap. 37, Lisboa, Fundação Calouste Gulbenkian, 1899–1977.
- Bech, Gunnar (¹1955–1957, ²1983), *Studien über das deutsche Verbum infinitivum*, Tübingen, Niemeyer.
- Bechara, Evanildo (¹1961, ³⁷1999), *Nova Gramática Portuguesa*, São Paulo, Companhia Editora Nacional.
- Bok-Bennema, Reineke (2006), *Clitic Climbing*, in: Henk van Riemsdijk et al. (edd.), *The Blackwell Companion to Syntax*, vol. I–V, *Blackwell Handbooks in Linguistics*, Malden, Massachusetts, Blackwell Publishing, 469–518.
- Bok-Bennema, Reineke/Kampers-Manhe, Brigitte (1994), *Transparency effects in the Romance languages*, in: Michael L. Mazzola (ed.), *Issues and theory in Romance linguistics*, Washington, D.C., Georgetown University Press, 199–217.
- Bordelois, Ivonne (1988), *Causatives: from Lexicon to Syntax*, *Natural Language and Linguistic Theory* 6, 57–93.
- Burzio, Luigi (1986), *Italian Syntax: a Government-Binding approach*, Dordrecht, Kluwer.
- Carrilho, Ernestina/Lobo, Maria (1999), *Variação sintáctica: alguns aspectos*, comunicação apresentada no ciclo de conferências *Conversas d’HorAl* organizado pelo CLUL, 15 Junho 1999.
- Carrilho, Ernestina/Sousa, Xulio (2010), *Embedded subjects of causative infinitival constructions in Galician and Portuguese*, *6th International Contrastive Linguistics Conference*, Freie University Berlin, October 2010, ms., http://www.clul.ul.pt/files/ernestina_carrilho/CarrilhoSousa2010tx.pdf (21.04.2016).
- CETEMPúblico – *Corpus de Extratos de Textos Eletrónicos MCT/Público*, <http://www.linguateca.pt/CETEMPUBLICO/> (21.04.2016).

- Chomsky, Noam (1981), *Lectures on Government and Binding*, Dordrecht, Foris.
- Cinque, Guglielmo (2006), *Restructuring and Functional Heads. The Cartography of Syntactic Structures*, vol. 4, New York, Oxford University Press.
- CORDIAL-SIN – *Syntax-oriented Corpus of Portuguese Dialects*, Lisboa, CLUL, <http://www.clul.ul.pt> (13.05.2016).
- CRPC – *Corpus de Referência do Português Contemporâneo*, Lisboa, CLUL, <http://alfclul.clul.ul.pt/CQPweb> (21.04.2016).
- Cunha, Celso/Cintra, Luís Filipe Lindley (1984), *Nova Gramática do Português Contemporâneo*, Lisboa, Sá da Costa.
- Cyrino, Sonia (2008), *On the lack of complex predicates in Brazilian Portuguese*, comunicação apresentada no *18th Colloquium on Generative Grammar*, Universidade de Lisboa.
- Cyrino, Sonia (2010), *On Romance syntactic complex predicates: why Brazilian Portuguese is different*, *Estudos da Língua(gem)* 8:1, 187–222.
- Declerck, Renaat (1991), *Tense in English: Its Structure and Use in Discourse*, Londres/Nova Iorque, Routledge.
- Duarte, Inês (1993), *Complementos infinitivos preposicionados e outras construções temporalmente defectivas em Português Europeu*, in: *Actas do VIII Encontro da Associação Portuguesa de Linguística*, Lisboa, APL, 145–158.
- Duarte, Inês (2003), *Subordinação completiva – as orações completivas*, in: Maria Helena Mira Mateus et al. (edd.), *Gramática da Língua Portuguesa*, Lisboa, Caminho, 593–651.
- Duarte, Inês (2013), *Construções ativas, passivas, incoativas e médias*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 1, cap. 13, Lisboa, Fundação Calouste Gulbenkian, 428–458.
- Duarte, Inês/Gonçalves, Anabela (2002), *Construções de subordinação funcionalmente defectivas: o caso das construções perceptivas em português europeu e em português brasileiro*, in: Anabela Gonçalves/Clara Nunes Correia (edd.), *Actas do XVII Encontro da Associação Portuguesa de Linguística*, Lisboa, APL, 161–173.
- Duarte, Inês/Gonçalves, Anabela/Santos, Ana Lúcia (2012), *Infinitivo flexionado, independência temporal e controlo*, in: Armanda Costa/Cristina Flores/Nélia Alexandre (edd.), *Textos Seleccionados, XXVII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 217–234.
- Epiphany da Silva Dias, Augusto (1918), *Syntaxe Histórica Portuguesa*, Lisboa, Livraria Clássica Editora.
- Fiéis, Alexandra/Madeira, Ana Maria (2012), *Predicados de controlo na diacronia do Português*, in: Armanda Costa/Cristina Flores/Nélia Alexandre (edd.), *Textos Seleccionados, XXVII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 271–284.
- Folli, Raffaella/Harley, Heidi (2007), *Causation, obligation and argument structure: On the nature of little «v»*, *Linguistic Inquiry* 38:2, 197–238.
- Gonçalves, Anabela (1992), *Para uma sintaxe dos verbos auxiliares em português europeu*, dissertação de Mestrado, Faculdade de Letras da Universidade Lisboa.
- Gonçalves, Anabela (1999), *Predicados Complexos Verbais em Contextos de Infinitivo não Preposicionado do Português Europeu*, tese de Doutoramento, Universidade de Lisboa.
- Gonçalves, Anabela (2001), *Predicados complexos com verbos causativos e perceptivos do Português europeu*, in: Clara Nunes Correia/Anabela Gonçalves (edd.), *Actas do XVI Encontro da Associação Portuguesa de Linguística*, Lisboa, APL, 227–239.
- Gonçalves, Anabela (2014), *Temporal properties of infinitival domains in EP: Data from inflected infinitives in obligatory subject control contexts*, comunicação convidada apresentada no *First Seminar on Control and Finiteness at USP*, Universidade de São Paulo, 30–31 de julho e 1 de agosto de 2014.

- Gonçalves, Anabela/Duarte, Inês (2001), *Construções Causativas em Português Europeu e em Português Brasileiro*, in: Anabela Gonçalves/Clara Nunes Correia (edd.), *Actas do XVI Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL/Colibri, 657–671.
- Gonçalves, Anabela/Cunha, Luís Filipe/Silvano, Purificação (2010), *Interpretação temporal dos domínios infinitivos na construção de reestruturação do Português europeu*, in: Ana Maria Brito et al. (edd.), *Textos Seleccionados do XXV Encontro Nacional da Associação Portuguesa de Linguística*, Porto, APL, 435–447.
- Gonçalves, Anabela/Santos, Ana Lúcia/Duarte, Inês (2014), *(Pseudo-)Inflected infinitives and Control as Agree*, in: Karen Lahousse/Stefania Marzo (edd.), *Romance Languages and Linguistic Theory 2012. Selected papers from «Going Romance» Leuven 2012*, Amsterdam/Philadelphia, Benjamins, 161–180.
- Gonçalves, Perpétua (2015), *Aspetos morfossintáticos da gramática do Português de Moçambique: a concordância nominal e verbal*, Cuadernos de la ALFAL 7, 9–16.
- Guasti, Maria Teresa (1993), *Causative and perception verbs*, Turin, Rosenberg and Sellier.
- Guasti, Maria Teresa (1997), *Romance causatives*, in: Liliane Haegeman (ed.), *The New Comparative Syntax*, London/New York, Longman, 124–144.
- Hernanz, Maria Lluïsa (1999), *El infinitivo*, in: Ignacio Bosque/Violeta Demonte (edd.), *Gramática Descriptiva de la Lengua Española*, Madrid, Espasa Calpe, 2197–2356.
- Hornstein, Norbert/Martins, Ana Maria/Nunes, Jairo (2008), *Perception and Causative Structures in English and European Portuguese: Φ -feature Agreement and the Distribution of Bare and Prepositional Infinitives*, *Syntax* 11:2, 205–229.
- Kamp, Hans/Reyle, Uwe (1993), *From Discourse to Logic – Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht, Kluwer.
- Kayne, Richard (1975), *French Syntax: the Transformational Cycle*, Cambridge MA, MIT Press (trad. francesa: *Syntaxe du Français. Le cycle transformationnel*, Paris, Éditions du Seuil, 1977).
- Magro, Catarina (2005), *Introduutores de orações infinitivas – o que diz a sintaxe dos clíticos*, in: Inês Duarte/Isabel Leiria (edd.), *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística, 649–664.
- Martins, Ana Maria (1995), *A Minimalist Approach to Clitic Climbing*, in: Audra Dainora et al. (edd.), *Papers from the 31st Regional Meeting of the Chicago Linguistic Society*, vol. 2: *Parasession on Clitics*, Chicago, Chicago Linguistic Society, 215–233.
- Martins, Ana Maria (2004), *Ambiguidade estrutural e mudança linguística: a emergência do infinitivo flexionado nas orações complemento de verbos causativos e perceptivos*, in: Ana Maria Brito/Olívía Figueiredo/Clara Barros (edd.), *Linguística Histórica e História da Língua Portuguesa. Actas do Encontro de Homenagem a Maria Helena Paiva*, Porto, Secção de Linguística do Departamento de Estudos Portugueses e de Estudos Românicos da Fac. Letras Univ. Porto, 197–225.
- Pereira, Sandra (2012), *Protótipo de um glossário dos dialetos portugueses com informação sintática*, tese de Doutoramento, Universidade de Lisboa.
- Pereira, Sandra (2015), *Causative and Perception Constructions in European Portuguese: the dialectal data*, *Dialectologia*, Special Issue V, Syntactic Variation in Western European Languages, 53–80.
- Raposo, Eduardo Paiva (1981), *A construção de «União de Orações» na Gramática do Português*, tese de Doutoramento, Universidade de Lisboa.
- Raposo, Eduardo Paiva (1989), *Prepositional Infinitival Constructions in European Portuguese*, in: Osvaldo Jaeggli/Ken Safir (edd.), *The Null Subject Parameter*, Dordrecht, Kluwer, 277–305.
- Rizzi, Luigi (1976), *Ristrutturazione*, *Revista di Grammatica Generativa* 1, 1–54.
- Rizzi, Luigi (1978), *A Restructuring rule in Italian Syntax*, in: Samuel J. Keyser (ed.), *Recent Transformational Studies in European Languages*, Cambridge MA, MIT Press, 113–158.
- Rizzi, Luigi (1982), *Issues in Italian Syntax*, Dordrecht, Foris.

- Roberts, Ian (1997), *Restructuring, Head Movement, and Locality*, *Linguistic Inquiry* 28, 423–460.
- Said Ali, Manuel (1908, ⁶1965), *Dificuldades da Língua Portuguesa*, Rio de Janeiro, Livraria Académica.
- Said Ali, Manuel (1931, ⁷1971), *Gramática Histórica da Língua Portuguesa*, Rio de Janeiro, Livraria Académica.
- Sapir, Edward (1921), *Language*, Nova Iorque, Harcourt Brace Jovanovich.
- Silva, Augusto Soares (1999), *A Semântica de Deixar. Uma contribuição para a Abordagem Cognitiva em Semântica Lexical*, Lisboa, Fundação Calouste Gulbenkian/MCT.
- Soares Barbosa, Jerónimo (1822, ⁵1871), *Grammatica Philosophica da Língua Portuguesa ou Principios da Grammatica Geral Aplicada Aplicados à Nossa Linguagem*, Lisboa, Typographia da Academia Real das Ciências.
- Sousa, Xulio (2004), *The expression of the causee in Galician, Portuguese, and Spanish*, in: Elsa González-Álvarez/Andrew Rollings (edd.), *Studies in contrastive linguistics*, Santiago de Compostela, Universidade de Santiago de Compostela, 585–591.
- Stowell, Tim (1982), *The tense of the infinitives*, *Linguistic Inquiry* 13, 561–570.
- Svenonius, Peter (2008), *Complex Predicates and the Functional Structure*, *Nordlyd* 35:1, 47–88.
- Torrego, Esther (1998), *The dependencies of objects*, Cambridge MA, MIT Press.
- Torrego, Esther (2010), *Variability in the Case Patterns of Causative Formation in Romance and Its Implications*, *Linguistic Inquiry* 41:3, 445–470.
- Vasconcelos, Sofia (2013), *O clítico SE: entre norma e variação*, dissertação de Mestrado, Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa.
- Villalba, Xavier (1992), *Incorporation, Case and Economy. A Principled Approach to Causative Constructions*, *Catalan Working Papers in Linguistics*, 345–389.
- Wurmbrand, Suzanne (2001), *Infinitives: Restructuring and clause structure*, Berlin/New York, Mouton de Gruyter.
- Wurmbrand, Suzanne (2004) *Two types of restructuring – Lexical vs. functional*, *Lingua* 114:8, 991–1014.
- Wurmbrand, Suzanne (2006), *Verb Clusters, Verb Raising, and Restructuring*, in: Martin Everaert/Henk van Riemsdijk (edd.), *The Blackwell Companion to Syntax*, Oxford, Blackwell, 227–341.
- Zubizarreta, Maria Luisa (1985), *The relation between Morphophonology and morphosyntax and the case of Romance Causatives*, *Linguistic Inquiry* 16, 247–289.
- Zubizarreta, Maria Luisa (1987), *Levels of Representation in the Lexicon and in the Syntax*, Dordrecht, Foris.

Maria Lobo

21 Sujeitos nulos: gramática do adulto, aquisição de L1 e variação dialetal

Abstract: Neste capítulo, descreve-se o estatuto do português europeu, comparativamente a outras línguas, relativamente ao parâmetro gramatical que legitima a omissão do sujeito em orações finitas. Pelas propriedades manifestadas, o português europeu padrão é caracterizado como uma língua de sujeito nulo clássica (ou consistente). Para além disso, apresenta diferenças interpretativas entre sujeitos nulos e sujeitos pronominais. São ainda referidos aspetos da aquisição desta propriedade que mostram que, embora as propriedades básicas das línguas de sujeito nulo sejam adquiridas cedo pelas crianças portuguesas, as diferenças interpretativas entre sujeitos nulos e sujeitos pronominais são de aquisição tardia. Finalmente, são referidos dados de variedades dialetais do português europeu, que mostram que estas variedades mantêm as propriedades características das línguas de sujeito nulo, apesar de, nalguns casos, haver um uso alargado de pronomes realizados.

Keywords: sujeito nulo, aquisição, variação, concordância, pronome

1 Introdução

Algumas línguas, entre as quais o português, permitem que, em orações finitas, o sujeito gramatical não esteja foneticamente realizado, como em (1), outras, como o francês, não permitem a omissão do sujeito (2):

- (1) a. Vamos ao parque.
b. Nós vamos ao parque.

- (2) a. *Allons au parc.
b. Nous allons au parc.

Esta é uma área de variação entre línguas que tem sido amplamente estudada nos estudos sintáticos e que, no modelo teórico de Princípios e Parâmetros da gramática generativa, tem sido atribuída a um parâmetro gramatical – o Parâmetro do Sujeito Nulo (Rizzi 1982; Jaeggli/Safir 1989a). Habitualmente, chamam-se línguas de sujeito nulo (ou línguas *pro-drop*)¹ as línguas que permitem a omissão do sujeito e línguas de sujeito obrigatório as línguas que não permitem a omissão do sujeito.

1 O termo *línguas pro-drop* surge por se considerar que nestas línguas se deixa cair (não se pronuncia) o sujeito pronominal.

Este fenómeno, que é aparentemente mais complexo do que aquilo que se pensava, tem sido considerado sob diferentes perspetivas: numa perspetiva sincrónica, comparando diferentes sistemas linguísticos; numa perspetiva diacrónica, procurando explicar mudanças entre sistemas; e do ponto de vista da aquisição da linguagem, procurando explicar de que forma o sistema gramatical da criança se aproxima do sistema da língua alvo. Também neste capítulo, iremos considerar o fenómeno sob diferentes perspetivas.

Para uma melhor compreensão do fenómeno, é necessário considerar, por um lado, diferentes tipos de sujeitos e, por outro lado, os contextos de alternância entre sujeito nulo e sujeito pronominal. Considerem-se as seguintes frases:

- (3) Parece que o tempo vai mudar.
- (4) Choveu durante toda a noite.
- (5) Estão a bater à porta.
- (6) Telefonei aos meus pais.

Todas estas frases têm sujeitos nulos. Repare-se que as frases correspondentes numa língua de sujeito obrigatório, como o francês, têm todas um sujeito pronominal realizado:

- (7) Il semble que le temps va changer.
- (8) Il a plu pendant toute la nuit.
- (9) On frappe à la porte.
- (10) J'ai téléphoné à mes parents.

Assim, podemos considerar que todas estas frases têm um sujeito,² ainda que apenas algumas línguas permitam que ele se encontre omissa. Contudo, ao contrário do que acontece em (1), nem todos os sujeitos nulos de (3)–(6) podem alternar com um pronome. Em (3), temos um sujeito não argumental, expletivo, sem conteúdo referencial. Em (4), temos um sujeito de um verbo meteorológico, que alguns autores consideram ser quase argumental, pelo facto de poder ocorrer um argumento cognato como *Choveu uma chuva miudinha*. Em (5), temos um sujeito argumental, mas arbitrário ou de interpretação indefinida. Em (6), temos um sujeito argumental com

² Nas gramáticas tradicionais portuguesas, considerava-se que frases como (3) e (4) não tinham sujeito.

referência específica. Em português europeu padrão, apenas em (6) pode haver alternância com um pronome, mantendo a mesma interpretação:³

- (11) (*Ele) parece que o tempo vai mudar.
 (12) (*Ele) choveu durante toda a tarde.
 (13) (*Eles) estão a bater à porta.⁴ Vai ver quem é.
 (14) (Eu) telefonei aos meus pais.

Estes diferentes tipos de sujeitos são relevantes para a descrição da variação interlinguística que caracteriza as línguas de sujeito nulo. Na realidade, nem todas as línguas que permitem a omissão do sujeito o fazem em todos os casos. Como mostra Rizzi (1986), não é possível estabelecer uma dicotomia simples entre línguas de sujeito nulo e línguas de sujeito obrigatório. Há línguas que permitem sujeitos nulos referenciais e expletivos, como é o caso do português, mas há também línguas que só permitem sujeitos nulos expletivos, como é o caso do islandês (Rizzi 1986) ou do cabo-verdiano (Costa/Pratas 2013):⁵

- (15) *(E) papia txeu.
 Ele fala muito
 (16) Sata txobe na Lisboa.
 PROG chove em Lisboa
 ‘Está a chover em Lisboa.’

Estas línguas são, por vezes, designadas como línguas de sujeito nulo parcial.⁶

Para além disso, algumas línguas que admitem sujeitos nulos parecem caracterizar-se mais como línguas de omissão de tópico do que como línguas de omissão de pronomes sujeito. É o caso do chinês, por exemplo, que, apesar de não ter morfologia flexional verbal, permite quer omissão de sujeitos expletivos (17a), quer omissão de sujeitos referenciais quando estes correspondem a um tópico (17b e 17c), quer omissão de objetos identificados por um tópico discursivo externo à oração (17d) (Huang 1984).⁷

³ No português dialetal é possível a presença do pronome *ele* em frases como (11) e (12) – cf. (24) e Carrilho (2005; 2009), entre outros.

⁴ A frase *Eles estão a bater à porta* é gramatical, mas deixa de ter uma leitura arbitrária.

⁵ Os exemplos (15) e (16) são retirados de Costa/Pratas (2013, 39–41).

⁶ Nas línguas de sujeito nulo parcial cabem línguas com comportamentos diferenciados: as línguas que só admitem sujeitos nulos em algumas formas do paradigma (*split pro-drop languages*), como parece ser o caso de alguns dialetos italianos e occitanos, e as línguas que só admitem sujeitos nulos em alguns contextos sintáticos restritos, como é o caso do cabo-verdiano (por vezes chamadas *línguas semi pro-drop*).

⁷ Alguns trabalhos referem-se a estas línguas como *radical pro-drop languages*.

- (17) a. (Tiān) kànshàngqù yào xiàyǔ-le (Wang et al. 1992)
(céu) parece vai chover-ASP
'Parece que vai chover.'
- b. lái-le (Huang 1984)
vem-ASP
'Ele/ela veio.'
- c. Zhāngsān_i, tā_i shuō [-]_i méi kànjiàn Lisi (Wang et al. 1992)
Zhangsan ele diz não vê Lisi
'Zhangsan, ele disse que não viu Lisi.'
- d. Zhāngsān_i shuō Lisi kànjiàn le (Huang 1984)
Zhangsan diz Lisi vê ASP
'Zhangsan_i disse que Lisi_j (o)_k viu.'

É, assim, possível identificar tipologicamente diferentes grupos de línguas em que o sujeito pode estar omissivo: línguas de sujeito nulo clássicas, línguas de sujeito nulo parcial, línguas de sujeito nulo radical. Na próxima secção, caracterizaremos de forma mais detalhada o português na tipologia das línguas de sujeito nulo.

2 O português na tipologia das línguas de sujeito nulo

As línguas de sujeito nulo clássicas,⁸ como mostram vários autores, são línguas que têm morfologia flexional rica (e.g. Taraldsen 1978; Huang 1984), em que se manifestam distinções de pessoa e número (e por vezes de género) concordando com o sujeito. Em Chomsky (1981), para além da possibilidade de omitir o sujeito (referencial ou expletivo), são associadas às línguas de sujeito nulo clássicas várias propriedades, entre as quais estão:

- i) a possibilidade de inversão do sujeito;
- ii) a ausência de efeitos *que-vestigio* (*that-trace effect*).⁹

⁸ Holmberg (2010) distingue quatro grupos tipológicos: i) língua de sujeito obrigatório; ii) língua de sujeito nulo radical (como o chinês); iii) língua de sujeito nulo consistente (como o italiano); e iv) língua de sujeito nulo parcial (como o cabo-verdiano). De acordo com esta tipologia, o português europeu será uma língua de sujeito nulo consistente e o português do Brasil será uma língua de sujeito nulo parcial. A generalização descritiva de Holmberg de que numa língua de sujeito nulo consistente não existe um sujeito pronominal arbitrário expresso é, contudo, contestável, uma vez que existem várias estratégias de indeterminação do sujeito em português europeu que recorrem a um pronome expresso, incluindo formas de segunda pessoa do singular (*tu*), de primeira pessoa do plural (*nós*) e a expressão pronominal *a gente*, geralmente associadas a contextos genéricos.

⁹ Chomsky (1981) refere ainda a possibilidade de extração longa do sujeito e a possibilidade de ter pronomes ressuntivos vazios em orações subordinadas. Contudo, alguns autores têm defendido que nem todas as correlações são igualmente válidas (Gilligan 1987).

Nas línguas de sujeito nulo clássicas, o sujeito pode ocorrer em posição pós-verbal em frases declarativas, por exemplo como resposta a uma interrogativa de sujeito (*Quem telefonou?*), como em (18), ao passo que nas línguas de sujeito obrigatório isso não é possível, nem mesmo com um sujeito expletivo realizado (19):

- (18) a. Telefonou a Ana.
 b. Ha telefonato Anna.
- (19) a. **(*There/It) has phoned Ann.*
 b. **(*Il) a téléphoné Anne.*

Nas línguas de sujeito nulo, pode ser extraído um sujeito do interior de uma oração completiva com o complementador *que* (20), ao passo que isso não é possível numa língua de sujeito obrigatório (21):

- (20) a. Quem pensas que __ escreveu este poema?
 b. Chi pensi che __ ha scritto questo poema?
- (21) a. **Who do you think that __ has written this poem?*
 b. **Qui crois-tu qu' __ a écrit ce poème ?*

O português europeu pode ser considerado uma língua de sujeito nulo clássica (ou consistente), uma vez que:

- i) tem morfologia de concordância rica (22);
- ii) tem sujeitos nulos referenciais (23a) e expletivos (23b);
- iii) admite inversão do sujeito (18a);
- iv) não manifesta efeitos *que*-vestigio (20a).¹⁰

(22) Vou / vais / vai / vamos / ides / vão

- (23) a. Estou em casa.
 b. Parece que vai chover.

Note-se que o pronome expletivo que ocorre em algumas variedades do português (24) não tem um funcionamento equivalente ao dos pronomes expletivos das línguas de sujeito obrigatório, podendo ser analisado como uma partícula discursiva, uma vez que pode coocorrer com sujeitos (24c) e pode ocupar posições periféricas (24d) (cf. Carrilho 2005; 2009):

10 Os efeitos de extração foram explicados segundo Rizzi (1982) por, numa língua de sujeito nulo, o sujeito poder ser extraído de uma posição pós-verbal. Barbosa (1995) defende que a posição básica do sujeito numa língua de sujeito nulo é a posição pós-verbal.

- (24) a. Ele há cada uma!
 b. Ele chovia que Deus a dava!
 c. Ele (a) folha do pinheiro é em bico (CORDIAL-SIN, ALC19)
 d. Ele quem foi que pôs aqui o arroz? (Carrilho 2005, 131)

Tal como nas línguas de sujeito nulo clássicas, em português europeu, a omissão do sujeito ocorre quer em frases principais quer em subordinadas (25):

- (25) Não sei se podes sair.

Múltiplas têm sido as análises para dar conta desta propriedade que distingue grupos de línguas (ver Jaeggli/Safir 1989b; Holmberg 2010; entre muitos outros) e é impossível retomá-las todas aqui. De acordo com as análises clássicas do Modelo de Princípios e Parâmetros, o sujeito nulo das línguas de sujeito nulo corresponde a uma categoria vazia pronominal sem realização fonética – *pro* – que se distingue do sujeito nulo das orações infinitivas. *pro* seria semelhante a um pronome salvo na matriz fonética (Chomsky 1981). *pro* seria legitimado apenas em línguas em que a Flexão tem propriedades morfológicas ricas ou, na hipótese de Jaeggli/Safir (1989a), em línguas com paradigmas morfológicamente uniformes. Distinguem-se geralmente mecanismos de legitimação e de identificação de *pro* (Rizzi 1986), o que permite dar conta das diferenças na possibilidade de uma língua ter sujeitos nulos referenciais e sujeitos nulos expletivos ou apenas sujeitos nulos expletivos. Em estudos mais recentes, o estatuto do sujeito nulo tem sido reavaliado. Alguns autores propõem que, nas línguas de sujeito nulo, a posição de sujeito não é projetada e é o morfema de concordância que funciona como pronome. É esta a posição adotada por Barbosa (2009), na sequência de outros autores (Manzini/Savoia 1997; Alexiadou/Anagnostopoulou 1998; entre outros).

Qualquer que seja o estatuto teórico do sujeito nulo e a explicação para a variação interlinguística, descritivamente podemos verificar que o português se enquadra nas línguas de sujeito nulo clássicas, em que é possível estabelecer uma relação entre legitimação de sujeito nulo e morfologia flexional rica.

Apesar de a omissão do sujeito nas línguas de sujeito nulo ser frequentemente referida como facultativa, observam-se diferenças interpretativas entre sujeitos nulos e sujeitos pronominais. Estas diferenças interpretativas serão tratadas na próxima secção.

3 Alternância sujeito nulo – sujeito pronominal

Embora as línguas de sujeito nulo permitam a omissão do sujeito, não é possível considerar que a omissão do sujeito é livre. Na realidade, há contextos que bloqueiam ou desfavorecem a omissão, e há contextos que exigem ou favorecem a omissão. Este fenómeno foi atribuído por Chomsky (1981) a um Princípio de Economia – o Princípio

Evite o Pronome –, que considera que a forma nula é mais económica do que a forma pronominal, sendo esta evitada sempre que possível.¹¹

Em primeiro lugar, contudo, é necessário distinguir entre primeira e segunda pessoas e terceira pessoa (Lobo 1995). A presença ou ausência de sujeitos pronominais de primeira e segunda pessoa não altera a identificação do referente. No caso de primeiras e segundas pessoas, o referente será sempre, respetivamente, o sujeito da enunciação e o interlocutor. Assim, com estas formas, a presença ou ausência do sujeito poderá estar condicionada por questões de atribuição de ênfase (Lobo 1995).

(26) (Eu) não me sinto bem.

(27) (Tu) deves estar a brincar.

(28) A minha avó acha que (eu) gosto de pantufas de lã.

(29) O médico acha que (tu) devias descansar.

Para além disso, se o sujeito for contrastado ou focalizado, terá necessariamente de estar realizado, o que é visível em (30).

- (30) a. – Quem fez os convites? / – Fiz *(eu).
 b. O André não arrumou a cozinha, arrumei-a *(eu).

O referente de uma terceira pessoa, contudo, não está dependente da mesma forma da situação de enunciação. Quando o pronome está inserido numa oração subordinada, a sua interpretação pode estar condicionada por diferentes fatores, incluindo o tipo de verbo principal, o modo da oração e o estatuto nulo ou pronominal do sujeito. Em muitos casos, sobretudo quando temos orações no indicativo (completivas ou adverbiais), existem diferenças interpretativas entre sujeito nulo e sujeito pronominal quanto ao antecedente preferencial.

- (31) a. O João disse ao Pedro que \emptyset tinha ganho a corrida.
 b. O João disse ao Pedro que ele tinha ganho a corrida.

- (32) a. O João cumprimentou o Pedro quando \emptyset chegou a casa.
 b. O João cumprimentou o Pedro quando ele chegou a casa.

Um sujeito nulo retoma preferencialmente o sujeito da oração principal (31a) e (32a), ao passo que um sujeito pronominal retoma preferencialmente o complemento (31b) e (32b).

Esta diferença entre sujeitos nulos e pronominais não existe em línguas que não permitem sujeitos nulos. Em (33) e (34), o pronome *il* ou *he* tanto pode ter como

¹¹ Estes efeitos são descritos em Montalbetti (1986).

antecedente o sujeito da oração principal (Jean/John) como o complemento (Pierre/Peter):

- (33) a. Jean a dit à Pierre qu'il avait gagné la course.
 b. John told Peter that he won the race.
- (34) a. Jean a salué Pierre quand il est arrivé.
 b. John greeted Peter when he arrived.

Nas línguas de sujeito nulo, há, contudo, diferentes fatores que podem condicionar a preferência por um determinado antecedente para o pronome. Enquanto um pronome nulo geralmente prefere tomar como antecedente o sujeito,¹² um pronome realizado, apesar de preferencialmente tomar como antecedente um constituinte diferente do sujeito, tem uma interpretação mais instável e é sensível à animacidade do antecedente. Morgado (2014) mostra que um pronome realizado toma preferencialmente como antecedente um referente animado, o que não acontece com um sujeito nulo:¹³

- (35) a. O bombeiro segurou o menino quando \emptyset /ele caiu.
 b. O bombeiro segurou o escadote quando \emptyset /ele caiu.
 c. O escadote magoou o bombeiro quando \emptyset /ele caiu.

Em (35b), o pronome *ele* retoma preferencialmente o sujeito, que é animado, e não o objeto inanimado. O sujeito nulo, em contrapartida, retoma sempre preferencialmente o sujeito, independentemente de ele ser animado ou inanimado. Assim, nas frases (35a), (35b) e (35c) o sujeito nulo retoma sempre preferencialmente o sujeito, ao passo que o pronome *ele* apenas retoma preferencialmente o objeto, animado, em (35a) e em (35c).

Importa ainda referir alguns contextos em que a presença do sujeito não é permitida. Entre esses encontram-se respostas a interrogativas que não incidem sobre o sujeito (36) e sujeitos arbitrários de terceira pessoa (37):

- (36) – Com quem namora a Ana? / – (?Ela) namora com o Nuno.
- (37) (*Eles) bateram à porta. Vai ver quem é.

12 De entre os trabalhos da área da Psicolinguística, ficou conhecido o de Carminati (2002) que propõe a Hipótese de Posição do Antecedente para dar conta da preferência que um sujeito nulo encaixado tem por tomar como antecedente o sujeito da oração principal. Para o português europeu, veja-se Luegi (2012) e Morgado (2011).

13 Holmberg (2010) também refere contrastes de animacidade associados a pronomes sujeito em português europeu. Barbosa/Duarte/Kato (2005) também observam que, num corpus de textos escritos de português europeu, quase todos os sujeitos pronominais retomam antecedentes animados, o que não acontece na variedade brasileira. Outros contrastes entre sujeitos nulos e pronominais incluem a impossibilidade de um sujeito pronominal ser ligado por um sujeito quantificado (Montalbetti 1984).

Na variedade europeia do português, a presença do pronome em (37) bloqueia a interpretação arbitrária e em (36) produz resultados marginais. Assim, podemos constatar que, nas línguas de sujeito nulo clássicas, como é o caso do português, em muitos casos não temos uma alternância neutra entre sujeitos nulos e sujeitos pronominais, mas antes contextos em que o sujeito tem de estar expresso, contextos em que tem de estar omissos e contextos em que a alternância pode condicionar diferenças interpretativas.

4 Aquisição de sujeito nulo e de sujeito pronominal

A aquisição do Parâmetro do Sujeito Nulo tem sido um dos tópicos mais estudados na área da aquisição da linguagem. Desde o trabalho pioneiro de Hyams (1986), inúmeros estudos têm considerado dados das produções infantis, procurando perceber de que forma é que as crianças estabelecem o valor do parâmetro da língua alvo que estão a adquirir. Na aquisição de línguas como o inglês, o francês e o alemão, verifica-se que as crianças omitem frequentemente os sujeitos em fases iniciais (Hyams 2011).

- (38) a. Want more apple.
 b. Tickles me.
 c. No play matches.
 d. Show Mommy that.
- (39) a. A tout tout tout mangé
 b. Oter tout ta. [=Ôter tout ça] (Hyams 2011, 16)

Estas omissões iniciais foram explicadas através de diferentes hipóteses. Segundo alguns autores, elas serão devidas a propriedades do desenvolvimento gramatical; de acordo com outros autores, elas terão origem em restrições não gramaticais, como limitações de processamento, por exemplo.

Alguns autores atribuem a omissão de sujeitos a restrições extragramaticais, como custos de processamento que limitam a extensão dos enunciados (Valian 1990; 1991; Bloom 1990) ou restrições métricas (Gerken 1991). Estas análises foram, contudo, criticadas por não darem conta de correlações existentes entre a omissão de sujeitos e a possibilidade de termos orações não finitas como frases principais (Hyams/Wexler 1993; Hyams 2011).

De entre os autores que propõem explicações gramaticais para o fenómeno, cabe destacar Hyams (1986). Neste trabalho, a autora defendeu que o Parâmetro do Sujeito Nulo seria inicialmente fixado por defeito num valor positivo. Com base em evidência como a presença de expletivos visíveis, as crianças de uma língua de sujeito obrigatório alterariam o valor do parâmetro para negativo. Entre outros trabalhos que seguem a ideia de que há mudança na fixação do valor do parâmetro estão Jaeggli/

Hyams (1988), que, seguindo a hipótese de Jaeggli/Safir (1989a), propõem que as crianças começam por assumir que a sua língua é morfologicamente uniforme e, portanto, legitimadora de sujeitos nulos. Os autores defendem que as crianças considerariam inicialmente a sua língua como uma língua de queda de tópico, semelhante ao chinês. Ao adquirirem as propriedades flexionais verbais, alterariam o valor do parâmetro. Contudo, a ideia de que as gramáticas iniciais de uma língua de sujeito obrigatório se aproximam das de línguas de sujeito nulo (clássicas ou radicais) tem problemas empíricos, como mostram vários autores (Valian 1990; 1991; Wang et al. 1992):

- i) há diferenças nas taxas de omissão de sujeitos de crianças falantes de inglês, italiano e chinês – as crianças inglesas têm taxas de omissão inferiores;
- ii) há diferenças nos contextos de omissão de sujeitos na aquisição do inglês e do italiano – as crianças inglesas apenas omitem sujeitos em frases principais, ao passo que as crianças italianas omitem sujeitos também em orações subordinadas;
- iii) há diferenças no tipo de orações produzidas por crianças que adquirem uma língua de sujeito nulo e por crianças que adquirem uma língua de sujeito obrigatório – apenas na aquisição de línguas de sujeito obrigatório se encontra com alguma frequência orações não finitas como frases principais;
- iv) as crianças inglesas ainda omitem sujeitos numa fase em que já produzem formas flexionadas e pronomes expletivos.

Investigação mais recente tem vindo a mostrar que a fixação de parâmetros é muito precoce e que há evidência para que, mesmo nas primeiras produções, as crianças já manifestam conhecimento das especificações paramétricas da língua alvo (Lorusso/Caprin/Guasti 2005; Guasti 2002). No caso do Parâmetro do Sujeito Nulo, é possível verificar que as crianças que estão a adquirir uma língua de sujeito nulo têm produções diferentes das que estão a adquirir uma língua de sujeito obrigatório:

- i) em línguas de sujeito nulo, produzem sujeitos nulos em taxas superiores às das línguas de sujeito obrigatório;
- ii) em línguas de sujeito nulo, produzem sujeitos nulos não só em orações principais, mas também em orações subordinadas;
- iii) em línguas de sujeito obrigatório, produzem orações não finitas como frases principais, o que raramente acontece nas línguas de sujeito nulo (e.g. Roeper/Weissenborn 1990; Guasti 2002).

Uma das explicações possíveis para o facto de as crianças omitirem sujeitos numa língua de sujeito obrigatório admite que a gramática das crianças permite que não seja projetada toda a estrutura funcional da frase, estando o nível a que se dá o corte dependente das propriedades morfossintáticas do verbo da oração (Rizzi 1994; 2005b). A diferença entre a gramática das crianças e a dos adultos estaria, assim, na maior permissividade de truncção de estrutura funcional na gramática das crianças, possivelmente como estratégia para evitar maiores custos computacionais num siste-

ma de produção imaturo.¹⁴ Disto decorre a restrição de ocorrência de sujeitos nulos na aquisição das línguas de sujeito obrigatório a contextos de frase principal.

Os dados de produção espontânea da aquisição do português mostram que as crianças têm produções compatíveis com as características de uma língua de sujeito nulo clássica:

- i) adquirem cedo a morfologia flexional de concordância (Gonçalves 2004);
- ii) produzem sujeitos nulos quer em orações principais, quer em subordinadas;
- iii) têm taxas de omissão do sujeito diferentes das crianças inglesas e compatíveis com as de uma língua de sujeito nulo (Valian/Eisenberg 1996; Gonçalves 2004).¹⁵

Gonçalves (2004) mostra que as crianças portuguesas produzem formas verbais flexionadas desde cedo e fazem poucos erros de flexão, embora produzam as formas de singular mais cedo do que as de plural. A autora mostra ainda, com base em dados de produção espontânea de quatro crianças portuguesas e de três crianças brasileiras, que as crianças portuguesas omitem sujeitos em taxas equiparáveis às de outras línguas de sujeito nulo clássicas (com valores próximos dos 70% quando se considera a percentagem de sujeitos nulos referenciais sobre a totalidade dos sujeitos).

Se considerarmos apenas a proporção de sujeitos nulos e de sujeitos pronominais, verificamos que as taxas estão entre os 80 e os 90% de sujeitos nulos para a maioria das crianças, à exceção de uma criança brasileira (RAB), com taxas bastante mais elevadas de sujeitos pronominais, como se pode ver na tabela 1, feita a partir dos dados fornecidos em Gonçalves (2004).

Tabela 1: Proporção de sujeitos nulos e sujeitos pronominais na produção espontânea de crianças portuguesas e brasileiras (a partir de Gonçalves 2004).

	Crianças	Idades	Sujeitos Nulos Referenciais	Sujeitos Pronominais
Crianças portuguesas	LUI	1;9.29–2;11.02	89,7%	10,3%
	RAP	1;10.02–2;11.22	83,2%	16,8%
	JOA	2;00.02–2;07.16	91,9%	8,1%
	PED	2;07.16–3;01.15	82,9%	17,1%
Crianças brasileiras	PAU	1;08.21–2;02.18	85,7%	14,3%
	RAB	1;11.12–3;00.07	62,1%	37,9%
	TIA	2;02.28–3;00.15	89,9%	10,1%

¹⁴ Na realidade, como mostram Haegeman (1990) e Rizzi (1994), podem ocorrer sujeitos nulos em línguas de sujeito obrigatório em registos discursivos específicos, como a escrita diarística.

¹⁵ Também Faria (1993) refere, a partir de um estudo transversal de 30 crianças entre os 1;0 e os 2;5, que o Parâmetro do Sujeito Nulo é adquirido cedo no português.

Gonçalves (2004) chama a atenção para a heterogeneidade dos dados das crianças brasileiras e para o facto de as crianças brasileiras terem taxas de sujeito nulo superiores às que se encontram na gramática do adulto, tal como descrito em trabalhos como o de Barbosa/Duarte/Kato (2001). As crianças portuguesas, pelo contrário, têm taxas de sujeito nulo homogêneas, próximas das taxas dos adultos portugueses.

Dados de produção espontânea mostram que as crianças portuguesas produzem sujeitos nulos quer em orações principais (40)–(43) e (45)–(46), quer em subordinadas (44)–(46), como acontece no italiano:

- (40) [-] Vês a lua acordada? (Marta, 2;0.26, Soares 2006)
- (41) [-] Queres ver que eu ando (Marta, 2;2.17, Soares 2006)
- (42) [-] Vês um gato? (Marta, 1;10.4, Soares 2006)
- (43) [-] Estão a tirar estas flores com os paus (Sandra, 2;8.22, Soares 2006)
- (44) depois eu vou fazer depois quand(o) [-] arruma(r) isso (Tomas, 2;6.6, l. 346, Corpus Santos)
- (45) (es)pera [-] vou ver se [-] tenho muitas jardineiras (Tomas, 3;7.29, Corpus Santos)
- (46) [-] vou buscar aqui uma escova pa(ra) [-] verem (Tomas, 3;7.29, Corpus Santos)

A par de sujeitos nulos, as crianças produzem sujeitos pronominais de diferentes pessoas gramaticais:

- (47) Eu estou a encontrar este biberon desta menina (Sandra, 2;8.22, Soares 2006)
- (48) Acho que eu vou arranjar (Sandra, 2;8.22, Soares 2006)
- (49) eu tenho uma escova nova. (Tomas, 3;7.29, Corpus Santos)
- (50) sabes qual é que eu vi? (Tomas, 3;0.22, l. 1411, Corpus Santos)
- (51) tu sabes qual é que eu vou contar? (Tomas, 3;0.22, l. 2163, Corpus Santos)
- (52) depois ele começou a cai(r) (Gonçalves 2004)
- (53) nós temos Ode [*] colar (Gonçalves 2004)¹⁶
- (54) ai qu(e) a gente caiu (Gonçalves 2004)

¹⁶ O símbolo «Ode [*]» no sistema de codificação do CHILDES indica que a preposição não foi produzida, embora a sua omissão seja agramatical na gramática adulta.

(55) pa(ra) vocês colarem (Gonçalves 2004)

(56) são peixes eles têm barbatanas (Tomas, 3;0.22, l. 1652, Corpus Santos)

É possível encontrar também casos de inversão do sujeito, mas geralmente com verbos inacusativos ou predicativos (Adragão 2001; Adragão/Costa 2004; Friedmann/Costa 2011):

(57) queres ver que aparece um dos balões? (Tomas, 3;0.22, l. 1034, Corpus Santos)

(58) olha (es)tão aqui mais balões (Tomas, 3;0.22, l. 1225, Corpus Santos)

(59) depois vem os chimpanzés (Tomas, 3;0.22, l. 1741, Corpus Santos)

Apesar de os dados de produção espontânea mostrarem que as propriedades associadas ao Parâmetro do Sujeito Nulo são adquiridas cedo, há, contudo, em estudos que visam a compreensão, alguma evidência de que as diferenças interpretativas decorrentes de alternâncias entre sujeito nulo e sujeito pronominal, características das línguas de sujeito nulo clássicas, são de aquisição mais tardia. Silva (2015) estuda a interpretação de sujeitos nulos e pronominais em orações completivas com indicativo (selecionadas pelo verbo *dizer*) e conjuntivo (selecionadas pelos verbos *querer* e *pedir*) em crianças entre os 3 e os 6 anos. A tabela seguinte (adaptada de Silva 2015) mostra os resultados obtidos numa tarefa de juízo de valor de verdade para frases como *O bombeiro disse ao polícia que ø/ele saltou*.

Tabela 2: Taxas de aceitação de leituras correferenciais e disjuntas de pronomes nulos e lexicais em orações completivas com o indicativo selecionadas pelo verbo *dizer* com dois antecedentes na oração principal (adaptado de Silva 2015).

Idade	Sujeito Nulo		Sujeito pronominal	
	Correferência	Ref. disjunta	Correferência	Ref. disjunta
3 anos	71%	78%	64%	91%
4 anos	76%	82%	67%	80%
5 anos	89%	81%	69%	89%
6 anos	90%	67%	54%	80%
Adultos	99%	18%	26%	94%

Nas orações com o indicativo selecionadas pelo verbo *dizer* com dois antecedentes na oração principal, os resultados do estudo de Silva (2015) mostram que:

- i) as crianças aceitam mais facilmente do que os adultos leituras em que um sujeito pronominal encaixado é correferente com o sujeito da oração principal;

- ii) a presença de dois antecedentes na oração principal é uma fonte de dificuldade, podendo o sujeito nulo ser interpretado mais frequentemente pelas crianças como correferente com um complemento e não com o sujeito da oração principal.¹⁷

Estes efeitos, embora estejam sujeitos a desenvolvimento, ainda persistem aos 6 anos.

Lobo/Silva (2015) estudaram a interpretação preferencial atribuída a sujeitos nulos e pronominais em orações adverbiais temporais com indicativo, antepostas e pospostas à oração principal (*Quando chegou a casa, o avô cumprimentou o menino / Quando ele chegou a casa, o menino cumprimentou o avô / A mãe fotografou a menina quando saiu da garagem / O menino viu o ladrão quando ele começou a correr*).

O teste, que consistia numa tarefa de seleção de imagens, foi aplicado a três grupos de crianças – 30 crianças de 5 anos (5;0–5;11; média de idades 5;6), 25 crianças de 6 e 7 anos (6;7–7;11; média de idades 7;1), 35 crianças de 8–9 anos (8;1–9;6, média de idades 8;8) – e a um grupo de controlo de 35 adultos (18–68 anos; média de idades 31 anos). Os resultados mostram que as crianças têm interpretações diferentes das dos adultos. Enquanto os adultos distinguem claramente sujeitos nulos de sujeitos pronominais, preferindo leituras em que o sujeito nulo retoma o sujeito da oração principal e o sujeito pronominal retoma o objeto, tal como previsto na hipótese de Carminati (2002), as crianças não têm interpretações tão definidas, havendo desenvolvimento na especialização de sujeitos nulos e de sujeitos pronominais:

Tabela 3: Taxas de preferência na interpretação de sujeitos pronominais e nulos catafóricos e anafóricos em orações adverbiais.

Condição	Pronome Nulo Catafórico		Pronome Lexical Catafórico		Pronome Nulo Anafórico		Pronome Lexical Anafórico	
	Sujeito	Objeto	Sujeito	Objeto	Sujeito	Objeto	Sujeito	Objeto
5 anos	52%	48%	57%	43%	61%	39%	48%	52%
6–7 anos	69%	31%	49%	51%	67%	33%	43%	57%
8–9 anos	76%	24%	54%	46%	71%	29%	40%	60%
Adultos	90%	10%	39%	61%	90%	10%	15%	85%

A interpretação de sujeitos pronominais é a que se desenvolve mais tarde, em particular nos contextos catafóricos, em que o antecedente segue o pronome. É também este o contexto em que os adultos têm interpretações menos definidas.

Os resultados da aquisição mostram, assim, que as crianças portuguesas adquirem cedo as propriedades associadas às línguas de sujeito nulo clássicas (morfologia rica, contextos de omissão do sujeito, possibilidade de inversão), mas levam algum

¹⁷ Quando apenas o sujeito está expresso, as crianças não têm a mesma dificuldade com a interpretação do sujeito nulo (Costa/Ambulate 2010; Silva 2015).

tempo a distinguir diferenças interpretativas entre sujeitos nulos e sujeitos pronominais em contextos em que existem vários antecedentes possíveis para o pronome.

5 Variação entre sujeito nulo e sujeito pronominal no português dialetal

Muitos trabalhos têm sido feitos sobre o estatuto do português brasileiro relativamente ao Parâmetro do Sujeito Nulo (Duarte 1993; 1995; Figueiredo-Silva 2000; Kato/Negrão 2000; Rodrigues 2004, entre outros). É sabido que esta é uma área em que existem diferenças entre as duas variedades do português. Como mostram diferentes autores, na variedade brasileira do português, há um muito maior preenchimento da posição de sujeito do que na variedade europeia, podendo inclusivamente haver sujeitos pronominais arbitrários (60), sujeitos pronominais correspondendo a referentes não animados (61) e duplicação de sujeito por um pronome (62) (Duarte 1995). Contudo, os sujeitos expletivos correspondem sempre a sujeitos nulos (63).

(60) Eles deveriam ensinar amor às crianças. (Cyrino/Duarte/Kato 2000, 62)

(61) A casa virou um filme quando ela teve de ir abaixo. (Duarte 2000, 22)

(62) A Clarinha ela cozinha que é uma maravilha. (Duarte 2000, 28)

(63) (*Ele) parece que (*ele) vai chover.

Estas propriedades têm levado alguns autores a caracterizar o português brasileiro como uma língua de sujeito nulo parcial (Duarte 1993; Figueiredo-Silva 2000; Rodrigues 2004; Holmberg 2010; entre outros). Para além disso, um pronome realizado, na variedade brasileira, nem sempre recebe a mesma interpretação que na variedade europeia. De facto, um pronome lexical não induz necessariamente efeitos de obviação, o que permite que no português brasileiro, ao contrário do português europeu, o sujeito da oração principal e o sujeito pronominal da oração subordinada possam ser correferentes:

(64) a. O João_i disse que ele_j gosta da Maria. (Rodrigues 2004, 87, *apud* Barbosa 2009)

b. [A Rainha da Inglaterra]_i disse que ela_i quer seu filho divorciado. (Cyrino/Duarte/Kato 2000, 59)

Apesar de este fenómeno ter sido amplamente investigado para a variedade brasileira em confronto com a variedade europeia (ver, por exemplo, Barbosa/Duarte/Kato 2001), não tem sido considerado de uma forma sistemática relativamente às variedades dialetais do português europeu. Excetua-se a dissertação de doutoramento de Carvalho (2009), sobre variação entre sujeitos pronominais e nulos em dez localida-

des do CORDIAL-SIN,¹⁸ e a dissertação de mestrado de Sória (2013), sobre a variação entre o uso dos pronomes *nós*, *a gente* e sujeito nulo de primeira pessoa nas localidades do CORDIAL-SIN.

Carvalho (2009), num estudo de natureza sociolinguística, partindo de dados do CORDIAL-SIN, investiga a proporção entre sujeitos pronominais e sujeitos nulos em dez localidades de Portugal continental e insular, contemplando diferentes variáveis: especificidade, correferencialidade, pessoa gramatical, localidade, tipo de oração, animacidade do sujeito, tempo verbal, entre outros. Considerando apenas os sujeitos de referência específica, a autora obtém os seguintes resultados para a totalidade dos sujeitos (Tabela 4) e para os sujeitos correferentes (Tabela 5):

Tabela 4: Proporção de sujeitos pronominais e nulos com referência específica em 10 localidades do CORDIAL-SIN, de acordo com Carvalho (2009).

Localidade	Sujeito pronominal	Sujeito nulo	Total
ALC	48 (69%)	22 (31%)	70
ALV	78 (76%)	25 (24%)	103
CBV	326 (92%)	29 (8%)	355
PST	85 (62%)	51 (38%)	136
MIG	288 (93%)	22 (7%)	310
SRP	19 (53%)	17 (47%)	36
MIN	167 (65%)	90 (35%)	257
OUT	86 (55%)	70 (45%)	156
FIG	70 (36%)	123 (64%)	193
COV	561 (67%)	276 (33%)	837
Total	1728	725	2453

Tabela 5: Proporção de sujeitos pronominais e nulos correferentes com referência específica em 10 localidades do CORDIAL-SIN, de acordo com Carvalho (2009).

Localidade	Sujeito pronominal	Sujeito nulo	Total
ALC	8 (32%)	17 (68%)	25
ALV	13 (37%)	22 (63%)	35

¹⁸ O CORDIAL-SIN (Syntax-oriented Corpus of Portuguese Dialects) é um corpus de variedades dialetais do português europeu, construído para permitir estudos sobre sintaxe dialetal. O corpus é desenvolvido no Centro de Linguística da Universidade de Lisboa, sob a orientação de Ana Maria Martins, e está disponível em <http://www.clul.ul.pt/en/research-teams/411-cordial-corpus>. Carvalho (2009) contempla dados das seguintes localidades/regiões: Alcochete, Alvor, Minho, Cabeço de Vide, Porto Santo, Covo, Figueiró, Outeiro, São Miguel, Serpa.

Localidade	Sujeito pronominal	Sujeito nulo	Total
CBV	77 (75%)	26 (25%)	103
PST	30 (45%)	37 (55%)	67
MIG	113 (88%)	16 (12%)	129
SRP	4 (25%)	12 (75%)	16
MIN	42 (40%)	62 (60%)	104
OUT	7 (12%)	51 (88%)	58
FIG	14 (15%)	79 (85%)	93
COV	112 (33%)	223 (67%)	335
Total	420	545	965

Embora estes dados tenham de ser interpretados com alguma cautela, como é possível observar, há duas localidades – nomeadamente CBV e MIG – que se destacam pela percentagem de sujeitos realizados. Não será possível, por conseguinte, tratar o português dialetal como uma unidade.

Ainda assim, as proporções de sujeitos nulos e pronominais encontradas contrariam aquilo que seria esperável numa língua de sujeito nulo, tendo em conta as descrições de autores como Barbosa/Duarte/Kato (2001). Estes resultados levam a autora a concluir que já não é possível caracterizar o português europeu dialetal como uma língua de sujeito nulo prototípica. Contudo, cremos que algumas das opções metodológicas da autora terão condicionado os resultados. De facto, desta análise, a autora exclui uma série de dados, incluindo sujeitos nulos com verbo na primeira pessoa do plural sem um antecedente próximo e sujeitos preenchidos em contextos de possível ambiguidade, o que nos parece enviesar os dados. Para além disso, não distingue, nos sujeitos correferentes, a terceira pessoa, potencialmente geradora de ambiguidades, da primeira e segunda pessoas, que geralmente não dão origem a ambiguidades por remeterem para os intervenientes diretos na situação discursiva.

Na impossibilidade de considerar os dados das diferentes localidades, vejamos, a título exemplificativo, a distribuição de sujeitos pronominais e nulos na região do Alto Alentejo do corpus CORDIAL-SIN, distinguindo as pessoas gramaticais:¹⁹

¹⁹ A análise foi feita a partir de uma pesquisa de sujeitos pronominais e sujeitos referenciais nulos nos ficheiros com anotação sintática do corpus, à exceção dos sujeitos pronominais *a gente*, pesquisados nos ficheiros com anotação morfológica por classe de palavras. Não considerei: i) sujeitos em orações não finitas sem morfologia de concordância; ii) sujeitos expletivos (e.g. *Então, é preciso magueiras*); iii) casos em que não havia concordância entre o grupo nominal e o verbo, interpretáveis como casos de topicalização (e.g. *eu agora é que me está a lembrar*); iv) orações com *se* impessoal (e.g. *fazia-se montes*) – ainda que dialetalmente possam coocorrer com sujeitos – ; v) repetições.

Tabela 6: Sujeitos nulos e pronominais em AAL.

	1S	3S/2S	3S	3S/1P	1P	3P	Total
Sujeitos nulos	76 58,9%	12 100%	149 91,4%	5 5,7%	14 48,3%	89 85,6%	345 65,8%
Sujeitos pronominais	53 41,1%	0	14 8,6%	82 94,3%	15 51,7%	15 14,4%	179 34,2%
Total	129	12	163	87	29	104	524

Nestes dados, para a maioria das pessoas gramaticais são mais frequentes os sujeitos nulos do que os sujeitos pronominais, havendo uma predominância clara de sujeitos nulos na terceira pessoa. Apesar de, na terceira pessoa (do singular ou do plural), serem mais frequentes os sujeitos nulos do que os sujeitos pronominais, a ocorrência de um ou de outro não é, contudo, aleatória. Os sujeitos nulos de terceira pessoa retomam geralmente um antecedente anterior, que pode ser o sujeito de uma frase anterior (65a-c), um complemento (65d), ou têm uma interpretação arbitrária (65e), ao passo que os sujeitos pronominais podem ser referencialmente distintos do sujeito da oração principal, como em (66a–b), onde um sujeito nulo seria referencialmente ambíguo, ser correferentes do sujeito da oração principal (66c), ou não ter um antecedente na frase em que ocorrem (66d). Nos dois últimos exemplos o sujeito pronominal encontra-se contrastado ou focalizado.

- (65) a. Mas eu encontro que aquilo em bravo que é melhor porque [-] fica mais arriegado e a cepa dura mais tempo.
 b. Mas estas vinhas aqui nossas, como [-] são muito antigas, não podem ser aramadas – está o senhor a compreender?
 c. Enfim, os homens lá vão ainda, porque [-] não têm tractores.
 d. Chamávamos nós os Ratinhos. E [-] vinham lá da Beira.
 e. É por causa disso é que é que [-] lhe chamavam o ladrão.
- (66) a. [...] mas [-] nem sei quanto elas pedem por isto.
 b. Aquilo, depois, aquilo foi para ver se [-] safava lá a vida ao meu irmão, porque ele andava assim também um bocado mal.
 c. Ele, este ano, até que me comprou a vinha a mim, porque ele tem uma vinha quatro ou cinco vezes maior do que a minha e [-] não teve um cacho para comer, está a compreender?
 d. E queria ele que lhe eu fizesse uma ponte em ferro [!]

Nas tabelas, uso as seguintes abreviaturas: 1S = primeira pessoa do singular (*eu*); 3S/2S = forma de terceira pessoa do singular usada para se dirigir a uma segunda pessoa/interlocutor singular (*você*); 3S = terceira pessoa do singular (*ele/ela*); 1P = primeira pessoa do plural (*nós*); 3S/1P = forma de terceira pessoa do singular usada para referir uma primeira pessoa do plural (*a gente*) – independentemente da concordância verbal, que pode ser de 3.^a pessoa, 1.^a pessoa do plural ou 3.^a pessoa do plural, nas variedades dialetais – ; 3P = terceira pessoa do plural (*eles/elas*).

Na primeira pessoa do singular, pode encontrar-se quer sujeitos nulos, sem antecedente (67a–b) ou retomando um sujeito de primeira pessoa anterior (67c), quer sujeitos pronominais que parecem ter um valor enfático (67c) e (68a) ou marcar a focalização contrastiva do sujeito (68b). Há, contudo, sujeitos pronominais que não têm um valor claramente contrastivo, como em (67d), embora não seja de excluir um valor enfático (de reforço assertivo):

- (67) a. [-] tenho ali os papéis das vacinas, tudo legal.
 b. É que, este ano, se não chove, [-] estou encravado com isto.
 c. Mas como eu semeio de dois arcos, [-] faço o arco mais pequeno, assim, ao contrário.
 d. Agora, esta que [-] tenho aqui no tanque, se me derem o dinheiro que eu entendo, eu vendo-a.
- (68) a. Pois, eu não concordo com a monda química.
 b. Dentro da razão, quero eu ficar com ela.

Na primeira pessoa do plural, encontram-se ocorrências de *a gente* e de *nós*. Em ambos os casos, podem ocorrer também sujeitos nulos (69d–e) e (70b–c), mas enquanto um sujeito nulo com verbo na primeira pessoa do plural não precisa de antecedente (69e), um sujeito nulo remetendo para uma primeira pessoa do plural com verbo na terceira pessoa do singular só é possível quando existe um antecedente no contexto linguístico anterior; os casos de omissão de *a gente* são todos eles casos em que o sujeito nulo estabelece uma relação anafórica com um sujeito pronominal anterior (70b–c).

- (69) a. Santo António das Areias fala precisamente como nós falamos.
 b. Se nós tivermos uma coelheira e que [-] deite para lá umas ervas destas, os coelhos que comem isto morrem.
 c. Bem, nós cá, no tempo da minha mãe, quando nós éramos miúdos, era sempre as papas do dia dos Santos.
 d. Ora então, nós estávamos aqui em Portalegre e daqui [-] fomos para Lisboa.
 e. Fazia-se ali dezassete, dezoito, vinte, quinze... Era conforme a quantidade que [-] amassávamos [...].
- (70) a. quando em depois [-] começa a cair é que a gente vai despejando.
 b. Por exemplos, é um negalho que a gente tira quando [-] anda a ceifar ... (AAL)
 c. E de forma que quando a gente tinha a mão cheia já – o feixe da mão cheia – [-] atavallhe o negalho. (AAL)

Na realidade, a distribuição de *nós*, *a gente* e sujeito nulo de primeira pessoa do plural nas variedades dialetais é outro fenómeno que é relevante considerar e que foi estudado por Sória (2013). A autora investiga a proporção destes três tipos de formas no CORDIAL-SIN, contrastando-a com uma amostra dialetal de variedades rurais do português brasileiro do interior do estado da Bahia. Havendo muitos trabalhos que referem a produtividade da forma *a gente* no português brasileiro como um dos

fatores que favorecem o enfraquecimento do paradigma verbal e a mudança no estatuto do português brasileiro na tipologia das línguas de sujeito nulo, seria de esperar que a frequência de *a gente* como pronome de primeira pessoa fosse baixa no português europeu. Contudo, a autora mostra que também nas variedades dialetais do português europeu há uma frequência elevada de *a gente*, a par de ocorrências de *nós* e de sujeitos nulos que remetem para primeira pessoa do plural. Assim, a ocorrência de *a gente* não determina o desaparecimento nem da primeira pessoa do plural, nem de sujeitos nulos. Na realidade, pode haver fatores independentes que favorecem o uso de uma forma pronominal explícita. Dada a natureza das situações discursivas que fazem parte do CORDIAL-SIN, um pronome de primeira pessoa é frequentemente contrastivo: o informante contrasta os costumes da sua terra ou da sua geração com os costumes de outras terras ou de outras gerações, como se vê, por exemplo, em (69c). Para além disso, a comparação entre a variedade brasileira e as variedades dialetais europeias levada a cabo no trabalho de Sória (2013) mostrou que as duas variedades se comportam de forma claramente distinta quanto à tipologia do sujeito nulo.

6 Conclusões

Neste capítulo, mostrámos que o português europeu se enquadra nas línguas de sujeito nulo clássicas (ou consistentes), uma vez que tem morfologia verbal rica, legitima sujeitos nulos expletivos, arbitrários e referenciais e admite inversão do sujeito de forma mais livre do que noutros grupos tipológicos de línguas, e, para além disso, em português europeu sujeitos nulos e sujeitos pronominais tendem a associar-se a diferenças interpretativas. Vimos ainda que, enquanto as propriedades típicas de uma língua de sujeito nulo são adquiridas precocemente, as diferenças interpretativas entre sujeitos nulos e pronominais são de desenvolvimento mais tardio. Finalmente, considerámos alguns dados de variedades dialetais do português e concluímos que é necessário interpretar com cautela os dados de frequência de sujeitos pronominais e nulos. Ainda que os sujeitos nulos sejam em algumas localidades do corpus CORDIAL-SIN globalmente mais frequentes do que os sujeitos pronominais, é necessário distinguir, por um lado, pessoas gramaticais e, por outro lado, os contextos de ocorrência de cada uma das formas. Com uma análise mais detalhada dos contextos, podemos verificar que o preenchimento do sujeito pode funcionar como uma estratégia de marcação de ênfase, de focalização ou de desambiguação e que isso não implica necessariamente que o português europeu dialetal esteja a perder propriedades de uma língua de sujeito nulo clássica, uma vez que mantém sujeitos nulos que não são correferentes com um sujeito numa oração superior.

7 Referências

- Adragão, Maria do Mar (2001), *Aquisição da Inversão Numa Criança entre os dois e os três Anos*, ms, Universidade Nova de Lisboa.
- Adragão, Maria do Mar/Costa, João (2004), *On the status of preverbal subjects in null subject languages: Evidence from acquisition*, in: Jacqueline van Kampen/Sergio Baauw (edd.), *Proceedings of GALA 2003*, Lot Publications, The Hague, Utrecht University, 69–80.
- Alexiadou, Artemis/Anagnostopoulou, Elena (1998), *Parametrizing Agr: Word order, V-movement and EPP-checking*, *Natural Language and Linguistic Theory* 16, 491–539.
- Barbosa, Pilar (1995), *Null Subjects*, tese de Doutorado, MIT.
- Barbosa, Pilar (2009), *Two kinds of subject «pro»*, *Studia Linguistica* 63:1, 2–58.
- Barbosa, Pilar/Duarte, Maria Eugênia/Kato, Mary (2001), *A distribuição do sujeito nulo no português europeu e no português brasileiro*, in: *Actas do XVI Encontro nacional da Associação Portuguesa de Linguística/Colóquio Português Europeu Português Brasileiro: Unidade e diversidade na passagem do milênio*, Lisboa, APL/Colibri, 539–550.
- Barbosa, Pilar/Duarte, Maria Eugênia L./Kato, Mary A. (2005), *Null Subjects in European and Brazilian Portuguese*, *Journal of Portuguese Linguistics* 4:2, 11–52.
- Bloom, Paul (1990), *Subjectless sentences in child language*, *Linguistic Inquiry* 21:4, 491–504.
- Carminati, M.N. (2002), *The processing of Italian subject pronouns*, tese de Doutorado, Amherst, University of Massachusetts.
- Carrilho, Ernestina (2005), *Expletive «ele» in European Portuguese dialects*, tese de Doutorado, Universidade de Lisboa.
- Carrilho, Ernestina (2009), *Sobre o expletivo «ele» em português europeu*, *Estudos de Linguística Galega* 1, 1–29.
- Carvalho, Gislaíne Aparecida de (2009), *A realização do sujeito na fala do português europeu*, tese de Doutorado, Universidade Estadual Paulista.
- Chomsky, Noam (1981), *Lectures on Government and Binding*, Dordrecht, Foris.
- CORDIAL-SIN: Corpus Dialectal para o Estudo da Sintaxe / Syntax-oriented Corpus of Portuguese Dialects*, Martins, Ana Maria (coord.), [2000–] 2010, Lisboa, Centro de Linguística da Universidade de Lisboa, <http://www.clul.ul.pt/en/resources/411-cordial-corpus> (21.10.2015).
- Corpus Santos – European Portuguese, *Corpus of child and child-directed speech* (Ana Lúcia Santos), Lisboa, CLUL, <http://www.clul.ul.pt/en/resources/546> (21.10.2015).
- Costa, João/Ambulate, Joana (2010), *The acquisition of embedded subject pronouns in European Portuguese*, in: Michael Iverson et al. (edd.), *Proceedings of the 2009 Mind/Context Divide Workshop*, Somerville MA, Cascadilla, 1–12.
- Costa, João/Pratas, Fernanda (2013), *Embedded Null Subjects in Capeverdean*, *Journal of Linguistics* 49:1, 33–53.
- Cyrino, Sonia/Duarte, Maria Eugênia/Kato, Mary (2000), *Visible subjects and invisible clitics in Brazilian Portuguese*, in: Mary A. Kato/Esmeralda V. Negrão (edd.), *Brazilian Portuguese and the Null Subject Parameter*, Frankfurt a.M./Madrid, Vervuert/Iberoamericana, 55–73.
- Duarte, M. Eugênia L. (1993), *Do pronome nulo ao pronome pleno: a trajetória do sujeito no português brasileiro*, in: Ian Roberts/Mary A. Kato (edd.), *Português Brasileiro: uma viagem diacrônica*, Campinas SP, Ed. da UNICAMP, 107–128.
- Duarte, Maria Eugênia L. (1995), *A Perda do Princípio «Evite pronome» no Português Brasileiro*, tese de Doutorado, UNICAMP.
- Duarte, Maria Eugênia (2000), *The loss of the «Avoid Pronoun» Principle in Brazilian Portuguese*, in: Mary A. Kato/Esmeralda V. Negrão (edd.), *The Null Subject Parameter in Brazilian Portuguese*, Frankfurt a.M./Madrid, Vervuert/Iberoamericana, 17–36.

- Faria, Isabel (1993), *A aquisição da noção Agente e a produção de sujeitos sintáticos por crianças portuguesas até aos dois anos e meio*, Revista Internacional de Língua Portuguesa 10, 16–50.
- Figueiredo-Silva, Cristina (2000), *Main and embedded null subjects in Brazilian Portuguese*, in: Mary Kato/Esmeralda V. Negrão (edd.), *The Null Subject Parameter in Brazilian Portuguese*, Frankfurt a.M./Madrid, Vervuert/Iberoamericana, 127–146.
- Friedmann, Naama/Costa, João (2011), *Acquisition of SV and VS order in Hebrew, European Portuguese, Palestinian Arabic and Spanish*, Language Acquisition 18:1, 1–38.
- Gerken, Lou Ann (1991), *The Metrical Basis for Children's Subjectless Sentences*, Journal of Memory and Language 30, 431–451.
- Gilligan, Gary (1987), *A crosslinguistic approach to the pro-drop parameter*, tese de Doutorado, University of Southern California.
- Gonçalves, Fernanda (2004), *Riqueza morfológica e aquisição da sintaxe em português europeu e brasileiro*, tese de Doutorado, Universidade de Évora.
- Guasti, Maria-Teresa (2002), *Language Acquisition. The Growth of Grammar*, Cambridge MA, The MIT Press.
- Haegeman, Liliane (1990), *Understood subjects in English diaries*, Multilingua 9, 157–199.
- Holmberg, Anders (2010), *Null Subject Parameters*, in: Theresa Biberauer et al. (edd.), *Parametric variation: Null subjects in minimalist theory*, Cambridge, Cambridge University Press, 88–124.
- Huang, Cheng-Teh James (1984), *On the distribution and reference of empty pronouns*, Linguistic Inquiry 15:4, 531–574.
- Hyams, Nina (1986), *Language Acquisition and the Theory of Parameters*, Dordrecht, Reidel.
- Hyams, Nina (1989), *The Null Subject Parameter in Language Acquisition*, in: Osvaldo Jaeggli/Ken Safir (edd.), *The Null Subject Parameter*, Dordrecht, Kluwer, 215–238.
- Hyams, Nina (2011), *Missing Subjects in Early Child Language*, in: Jill de Villiers/Tom Roeper (edd.), *Handbook of Generative Approaches to Language Acquisition*, Dordrecht, Springer, 13–52.
- Hyams, Nina/Wexler, Kenneth (1993), *On the grammatical basis of null subjects in child language*, Linguistic Inquiry 24:3, 421–459.
- Jaeggli, Osvaldo/Hyams, Nina (1988), *Morphological Uniformity and the Setting of the Null Subject Parameter*, in: *Proceedings of NELS 18*, Amherst MA, GLSA/University of Massachusetts, 238–253.
- Jaeggli, Osvaldo/Safir, Ken (1989a), *The Null Subject Parameter and Parametric Theory*, in: Osvaldo Jaeggli/Ken Safir (edd.), *The Null Subject Parameter*, Dordrecht, Kluwer, 1–44.
- Jaeggli, Osvaldo/Safir, Ken (edd.) (1989b), *The Null Subject Parameter*, Dordrecht, Kluwer.
- Kato, Mary Aizawa/Negrão, Esmeralda Vailati (edd.) (2000), *Brazilian Portuguese and the Null Subject Parameter*, Frankfurt a.M./Madrid, Vervuert/Iberoamericana.
- Lobo, Maria (1995), *Para uma redefinição do Parâmetro do Sujeito Nulo*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Lobo, Maria/Silva, Carolina (2015), *Resolving pronoun ambiguity in European Portuguese: adults vs. children*, ms., Universidade Nova de Lisboa.
- Lorusso, Paolo/Caprin, Claudia/Guasti, Maria Teresa (2005), *Overt Subject Distribution in Early Italian Children*, in: Alejna Brugos/Manuella R. Clark-Cotton/Seungwan Ha (edd.), *A Supplement to the Proceedings of the Twenty-ninth Boston University Conference on Language Development*, Somerville MA, Cascadilla Press, <http://www.bu.edu/buclid/files/2011/05/29-LorussoBUCLD2004.pdf> (30.09.2015).
- Luegi, Paula (2012), *Processamento de sujeitos pronominais em Português: efeito da posição estrutural dos antecedentes*, tese de Doutorado, Universidade de Lisboa.
- Manzini, Rita/Savoia, Leonardo M. (1997), *Null Subjects without «pro»*, UCL Working Papers in Linguistics 9, 301–313.

- Montalbetti, Mario (1984), *After binding: on the interpretation of pronouns*, tese de Doutorado, Massachusetts Institute of Technology.
- Montalbetti, Mario (1986), *How pro is it?*, in: Osvaldo Jaeggli/Carmen Silva-Corvalán (edd.), *Studies in Romance Linguistics*, Dordrecht, Foris, 137–152.
- Morgado, Sara (2011), *Processamento da co-referência pronominal. Informação sintáctica e semântica*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Morgado, Sara (2014), *O processamento de pronomes nulos e plenos em Português Europeu. Efeito de animacidade do antecedente*, comunicação apresentada ao IX Fórum de Partilha Linguística, Universidade Nova de Lisboa.
- Rizzi, Luigi (1982), *Issues in Italian Syntax*, Dordrecht, Foris.
- Rizzi, Luigi (1986), *Null Objects in Italian and the Theory of «pro»*, *Linguistic Inquiry* 17:3, 501–557.
- Rizzi, Luigi (1994), *Early Null Subjects and Root Null Subjects*, in: Teun Hoekstra/Bonnie D. Schwartz (edd.), *Language Acquisition Studies in Generative Grammar. Papers in Honor of Kenneth Wexler from the 1991 Glow Workshops*, Amsterdam/Philadelphia, Benjamins, 151–176.
- Rizzi, Luigi (2005a), *On the grammatical basis of language development: A case study*, in: Guglielmo Cinque/Richard Kayne (edd.), *The Oxford Handbook of Comparative Syntax*, New York, Oxford University Press, 70–109.
- Rizzi, Luigi (2005b), *Grammatically-based target-inconsistencies in child language*, in: *Proceedings of the Inaugural Conference of GALANA*, Cambridge MA, The MIT Press, 19–49.
- Rodrigues, Cilene (2004), *Impoverished Morphology and A-Movement out of Case Domains*, tese de Doutorado, University of Maryland at College Park.
- Roeper, Thomas/Weissenborn, Jürgen (1990), *How to make parameters work: Comments on Valian*, in: Lyn Frazier/Jill de Villiers (edd.), *Language Processing and Language Acquisition*, Dordrecht, Kluwer, 147–162.
- Silva, Carolina (2015), *Interpretation of Clitic, Strong and Null Pronouns in the Acquisition of European Portuguese*, tese de Doutorado, Universidade Nova de Lisboa.
- Soares, Carla (2006), *La Syntaxe de la Périphérie Gauche en Portugais Européen et son Acquisition*, tese de Doutorado, Université Paris 8.
- Sória, Maíra (2013), *«Nós», «a gente» e o sujeito nulo de primeira pessoa do plural*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Taraldsen, Knut Tarald (1978), *On the NIC, vacuous application and the «that»-trace filter*, ms. MIT. (Distributed by the Indiana University Linguistics Club, Bloomington IN, 1980).
- Valian, Virginia (1990), *Null Subjects: A Problem for Parameter-Setting Models of Language Acquisition*, *Cognition* 35, 105–122.
- Valian, Virginia (1991), *Syntactic Subjects in the Early Speech of American and Italian Children*, *Cognition* 40, 21–81.
- Valian, Virginia/Eisenberg, Zena (1996), *Syntactic subjects in the spontaneous speech of Portuguese-speaking children*, *Journal of Child Language* 23:1, 103–128.
- Wang, Qi, et al. (1992), *Null Subjects versus Null Object: Some Evidence from the Acquisition of Chinese and English*, *Language Acquisition* 2, 221–254.

Ana Maria Martins

22 O sistema responsivo: padrões de resposta a interrogativas polares e a asserções

Abstract: Neste capítulo descrevem-se os padrões de resposta a interrogativas polares e asserções característicos da gramática do português europeu, considerando como traços de polaridade os que subjazem às oposições afirmação/negação e confirmação/inversão. Identificam-se as questões a explorar e mostra-se que é um domínio em que o português apresenta especificidades interessantes quando comparado com as outras línguas românicas. No campo das respostas mínimas afirmativas, dispõe de respostas verbais, respostas com *sim*, respostas com *ser* e respostas adverbiais, que se mostra terem propriedades diferentes e não estarem em variação livre entre si. No campo das respostas alargadas, usa a posição na frase dos marcadores de polaridade para dar expressão gramatical à oposição confirmação/inversão. Para exprimir a afirmação inversiva enfática, o português europeu dispõe de uma estratégia de reduplicação do verbo (em estruturas monoracionais) que o separa do português brasileiro e aproxima do galego.

Keywords: afirmação/negação, confirmação/inversão, respostas mínimas (verbais, adverbiais, com *sim*, com *ser*), posição inicial/final dos marcadores de polaridade, reduplicação enfática do verbo

1 Introdução

A sintaxe das frases interrogativas tem sido objeto de vasta investigação linguística tanto no que diz respeito a línguas particulares como numa perspectiva comparativa. Em contraste, o modo como em diferentes línguas se responde às interrogativas polares (também chamadas interrogativas totais ou interrogativas *sim-não*) tem ainda hoje um espaço limitado quer nas obras essencialmente descritivas (como as gramáticas) quer nos trabalhos de orientação mais teórica. Nas gramáticas de referência de línguas particulares produzidas nas últimas décadas não falta uma secção sobre as frases interrogativas (especialmente as interrogativas *qu-*, também chamadas interrogativas parciais) nem sobre a negação e as palavras negativas, mas não se aborda de uma forma mais abrangente o tema da *polaridade* (considerando o par *afirmação/negação*) nem, especificamente, o das *respostas polares*. O tratamento deste tópico na literatura linguística tem demonstrado, no entanto, que se reveste de grande interesse quer no plano teórico-descritivo (Pope 1976; Jones 1999; Holmberg 2001; 2003; 2007; 2013; 2015; Farkas/Bruce 2010; Costa/Martins/Pratas 2012; Authier 2013; Martins 2013;

Martins, no prelo) quer no âmbito da linguística histórica e das línguas em contacto (Martins 1994a; 1994b; 2005; 2006; Vennemann 2009), quer ainda na investigação sobre aquisição da linguagem (Santos 2003; 2009). As propriedades do sistema responsivo relacionam-se de forma não trivial com outros aspetos da gramática de cada língua e a comparação entre línguas neste domínio permite adicionar novos traços linguísticos aos que têm sido escrutinados pela tipologia linguística (cf. Dryer/Haspelmath 2013) e daí avançar para interpretações teóricas produtivas.

Neste capítulo apresentaremos uma descrição do sistema responsivo do português europeu, analisando em detalhe os padrões de resposta a interrogativas polares e asserções. No quadro das línguas românicas, o português (juntamente com o galego) tem a particularidade de permitir, e preferir, respostas verbais para expressar a afirmação, uma propriedade herdada do latim,¹ mas que hoje, numa perspetiva tipológica, aproxima o português de línguas exteriores à família românica, como o finlandês (Holmberg 2001), as línguas célticas (Jones 1999; Thoms 2016), algumas línguas eslavas, asiáticas e africanas (cf. Holmberg 2015).

O tratamento deste tema implica que comecemos por considerar questões con-/ceptuais e terminológicas. O termo *interrogativas sim-não* (do inglês *yes-no questions*), consagrado na literatura linguística, sugere uma oposição binária que teria paralelo no par *respostas afirmativas/respostas negativas*. O que se conhece sobre os sistemas responsivos das línguas do mundo revela, contudo, que juntamente com a oposição afirmação/negação é necessário contar com a oposição confirmação/inversão. No trabalho pioneiro de Pope (1976), as quatro opções a considerar ao descrever-se o sistema responsivo de uma língua são identificadas como *positive agreement*, *positive disagreement*, *negative agreement* e *negative disagreement*. Em Farkas/Bruce (2010), a oposição confirmação/inversão é formalizada pelos traços [same] / [reverse] e a oposição afirmação/negação pelos traços [+] / [-], como se mostra em (1) com

1 Tomem-se como exemplo os seguintes pares de pergunta-resposta retirados de textos latinos:

(i) Clodius insidias fecit Miloni? – Fecit.

‘Clodius fez uma emboscada a Milo? – Fez.’ (Cícero. Exemplo extraído de Pinkster 1990, 191).

(ii) Fuiſtis liber? – Fui.

‘Fostes livre? – Fui.’

(Plauto. Exemplo extraído de Hale/Buck⁵1987, 137).

Ao mesmo tempo que mantém o padrão de respostas verbais herdado do latim, o português admite, tal como o latim e diferentemente da maioria das línguas românicas, a chamada *ellipse do predicado*, que legitima o apagamento dos complementos do verbo tanto nos contextos de pergunta-resposta como em contextos de coordenação (cf. Matos 2003; 2013; Martins no prelo). O contraste entre o português e línguas irmãs está ilustrado em (iii), usando como termos de comparação o espanhol e o catalão. As respostas só com o verbo ou sem realização dos complementos verbais não são uma opção gramatical nestas línguas.

(iii) – Deste-lhe o livro? (português) / Le diste el libro? (espanhol) / Li has donat el llibre? (catalão)

– Dei. / Sim. / Sim, dei. / Sim, dei-lho. (português)

– *Di. / Sí. / *Sí, di. / Sí, se lo di. (espanhol)

– *He donat. / Sí. / *Sí, he donat. / Sí, l’hi he donat. (catalão)

exemplos dos autores. O primeiro tipo de traços (designados por Farkas/Bruce *relative polarity features*) é específico das frases assertivas responsivas (*responding assertions*), enquanto o segundo tipo de traços (designados por Farkas/Bruce *absolute polarity features*) é comum a todos os tipos de frases e define só por si a polaridade das asserções não responsivas (*initiating assertions*).

- (1) a. Anne: Sam is home. / Is Sam home?
 Ben: Yes he is. [same, +]
 Connie: No, he isn't. [reverse, -]
- b. Anne: Sam is not home. / Is Sam not home?
 Ben: Yes, he is. [reverse, +]
 Connie: No, he isn't. [same, -]

Neste capítulo usaremos os termos *afirmação concordante* (equivalente a [same, +]), *afirmação discordante* (equivalente a [reverse, +]), *negação concordante* (equivalente a [same, -]) e *negação discordante* (equivalente a [reverse, -]) para identificar os quatro tipos relevantes de respostas polares e observaremos quais os padrões lexicais e sintáticos que correspondem a cada tipo.

Algumas línguas do mundo (como o inglês, por exemplo) parecem ter gramáticas mais centradas na oposição afirmação/negação, enquanto outras (como o japonês, por exemplo) parecem privilegiar a oposição confirmação/inversão, conforme se exemplifica respetivamente em (2)–(3) e (4)–(5). Como se pode ver nos exemplos, o inglês usa a mesma palavra (*yes*) para a afirmação concordante e discordante e a mesma palavra (*no*) para a negação concordante e discordante. O japonês, por seu lado, usa a mesma palavra (*hai*) para a afirmação e negação concordantes e a mesma palavra (*ie*) para a afirmação e negação discordantes.

Inglês

- (2) A: It's hot today, isn't it?
 B: a. Yes, it is. *afirmação concordante*
 b. No, it isn't. *negação discordante*
- (3) A: It isn't hot today, is it?
 B: a. Yes, it is. *afirmação discordante*
 b. No, it isn't. *negação concordante*

Japonês

- (4) A: Kyoo-wa atu-I des-u ne.
 'Está calor hoje, não está?'
- B: a. Hai, soo des-u ne. *afirmação concordante*
 '[confirmação], está calor hoje.'
- b. Ie, atuku-waarimasen *negação discordante*
 '[inversão], não está calor hoje.'
- (5) A: Kyoo-wa atuku-na-I des-u ne.
 'Não está calor hoje, pois não?'

- B: a. Iie, kyoo-wa atu-I des-u afirmação *discordante*
 '[inversão], está calor hoje.'
- b. Hai, soo des-u ne. negação *concordante*
 '[confirmação], não está calor hoje.'

A par das línguas que têm sistemas lexicalmente binários, como ilustrado acima pelo inglês (*yes-no*) e o japonês (*hai-ii*), existem outras que têm sistemas lexicalmente ternários, como exemplificado em (6)–(7) com o francês (*oui-non-si*) e em (8)–(9) com o alemão (*ja-nein-doch*). Mais raramente encontram-se também sistemas com quatro itens lexicais distintos para cada um dos quatro tipos de resposta polar.²

Francês

- (6) A: Est-ce qu'il viendra? 'Ele virá?'
 B: a. Oui, il viendra. afirmação *concordante*
 'Sim, virá.'
 b. Non, il ne viendra pas. negação *discordante*
 'Não, não virá.'
- (7) A: Est-ce qu'il ne viendra pas?
 'Ele não virá?'
 B: a. Si, il viendra. afirmação *discordante*
 '[inversão], virá.'
 b. Non, il ne viendra pas. negação *concordante*
 'Não, não virá.' *Alemão*
- (8) A: Ist er da? 'Ele está aqui?'
 B: a. Ja, er ist da. afirmação *concordante*
 'Sim, ele está aqui.'
 b. Nein, er ist nicht da. negação *concordante*
 'Não, ele não está aqui.'
- (9) A: Ist er nicht da?
 'Ele não está aqui?'
 B: a. Doch, er ist da. afirmação *discordante*
 '[inversão], ele está aqui.'
 b. Nein, er ist nicht da. negação *concordante*
 'Não, ele não está aqui.'

A organização bipartida, tripartida ou quadripartida dos sistemas responsivos das línguas do mundo é, no entanto, aparente. De facto, todas as línguas têm como exprimir, no plano gramatical, os quatro valores distintos das respostas polares, embora a maior parte das línguas não possa fazê-lo numa base puramente lexical.

² É o caso do Chaha, língua semítica falada na Etiópia. Cf. Pope (1976, 123) e Jones (1999, 307, 315).

Neste capítulo mostraremos como esses quatro valores distintos são gramaticalmente expressos no português europeu contemporâneo. Tal como observado por Pope (1976) e por Farkas/Bruce (2010), estratégias sintáticas, e não puramente lexicais, são usadas por diferentes línguas para dar expressão às opções pragmaticamente mais marcadas, em particular a afirmação discordante.³ Exemplificando, por agora, com o inglês, a palavra *yes*, por si só, pode ser uma resposta natural à pergunta em (10), mas para muitos falantes do inglês não é uma resposta possível a (11). A diferença relevante entre os dois exemplos é que no primeiro temos um contexto de afirmação concordante e no segundo um contexto de afirmação discordante, caso em que a resposta mínima, com *yes*, não parece adequada (cf. Pope 1976; Jones 1999; Holmberg 2003).

- (10) A: Did she go to the party? / She went to the party, didn't she?
 B: a. Yes. *afirmação concordante*
 b. Yes, she did.

- (11) A: Didn't she go to the party? / She didn't go to the party, did she?
 B: a. *Yes *afirmação discordante*
 b. Yes, she did.

Um último ponto a considerar antes de nos centrarmos no português é o próprio conceito de *resposta*. O artigo de Farkas/Bruce (2010) intitula-se *On Reacting to Assertions and Polar Questions* e os autores demonstram que o mesmo quadro conceptual e formal é necessário para descrever respostas a interrogativas polares e respostas a asserções. Uma e outras são gramaticalmente afins, embora existam padrões de resposta mais neutros e outros mais marcados. Numa perspetiva semântica e pragmática, as respostas a interrogativas polares são esperadas enquanto as respostas a asserções podem não o ser e tenderão a emergir, sobretudo, em contextos de discordância e ênfase.

Este capítulo está organizado em 3 secções, além desta introdução. A secção 2 é dedicada às respostas mínimas a interrogativas polares e considera as seguintes questões: em que casos as respostas verbais não repetem o verbo que ocorre na frase interrogativa; em que casos as respostas verbais e as respostas com *sim* estão em distribuição complementar e não em variação entre si; de que outros tipos de resposta mínima afirmativa dispõe o português (respostas adverbiais e respostas com *ser*); em que medida a oposição confirmação/inversão (em interação com a oposição afirmação/negação) é importante para caracterizar os padrões de resposta mínima afirmativa. A secção 3 ocupa-se das respostas alargadas (ou seja, as palavras *sim/não* associadas a frases assertivas) e mostra como a posição frásica inicial ou final das palavras *sim/não* funciona como uma estratégia sintática que dá expressão à oposição

³ As estratégias prosódicas são igualmente relevantes mas não são o objeto deste capítulo.

confirmação/inversão. A secção 4 centra-se na expressão gramatical da afirmação discordante como reacção a asserções. Neste contexto, pragmaticamente marcado, a afirmação discordante tem sido designada na literatura por «afirmação enfática» e exhibe no português europeu um padrão de reduplicação do verbo que não tem paralelo na generalidade das línguas românicas, à exceção do galego. As propriedades da construção de reduplicação verbal enfática descrevem-se na secção 4.2.

2 Respostas mínimas a interrogativas polares

Como resposta à pergunta «Já viste o último filme do Woody Allen?», o português tem como opções de resposta afirmativa «Vi.», «Sim.» e «Já.», a par da resposta negativa «Não.». A possibilidade da resposta adverbial com *já* (e advérbios similares) depende da presença deste advérbio na frase interrogativa, a resposta verbal também parece repetir a forma verbal presente na pergunta e a resposta com *sim* parece menos condicionada pela forma da frase interrogativa. Estas primeiras impressões são em parte justificadas, em parte enganadoras. Nesta secção começaremos por mostrar que nas respostas verbais nem sempre há transposição da forma verbal finita presente na pergunta. Observaremos depois que *sim*, só por si, não é capaz de exprimir afirmação discordante, em contraste com as respostas verbais, mas veremos também que em contextos particulares é necessário responder com *sim* (ainda que as respostas verbais constituam o padrão preferencial de resposta afirmativa em português). Antes de nos debruçarmos sobre as respostas adverbiais (i.e. respostas que repetem um advérbio que ocorre na frase interrogativa), consideraremos o caso das respostas com *ser*. Embora claramente marginais para muitos falantes do português europeu (diferentemente do que acontece no português brasileiro), as respostas com *ser* são interessantes por exprimirem sempre confirmação (i.e. afirmação ou negação concordantes). Isto mostra-nos que dentro de um mesmo sistema gramatical podem coexistir, produtivamente, palavras responsivas cuja relação com os binómios afirmação/negação e confirmação/inversão não é uniforme.

2.1 Quando a resposta verbal não é repetição

Em geral, as respostas verbais repetem o verbo finito que ocorre na frase interrogativa, parecendo corresponder por isso ao que alguns autores classificam por respostas-eco (Jones 1999). No entanto, a dependência das respostas verbais em relação ao seu antecedente não é tão estrita quanto o termo resposta-eco sugere já que a forma verbal na resposta pode divergir da forma verbal no antecedente em mais do que os traços de pessoa-número.

De facto, as respostas verbais apresentam um certo grau de autonomia em relação ao seu antecedente, o que permite, por exemplo, que ocorram no contexto de interro-

gativas truncadas: «– Para a praia? – Vamos.». Além disso, exprimindo sempre asserções afirmativas, afastam-se necessariamente da forma verbal presente no antecedente interrogativo quando esta é incompatível com a interpretação assertiva. Isso acontece quando a frase interrogativa inclui o futuro ou condicional morfológicos exprimindo modalidade epistêmica (com valores de dúvida, incerteza, probabilidade). A resposta afirmativa a uma frase interrogativa com futuro/condicional epistêmico implica obrigatoriamente a alteração do tempo do verbo, substituindo-se o futuro/condicional pelo presente ou pretérito.

- (12) A: O João teria passado no exame?!
 B: a. Passou.
 b. *Teria.
- (13) A: Eles (não) estarão em casa?
 B: a. Estão.
 b. *Estarão.

Mesmo quando o futuro morfológico não tem valor modal (\nearrow 12 Semântica e pragmática), a resposta com o presente do indicativo, em substituição do futuro, parece mais natural, por ser mais forte o grau de assertividade que o presente do indicativo imprime à resposta afirmativa.⁴

- (14) A: Farás isso por mim?
 B: a. Faço.
 b. ?Farei.
- (15) A: Mentirás para o encobrir?
 B: a. Minto.
 b. ?Mentirei.

As frases interrogativas que incluem o advérbio modal *talvez* apresentam o verbo no conjuntivo. Uma vez que o conjuntivo precisa de legitimação sintática (não podendo ocorrer isolado), a resposta verbal a uma interrogativa com *talvez* envolve sempre a substituição do conjuntivo pelo indicativo. Em alternativa, a resposta mínima afirmativa pode construir-se repetindo o advérbio *talvez*, mas os dois padrões de resposta não são equivalentes interpretativamente, já que o valor modal de *talvez* é eliminado na resposta verbal.

- (16) A: O João talvez saia do hospital hoje?
 B: a. Talvez. (Possível continuação: Se sair, telefonam-nos do hospital.)
 b. Sai. (Possível continuação: Acabaram de telefonar do hospital.)

⁴ Em português, o presente morfológico pode exprimir semanticamente o futuro (\nearrow 12 Semântica e pragmática).

Quando o verbo auxiliar *ter*, um auxiliar modal como *poder* ou outros verbos auxiliares integram a frase interrogativa, a resposta mínima afirmativa pode construir-se repetindo a forma finita do verbo auxiliar ou alterando a forma de particípio passado ou infinitivo do verbo principal para uma forma finita. Dependendo das propriedades semânticas do verbo auxiliar e de fatores pragmáticos, os dois padrões de resposta podem ser ou não interpretativamente equivalentes, como se exemplifica em (17) a (19). O facto de o português admitir elipse do predicado (cf. nota 1) explica que nas respostas em que ocorre apenas o verbo auxiliar seja interpretativamente recuperado o verbo principal.

- (17) A: Eu não vos tinha contado isto?
 B: a. Tinhas.
 b. Contaste.
- (18) A: Ele tem tomado os comprimidos?
 B: a. Tem.
 b. Toma.
- (19) A: Podes emprestar-me uma caneta?
 B: a. Posso. (Continuação possível, embora improvável: Mas não empresto.)
 b. Empresto. (A continuação em a. é impossível.)

2.2 Respostas verbais, respostas com *sim* e respostas com *ser*

Enquanto a resposta mínima negativa a uma interrogativa polar apresenta quase invariavelmente a palavra *não* (cf. respostas com *ser* no final desta secção), várias opções existem no português para exprimir uma resposta mínima afirmativa, conforme se ilustra em (20). Nesta secção começaremos por analisar o que distingue as respostas verbais (cf. 20a) das respostas com *sim* (cf. 20b). Depois compararemos estes padrões de resposta, mais comuns, com o padrão mais marginal (por não ser produtivo para todos os falantes do português europeu) de respostas com *ser* (cf. 20c). As respostas adverbiais serão consideradas na secção 2.3 (cf. 20d).

- (20) A: Ele já terminou o curso?
 B: a. Terminou. *afirmação concordante*
 b. Sim. *idem*
 c. Foi. *idem*
 d. Já. *idem*
 e. Não. *negação discordante*

Uma primeira observação importante relativamente à distinção entre respostas verbais e respostas com *sim* é que só as primeiras expressam afirmação discordante. Os dados em (21) ilustram este contraste. O sinal # em (21b) indica que a resposta com *sim* não é apropriada (nem interpretável), por razões que esclareceremos mais adiante.

Em resposta a uma interrogativa negativa, *sim* poderia apenas ser interpretado como marcador de negação concordante, mas a existência de *não* como marcador de negação parece bloquear o recurso a *sim* neste contexto.⁵ A negação discordante, tal como a negação concordante, é expressa por *não* nas respostas polares mínimas (i.e. aquelas em que a palavra que fixa o valor da polaridade ocorre isolada).

- (21) A: Ele hoje não vai sair, pois não?
 B: a. Vai. *afirmação discordante*
 b. #Sim.
 c. Não. *negação concordante*

O Quadro 1 sintetiza o funcionamento dos principais marcadores de polaridade nas respostas mínimas, indicando que as respostas verbais e as respostas com *sim* não estão em variação livre em português. Enquanto as respostas verbais traduzem, intrinsecamente, a polaridade afirmativa, seja concordante seja discordante, as respostas com *sim* expressam concordância com um antecedente preferencialmente afirmativo (com exceções pontuais). A interpretação da resposta com *sim* depende da recuperação do antecedente interrogativo, incluindo o seu valor polar. A seguir identificaremos outros contextos de pergunta-resposta que ajudam a clarificar a diferença entre respostas verbais e respostas com *sim*.⁶

Quadro 1: Respostas mínimas a interrogativas polares.

	concordante / confirmação	discordante / inversão
afirmação	verbo / <i>sim</i>	verbo
negação	<i>não</i>	<i>não</i>

Uma vez que a interpretação das respostas com *sim* depende da recuperação do antecedente como um todo, o que não acontece nas respostas verbais, os dois padrões de resposta não são interpretativamente equivalentes em contexto de interrogativa

⁵ As frases interrogativas negativas que não são perguntas típicas mas uma forma cortês de fazer um pedido, oferta, convite ou comentário permitem respostas com *sim* porque antecipam uma resposta positiva. Neste contexto, as respostas com *sim* são instâncias de afirmação concordante.

- (i) A: Não me podes emprestar uma caneta, por favor?
 B: Sim (claro).
 (ii) A: Não queres um chocolate?
 B: Sim (por favor).

⁶ Em línguas que não dispõem de respostas verbais, como o espanhol e o catalão, os correlatos do português *sim* podem exprimir tanto afirmação concordante como discordante:

- (i) A: ¿(No) va a ir al cine hoy Juan? (espanhol) / Avui (no) anirà al cine en Joan? (catalão)
 ‘O João (não) vai hoje ao cinema?’
 B: Sí. (= Sí, Juan va a ir al cine hoy) (espanhol) / Sí. (= Sí, en Joan anirà al cine avui) (catalão)
 ‘Sim, vai.’ / ‘Vai sim.’

indireta. Como se exemplifica em (22), *sim* recupera necessariamente o verbo superior (e não, diretamente, a oração subordinada), o que torna (22-B-a) uma resposta pouco natural à pergunta em (22-A). Pelo contrário, a resposta verbal tem um certo grau de autonomia relativamente ao antecedente, correspondendo a uma frase declarativa com traços de polaridade afirmativa especificados. A sua interpretação só depende do antecedente relativamente aos complementos verbais não realizados (por elipse do predicado). No exemplo (22), a resposta verbal é a resposta natural à interrogativa indireta.⁷

- (22) A: Sabes se o João foi à festa?
 B: a. # Sim. (= 'Sim, sei.')

b. Foi. (= 'Sim, foi.')

O facto de as respostas com *sim* recuperarem necessariamente o antecedente na sua totalidade, também explica os dados em (23) e (24). Quando na frase interrogativa ocorre um auxiliar modal seguido de infinitivo, a resposta verbal pode incluir ou excluir o predicado modal, mas a resposta com *sim* inclui-o obrigatoriamente, no plano interpretativo. Daí decorre que a resposta verbal com o verbo modal é equivalente à resposta com *sim*, enquanto a resposta verbal com o verbo principal difere na sua interpretação da resposta com *sim*.

- (23) A: Podes emprestar-me uma caneta?
 B: a. Posso. / Sim. (Continuação possível, embora improvável: Mas não empresto)
 b. Empresto. (A continuação em a. é impossível)

- (24) A: Podes emprestar-me dinheiro?
 B: a. Empresto, embora não possa.
 b. #Sim, embora não possa.

Do mesmo modo, se o antecedente interrogativo incluir o advérbio *talvez*, a resposta com *sim* é equivalente à resposta adverbial com *talvez* e não à resposta verbal (cf., acima, o exemplo (16)).

Por outro lado, o carácter intrinsecamente afirmativo das respostas verbais faz com que sejam incompatíveis com a presença no antecedente interrogativo de elementos que introduzam uma implicatura negativa (como os advérbios *só*, *quase*, *mal*, este

7 No espanhol e no catalão a resposta com *sí* constitui uma resposta adequada a uma interrogativa indireta:

- (i) A: ¿Sabes si Juan fue a la fiesta? (espanhol) / ¿Saps si en Joan va anar al cine? (catalão)
 'Sabes se o João foi à festa?'
 B: a. Sí. (= Sí, lo sé. / Sí, Juan fue a la fiesta.) (espanhol)
 (= 'Sim, sei.' / 'Sim, o João foi à festa')
 b. Sí. (= Sí, ho sé. / Sí, en Joan va anar al cine.) (catalão)
 (= 'Sim, sei.' / 'Sim, o João foi ao cinema.')

último com a interpretação de ‘quase não’).⁸ Pelo contrário, as respostas com *sim* são adequadas neste contexto porque a palavra *sim* recupera o contributo semântico do advérbio e exprime concordância com o valor negativo da implicatura. Se o advérbio puder constituir uma resposta adverbial (casos de *só* e *quase*, mas não de *mal*), a resposta com *sim* e a resposta adverbial são equivalentes e igualmente naturais, enquanto as respostas verbais são excluídas.

(25) A: Ele mal tocou na comida?

- B: a. Sim.
b. *Tocou.

(26) A: Ele só dorme?!

- B: a. Sim.
b. Só.
c. *Dorme.

Tomemos como exemplo o advérbio *só*. A resposta a uma interrogativa polar que inclua *só* decompõe-se, interpretativamente, em duas partes: a confirmação de uma pressuposição positiva e a confirmação de uma implicatura negativa (cf. Horn 1969; von Stechow 2007). O primeiro componente é uma instância de afirmação concordante, enquanto o segundo é uma instância de negação concordante. A palavra *sim* pode expressar ambos os valores polares (tratando-se de uma palavra essencialmente concordante), mas o verbo *só* pode expressar a afirmação concordante, sendo por isso excluído dos contextos em que uma resposta semântica e pragmaticamente adequada exige que se expressem simultaneamente afirmação e negação concordantes. O exemplo (27) clarifica os dois componentes da resposta a uma interrogativa polar que transporte uma implicatura negativa.

(27) A: Ele só comeu a sopa?

(pressuposição: ele comeu a sopa)

B: Sim:

- (i) Sim, ele comeu a sopa. *afirmação concordante*
(ii) Sim, ele não comeu senão a sopa. *negação concordante*

Há um contraste muito claro entre o marcador de foco exclusivo *só* e os marcadores de foco inclusivo *até* e *também* relativamente ao assunto em discussão. A total naturalidade de respostas verbais a interrogativas polares com *até* e *também* a preceder o verbo, em contraste com *só* e advérbios similares, mostra que não é a posição estrutural do advérbio *só* por si (i.e. ser externo ou interno ao sintagma verbal) que importa.⁹ Mais

⁸ Para uma abordagem diferente destes factos, veja-se Santos (2002; 2003; 2009).

⁹ Quando o advérbio *só* é interno ao sintagma verbal, pode ser sintaticamente recuperado numa resposta verbal, juntamente com os complementos do verbo (cf. nota 1).

(i) A: Ele comeu só as batatas?

B: a. Comeu.
b. Sim.

determinantes são as suas propriedades polares, especificamente se o advérbio introduz ou não uma implicatura negativa. Nas respostas em (28) e (29), embora os advérbios *até* e *também* não sejam sintaticamente recuperados através dos processos gramaticais associados à elipse do predicado, a sua interpretação pode ser pragmaticamente recuperada, o que não acontece com o advérbio *só* no exemplo (30).

- (28) A: Ele até/também ressona?
 B: a. Ressona.
 b. Sim.
- (29) A: Até/também ele sabe a verdade?
 B: a. Sabe.
 b. Sim.
- (30) A: Só ele sabe a verdade?
 B: a. *Sabe.
 b. Sim.

Algumas frases complexas interrogativas possibilitam facilmente uma interpretação de foco contrastivo associada à oração subordinada, como se exemplifica em (31). Com a interpretação relevante, o foco contrastivo introduz uma implicatura negativa e uma resposta verbal mostra-se de novo inadequada. Este facto é uma consequência lógica das observações anteriores e suporta a generalização de que as respostas mínimas verbais são estritamente afirmativas e incapazes de expressar qualquer forma de negação.

- (31) A: Ele está preso porque contou tudo?
 (interpretação relevante: 'É porque contou tudo que ele está preso?')
 B: a. #Está.¹⁰
 b. #Contou.
 c. Sim.
 (i) Sim, ele está preso. (confirmação da pressuposição) *afirmação concordante*
 (ii) Sim, não está preso senão porque contou tudo. *negação concordante*

10 A resposta com o verbo da oração principal seria adequada se o facto de ele estar preso fosse desconhecido, i.e. não pressuposto. Excetuando o caso das interrogativas indiretas, as respostas verbais a frases complexas normalmente retomam o verbo superior:

- (i) A: Ele estava preso quando contou tudo?
 B: a. Estava.
 b. *Contou.
- (ii) A: Ela disse que ele contou tudo?
 B: a. Disse.
 b. #Contou. (Só possível se corresponder à intenção de dar proeminência informacional ao facto de ele ter contado tudo.)

Em conclusão, os traços de polaridade das respostas verbais são necessariamente especificados com o valor [afirmação]. Por isso este tipo de resposta não permite a recuperação de implicaturas negativas com escopo sobre o predicado verbal, sejam elas introduzidas por advérbios como o marcador de foco exclusivo *só* sejam introduzidas por outros mecanismos estruturais associados à expressão de foco contrastivo.

Um último caso a considerar para descrever as diferenças entre respostas verbais e respostas com *sim* é o das estruturas de coordenação. As interrogativas polares que integram uma sequência de orações coordenadas admitem uma resposta com *sim* mas não com o verbo. Uma vez que a interpretação de *sim* depende do antecedente no seu todo, a estrutura coordenada é integralmente recuperada pela resposta com *sim*. Nas respostas verbais, pelo contrário, normalmente recupera-se do antecedente apenas a parte da estrutura que corresponde ao complemento do verbo retomado na resposta. Deste modo, em (32), *sim* responde globalmente às duas componentes oracionais da pergunta, o que a resposta verbal não permite. A resposta (b) será possível apenas se a informação associada ao primeiro membro da estrutura coordenada for considerada parte do conhecimento partilhado.

- (32) A: A Joana arranjou emprego e comprou uma casa?
 B: a. *Arranjou.
 b. #Comprou.
 c. Sim.

O Quadro 2 sintetiza as principais diferenças entre as respostas verbais e as respostas com *sim*, ao mesmo tempo que mostra em que contextos os dois padrões de resposta estão em variação livre (descontadas questões de preferência ou de grau de adequação pragmática em situações que não são objeto de discussão neste capítulo). Lembre-se que embora as respostas com *sim* não possam expressar negação concordante em contexto de interrogativas negativas, revelam contudo compatibilidade com a negação quando o antecedente interrogativo é uma frase afirmativa mas contém um elemento que associa a uma pressuposição positiva uma implicatura negativa.

Quadro 2: Respostas mínimas afirmativas a interrogativas polares: verbo vs. *sim*.

	verbo	<i>sim</i>
expressa afirmação concordante	+	+
expressa afirmação discordante	+	-
expressa negação concordante em respostas a perguntas que introduzem implicaturas negativas (com escopo sobre o predicado verbal)	-	+
responde adequadamente a interrogativas indiretas	+	-
responde adequadamente a interrogativas polares que incluem uma sequência de orações ligadas por coordenação	-	+

Em conclusão: as respostas mínimas com *sim* estão associadas a traços de polaridade absoluta (afirmação/negação) não especificados e extraem a sua interpretação do antecedente como um todo; as respostas mínimas verbais têm uma certa autonomia em relação ao antecedente, estão independentemente associadas a um traço de polaridade afirmativa e têm a força assertiva típica das frases declarativas.

Além das respostas verbais e das respostas com *sim*, o português admite um padrão de resposta mínima que usa formas da terceira pessoa do singular do verbo *ser*. Este padrão oscila entre recorrer invariavelmente à forma de terceira pessoa do singular do presente do indicativo (*é*) ou exibir distinções de tempo e aspeto entre presente do indicativo, pretérito perfeito e pretérito imperfeito, sempre na terceira pessoa do singular (*é/foi/era*).

(33) A: Eles encontraram as chaves?

- B: a. Encontraram.
 b. Sim.
 c. Foi.
 d. É.
 e. *Foram.

(34) A: Ela tinha uma avó belga?

- B: a. Era.
 b. É.

As respostas com *ser* não são usadas, ou sentidas como naturais, por todos os falantes do português europeu. Para alguns falantes, as respostas com *ser* a interrogativas polares vão de pouco naturais (quando permitem flexão de tempo e aspeto) a agramaticais (quando envolvem a forma invariável *é*). Para os mesmos falantes, no entanto, as confirmações de um antecedente assertivo usando uma forma do verbo *ser* são naturais, quer a forma verbal ocorra sozinha quer ocorra associada à palavra confirmativa *pois*. Isto sugere que as confirmações com a palavra *pois* associada à cópula *ser* podem ser a origem diacrónica das respostas com *ser* a interrogativas polares e asserções.¹¹

11 A natureza confirmativa das respostas com *ser* explica que não possam ocorrer como reação a interrogativas que são atos de fala indiretos, correspondendo a uma forma cortês de fazer um pedido, oferta, convite ou comentário.

(i) A: Podes vir aqui?

- B: a. Posso./Vou.
 b. Sim.
 c. *É.

(ii) A: Vai um cafezinho?

- B: a. Vai.
 b. Sim.
 c. *É.

- (35) A: Eles já encontraram as chaves.
 B: (Pois) foi.

Diferentemente das respostas verbais e mais generalizadamente do que as respostas com *sim*, as respostas com *ser* expressam concordância com o antecedente. Ou seja, não são especificadas para nenhum dos valores dos traços de polaridade absoluta (afirmação/negação), mas são especificadas para o valor confirmativo dos traços de polaridade relativa (confirmação/inversão). Podem portanto exprimir, sem restrições, negação concordante, quer em relação a uma interrogativa negativa (cf. 36) quer em relação a uma interrogativa positiva que introduza uma implicatura negativa (cf. 37). Expressam também afirmação concordante, mas em contraste com as respostas verbais são incompatíveis com a expressão da afirmação discordante.

- (36) A: Ele não vem?
 B: a. Vem. *afirmação discordante*
 b. É. (= 'Ele não vem.'). *negação concordante*
- (37) A: Ele só viaja amanhã?
 B: a. É. (= 'Ele só viaja amanhã.')

Além disso, as respostas com *ser* divergem das respostas verbais pela impossibilidade de ocorrerem quando o antecedente é uma interrogativa indireta. Neste contexto as respostas com *ser* distinguem-se também das respostas com *sim*, pois enquanto estas são apenas pragmaticamente inadequadas as respostas com *ser* são claramente agramaticais.

- (38) A: Sabem se ele telefonou ao pai?
 B: a. Telefonou.
 b. #Sim. (= 'Sabemos.')

c. *É./*Foi. (sem interpretação possível)

Por fim, as respostas com *ser* comportam-se como as respostas com *sim* quando a interrogativa polar que as legitima corresponde a uma estrutura de coordenação oracional.

- (39) A: A Joana arranhou emprego e comprou uma casa?
 B: a. Foi.
 b. É.
 c. Sim.
 d. *Arranhou.
 e. #Comprou.

As diferentes propriedades das respostas verbais, respostas com *sim* e respostas com *ser*, como resposta mínima a interrogativas polares, estão sintetizadas no Quadro 3.¹²

Quadro 3: Respostas mínimas a interrogativas polares: verbo vs. *sim* vs. *ser*-3SG.

	verbo	<i>sim</i>	<i>ser</i>
Expressa afirmação concordante	+	+	+
Expressa afirmação discordante	+	-	-
Expressa negação concordante	-	+/-	+
Responde adequadamente a interrogativas indiretas	+	-	-
Responde ao predicado superior que introduz uma interrogativa indireta	+	+	-
Responde adequadamente a interrogativas polares que incluem uma sequência de orações ligadas por coordenação	-	+	+

2.3 Respostas adverbiais

As respostas adverbiais características do sistema responsivo do português (as quais diferem das respostas com advérbios evidenciais como *claro*, *evidentemente*, comuns a todas as línguas) caracterizam-se, tipicamente, pela repetição na resposta de um advérbio presente na interrogativa polar. Pertencem a este grupo os advérbios *também*, *só*, *já*, *ainda*, *sempre*, *quase*, *talvez*. Os três últimos, e apenas estes, podem constituir uma resposta polar sem que ocorram no antecedente interrogativo. Por outro lado, *já* e *ainda* permutam entre si quando expressam discordância, como mostraremos mais adiante.

(40) A: Tu também queres morangos?
B: Também.

(41) A: O Pedro pediu morangos. E tu, queres morangos?
B: *Também.

(42) A: Ele só gosta de morangos?
B: Só.

(43) A: Ele gosta de morangos?
B: *Só.

¹² No português brasileiro as respostas com *ser* adquiriram um papel mais central do que no português europeu como resultado do declínio do uso das respostas mínimas com *sim* (cf. Kato/Tarallo 1992; Oliveira 1996). De facto, no português brasileiro coloquial as respostas com *ser* substituíram as respostas com *sim* nos contextos em que as respostas verbais não são uma opção gramatical. Mas as duas variedades do português são idênticas na preferência geral pelas respostas mínimas verbais.

(44) A: Ele tem o perfil certo? / Ele tem quase o perfil certo?

B: Quase.

(45) A: Lês poesia? / Lês sempre poesia?

B: Sempre.

O comportamento do conjunto de advérbios responsivos que estamos a considerar revela alguma variação mas também um espectro largo de similitude. Uma propriedade comum a todos estes advérbios é a possibilidade de exprimirem confirmação, quer sob a forma de afirmação concordante quer de negação concordante. Em geral associam-se a *não* (ou outro marcador de negação) quando traduzem negação discordante, mas os advérbios *talvez*, *quase* e *só* são capazes de exprimir este valor isoladamente, dispensando o marcador de negação. Lembre-se que estes são os advérbios que introduzem implicaturas negativas, como discutimos na secção 2.2 (cf. Horn 1969; von Stechow 2007).

(46) A: A Rita também vem jantar?

B: Também. *afirmação concordante*

(47) A: Tu também não queres que ele seja eleito?

B: Também não. *negação concordante*

(48) A: As roseiras já foram podadas?

B: Já. *afirmação concordante*

(49) A: Ele já não vai viajar?

B: Já não. *negação concordante*

(50) A: Só esse médico poderá operar o golfinho?

B: Só. *afirmação concordante*

(51) A: A tua filha só não gosta de bacalhau?

B: Só. *negação concordante*

(52) A: Ele quase morreu?

B: Quase. *afirmação concordante*

(53) A: Ele quase não morreu?

B: Quase. *negação concordante*

Em associação com o marcador de negação, a maior parte dos advérbios relevantes podem também exprimir negação discordante (*ainda, já, talvez, sempre, só*), enquanto alguns expressam apenas confirmação, ocorrendo exclusivamente nos contextos polares associados à afirmação e negação concordantes (*também, quase*). *Já* e *ainda* alternam entre si para exprimir negação discordante.

- (54) A: Só ele disse a verdade?
B: Não só. *negação discordante*
- (55) A: Ele sempre usou chapéu?
B: Nem sempre. *negação discordante*
- (56) A: Ele já chegou a casa?
B: Ainda não. *negação discordante*
- (57) A: Tu ainda vais a casa?
B: Já não. *negação discordante*

Uma importante generalização relativa aos advérbios responsivos é que, em geral, não podem exprimir afirmação discordante. Neste aspeto assemelham-se às respostas com *sim* e diferem das respostas verbais, como se ilustra em (58). Assemelham-se também às respostas com *sim* noutros aspetos, como se mostra em (59) e (60), correspondendo o primeiro exemplo a um caso de coordenação oracional no antecedente interrogativo e o segundo à presença na interrogativa polar do advérbio *quase*, que introduz uma implicatura negativa (cf. acima o Quadro 2). A exceção a este padrão caracterizador dos advérbios responsivos é o advérbio *já*, que pode exprimir afirmação discordante quando no antecedente interrogativo ocorre o advérbio *ainda*, como se vê em (61).¹³

- (58) A: Não compras sempre produtos biológicos, pois não?
B: a. Compro. *afirmação discordante*
b. *Sempre.
c. *Sim.
- (59) A: Ele já lavou os dentes e foi-se deitar?
B: a. Já.
b. Sim.
c. *Lavou.
- (60) A: Ele quase morreu envenenado?
B: a. Quase.
b. Sim.
c. *Morreu.
- (61) A: Tu ainda não foste a casa?
B: a. Já. *afirmação discordante*
b. Fui.
c. *Sim.

13 O advérbio *já* é também o único que pode exprimir afirmação discordante através de um padrão de reduplicação (cf. secção 3) e ocorrer em interrogativas-*tag* («– Já viste este filme, não já? – Já vi, já.»). Nos dois casos alterna com o verbo («– Já viste este filme, não viste? – Vi, vi.»).

3 Respostas polares alargadas

Nas respostas alargadas a interrogativas polares, as palavras *sim/não* ocorrem na periferia (inicial ou final) de uma frase com elipse dos complementos do verbo.

- (62) A: O João recebeu os livros que encomendou?
 B: a. Sim, recebeu.
 b. Não, não recebeu.

Dois factos novos e interessantes podem observar-se quando se consideram os padrões de resposta polar alargada (comparativamente às respostas mínimas). Por um lado, comprova-se que a palavra *sim* é essencialmente confirmativa, pois em respostas alargadas pode associar-se a uma frase negativa para exprimir negação concordante, como se mostra em (63-B-a), embora a alternativa em (63-B-b), com *não* periférico inicial, em lugar de *sim*, seja a opção não marcada.

- (63) A: O João não vai sair hoje?
 B: a. Sim, não vai.
 b. Não, não vai.

Por outro lado, as respostas alargadas mostram-nos que a posição (inicial/final) dos marcadores de polaridade na frase desempenha um papel muito relevante no sistema responsivo do português. Assim, enquanto a posição inicial das palavras *sim/não* é típica das respostas pragmaticamente mais neutras (ou seja, as que não contradizem expectativas do interlocutor), a posição final é típica das respostas pragmaticamente mais marcadas, na medida em que contradizem uma asserção prévia expressa por uma frase declarativa (que pode ser parte de uma interrogativa-*tag*). Como se vê comparando as respostas a (64-A) com as respostas a (65-A), a oposição entre posição inicial e posição final de *sim/não* periféricos marca a dicotomia entre confirmação e inversão. Esta marcação é claramente preferencial para traduzir a distinção entre afirmação concordante e afirmação discordante e opcional no que diz respeito à distinção entre negação concordante e negação discordante. A probabilidade de as palavras *sim/não* ocorrerem no final da frase será tanto maior quanto mais o falante quiser enfatizar a discordância em relação à asserção ou expectativa do interlocutor.

- (64) A: O João hoje janta em casa. / O João hoje janta em casa, não janta?
 B: a. Sim, janta. *afirmação concordante*
 b. Não janta, não. *negação discordante*
 c. Não, não janta. *idem*
- (65) A: O João hoje não janta em casa. / O João hoje não janta em casa, pois não?
 B: a. Janta sim. *afirmação discordante*
 b. Não, não janta. *negação concordante*

No que diz respeito à palavra *sim*, os dois factos acima descritos poderão parecer contraditórios. Se *sim* é uma palavra essencialmente confirmativa (cf. 63-B-a), como se explica que possa contribuir para a expressão da afirmação discordante (cf. 65-B-a)? A contradição é apenas aparente. Como mostrámos na secção 2, a interpretação da palavra *sim* é fixada pelo seu antecedente. Quando *sim* ocorre em posição inicial, como em (63-B-a), o antecedente é a interrogativa polar ou a asserção a que responde; já quando aparece em posição final, o antecedente é a frase declarativa que ocorre à sua esquerda. Neste caso, *sim* reforça o valor assertivo da resposta.

O português dispõe de outras estratégias para exprimir a afirmação discordante. Além da colocação de *sim* em posição final, como em (66-B-a), são também comuns na língua falada a reduplicação do verbo, como em (66-B-b), e a colocação da palavra confirmativa *pois* em posição final, como em (66-B-c).¹⁴ A colocação do elemento de reforço da asserção afirmativa no final da frase é pois comum às três estratégias. Prosodicamente, todas são marcadas por uma entoação ascendente, tanto mais acentuada quanto maior o grau de ênfase atribuído à resposta discordante.

- (66) A: O João hoje não vai trabalhar, pois não? / O João hoje não vai trabalhar.
 B: a. Vai (trabalhar) sim.
 b. Vai (trabalhar) vai.
 c. Vai (trabalhar) pois.

Quando consideramos os padrões de respostas polares alargadas, obtemos um quadro mais rico relativamente ao modo como o sistema responsivo do português expressa os diferentes valores de polaridade definidos, conjuntamente, pelas oposições afirmação/negação e confirmação/inversão. Torna-se então claro que à distinção conceptual quadripartida corresponde uma codificação gramatical também quadripartida, como mostra o Quadro 4.¹⁵

14 Ao contrário de *sim*, a palavra *pois* não pode ocorrer isolada em resposta a interrogativas polares. Associada ao verbo, confirma asserções afirmativas; associada ao marcador de negação predicativa, confirma asserções negativas.

- (i) A: O João comprou um carro?
 B: *Pois.
 (ii) A: O João comprou um carro.
 B: Pois comprou.
 (iii) A: O João não comprou um carro.
 B: Pois não.

15 O português europeu dispõe de outras formas de exprimir a negação enfática que não são descritas neste capítulo (cf. Hagemeyer/Santos 2004; Pinto 2010; Martins 2013). Também as especificidades da gramática do português europeu relativamente à expressão da negação metalinguística não são aqui abordadas (cf. Martins 2010; 2014; Pereira 2010; Pinto 2010).

Quadro 4: Respostas alargadas a interrogativas polares.

	concordante / confirmação	discordante / inversão
afirmação	<i>sim</i> verbo	verbo <i>sim</i> verbo verbo verbo <i>pois</i>
negação	<i>não não</i> verbo <i>sim não</i> verbo (opção marcada)	<i>não</i> verbo <i>não</i> não não verbo (opção não enfática)

Na secção 2.3, dedicada às respostas mínimas adverbiais, vimos que só o advérbio *já* pode exprimir afirmação discordante em resposta a uma interrogativa polar. Outra particularidade deste advérbio é admitir um padrão de reduplicação nas respostas alargadas, também como forma de exprimir (e enfatizar) a afirmação discordante. Na verdade, a presença de *já* numa resposta polar bloqueia a possibilidade de reduplicação do verbo, ocorrendo, em alternativa, a reduplicação do advérbio na mesma posição em que ocorreria o verbo.

- (67) A: O João ainda não saiu, pois não?
 B: a. Saiu, saiu.
 b. *Já saiu, saiu.
 c. Já saiu, já.

4 Afirmação enfática (afirmação discordante como reação a asserções)

Nesta secção final começaremos por mostrar, de forma abreviada, que estratégias o português europeu não partilha com outras línguas românicas para exprimir, com ênfase, a afirmação discordante. Depois centrar-nos-emos numa dessas estratégias, a reduplicação do verbo, para descrevermos as suas propriedades. Para as questões que são deixadas fora do âmbito deste capítulo, vejam-se as referências bibliográficas na nota 15.

4.1 O português europeu, o português brasileiro e outras línguas românicas

O português brasileiro só dispõe de uma das estratégias identificadas no Quadro 4 (cf. secção 3) para expressar a afirmação discordante, a estratégia que recorre à colocação da palavra *sim* no final da frase (e atribui à frase uma entoação particular). A palavra confirmativa *pois* sobrevive no português brasileiro apenas em expressões fixas e não

tem correlatos nas outras línguas românicas.¹⁶ A reduplicação verbal, com as propriedades de estrutura mono-oracional que descreveremos na secção 4.2, é uma especificidade que o português europeu partilha apenas com o galego. Por outro lado, a maioria das línguas românicas dispõe de uma estratégia sintática para expressar a afirmação discordante (enfática) que consiste em colocar o marcador de polaridade afirmativa *sí* antes do verbo, opcionalmente seguido do complementador *que*, como se mostra em (68) e (69) com exemplos, respetivamente, do espanhol e do catalão. O português desconhece esta estratégia (sendo neste aspeto idênticas as variedades europeia e brasileira), enquanto o galego a permite a par das estratégias que partilha com o português, como se exemplifica em (70).

- (68) A: Juan no se compró el coche.
 'J. não comprou o carro.'
 B: Juan sí (que) se compró el coche.
 'Comprou o carro {sim/comprou}.'
- (69) A: En Joan no sopa aquí els dijous.
 'O J. não janta aqui à quinta-feira.'
 B: Sí que sopa aquí.
 'Janta aqui {sim/janta}.'
- (70) A: Aníbal non sabe destes assuntos.
 B: a. Aníbal sabe destes assuntos sabe.
 b. Aníbal si que sabe destes assuntos.
 c. Aníbal sabe destes assuntos si.

O Quadro 5 mostra as semelhanças e diferenças entre português europeu, português brasileiro, galego, espanhol, catalão e italiano no que diz respeito às estratégias usadas para expressar a afirmação enfática (i.e., a afirmação discordante como reação a asserções), considerando três tipos de estruturas sintáticas: com reduplicação do verbo (cf. 70-B-a), com *sí que* (cf. 70-B-b) e com *sim/si* final (cf. 70-B-c). Para uma descrição detalhada dos dados relevantes, cf. Martins (2013).

16 No português brasileiro não existe a palavra *pois* como forma isolada. *Pois* sobrevive apenas em duas expressões gramaticalizadas: *pois não?*, com o significado de 'posso ajudá-lo?', e *pois é* (com a cópula invariavelmente na terceira pessoa do presente do indicativo), guardando neste caso o valor confirmativo que também tem no português europeu:

(i) A: Isto vai acabar mal.
 B: Pois é.

Quadro 5: A expressão da afirmação enfática em seis variedades/línguas românicas.

	Português brasileiro	Português europeu	Galego	Espanhol	Catalão	Italiano
Reduplicação do verbo	*	√	√	*	*	*
<i>sí que</i>	*	*	√	√	√	√
<i>sim/sí final</i>	√	√	√	√	√	√

4.2 Propriedades da construção de reduplicação verbal enfática

Em geral, as línguas permitem recorrer à repetição, incluindo a repetição de frases (elíticas ou não), como recurso discursivo cujo objetivo pode ser enfatizar. Nas próximas secções mostraremos que a reduplicação verbal enfática do português europeu não é um caso de simples repetição. Antes corresponde a uma estrutura mono-oracional com propriedades prosódicas, sintáticas e morfológicas próprias.

4.2.1 Padrão prosódico (ausência de pausa; entoação ascendente)

Enquanto a repetição frásica envolve normalmente uma pausa prosódica a separar as duas frases (que sendo declarativas terão entoação descendente), nas estruturas de reduplicação verbal enfática não existe pausa (sendo a vírgula, na escrita, uma mera convenção ortográfica) e as frases têm globalmente uma entoação ascendente. A impossibilidade de produzir a estrutura de reduplicação com uma pausa antes da segunda instância do verbo pode demonstrar-se observando a forma como se realiza o /s/ em coda. No português europeu, uma regra fonológica sem exceções determina que a sibilante surda em coda se realize como palatal surda [ʃ] antes de pausa mas como dental sonora [z] antes de vogal. O exemplo (71-B-a) mostra-nos que quando há repetição frásica a realização da sibilante em coda é [ʃ] ainda que preceda imediatamente a segunda instância de uma forma verbal iniciada por vogal. Pelo contrário, quando há reduplicação verbal enfática e a forma verbal se inicia por vogal, a realização de /s/ em coda é obrigatoriamente [z] no contexto relevante, como mostram os exemplos (71-B-b/c), o que prova que não pode haver uma pausa a separar as duas instâncias do verbo. Outra importante diferença prosódica a assinalar é que a entoação de cada uma das unidades frásicas de (71-B-a) é descendente, enquanto a entoação de (71-B-b) é globalmente ascendente. Com a mesma entoação ascendente, (71-B-c) é agramatical.

- (71) A: Eu não abraço o diretor.
 B: a. Abraça[f]. Abraça[f] sim.
 b. Abraça[z] abraça[f].
 c. *Abraça[f] abraça[f].

4.2.2 Incompatibilidade com advérbios evidenciais

As frases com reduplicação verbal enfática não podem incluir advérbios evidenciais como *realmente*, *certamente*, *efetivamente*, *obviamente* ou a expressão adverbial *de facto* (cf. 72-B-a). Pelo contrário, a repetição frásica reiterativa é totalmente compatível com a presença de advérbios evidenciais (cf. 72-B-b).

- (72) A: O teu filho não gosta de ler.
 B: a. *O meu filho de facto gosta de ler, gosta. (entoação ascendente)
 b. O meu filho (de facto) gosta de ler. (De facto) gosta.

4.2.3 Reduplicação limitada ao verbo finito

É possível repetir uma sequência frásica constituída por um verbo finito seguido de infinitivo ou particípio passado, como se exemplifica em (73) e (74). No entanto, na reduplicação verbal enfática só o verbo finito pode ser repetido, como se mostra em (75) e (76), quer se trate de um verbo auxiliar ou não. Ou seja, nas estruturas de reduplicação nenhum elemento além do verbo temporalmente finito pode reaparecer no final da frase, como se confirmará nas secções seguintes. As frases em (75-B) e (76-B) devem ser lidas com a entoação globalmente ascendente característica da reduplicação verbal enfática. É com essa entoação que as frases (B-b) são agramaticais.

- (73) A: Eu não fui consultada.
 B: Tu foste consultada. Foste consultada. Não digas que não.
- (74) A: O João não vai conseguir.
 B: Tem calma. O João vai conseguir. Vai conseguir.
- (75) A: Eu não fui consultada.
 B: a. Tu foste consultada, foste.
 b. *Tu foste consultada, foste consultada.
 c. *Tu foste consultada, consultada.
- (76) A: O João não vai conseguir.
 B: a. O João vai conseguir, vai.
 b. *O João vai conseguir, vai conseguir.
 c. *O João vai conseguir, conseguir.

4.2.4 Incompatibilidade com a negação

A reduplicação verbal enfática exprime exclusivamente afirmação, sendo incompatível com a negação predicativa. Uma frase negativa com duplicação do verbo é ininterpretável e agramatical (cf. 77-B-a). Uma frase negativa com duplicação da sequência *não*-verbo é igualmente agramatical (ainda que interpretável), desde que produzida com a característica entoação ascendente (cf. 77-B-b). A repetição frásica de negação e verbo não é problemática, como se vê em (77-B-c).

- (77) A: O teu partido ganhou as eleições.
 B: a. *O meu partido não ganhou as eleições, ganhou.
 b. *O meu partido não ganhou as eleições, não ganhou.
 c. O meu partido não ganhou as eleições. Não ganhou. Infelizmente.

4.2.5 Exclusão de pronomes clíticos

Um pronome clítico pode ocorrer nas respostas a interrogativas polares, como em (78-B-b) ainda que seja mais natural a elipse do predicado, que permite a ausência dos complementos verbais, como em (78-B-a).

- (78) A: Devolveste-me o livro que te emprestei?
 B: a. Devolvi.
 b. Devolvi-to.

Mas um pronome clítico não pode ser duplicado juntamente com o verbo nas frases com reduplicação enfática, como mostra o contraste de gramaticalidade entre (79-B-a/b) e (79-B-c/d/e). A repetição frásica não apresenta este tipo de restrição, como se exemplifica em (80).

- (79) A: Não me devolveste o livro que eu te emprestei, pois não?
 B: a. Devolvi-te o livro que me emprestaste, devolvi.
 b. Devolvi, devolvi.
 c. *Devolvi-te o livro que me emprestaste, devolvi-te.
 d. *Devolvi-te, devolvi-te.
 e. *Devolvi-to, devolvi-to.

- (80) A: Não me devolveste o livro que eu te emprestei, pois não?
 B: Devolvi-to. Devolvi-to sim. Tenho a certeza.

4.2.6 As formas morfológicamente complexas resistem à reduplicação

As formas verbais morfológicamente complexas tornam a reduplicação enfática marginal. A resistência à reduplicação é visível com verbos compostos, como *fotocopiar*, ou com verbos derivados que integram prefixos acentuados como, por exemplo, o

prefixo *contra*. Daqui decorrem os contrastes de aceitabilidade entre (81) e (82), por um lado, e (83) e (84) por outro. Enquanto as frases com formas verbais morfologicamente simples (*copiar, atacar*) são perfeitas, as frases com as formas verbais morfologicamente complexas (*fotocopiar, contra-atacar*) são pouco naturais ou marginais.

- (81) A: Ele não copiou o livro sem autorização, pois não?
B: Copiou, copiou.
- (82) A: Ele não fotocopiou o livro sem autorização, pois não?
B: ?Fotocopiou, fotocopiou.
- (83) A: Ele não atacou o candidato, pois não?
B: Atacou, atacou.
- (84) A: O candidato não contra-atacou, pois não?
B: ?Contra-atacou, contra-atacou.

Também as formas verbais de futuro e condicional são morfologicamente complexas (cf. Roberts 1992; Duarte/Matos 2000; Arregi 2000; Roberts/Roussou 2002; 2003; Villalva 2003; Oltra-Massuet/Arregi 2005), o que se relaciona com o facto de serem as únicas formas verbais com as quais ocorre mesóclise, ou seja, a colocação de um pronome clítico no interior da forma verbal, como se vê em (85). Sendo morfologicamente complexas, as formas de futuro e condicional têm em relação à reduplicação enfática o mesmo tipo de efeito que a composição e a derivação com prefixos acentuados. Quer dizer, tornam também a reduplicação marginal, como se exemplifica em (86), que contrasta com a frase totalmente natural em (87).

- | | | |
|---------|-----------------------------|----------------------------------|
| (85) a. | Ele ataca-o se puder. | Presente do Indicativo/Ênclise |
| b. | Ele atacá-lo-á se puder. | Futuro do Indicativo/Mesóclise |
| c. | Ele atacava-o se pudesse. | Imperfeito do Indicativo/Ênclise |
| d. | Ele atacá-lo-ia se pudesse. | Condicional/Mesóclise |
- (86) A: Ele não {atacará/atacaria} o candidato, pois não?
B: ?Atacará, atacará. / ?Atacaria, atacaria.
- (87) A: Ele amanhã não ataca o candidato, pois não?
B: Ataca, ataca.

Como a complexidade morfológica pode ser cumulativa, quando se constrói uma forma de futuro ou condicional a partir de um composto ou de uma forma derivada com prefixo acentuado, o resultado são frases agramaticais como as que se apresentam em (88) e (89).

- (88) A: O candidato não contra-atacará, pois não?
B: *Contra-atacará, contra-atacará.

- (89) A: Ele não fotocopiaria o livro sem a tua autorização, pois não?
 B: *Fotocopiaria, fotocopiaria.

Em síntese, a reduplicação verbal enfática tem a particularidade de ser uma construção que apresenta uma dupla realização do verbo finito numa estrutura monooracional. A segunda realização do verbo ocorre sempre na posição final absoluta e exclui formas verbais não finitas, negação, clíticos e formas morfológicamente complexas. A construção está associada a um padrão prosódico caracterizado por entoação ascendente e ausência de pausa prosódica antes da forma verbal que ocorre no final da frase.

O português brasileiro tem em comum com o português europeu um sistema responsivo em que as respostas verbais são a opção não marcada para responder afirmativamente a interrogativas polares. No entanto, o português brasileiro não permite a reduplicação verbal enfática (como estrutura monooracional), o que indica que a existência de respostas verbais no sistema responsivo de uma língua não é condição suficiente para que seja gramaticalmente possível a construção de reduplicação verbal enfática.

5 Referências

- Arregi, Karlos (2000), *How the Spanish verb works*, comunicação apresentada em: *30th Linguistic Symposium on the Romance Languages*, University of Florida, Gainesville, <http://home.uchicago.edu/~karlos/Arregi-theme.pdf> (07.01.2016).
- Authier, Jean-Marc (2013), *Phase-Edge Features and the Syntax of Polarity Particles*, *Linguistic Inquiry* 44:3, 345–389.
- Costa, João/Martins, Ana Maria/Pratas, Fernanda (2012), *VP Ellipsis: New Evidence from Capeverdean Creole*, in: Irene Franco/Sara Lusini/Andrés Saab (edd.), *Romance Languages and Linguistic Theory 2010: Selected Papers from «Going Romance» Leiden 2010*, Amsterdam/Philadelphia, Benjamins, 155–175.
- Dryer, Matthew S./Haspelmath, Martin (edd.) (2013), *The World Atlas of Language Structures Online*, Leipzig, Max Planck Institute for Evolutionary Anthropology, <http://wals.info> (07.01.2016).
- Duarte, Inês/Matos, Gabriela (2000), *Romance clitics and the minimalist program*, in: João Costa (ed.), *Portuguese Syntax: New Comparative Studies*, Oxford/New York, Oxford University Press, 116–142.
- Farkas, Donka F./Bruce, Kim B. (2010), *On Reacting to Assertions and Polar Questions*, *Journal of Semantics* 27, 81–118.
- von Stechow, Kai/Iatridou, Sabine (2007), *Anatomy of a Modal Construction*, *Linguistic Inquiry* 38, 445–483.
- Hagemeyer, Tjerk/Santos, Ana Lúcia (2004), *Elementos polares na periferia direita*, in: Clara Nunes Correia/Anabela Gonçalves (edd.), *Actas do XIX Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística, 465–476.
- Hale, William Gardner/Buck, Carl Darling (c.1987, 1903), *A Latin Grammar*, Boston/London, Ginn & Company/The Athenaeum Press.
- Holmberg, Anders (2001), *The syntax of Yes and No in Finnish*, *Studia Linguistica* 55, 141–175.

- Holmberg, Anders (2003), *Questions, answers, polarity and head movement in Germanic and Finnish*, in: Anne Dahl/Kristine Bentzen/Peter Svenonius (edd.), *Proceedings of the 19th Scandinavian Conference of Linguistics* (= Nordlyd, University of Tromsø Working Papers on Language and Linguistics 31), 88–115.
- Holmberg, Anders (2007), *Null subjects and polarity focus*, *Studia Linguistica* 61, 212–236.
- Holmberg, Anders (2013), *The syntax of answers to polar questions in English and Swedish*, *Lingua* 128, 31–50.
- Holmberg, Anders (2015), *The Syntax of «Yes» and «No»*, Oxford/New York, Oxford University Press.
- Horn, Laurence R. (1969), *A Presuppositional Analysis of «only» and «even»*, in: Robert I. Binnick et al. (edd.), *Papers from the Fifth Regional Conference Meeting of the Chicago Linguistic Society*, Chicago, IL, Chicago Linguistic Society, 98–107.
- Jones, Bob Morris (1999), *The Welsh Answering System*, Berlin/New York, Mouton de Gruyter.
- Kato, Mary/Tarallo, Fernando (1992), «*Sim*»: *respondendo afirmativamente em português*, in: Mara Sofia Zanotto Paschoal/Maria Antonieta Alba Celani (edd.), *Linguística Aplicada: Da Linguística Aplicada para uma Linguística Transdisciplinar*, São Paulo, EDUC, 259–278.
- Kayne, Richard (1994), *The Antisymmetry of Syntax*, Cambridge MA, MIT Press.
- Laka, Itziar (1990), *Negation in Syntax: On the Nature of Functional Categories and Projections*, tese de Doutorado, Massachusetts Institute of Technology.
- Martins, Ana Maria (1994a), *Enclisis, VP-Ellipsis and the Nature of Sigma*, *Probus* 6, 173–205.
- Martins, Ana Maria (1994b), *Clíticos na História do Português*, tese de Doutorado, Universidade de Lisboa.
- Martins, Ana Maria (2005), *Clitic Placement, VP-ellipsis and scrambling in Romance*, in: Montserrat Batllori et al. (edd.), *Grammaticalization and Parametric Change*, Oxford/New York, Oxford University Press, 175–193.
- Martins, Ana Maria (2006), *Emphatic Affirmation and Polarity: Contrasting European Portuguese with Brazilian Portuguese, Spanish, Catalan and Brazilian*, in: Jenny Doetjes/Paz González (edd.), *Romance Languages and Linguistic Theory 2004: Selected Papers from «Going Romance» Leiden 2004*, Amsterdam/Philadelphia, Benjamins, 197–223.
- Martins, Ana Maria (2007), *Double realization of verbal copies in European Portuguese emphatic affirmation*, in: Norbert Corver/Jairo Nunes (edd.), *The Copy Theory of Movement*, Amsterdam/Philadelphia, Benjamins, 77–118.
- Martins, Ana Maria (2010), *Negação metalinguística («lá», «cá» e «agora»)*, in: Ana Maria Brito et al. (edd.), *Actas do XXIV Encontro Nacional da Associação Portuguesa de Linguística*, Porto, Associação Portuguesa de Linguística, 567–587.
- Martins, Ana Maria (2013), *Emphatic polarity in European Portuguese and beyond*, *Lingua* 128, 95–123, doi: 10.1016/j.lingua.2012.22.002.
- Martins, Ana Maria (2014), *How much syntax is there in Metalinguistic Negation?*, *Natural Language and Linguistic Theory* 32:2, 635–672, doi: 10.1007/s11049–013–9221–9.
- Martins, Ana Maria (no prelo), *VP and TP Ellipsis: sentential polarity and information structure*, in: Susann Fischer/Christoph Gabriel (edd.), *Manual of Grammatical Interfaces in Romance*, Berlin/Boston, De Gruyter.
- Matos, Gabriela (2003), *Construções Elípticas*, in: Maria Helena Mira Mateus et al. (edd.), *Gramática da Língua Portuguesa*, Lisboa, Caminho, 869–913.
- Matos, Gabriela (2013), *Ellipse*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática do Português*, vol. 2, cap. 45, Lisboa, Fundação Calouste Gulbenkian, 2351–2407.
- Nunes, Jairo (2001), *Sideward Movement*, *Linguistic Inquiry* 32, 303–344.
- Nunes, Jairo (2004), *Linearization of Chains and Sideward Movement*, Cambridge MA, MIT Press.
- Oliveira, Marilza de (1996), *Respostas Assertivas e sua Variação nas Línguas Românicas: O seu papel na aquisição*, tese de Doutorado, Campinas, SP, UNICAMP.

- Oltra-Massuet, Isabel/Arregi, Karlos (2005), *Stress-by-structure in Spanish*, Linguistic Inquiry 36, 43–84.
- Pereira, Sílvia (2010), *O Marcador de Negação Metalinguística «Agora» nos Dialeto do Português Europeu*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Pinkster, Harm (1990), *Latin Syntax and Semantics*, London, Routledge.
- Pinto, Clara (2010), *Negação Metalinguística e Estruturas com «Nada» no Português Europeu*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Pope, Emily Norwood (1976), *Questions and Answers in English*, The Hague/Paris, Mouton.
- Roberts, Ian (1992), *A formal account of grammaticalization in the history of Romance futures*, Folia Linguística Historica 13:1–2, 219–258.
- Roberts, Ian/Roussou, Anna (2002), *The history of the future*, in: David Lightfoot (ed.), *Morphological Effects of Syntactic Change*, Oxford/New York, Oxford University Press, 23–56.
- Roberts, Ian/Roussou, Anna (2003), *Syntactic Change: A minimalist approach to grammaticalization*, Cambridge, Cambridge University Press.
- Santos, Ana Lúcia (2002), *Answers to yes/no questions and clitic placement: the question of adverbs*, in: Anabela Gonçalves/Clara Nunes Correia (edd.), *Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, Associação Portuguesa de Linguística, 445–455.
- Santos, Ana Lúcia (2003), *The acquisition of answers to yes/no questions in Portuguese: syntactic, discourse and pragmatic factors*, Journal of Portuguese Linguistics 2:1, 61–91.
- Santos, Ana Lúcia (2009), *Minimal Answers: Ellipsis, Syntax and Discourse in the Acquisition of European Portuguese*, Amsterdam/Philadelphia, Benjamins.
- Thoms, Gary (2016), *Short answers in Scottish Gaelic and their theoretical Implications*, Natural Language and Linguistic Theory 34, 351–391, doi:10.1007/s11049–015–9304-x.
- Vennemann, Theo (2009), *Celtic Influence in English? Yes and no*, English Language and Linguistics 13:2, 309–334.
- Villalva, Alina (2003), *Estrutura Morfológica Básica*, in: Maria Helena Mira Mateus et al. (orgs.), *Gramática da Língua Portuguesa*, Lisboa, Caminho, 917–938.

Rui Marques

23 O modo conjuntivo

Abstract: Apresentam-se neste capítulo as principais linhas exploradas na literatura para explicar a distribuição do modo conjuntivo em português. A ideia tradicional de que este modo está associado à não realidade não explica a sua obrigatoriedade nalguns contextos em que é feita referência à realidade nem a impossibilidade de ocorrer nalguns contextos de irrealidade. O conceito de veridicidade resolve o segundo destes problemas, mas não o primeiro. Uma outra linha assume uma relação entre conjuntivo e não asserção. Dois problemas apontados a esta proposta são o facto de o conjuntivo não ocorrer em frases interrogativas e de os dados não se coadunarem com a noção clássica de asserção. Uma terceira análise relaciona modo com atitudes proposicionais, ocorrendo o conjuntivo nos casos em que a atitude não é conjuntamente epistémica e verídica. Esta hipótese pode ser formalizada numa semântica de mundos possíveis, que permite também descrever o significado dos tempos do conjuntivo.

Keywords: conjuntivo, modo, tempos, veridicidade, atitudes proposicionais

1 Introdução

Em português, os verbos flexionam em função de várias categorias, uma das quais é o modo. Um dos principais modos do português é, juntamente com o indicativo, o conjuntivo, que tem sido objeto de várias análises, quer de pendor sintático quer de pendor semântico ou pragmático. No quadro da sintaxe generativa (cf., e.g., Raposo 1987; Ambar 1992), tem sido explorada a ideia de que uma frase com conjuntivo, ao contrário de uma frase com indicativo, é deficitária na especificação de traços de tempo, o que faz com que tenha de se ligar a uma estrutura sintática superior. Esta assunção de que o conjuntivo é um modo deficitário conjuga-se com a ideia tradicional de que o conjuntivo é, por excelência, o modo da subordinação, ocorrendo sobretudo em orações subordinadas, ideia que também tem reflexo em análises de pendor semântico. Estas têm tentado perceber a que fator semântico é sensível o conjuntivo. Isto é, o conjuntivo será uma marca morfológica que assinala a presença, no contexto sintático em que ocorre o verbo, de um dado valor semântico. Nesta perspetiva, o conjuntivo é uma marca sensível a propriedades do contexto, tal como os itens de polaridade negativa ou de polaridade positiva são expressões sensíveis à polaridade (negativa ou positiva) da frase, que não atribuem à frase em que ocorrem um valor de polaridade, mas são sensíveis à sua presença. Uma outra perspetiva é que o próprio conjuntivo marca um valor semântico.

Neste capítulo, apresentam-se as principais linhas que têm sido exploradas na tentativa de perceber o que caracteriza o modo conjuntivo em português. Na secção 2,

identificam-se sumariamente os contextos sintáticos em que este modo ocorre, após o que, na secção 3, se consideram análises propostas na literatura para dar conta da sua distribuição em português. A principal questão que tem sido debatida a propósito deste modo é relativa à sua distribuição: porque é que o conjuntivo não pode ocorrer nalgumas orações finitas, é opcional noutras e obrigatório noutras ainda. Naturalmente, a resposta a esta questão tem de entrar em linha de conta com outras formas que concorrem com o conjuntivo, em particular o indicativo e o infinitivo. Uma questão menos tratada na literatura é relativa aos tempos do conjuntivo. É este o tema da secção 4, onde se discute a interpretação dos tempos do conjuntivo em português.

2 Distribuição do conjuntivo em português europeu

Em português, o conjuntivo ocorre nos diferentes tipos de frases, subordinadas ou não. Vejamos, separadamente, cada caso.

Em frases não subordinadas, o conjuntivo ocorre em frases imperativas negativas – e.g. *não tropeces!* –, em frases imperativas afirmativas, nos casos em que o modo imperativo é deficitário – e.g. *saiam!* –, em expressões fixas ou semifixas, como *Deus queira, assim seja*, etc., tal como em frases independentes que expressam desejo, como *Desse-me ele presuntos em vez de rezas!* e ainda em frases declarativas sob o escopo de *talvez* ou *oxalá* – e.g. *talvez esteja alguém em casa, oxalá cheguemos a tempo!* O conjuntivo ocorre também em orações coordenadas disjuntivas como (1), coordenadas copulativas com interpretação condicional, como (2), e em orações introduzidas pelo conector *daí que*:

- (1) a. Fosse pelo custo elevado, fosse pela pressão da opinião pública, o certo é que a dimensão da obra foi reduzida.
b. Quer chova quer faça sol, ele faz sempre o caminho a pé.
- (2) a. Tivesse eu dinheiro e quem teria comprado a casa era eu.
- (3) «Os Ghymes procuram, pois, as afinidades entre os homens, não as diferenças que os dividem, daí que os seus espectáculos sejam sempre bem recebidos em todo o mundo.»
(CETEMPúblico, par=ext50344-soc-98a-2)

Em orações subordinadas adverbiais, o conjuntivo é obrigatório nas frases finitas que sejam subordinadas adverbiais concessivas, como (4), condicionais-concessivas, ou incondicionais, como (5), finais, como (6), ou orações introduzidas por *sem*, como (7):

- (4) Embora fosse verão, o tempo estava frio.
- (5) Mesmo que chova, ele vem!

- (6) Ele faltou à reunião para que a sua presença não influenciasse a decisão.
- (7) Ela saiu sem que alguém desse por isso.

Em orações subordinadas adverbiais temporais e em orações condicionais, pode ser usado o conjuntivo ou o indicativo, como se descreve mais detalhadamente abaixo. A oração subordinada adverbial pode também ser infinitiva (ou gerundiva nalguns casos, como em *mesmo chovendo, ele vem*), sem que, na generalidade dos casos, haja uma diferença evidente de significado em relação à correspondente finita. No entanto, esse não é o caso de orações temporais introduzidas por *antes* e *até*, que introduzem uma oração infinitiva ou uma oração finita, cujo verbo flexiona obrigatoriamente no conjuntivo:

- (8) a. A Maria emigrou antes de a guerra começar.
 b. A Maria emigrou antes que a guerra começasse.
- (9) a. Fica aqui até alguém te chamar.
 b. Fica aqui até que alguém te chame.

Como observado por Lobo (2013) ou Marques/Alves (2014), construções com *antes de* expressam apenas uma ordenação temporal das situações, enquanto o significado de construções com *antes que* envolve a consideração de um valor de finalidade. A frase (8b) pode ser parafraseada por *a Maria emigrou para evitar a situação, e suas consequências, de a guerra começar e ela estar no país*. Além disso, (8a), mas não (8b), tem a pressuposição de que a guerra começou. Esta pressuposição da verdade da oração introduzida por *antes de* não é sistemática (cf. Anscombe 1964 ou Heinämäki 1972). Existe em frases como (8a), mas não em exemplos como *ele desistiu da corrida antes de cortar a meta*, em que a oração subordinada é contrafactual, ou *ele saiu da sala antes de alguém ter entrado*, em que a aceitação da verdade da frase complexa não obriga a aceitar nem a verdade nem a falsidade de que alguém entrou. Também em construções com *até se* se pode observar uma diferença discursiva entre os casos em que este operador introduz uma oração infinitiva e aqueles em que introduz uma oração finita: frases como (9a) permitem inferir que o enunciador acredita na verdade da oração subordinada, enquanto em frases como (9b) a oração subordinada descreve uma possibilidade, mas não uma certeza do enunciador.

Em orações temporais introduzidas por outro conector, o conjuntivo está em distribuição complementar com o indicativo, verificando-se que o primeiro modo ocorre apenas em orações que referem situações futuras em relação ao ponto de perspetiva temporal relevante, o tempo de enunciação nos exemplos seguintes:

- (10) a. Aviso-te logo que receba notícias!
 b. Avisei-te logo que recebi notícias!

Quanto às orações condicionais, à primeira vista, há uma associação entre a tripartição condicionais factuais, hipotéticas e contrafactuais¹ e o uso de um tempo do indicativo, do futuro do conjuntivo e do pretérito (imperfeito ou mais-que-perfeito) do conjuntivo, respetivamente:

- (11) a. Se estás cansado (como dizes), devias tirar umas férias. [condicional factual]
 b. Se eu encontrar a Ana, falo-lhe desse assunto. [condicional hipotética]
 c. Se ele tivesse estado atento, saberia o que se passa. [condicional contrafactual]

No entanto, esta associação não é sistemática. O indicativo pode também ocorrer em condicionais hipotéticas e contrafactuais, como mostram os exemplos (12a) e (12b), respetivamente, e o pretérito (imperfeito ou mais-que-perfeito) do conjuntivo não marca necessariamente a contrafactualidade, como mostram exemplos clássicos como (13):

- (12) a. Se há uma inundação nos próximos meses, nada disto se salva!
 b. Se ele perde o comboio, já não conseguia sair da cidade!
- (13) Se ela fosse alérgica à penicilina, coisa que não sabemos se acontece (ou não), teria estes sintomas que apresenta...

Considerando agora orações relativas, a primeira observação é que o conjuntivo pode ocorrer em relativas restritivas (tal como o indicativo), mas não em relativas explicativas (a menos, naturalmente, que seja legitimado por um operador como *talvez* ou *oxalá*, à semelhança do que se verifica em frases independentes declarativas):

- (14) a. Preciso de um hotel que tenha canil.
 b. encontrei um livro, que te {deve / *deva} interessar, bastante raro
 c. Encontrei um livro, que talvez te interesse, bastante raro.

Em orações relativas restritivas, a ideia generalizada é a de que sintagmas nominais com relativas no indicativo identificam entidades concretas que existem de facto, enquanto sintagmas nominais com relativas no conjuntivo têm uma interpretação intensional, referindo entidades que podem não existir na realidade. É o que mostram os seguintes exemplos:

- (15) a. Gostava de comprar uma casa que tem dez quartos.
 b. Gostava de comprar uma casa que tivesse dez quartos.

¹ Esta tripartição diz respeito ao valor de verdade do antecedente da construção condicional (nos exemplos apresentados, o antecedente corresponde à oração subordinada adverbial condicional). As condicionais são factuais se o antecedente for tido como verdadeiro; são contrafactuais se for tido como falso e são hipotéticas se a proposição correspondente ao antecedente não for dada como verdadeira ou falsa.

No entanto, o conjuntivo pode também ocorrer em relativas dentro de sintagmas nominais que identificam entidades cuja existência é assumida pelo enunciador:

- (16) a. Diz-me o nome de um livro que tenha sido escrito por José Saramago (qualquer um serve).
 b. Houve quem se insurgisse contra a proposta.
 c. Eles têm lá muitos pães. Traz-me o que estiver mais bem cozido

Simplificadamente, a diferença entre relativas com conjuntivo e relativas com indicativo é que, ao usar o indicativo, o enunciador identifica entidades concretas da realidade, enquanto que, ao usar o conjuntivo na relativa, o enunciador não se compromete com a capacidade de identificar as entidades identificadas pelo sintagma nominal, mesmo que estas existam na realidade (cf. Marques 2013). Ou seja, um sintagma nominal com conjuntivo na relativa tem um menor grau de referencialidade do que um sintagma nominal com uma relativa no indicativo.

Em orações completivas, a ocorrência do conjuntivo é determinada fundamentalmente pelo verbo, nome ou adjetivo de que a oração é argumento. Alguns predicadores de complementação frásica são regentes de conjuntivo, outros de indicativo e alguns aceitam quer o conjuntivo quer o indicativo no seu argumento relevante.

O conjuntivo é obrigatório nas completivas finitas dependentes de expressões predicativas dos seguintes tipos:

- (i) associadas à expressão de valores de desejo (e.g. *querer, preferir, esperar, rezear, desejo, ansioso*, etc.);
- (ii) associadas à expressão de uma avaliação (e.g. *lamentar, admirar, surpreender, agradecer*, etc.);
- (iii) associadas à expressão de valores de permissão, obrigação e afins (e.g. *mandar, pedir, ordem, proibido*, etc.);
- (iv) associadas à expressão de causa, condição necessária ou condição suficiente (e.g. *fazer com que, levar a que, ser necessário, bastar, impedir*, etc.);
- (v) associadas à expressão da dúvida (e.g. *duvidar, duvidoso, ter dúvidas*);
- (vi) verbos declarativos negativos (*negar, recusar*).

São regentes de indicativo expressões predicativas dos seguintes tipos:

- (i) associadas à expressão de conhecimento (e.g. *saber, descobrir, verificar, ciente, constatação*, etc.);
- (ii) associadas à expressão de um compromisso (e.g. *prometer, juramento*, etc.);
- (iii) verbos declarativos (e.g. *afirmar, dizer, garantir*, etc.);
- (iv) verbos de ficção (e.g. *fingir, sonhar*, etc.).

Em orações completivas dependentes de expressões predicativas que expressam crença pode ocorrer o indicativo ou o conjuntivo, dependendo de o grau de crença que é veiculado ser, respetivamente, forte ou fraco:

- (17) a. Acredito (convictamente) que estava alguém em casa.
 b. Acredito (pouco) que estivesse alguém em casa.

Com as expressões predicativas que expressam (des)crença (e.g. *acreditar*, *duvidar*, *pensar*), bem como com verbos declarativos (e.g. *dizer*, *garantir*, *assegurar*) e verbos compromissivos (e.g. *prometer*, *jurar*), a negação na frase matriz influencia a possibilidade de o conjuntivo ocorrer na oração completiva finita, já que inverte o grau de crença expresso. Assim, verbos como *duvidar*, em cuja oração completiva finita ocorre o conjuntivo no caso de a frase matriz ser afirmativa, admitem o indicativo em frases negativas (cf. *duvido que consiga acabar o trabalho a tempo* vs. *não duvido que consigo acabar o trabalho a tempo*), enquanto verbos como *achar*, *dizer* ou *prometer* têm o comportamento inverso: em frases afirmativas, é usado o indicativo no seu complemento, mas em frases negativas ocorre o conjuntivo (cf., e.g. *garanto que chego a tempo* vs. *não garanto que chegue a tempo*). Ainda assim, o modo que é usado no complemento destes verbos quando ocorrem em frases negativas pode ser o mesmo que em frases afirmativas, o que se pode dever a duas razões: (i) a negação ser metalinguística (cf. Horn 1989), pelo que não influencia a seleção de modo, já que se tem presente a frase afirmativa correspondente; (ii) a opção pelo indicativo em detrimento do conjuntivo ser uma forma de o enunciador indicar que a oração completiva é verdadeira (é o que se observa no contraste entre *o Pedro não acredita que a Maria está doente* e *o Pedro não acredita que a Maria esteja doente*; ambas as frases indicam que o Pedro não acredita que a situação descrita pela oração completiva se verifique, mas a primeira, contrariamente à segunda, indica que, de acordo com o enunciador, essa situação se verifica de facto).

Para além da negação, também o tempo da frase matriz pode influenciar a seleção de modo nas orações completivas de verbos associados a valores de crença. É o caso do verbo *pensar*. Este verbo, quando usado no pretérito perfeito simples, facilmente aceita na sua completiva quer o indicativo quer o conjuntivo, mesmo em frases afirmativas:

- (18) a. (Sempre) pensei que a Maria era espanhola.
 b. (Sempre) pensei que a Maria fosse espanhola.

Note-se que, nestes exemplos, o contraste entre conjuntivo e indicativo não está associado à expressão de diferentes graus de crença. Tanto (18a) como (18b) indicam que o enunciador tinha um forte grau de crença na verdade da proposição completiva. Pode-se pensar que a opção pelo conjuntivo nestes casos indica que no tempo de enunciação o enunciador já não acredita na verdade da proposição. No entanto, nem (18a), com indicativo, nem (18b), com conjuntivo, indicam qual é o grau de crença no tempo de enunciação. Ambas as frases podem ser seguidas por uma indicação de descrença na verdade da proposição (como *mas estava enganado*) ou por uma indicação de manutenção da crença (e.g. *e não me enganei*). Ainda assim, num contexto em que se sabe à partida que a proposição completiva é verdadeira, seria

mais normal enunciar (18a), com indicativo, do que (18b), com conjuntivo. Pelo contrário, num contexto em que se sabe de antemão que a proposição completiva é falsa, tal como num contexto em que não se sabe se essa proposição é verdadeira ou falsa, será adequado enunciar tanto (18a) como (18b).

Descritos sumariamente os contextos de ocorrência do conjuntivo em português europeu, vejamos agora as principais linhas que foram exploradas na literatura para explicar a distribuição deste modo.

3 Análises do modo conjuntivo

Como referido no início deste texto, o modo conjuntivo em português europeu tem sido alvo de análises de pendor semântico ou pragmático e também no quadro da sintaxe generativa. Estas últimas exploraram aprofundadamente a ideia de que o conjuntivo é temporalmente deficitário, do que resulta a necessidade de se ligar ao tempo da frase matriz, obtendo-se uma sequência de concordância de tempos. Esta assunção de que o conjuntivo é temporalmente deficitário será retomada na secção 4. Na presente secção, ter-se-ão em conta análises do conjuntivo de cariz semântico ou pragmático.

3.1 Conjuntivo e valor de verdade

Duas ideias básicas bastante difundidas são as de que o conjuntivo é, por excelência, o modo da subordinação e que a oposição entre indicativo e conjuntivo reflete a oposição entre *realis* e *irrealis*. Ou seja, o indicativo ocorrerá em frases que descrevem a realidade e o conjuntivo em frases que não descrevem factos reais. Uma vez que, por defeito, a interpretação de uma frase é feita tendo por base a realidade, o conjuntivo só poderá ocorrer sob o escopo de um operador que assinale um «deslocamento da realidade». Por esta razão, tipicamente o indicativo é considerado o modo que ocorre por defeito, sendo o conjuntivo o modo marcado. É esta, basicamente, a ideia que se encontra expressa nas gramáticas tradicionais quer do português quer de outras línguas românicas (cf., e.g. Soares Barbosa 1822; Cunha/Cintra 1984; Grevisse ⁵1953; Grevisse/Goosse ¹⁵2011).

No entanto, esta ideia tradicional de que o conjuntivo é o modo do não real ou do não certo, apesar de ser coerente com muitos dos contextos em que este modo ocorre, enfrenta dois problemas: em primeiro lugar, em várias construções o conjuntivo é obrigatório, mas a oração relevante descreve uma situação real. É este o caso de orações completivas de várias expressões predicativas, como *lamentar*, *conseguir*, *fazer com que*, *ter sido suficiente para*, entre outras, e também de orações subordinadas adverbiais concessivas introduzidas por *embora* ou *se bem que* e de orações não subordinadas introduzidas por *daí que*. Em segundo lugar, como observado inicial-

mente por Farkas (1992) para outras línguas românicas, há construções em que o conjuntivo não pode ocorrer, apesar de a situação descrita não ser um facto da realidade. É este o caso de orações completivas de verbos de ficção, como *sonhar*, ou expressões predicativas como *fazer de conta*. Além disso, como visto acima, se bem que em orações subordinadas adverbiais condicionais hipotéticas ou contrafactuais possa ocorrer o modo conjuntivo, o indicativo também pode ocorrer. Em suma, não é verdade que a referência a um plano diferente do da realidade determine o uso do conjuntivo, tal como nem sempre que a oração descreve um facto é excluído o conjuntivo.

De entre os casos problemáticos para a ideia tradicional de que o conjuntivo é o modo da não realidade, o mais debatido é a obrigatoriedade deste modo em orações completivas de verbos como *lamentar* e outras expressões predicativas que expressam igualmente um comentário acerca de um facto, ou uma avaliação (cf. secção 2). Estes verbos são classificados como factivos, termo que é usado com aceções diferentes na literatura, mas que têm em comum a observação de que o complemento de verbos como *lamentar* descreve um facto da realidade. Uma definição técnica de verbo factivo encontra-se em Karttunen (1971). De acordo com esta definição, uma sequência formada por um verbo factivo e a sua oração complemento permite inferir que a proposição completiva é verdadeira tanto quando a frase matriz é afirmativa como quando ela é negativa; isto é, são estruturas pressuposicionais, associadas à pressuposição da verdade da proposição completiva:

- (19) a. Lamento que a Ana tenha perdido as eleições. → A Ana perdeu as eleições.
 b. Não lamento que a Ana tenha perdido as eleições. → A Ana perdeu as eleições.

Esta definição de factividade permite classificar como factivas várias outras expressões predicativas, algumas regentes de conjuntivo (e.g. *surpreender*, *ter pena*, *achar bem/mal...*), outras regentes de indicativo (e.g. *descobrir*, *saber*, *esquecer-se...*).

Para além dos verbos factivos, há outras expressões predicativas que permitem também inferir que a sua proposição complemento é verdadeira, embora apenas em frases afirmativas ou em frases negativas. É o caso, respetivamente, de verbos como *conseguir* ou *levar a*, uma classe que Karttunen (1971) designa de «verbos implicativos», e de verbos como *impedir* ou *evitar*, «verbos implicativos negativos» na designação do mesmo autor. Independentemente de a frase matriz ser afirmativa ou negativa, é o conjuntivo que ocorre na oração completiva finita dos verbos de uma e de outra classe. No que respeita ao valor de verdade da sua proposição complemento, outros verbos e expressões predicativas não permitem fazer qualquer inferência acerca da verdade ou falsidade do seu complemento. É o que se verifica, por exemplo, com verbos como *querer*, *pedir* ou *duvidar*, regentes de conjuntivo, com verbos como *dizer* ou *prometer*, regentes de indicativo, e também com verbos como *pensar* ou *acreditar*, que admitem ambos os modos na sua oração completiva. Em suma, tendo em conta o valor de verdade da proposição relevante, pode dizer-se que em português o conjunti-

vo ocorre nalguns casos em que a proposição é tida como verdadeira (e.g. orações completivas de verbos factivos como *lamentar*), noutros em que a proposição é tida como falsa (e.g. orações completivas de verbos como *impedir* em frases afirmativas) e noutros ainda em que a proposição não é tida nem como verdadeira nem como falsa (e.g. orações completivas de verbos volitivos). O mesmo se pode dizer acerca do modo indicativo. Este é o modo obrigatório em orações completivas de verbos como *saber*, que permitem inferir que a sua proposição complemento é verdadeira, de verbos como *fingir*, que permitem inferir a falsidade da sua proposição complemento, e de verbos como *dizer* ou *prometer*, que não permitem inferir nem a verdade nem a falsidade da sua proposição completiva. Assim, não parece existir em português europeu uma relação entre a (não) assunção da verdade de uma proposição e os modos indicativo ou conjuntivo.

Uma noção relacionada com a verdade da proposição, mas que não se confunde com a questão de a proposição ser, de facto, verdadeira ou falsa na realidade, é a noção de *veridicidade (relativizada)* de Giannakidou (1999) (*veridicidade subjetiva*, em Giannakidou 2013). Simplificadamente, a ideia-base é a de que uma frase não é verdadeira ou falsa sem mais. É-o sempre relativamente a um modelo em relação ao qual a frase é interpretada. Por exemplo, uma frase como *a Ana pensa que o Pedro é espanhol* indica que a proposição *o Pedro ser espanhol* é verdadeira para a Ana, independentemente de o ser ou não na realidade. A interpretação desta oração completiva é feita relativamente a um modelo que corresponde às crenças da Ana. Em orações completivas de verbos como *sonhar*, o modelo relevante é o que corresponde ao do sonho em questão; em orações completivas de verbos como *querer*, é considerado um modelo relacionado com os desejos de uma entidade, etc.

Dada esta noção de veridicidade, um operador é verídico se permitir inferir que a sua proposição complemento é tida como verdadeira no modelo relevante; caso contrário, será um operador não verídico. Posto isto, Giannakidou observa que em grego moderno (tal como se verificará noutras línguas) o conjuntivo ocorre em contextos não verídicos, isto é, sob o escopo de um operador não verídico. No que respeita ao português, pode observar-se que todos os verbos regentes de indicativo são verídicos, mas não se verifica que em todos os contextos de veridicidade é o indicativo o modo que ocorre. Por outras palavras, em português a não veridicidade conduz à seleção do conjuntivo, mas não é condição necessária para a ocorrência deste modo. O conjuntivo não assinala necessariamente não veridicidade.

Em contraste com a ideia tradicional de que o conjuntivo assinala a não realidade, a noção de veridicidade permite uma análise mais precisa dos dados e resolve um dos problemas apontados acima: o facto de o conjuntivo não ocorrer nalgumas orações que não descrevem a realidade, como é o caso de completivas de verbos de ficção, como *sonhar*. Estes verbos são operadores verídicos, já que permitem inferir que a sua proposição complemento é verdadeira no modelo relevante: o cenário fictício que introduzem. No entanto, o outro problema apontado acima à ideia tradicional – o de que o conjuntivo ocorre nalgumas orações que descrevem a

realidade – não é resolvido pela hipótese da veridicidade. De facto, um operador que introduz uma proposição que descreve um facto real é um operador verídico, pelo que, se em português o conjuntivo também assinalasse não veridicidade, este modo não ocorreria nestas orações. No entanto, em vários tipos de orações que descrevem a realidade, o conjuntivo é obrigatório em português. É o caso de orações introduzidas por *embora*, por *daí que*, por verbos avaliativos, como *lamentar*, por verbos implicativos, como *conseguir*, e por verbos implicativos negativos, como *impedir*, que, quando ocorrem sob o escopo da negação, criam um contexto verídico (uma frase como *a chuva não impediu que a prova se realizasse* permite inferir a verdade da oração completiva).

De todos estes casos em que a ocorrência do conjuntivo é problemática para as análises que relacionam o modo com o valor de verdade da proposição, o mais debatido é o das orações completivas de verbos como *lamentar*. Uma proposta de explicação para o facto de, em português e noutras línguas, verbos como este serem regentes de conjuntivo é apresentada de seguida.

3.2 Conjuntivo e não asserção

Vários autores (e.g. Bybee/Terrell 1990; Santos 2000; Panzeri 2003; Mateus et al. 2003), a partir pelo menos de Hooper (1975), que se baseou em dados do espanhol, idênticos aos que se verificam em português, defendem que o uso do conjuntivo ou do indicativo é condicionado, respetivamente, pelo caráter assertivo ou não assertivo do enunciado. Um argumento apresentado em defesa desta hipótese é o facto de a interpretação de alguns verbos ser diferente consoante a sua oração completiva ocorra no indicativo ou no conjuntivo:

- (20) a. «Depois, o vice-rei insistiu que se assinassem novas pazes, mas o embaixador respondeu que só tinha ordem do seu rei para as confirmar oralmente.»
(CETEMPúblico, par=ext118033-des-95a-1)
- b. «O vereador João Silva bem insistiu que a questão era simples.»
(CETEMPúblico, par=ext384064-soc-98a-1)
- (21) a. Ele não disse que saíesses a correr.
b. Ele (não) disse que saíste a correr.

Com o conjuntivo na oração completiva, o verbo matriz tem interpretação deôntica, expressando um ato de fala diretivo, sendo *grosso modo* equivalente a *pedir* ou a *ordenar*, enquanto em (20b) e (21b), com indicativo, a frase matriz refere-se a um ato de fala assertivo.

Esta hipótese explica também porque é que em frases independentes é o indicativo que ocorre em frases declarativas e o conjuntivo em frases imperativas ou associadas à expressão de valores de desejo, já que só as primeiras correspondem a atos de fala assertivos.

Por fim, o facto de verbos como *lamentar* serem regentes de conjuntivo é explicado por esta hipótese como decorrendo do carácter pressuposicional destes verbos. Concretamente, a ideia defendida é a de que só faz sentido asserir uma frase com este verbo num contexto em que o falante assume já se saber que a sua oração completiva é verdadeira. Ou seja, uma frase como *o Governo lamentou que as buscas tenham sido inconclusivas* asseire que o Governo lamentou um facto, mas não que as buscas foram inconclusivas, informação que se assume ser conhecida antes da asserção da frase. Verbos como *lamentar* são, então, classificados como não assertivos, tal como o são outros verbos também regentes de conjuntivo, mas que não estão associados à pressuposição da verdade do seu complemento.

Em suma, esta hipótese defende que o conjuntivo é o modo da não asserção (ou não afirmação, em Santos 2000), sendo que, no que respeita a verbos de complementação frásica, é observada uma distinção entre verbos assertivos e não assertivos.

O principal problema apontado a este tipo de análise relaciona-se com a indefinição do termo «asserção». Como observa Palmer (1986), dizer que alguns verbos são assertivos, pelo que selecionam indicativo, e outros são não assertivos, pelo que selecionam conjuntivo, não é mais do que usar uma etiqueta para classificar verbos, no caso de não se definir o que é asserção. Além disso, como observa o mesmo autor, frases interrogativas constituem enunciados não assertivos, qualquer que seja o conceito de asserção, mas é o indicativo, não o conjuntivo, que ocorre neste tipo de frases.

Em relação à definição do conceito de asserção, a Teoria dos Atos de Fala (Austin 1962; Searle 1969) observa a existência de atos de fala assertivos, que correspondem a enunciados cujo objetivo é o de o enunciador expressar a sua crença na proposição que asseire. Assim, pode considerar-se a hipótese de que o que caracteriza enunciados assertivos, em que, de acordo com a hipótese em análise, ocorre o indicativo, é o facto de o enunciador expressar a crença no enunciado que profere. No entanto, a ideia de que o indicativo ocorre nos casos que indicam crença na proposição e o conjuntivo nos casos em que se expressa crença negativa (como em *não acredito que eles já tenham chegado*) ou um grau fraco de crença (como em *talvez eles já estejam em casa*), hipótese defendida por Palmer (1986) e Bell (1990), entre outros, enfrenta o mesmo problema que as hipóteses analisadas em 3.1: o facto de o conjuntivo ocorrer em orações que descrevem factos tidos como verdadeiros.

Uma outra possibilidade, mais próxima da intuição de que o conjuntivo ocorre em enunciados não assertivos, é recorrer à definição de asserção de Stalnaker (1979), segundo a qual, simplificada, asserir é introduzir informação nova no discurso. Com base nesta conceção de asserção, a hipótese de que o conjuntivo ocorre em enunciados não assertivos corresponderá à constatação de que uma frase com conjuntivo não introduz no discurso informação nova, seja porque a proposição em causa não é dada como verdadeira, seja porque, sendo dada como verdadeira, é informação pressuposta, que já faz parte do contexto discursivo antes da enunciação. No entanto,

por um lado, o indicativo é obrigatório nalguns contextos em que a oração não introduz informação nova no discurso, como no seguinte exemplo:

(22) Já todos sabíamos que a Ana gosta de Londres; a novidade é que se vai casar lá.

Por outro lado, orações em que o conjuntivo é obrigatório podem introduzir informação nova no discurso, como se pode verificar em exemplos como os que se seguem:

- (23) a. Os engenheiros descobriram que a chuva forte dos últimos dias levou a que as fissuras do edifício se tenham agravado.
 b. Vai chover fortemente nos próximos dias. Daí que a proteção civil esteja a pedir à população para se manter em casa.

3.3 Conjuntivo e atitudes proposicionais

Em Marques (1995), é proposto que a distribuição do conjuntivo e do indicativo em português é condicionada pelo tipo de atitude que é expressa para com a proposição. Simplificadamente, o indicativo é selecionado nos casos em que se expressa uma atitude de conhecimento ou crença forte e o conjuntivo nos restantes casos. Assim, o conjuntivo não assinalaria nenhum valor particular. Pelo contrário, é o indicativo que assinala um valor específico, ocorrendo em contextos que são conjuntamente epistémicos (i.e., contextos em que se expressa uma atitude de conhecimento ou crença) e verídicos, sendo o conjuntivo o modo complementar, que ocorre nos casos em que o contexto é não epistémico ou não verídico. Esta hipótese dá conta do facto de o conjuntivo ser obrigatório em contextos não verídicos (e.g. orações completivas de verbos como *querer*, *ordenar*, *impedir*, orações finais, orações introduzidas por *sem*, etc.), já que a veridicidade é condição necessária para a ocorrência do indicativo. Dá conta também do facto de o conjuntivo ocorrer em orações completivas de expressões predicativas avaliativas, como o verbo *lamentar* ou a expressão *achar bem/mal*, e de expressões predicativas causativas, como *levar a*, uma vez que a atitude que estas expressões predicativas expressam não é epistémica. De facto, expressões predicativas como *lamentar* ou *achar bem/mal* expressam primariamente uma atitude de avaliação de um facto, não uma atitude de conhecimento desse facto, ao contrário de, por exemplo, expressões predicativas como *saber* ou *ter conhecimento de*. Também as expressões predicativas causativas não expressam primariamente uma atitude de conhecimento ou crença. Indicam uma relação de causalidade entre os eventos descritos pelos seus argumentos externo e interno, não uma relação de conhecimento ou crença entre uma entidade e a proposição completiva.

Em estruturas de complementação frásica, a atitude proposicional que é expressa corresponde a uma relação entre a entidade identificada pelo sujeito da frase matriz e a proposição completiva. No entanto, em frases negativas com alguns verbos, como

acreditar e outros de sentido idêntico, é possível serem expressas duas atitudes, como se verifica nos seguintes exemplos:

- (24) a. A Ana (ainda) não acredita que o Pedro é licenciado em Física.
 b. Na altura, não acreditei que a Maria era italiana.

Ambas as frases indicam uma atitude de não crença no estado-de-coisas descrito pela proposição completiva, atitude que envolve a Ana, no exemplo (24a), e o enunciador num período anterior ao tempo da enunciação em (24b), mas indicam também uma atitude de crença positiva da parte do enunciador no tempo da enunciação na mesma proposição. Assim, a ocorrência do indicativo nestas orações completivas deve-se à marcação de uma atitude de crença positiva: a proposição encaixada é tida como verdadeira para o enunciador no tempo de enunciação e é expressa para com essa proposição uma atitude de crença. Ou seja, o contexto é verídico e epistémico, o que leva à seleção do indicativo.

Em construções como estas, não é obrigatório o uso do indicativo na oração completiva; o conjuntivo pode igualmente ocorrer:

- (25) a. A Ana (ainda) não acredita que o Pedro seja licenciado em Física.
 b. Na altura, não acreditei que a Maria fosse italiana.

Neste caso, as frases não dão qualquer indicação sobre se, no tempo de enunciação, o enunciador acredita ou não que a proposição completiva é verdadeira. Ou seja, estas frases expressam uma única atitude – de não crença – para com a proposição encaixada. Embora esta atitude seja epistémica, a oração completiva ocorre num contexto não verídico (não se pode inferir que a proposição completiva é verdadeira para alguma entidade), o que leva à seleção do conjuntivo.

A hipótese da relação entre modo e atitudes proposicionais pode ser formalizada no quadro de uma semântica de mundos possíveis, como se descreve de seguida.

Numa semântica de mundos possíveis, uma frase declarativa denota um conjunto de mundos possíveis: o conjunto dos mundos possíveis em que se verifica o estado-de-coisas descrito pela proposição. A interpretação da frase é feita relativamente a um contexto. No caso de frases declarativas, este contexto – o *context set* – pode ser descrito como o conjunto de mundos possíveis compatíveis com o que é assumido no contexto conversacional (cf., e.g., Heim 1992). No caso de orações sob o escopo de um operador, o contexto relativamente ao qual são interpretadas é o que é criado por este operador. Por exemplo, a oração completiva de uma expressão predicativa como *fazer de conta* é interpretada relativamente ao cenário fictício que esta expressão introduz e não tendo em conta o conjunto de mundos possíveis compatíveis com o que é assumido no contexto conversacional.

Posto isto, como é observado em Marques (2009) para as orações completivas, em português, as orações com conjuntivo envolvem na computação do seu significado a consideração de pelo menos um mundo não-p (i.e., um mundo possível em que não se verifica o estado-de-coisas descrito pela proposição), enquanto que a interpretação

de orações completivas com indicativo envolve a consideração apenas de mundos-p. Assim, a essência do conjuntivo em português será a de assinalar um contexto sintático cuja interpretação envolve a consideração de pelo menos um mundo não-p. Vejamos, brevemente, esta análise.

Como observado acima, o conjuntivo está excluído dos casos em que a atitude proposicional expressa para com a proposição relevante é uma atitude de conhecimento ou crença positiva. É esse o caso de exemplos como *a Ana pensa que os tubarões são mamíferos*. Esta frase indica que em todos os mundos possíveis compatíveis com as crenças da Ana os tubarões são mamíferos. Assim, a interpretação do significado da oração completiva envolve a consideração apenas de mundos-p (mundos em que se verifica o estado-de-coisas descrito pela proposição encaixada). Em orações completivas de verbos como *acreditar* pode ocorrer quer o indicativo quer o conjuntivo:

- (26) a. Acredito que ainda há sobreviventes nos escombros.
 b. Acredito que ainda haja sobreviventes nos escombros.

A frase (26a), com indicativo, expressa um grau de crença forte, sendo o seu significado próximo de *tenho a convicção de que ainda há sobreviventes nos escombros*, enquanto (26b), com conjuntivo, expressa um grau de crença mais fraco, sendo a construção parafraseável por *não excluo a hipótese de que ainda haja sobreviventes nos escombros, mas talvez não haja*. O mesmo é dizer que (26a) indica que todos os mundos possíveis compatíveis com as crenças do enunciador são mundos-p e (26b) indica que, no sistema de crenças do enunciador, há mundos-p e mundos não-p (i.e., há pelo menos um mundo possível compatível com as suas crenças em que ainda há sobreviventes nos escombros e pelo menos um mundo possível em que não há sobreviventes nos escombros).

De igual modo, predicados como *duvidar que p* e frases como *talvez p* indicam que no sistema de crenças relevante (o da entidade referida pelo sujeito da frase matriz no primeiro caso e o do enunciador no segundo caso) há mundos-p e mundos não-p. Por outras palavras, estas construções não permitem inferir que *p* é uma proposição verdadeira; o contexto em que *p* ocorre é não verídico. Em suma, um contexto não verídico envolve no seu significado pelo menos um mundo não-p. Como observado acima, em português e noutras línguas, nestes contextos ocorre o conjuntivo, mas não o indicativo.

Para além de contextos não verídicos, o conjuntivo em português ocorre também em orações tidas como verdadeiras na realidade; portanto, em contextos verídicos. É o caso, em orações completivas, de predicados avaliativos, como *lamentar p*, *ter pena de que p* ou *achar bem/mal que p*, por exemplo, e de predicados causativos, como *levar a que p*. Também no que respeita a estes predicados é defensável que o seu significado envolva a consideração de mundos não-p. No que respeita a predicados avaliativos, Heim (1992) propõe uma descrição do seu significado, que aqui se adota, que é próxima da dos predicados volitivos (estes não verídicos):

- (27) *C(ontext) Change Potential for want:*
 $c + a \text{ wants that } \varphi = \{w \in c: \text{for every } w' \in \text{Dox}_\alpha(w):$
 $\text{Sim}_{w'}(\text{Dox}_\alpha(w) + \varphi) <_{\alpha,w} \text{Sim}_{w'}(\text{Dox}_\alpha(w) + \text{not } \varphi)\}$ (Heim 1992, 197)
- (28) *C(ontext) Change Potential for be glad:*
 $c + a \text{ is glad that } \varphi = \{w \in c: \text{for every } w' \in \text{Dox}_\alpha(w):$
 $w' <_{\alpha,w} \text{Sim}_{w'}(\text{rev}\varphi(\text{Dox}_\alpha(w) + \text{not } \varphi)\}$ (Heim 1992, 206)

Informalmente, estas definições indicam que o significado de *querer* é o de que, de entre as possibilidades compatíveis com as crenças do agente da atitude, aquelas em que se verifica o estado-de-coisas descrito pela proposição completiva são preferíveis àquelas em que não se verifica. Naturalmente, esta definição é extensível a outros verbos volitivos. Por exemplo, o significado do verbo *receptar* será idêntico ao de *querer*, com diferença na ordenação dos mundos possíveis (i.e. *a receptar que φ* significa que, para *a*, há a possibilidade de φ e de *não φ* e os mundos *não φ* são preferíveis, para *a*, relativamente aos mundos- φ). Em suma, o significado de predicados volitivos envolve uma comparação entre mundos possíveis (compatíveis com o sistema de crenças do agente da atitude) em que se verifica o estado-de-coisas descrito pela oração completiva e mundos possíveis em que não se verifica. Quanto a predicados avaliativos, a definição em (28) capta a intuição de que também estes predicados envolvem uma comparação de alternativas. Informalmente, esta definição significa que em todos os mundos compatíveis com as crenças de *a* se verifica φ e que esses mundos são preferíveis aos mundos idênticos, mas em que φ não se verifica. Por outras palavras, uma frase como *agrada-me que ele tenha dito isso* significa que é melhor, para o enunciador, ele ter dito isso do que não o ter dito. Assim, a diferença entre predicados volitivos, como *querer*, e predicados avaliativos é a de que estes últimos envolvem um raciocínio contrafactual, contrariamente aos primeiros. Fora esta distinção, o significado de uns e outros envolve uma comparação entre mundos possíveis em que se verifica o estado-de-coisas descrito pela proposição completiva (mundos-p) e mundos possíveis em que não se verifica (mundos não-p). Assim, também o significado de predicados avaliativos envolve a consideração de mundos não-p, o que justificará a ocorrência do conjuntivo na oração completiva destes predicados.

Quanto a predicados causativos, como *levar a que p* ou *fazer com que p* , e também ao verbo *conseguir*, pode igualmente defender-se que o seu significado envolve um raciocínio contrafactual e, por conseguinte, a consideração de mundos não-p, o que explica que estas expressões predicativas sejam regentes de conjuntivo em português. Em relação a expressões causativas, a hipótese de que o seu significado envolve um raciocínio contrafactual encontra suporte na ideia, explorada por exemplo por Salmon (1998), de que indicar que *p* é a causa de *q* significa que, se *p* não tivesse ocorrido, sendo tudo o resto igual, *q* não se verificaria também. Assim, construções como (29a) e (29b) indicarão que, se o mar não estivesse em mau estado, sendo tudo o resto idêntico à realidade, as operações de busca não teriam sido interrompidas:

- (29) a. O mau estado do mar fez com que as operações de busca fossem interrompidas.
 b. O mar estava em mau estado; daí que as operações de busca tivessem sido interrompidas.

Analogamente, (30) indica que, se não fosse o esforço dos sindicatos, sendo tudo o resto idêntico à realidade, o Governo não teria adiado a privatização:

- (30) Os sindicatos conseguiram que o Governo adiasse a privatização.

As orações concessivas introduzidas por *embora* ou *se bem que* são outro contexto verídico em que o conjuntivo ocorre em português. Também para estes casos se pode assumir que a obrigatoriedade do conjuntivo se deve a um raciocínio contrafactual, já que os operadores concessivos expressam negação de expectativas: o significado de uma frase como (31a), abaixo, envolve a proposição implícita de que se esperava que a Ana falasse japonês, dado que vive no Japão. Assim, o significado destas construções também envolve a consideração de mundos não-p, o que poderá explicar a ocorrência do conjuntivo. No entanto, esta explicação, só por si, não é suficiente, dado que em construções adversativas, que expressam igualmente uma negação de expectativas, é o indicativo que ocorre, como mostra (31b):

- (31) a. A Ana não fala japonês, embora viva no Japão.
 b. A Ana vive no Japão, mas não fala japonês.

A razão para que o conjuntivo ocorra em orações concessivas, mas não em orações adversativas, apesar da semelhança de significado entre os dois tipos de construção, carece de maior aprofundamento. Muito provavelmente, há razões de natureza discursiva na base desta distribuição do modo. De acordo com a hipótese que relaciona modo e atitudes proposicionais, a obrigatoriedade do indicativo em orações adversativas dever-se-á à mesma razão pela qual é este o modo de frases declarativas simples, como *está a chover* ou *Lisboa é a capital de Portugal*: ao asserir estas frases, o enunciador expressa a sua crença de que as mesmas são verdadeiras. Além disso, a asserção de uma frase declarativa tem o potencial para atualizar a informação partilhada no contexto conversacional. No modelo de Heim (1992), esta atualização consiste em retirar do *context set* – o conjunto de mundos possíveis compatíveis com o que é assumido pelos participantes na interação discursiva – todos os mundos possíveis em que não se verifica o estado-de-coisas descrito pela proposição asserida. Claramente, é essa também a função de orações adversativas. Já no que respeita a orações concessivas, embora a sua asserção comprometa o enunciador com a verdade das mesmas, pode-se considerar que o seu efeito discursivo é diferente. De algum modo, o objetivo básico da asserção de uma oração concessiva não é o de atualizar o contexto conversacional com a informação expressa pela oração concessiva, mas mais o de relativizar a importância, no discurso, da informação expressa por essa oração. Se assim for, parece razoável que, na interpretação de orações concessivas,

sejam considerados mundos possíveis em que se verifica a expectativa criada pela oração concessiva (i.e., que a Ana fale japonês, no exemplo (31a)), o que justifica a ocorrência do conjuntivo, e em orações adversativas se considerem apenas mundos-p, dado o caráter assertivo destas orações; i.e., a função que têm de retirar do *context set* os mundos possíveis em que não se verifica o estado de coisas descrito pela oração.

Os restantes tipos de oração subordinada adverbial ou oração não subordinada em que o conjuntivo ocorre em português (e.g. orações finais, frases imperativas, etc.) são contextos não verídicos. Ainda assim, como observado na secção 2, em orações condicionais hipotéticas ou contrafactuais, contextos não verídicos, pode ocorrer o conjuntivo ou o indicativo:

- (32) a. Se ele encontrar alguém pelo caminho, apanha um susto!
 b. Se ele chegasse cinco minutos mais tarde, perdia o comboio!
 c. Se ele tivesse vencido aquela corrida, estaria agora milionário!
- (33) a. Se ele encontra alguém pelo caminho, apanha um susto!
 b. Se ele chega cinco minutos mais tarde, perdia o comboio!
 c. Se ele tem vencido aquela corrida, estava agora milionário!

Embora o significado das frases em (32) seja idêntico ao das suas correspondentes em (33), pode-se considerar que a interpretação dos exemplos em (33) é idêntica à de orações completivas de verbos de ficção, sendo a oração subordinada em (33a) parafraseável por *imagina que ele encontra alguém pelo caminho*. Ou seja, ao usar o indicativo, o enunciador dá a instrução para se considerarem apenas mundos-p (no exemplo (33a), mundos em que «ele» encontra alguém pelo caminho). Por contraste, em (32a), o contexto relativamente ao qual a frase é assertiva inclui mundos em que «ele» encontra alguém pelo caminho e mundos em que não encontra. Assim, a hipótese de que o conjuntivo ocorre em contextos que envolvem a consideração de mundos não-p, enquanto o indicativo ocorre em contextos que envolvem apenas mundos-p, permitirá dar conta de dados como (32) e (33).

Finalmente, esta hipótese permite também explicar a possibilidade de se usar o conjuntivo ou o indicativo em exemplos como os seguintes:

- (34) a. Estás a trabalhar?! Pensei que estavas doente!
 b. Estás a trabalhar?! Pensei que estivesses doente!

Ao asserir uma ou outra frase, o enunciador não indica que espera da parte do interlocutor a confirmação de que ele está doente nem a de que não está: quer para (34a) quer para (34b) é adequada uma réplica do interlocutor como *e estou / estive [doente]* ou *mas não estou / estive [doente]*. Ao usar o indicativo, o enunciador descreve a sua crença numa altura anterior ao tempo da enunciação, altura em que tinha a convicção de que o interlocutor estava doente. Ou seja, nessa altura, o seu sistema de crenças continha apenas mundos-p e é esse sistema de crenças que o

enunciador refere ao asserir (34a). Ao usar o conjuntivo, o enunciador dá a indicação que considera também mundos não-p; ou seja, indica que admite que estivesse enganado e que o interlocutor não estivesse de facto doente.

Em síntese, a análise do conjuntivo e do indicativo no quadro de uma semântica de mundos possíveis permite dizer que o que condiciona a ocorrência de um ou de outro modo é a questão de o contexto relativamente ao qual a frase é interpretada envolver a consideração apenas de mundos-p ou de pelo menos um mundo não-p.

4 Os tempos do conjuntivo

Em português, o modo conjuntivo tem seis formas (ou tempos), cujas designações tradicionais são indicadas na Tabela 1, acompanhadas por exemplos relativos à flexão do verbo *ser*:

Tabela 1: Formas do conjuntivo em português.

Pretérito imperfeito	Presente	Futuro imperfeito
<i>fosse</i>	<i>seja</i>	<i>for</i>
Pretérito mais-que-perfeito	Pretérito perfeito	Futuro perfeito
<i>tivesse sido</i>	<i>tenha sido</i>	<i>tiver sido</i>

Se para os tempos compostos for assumido o tratamento proposto por Peres (1993) para as formas do indicativo, baseado na análise de Kamp/Reyle (1993) para o inglês, o conjuntivo terá três tempos: pretérito, presente e futuro.

4.1 Conjuntivo e dependência temporal

Uma proposta bastante divulgada, sobretudo em estudos de sintaxe generativa (cf., e.g., Picallo 1984; Ambar 1992), explora a ideia de que os morfemas de conjuntivo são tempos dependentes, ligados pelo tempo da frase matriz. A favor desta proposta encontram-se dados como (35), onde tem de existir concordância de tempos entre a frase matriz e a subordinada, por contraste com estruturas de complementação frásica com indicativo:

- (35) a. Ele quer que o filho {seja / *fosse} médico.
 b. Ele quis que o filho {*seja / fosse} médico.
- (36) a. Lembro-me de que ele {é / era} médico.
 b. Lembrei-me de que ele {é / era} médico.

Estes dados podem ser explicados pela hipótese, explorada em vários trabalhos de sintaxe generativa, de que o conjuntivo é um tempo anafórico, ligado pelo tempo da frase matriz, enquanto o indicativo é um tempo independente. Embora atraente, a hipótese de que o conjuntivo é um tempo anafórico é questionável.

Em primeiro lugar, o conjuntivo pode ocorrer em orações independentes e, pelo menos nalguns casos, como em (37b), não parece defensável que o conjuntivo seja ligado por algum operador temporal presente na estrutura sintática ou no discurso:

- (37) a. Talvez ele {esteja / estivesse} doente.
 b. {Seja / fosse} ele capaz de acabar o curso!

Em segundo lugar, noutros tipos de orações subordinadas com conjuntivo, que não orações completivas, pode não existir concordância de tempos entre frase matriz e frase subordinada (cf., e.g., Vogel 1997). Vejam-se os seguintes exemplos, com orações subordinadas concessivas:

- (38) a. Naquela altura, defendi a proposta, embora agora me arrependa.
 b. Agora acho que tens razão, embora antes eu pensasse de outra forma.

Finalmente, mesmo em orações completivas com conjuntivo, nem sempre existe concordância de tempos entre a oração subordinada e a frase matriz. Um exemplo claro deste facto é dado por orações completivas de verbos como *pedir*. Como observado, por exemplo por Oliveira (2003), se a frase matriz identificar uma situação passada, na oração completiva de *pedir* pode ocorrer quer o pretérito imperfeito do conjuntivo quer o presente do conjuntivo. O mesmo se observa com outros verbos regentes de conjuntivo:

- (39) a. Ele pediu que lhe telefonasses / telefonas.
 b. O general {ordenou / exigiu / permitiu} que o portão {estivesse / esteja} sempre aberto.

Exemplos como os que se seguem ilustram a possibilidade de o verbo da frase matriz flexionar no presente e o da oração subordinada, no conjuntivo, no pretérito:

- (40) a. [Não vi se as luzes estavam apagadas.] Espero que estivessem.
 b. Lamento que ele estivesse tão deprimido.

Em suma, nalgumas estruturas de complementação verbal com conjuntivo, verifica-se que tem de existir concordância entre o tempo da frase matriz e o da encaixada, mas noutras não. Além disso, o mesmo se pode dizer acerca de estruturas de complementação verbal com o indicativo:

- (41) a. Sei que ele {morava / mora} em Coimbra.
 b. Disse que ela {estava / está} grávida.
 (42) a. Achei que ele {morava / *mora} em Coimbra.
 b. Supus que ela {estava / *está} grávida.

Assim, a questão de existir ou não concordância de tempos em estruturas de complementação frásica não depende do modo da oração subordinada, mas é uma questão lexical: há verbos de complementação frásica que impõem restrições de sequências de tempos e outros que não o fazem, sendo que num e noutra tipo se encontram verbos regentes de conjuntivo e verbos regentes de indicativo (cf. Gonçalves et al. 2013 ou Marques 2014 para uma descrição e análise mais detalhadas).

4.2 Conjuntivo e referência temporal

As designações das formas simples do conjuntivo – *pretérito (imperfeito)*, *presente e futuro* – sugerem que estas indicam que o intervalo de tempo em que se localiza o estado-de-coisas descrito pelas frases em que ocorrem é anterior, sobreposto ou posterior ao tempo da enunciação, respetivamente, assunção que é consistente com dados como os seguintes:

- (43) a. Ontem, talvez ele estivesse em casa (, mas hoje não está).
 b. Talvez ele esteja em casa (agora).
 c. Se amanhã ele estiver em casa, vou visitá-lo.

No entanto, em diversas construções pode ocorrer o presente ou o pretérito (imperfeito) do conjuntivo sem que existam necessariamente diferenças de interpretação temporal:

- (44) a. Seja ele capaz de acabar o curso!
 b. Fosse ele capaz de acabar o curso!
- (45) a. Com esta segurança, duvido que alguém consiga assaltar o edifício.
 b. Com esta segurança, duvido que alguém conseguisse assaltar o edifício.
- (46) a. Caso tenhas sede, podes beber água deste chafariz.
 b. Caso tivesses sede, podias beber água deste chafariz.
- (47) a. Com a distância que ela já leva, mesmo que grites, ela não te ouve.
 b. Com a distância que ela já leva, mesmo que gritasses, ela não te ouvia.
- (48) a. Com o frio que está, logo que saias, constipas-te!
 b. Com o frio que está, logo que saísse, constipavas-te!
- (49) a. Com esta tempestade, um barco que esteja no mar está em apuros.
 b. Com esta tempestade, um barco que estivesse no mar estaria em apuros.

Se o ponto de perspetiva temporal (cf. Kamp/Reyle 1993) for anterior ao tempo de enunciação, podem ser usadas as formas no pretérito imperfeito do conjuntivo, mas não as do presente do conjuntivo. No entanto, se o ponto de perspetiva temporal for o tempo de enunciação, podem ser usadas quer as formas com o presente do conjuntivo

vo, quer as formas com o pretérito imperfeito do conjuntivo. Neste caso, a diferença entre o primeiro e o segundo membro de cada par de frases prende-se com a consideração de hipóteses mais realistas ou menos realistas. Os casos com presente do conjuntivo apontam para hipóteses que estão disponíveis no contexto de enunciação da frase, enquanto os casos com pretérito do conjuntivo ou têm uma interpretação contrafactual ou apontam para hipóteses mais distantes do que é esperado no contexto conversacional. Por exemplo (44b) pode ser asserida com felicidade num contexto em que o enunciador duvida da capacidade de «ele» acabar o curso que está a frequentar, contrariamente a (44a), que só pode ser asserida com felicidade num contexto em que a capacidade de «ele» acabar o curso que frequenta não está em causa. Por outras palavras, se o ponto de perspetiva temporal for o tempo de enunciação, o pretérito imperfeito do conjuntivo remete para possibilidades menos realistas no contexto de enunciação.

Quanto ao futuro do conjuntivo, expressa a mesma informação temporal que o presente do conjuntivo. Nos exemplos (50) e (51), a relação temporal associada às formas de conjuntivo é de sobreposição ao tempo de enunciação e em (52) e (53) é de posterioridade ao tempo de enunciação:

- (50) a. Se souberes a resposta, é melhor dizeres já.
 b. Caso saibas a resposta, é melhor dizeres já.
- (51) a. Eles querem fotografar os navios que estiverem afundados no porto.
 b. Eles querem fotografar um navio que esteja afundado no porto.
- (52) a. Entrevista as pessoas que encontrares no caminho.
 b. Entrevista uma pessoa que encontre no caminho.
- (53) a. Telefona-me quando chegares.
 b. Telefona-me logo que chegues.

Como estes dados evidenciam, quer o presente do conjuntivo quer o futuro do conjuntivo podem ter o tempo de enunciação como ponto de perspetiva temporal. Em orações temporais, ambos os tempos expressam posterioridade em relação ao tempo de enunciação, enquanto em orações condicionais ou relativas expressam sobreposição se o predicado for estativo e posterioridade se for não estativo. Assim, não parece verificar-se que entre o presente e o futuro do conjuntivo se estabeleça uma oposição de natureza temporal. Ambas as formas podem ter o tempo de enunciação como ponto de perspetiva temporal (o ponto de perspetiva temporal do presente e do futuro do conjuntivo pode também ser posterior ao tempo de enunciação, mas não anterior) e a informação respeitante a localização relativa varia consoante o tipo de predicado e o tipo de construção, mas é independente dos tempos do conjuntivo.

Em síntese, no que respeita à informação temporal associada aos tempos do conjuntivo, verifica-se que:

- (i) todos os tempos podem ocorrer em contextos em que o ponto de perspectiva temporal é o tempo da enunciação;
- (ii) o pretérito imperfeito do conjuntivo pode ter como ponto de perspectiva temporal um tempo do passado (i.e., um ponto da linha do tempo anterior ao tempo de enunciação), contrariamente ao presente do conjuntivo e ao futuro do conjuntivo;
- (iii) com exceção das orações temporais e das completivas de verbos que impõem restrições temporais aos seus complementos (como *pedir*), a relação entre o estado-de-coisas descrito pela oração com conjuntivo e o ponto de perspectiva temporal é de sobreposição ou de posterioridade, consoante o predicado seja estativo ou não, respetivamente;
- (iv) em orações temporais, a leitura temporal é de posterioridade, independentemente de o predicado ser ou não ser estativo.

Assim, considerando apenas a informação temporal associada aos tempos do conjuntivo, o sistema parece bastante desordenado. No entanto, há a ter em conta que todos estes tempos são formas do mesmo modo e, portanto, estão associadas ao significado do conjuntivo.

Como foi proposto em 3.3, o conjuntivo caracteriza-se por envolver referência a mundos não-p. Ou seja, uma frase com conjuntivo requer que se considerem mundos possíveis alternativos ao mundo de referência. Naturalmente, a informação temporal é comum aos diferentes mundos possíveis considerados. Posto isto, a oposição entre o pretérito do conjuntivo, por um lado, e o presente e o futuro do conjuntivo, por outro, é bastante simples (cf. Marques 2010). Estas duas últimas formas apontam para possibilidades em aberto no contexto de enunciação, pelo que são formas dêiticas, enquanto o pretérito imperfeito do conjuntivo aponta para outras possibilidades: possibilidades acessíveis a partir de um ponto anterior ao tempo de enunciação (em casos como, e.g. *há dois anos atrás, eu duvidava que ele fosse capaz*) ou possibilidades que não estão em aberto no contexto de enunciação (em casos como, e.g. *se eu fosse rico, comprava uma quinta*).

Quanto à oposição entre o presente e o futuro do conjuntivo, como mostram os dados (50)–(53), acima, estas formas não se distinguem entre si por diferenças de natureza temporal. Além disso, estão em distribuição complementar: o futuro do conjuntivo ocorre em orações relativas com quantificador universal (cf. (54a)), em orações condicionais introduzidas pela conjunção *se* (cf. (55a)) e em orações temporais introduzidas por *quando*, *enquanto* ou *assim que* (cf. (56a)). O presente do conjuntivo está excluído desses contextos (embora pelo menos em relativas com quantificador universal haja alguma aceitabilidade), mas pode ocorrer noutros tipos de orações relativas, condicionais e temporais:

- (54) a. Todas as pessoas que {estiverem / ?estejam} inscritas podem concorrer.
- b. Uma pessoa que {esteja / *estiver} inscrita pode concorrer.

- (55) a. Se eu {encontrar / *encontre} a Ana, falo-lhe do assunto.
 b. Caso {concordem / *concordarem}, a reunião fica adiada para a próxima semana.
- (56) a. Enquanto {estiveres / *estejas} com gripe, não saias à rua!
 b. Logo que {possas / */?puderess}, telefona-me!

Se, seguindo Kratzer (2013) e von Stechow (1994), entre outros, for assumido que orações introduzidas por operadores temporais como *quando*² ou pela conjunção *se* envolvem quantificação universal, torna-se evidente que o futuro do conjuntivo só pode ocorrer em orações que correspondam a restritores de quantificadores universais. Assim, o futuro do conjuntivo parece ocorrer nas orações que denotam um conjunto de mundos possíveis (mundos possíveis em que se verifica o estado-de-coisas descrito pela proposição) que é apresentado como um subconjunto de outro conjunto de mundos possíveis (o que é denotado pela frase matriz). Por exemplo, (55a) indica que todos os mundos possíveis em que o enunciador encontra a Ana são mundos possíveis em que lhe fala do assunto em questão.

Finalmente, o facto de nas orações temporais com conjuntivo existir sempre leitura de futuridade, enquanto nas outras orações a leitura de posterioridade só se verifica com predicados não estativos, pode ser explicada como resultando da conjugação de dois dados, para os quais existe evidência independente (cf. Marques 2010):

- (i) a interpretação de orações temporais é feita tendo em conta apenas o mundo real (i.e., conectores temporais requerem uma base modal histórica, como definida por Werner 2006), enquanto noutras construções com conjuntivo se consideram mundos possíveis alternativos à realidade;
- (ii) o conjuntivo exige que se considerem mundos possíveis alternativos (i.e., mundos não-p).

A única forma de satisfazer ambas estas condições, aparentemente contraditórias, na interpretação de uma oração temporal, é considerar desenvolvimentos do mundo real posteriores ao tempo de enunciação, já que o futuro é, pela sua natureza, incerto (i.e., nalguns desenvolvimentos futuros do mundo real verifica-se o estado-de-coisas descrito pela oração e noutros não se verifica).

2 O mesmo se pode dizer, naturalmente, a respeito de outros operadores temporais, como *enquanto*. Em orações como *quando cheguei*, *a reunião já tinha começado*, pode-se assumir que há quantificação universal sobre um conjunto singular (i.e., *quando* é equivalente, *grosso modo*, a *no tempo em que*). Também em orações introduzidas por *enquanto* e *assim que* se pode assumir que existe quantificação universal, sobre intervalos de tempo.

5 Conclusão

Pretendeu-se neste capítulo apresentar as principais linhas de natureza semântica e pragmática que foram exploradas na literatura para a análise do conjuntivo em português. Duas assunções clássicas são as de que existe uma relação entre conjuntivo e não realidade e a de que este é o modo da não asserção. A análise do conjuntivo com base numa semântica de mundos possíveis permite explicar a distribuição deste modo sem os problemas apontados a análises tradicionais e, simultaneamente, captar estas duas intuições clássicas. Permite também descrever de uma forma simples o sistema dos tempos do conjuntivo. Por último, refira-se que a noção de mundo possível é clássica na análise dos verbos modais (e outros operadores de modalidade) e uma análise do conjuntivo nestas linhas permite tratar conjuntamente modo e modalidade.

6 Referências

- Ambar, Manuela (1992), *Temps et structure de la phrase en portugais*, in: Hans Obehauer/Anne Zribi-Hertz (edd.), *Structure de la phrase et Théorie du Liage*, Saint-Denis, Presses Universitaires de Vincennes, 29–49.
- Ancombe, Elizabeth (1964), *Before and After*, *The Philosophical Review* 73, 3–24.
- Austin, John L. (1980, ¹⁹⁶²), *How to Do Things with Words*, Oxford, Oxford University Press.
- Bell, Anthony (1990), *El modo en español: consideración de algunas propuestas recientes*, in: Ignacio Bosque (ed.), *Indicativo y subjuntivo*, Madrid, Taurus Universitaria, 81–106.
- Bybee, Joan/Terrell, Tracy David (1990), *Análisis semántico del modo en español*, in: Ignacio Bosque (ed.), *Indicativo y subjuntivo*, Madrid, Taurus Universitaria, 145–163.
- CETEMPúblico – *Corpus de Extratos de Textos Eletrónicos MCT/Público*, <http://www.linguateca.pt/CETEMPUBLICO/> (21.10.2015).
- Cunha, Celso/Cintra, Luis F. Lindley (1984, ¹⁹⁹²), *Nova Gramática do Português Contemporâneo*, Lisboa, Edições Sá da Costa.
- Farkas, Donka (1992), *On the semantics of subjunctive complements*, in: Paul Hirschbühler/Konrad Koerner (edd.), *Romance Languages and Modern Linguistic Theory*, Amsterdam/Philadelphia, Benjamins, 71–104.
- von Fintel, Kai (1994), *Restrictions on Quantifier Domains*, tese de Doutorado, Amherst MA, University of Massachusetts.
- Giannakidou, Anastasia (1999), *Affective dependencies*, *Linguistics & Philosophy* 22:4, 367–421.
- Giannakidou, Anastasia (2013), *The subjunctive as evaluation and nonveridicality*, in: Joanna Blaszack et al. (edd.), *Tense, Mood, and Modality: New Perspectives on Old Questions*, <http://home.uchicago.edu/~giannaki/papers.html> (30.09.2015).
- Gonçalves, Anabela, et al. (2013), *Sequências de tempos em completivas finitas: restrições semânticas e efeitos na aquisição*, in: *XXVIII Encontro Nacional da Associação Portuguesa de Linguística, Textos Seleccionados*, Coimbra, APL, 433–452.
- Grevisse, Maurice (⁵1953, ⁹1969), *Le Bon Usage. Grammaire française*, Paris, Duculot.
- Grevisse, Maurice/Goosse, André (¹⁵2011), *Le Bon Usage*, Paris/Louvain-la-Neuve, Duculot/De Boeck.
- Heim, Irene (1992), *Presupposition Projection and the Semantics of Attitude Verbs*, *Journal of Semantics* 9:3, 183–221.

- Heinämäki, Orvokki (1972), *Before*, in: Paul Peranteau/Judith N. Levi/Gloria C. Phares (edd.), *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society*, Chicago, CLS, 139–151.
- Hooper, Joan B. (1975), *On assertive predicates*, in: P. Kimball (ed.), *Syntax and Semantics 4*, New York, Academic Press, 91–124.
- Horn, Laurence R. (1989), *A Natural History of Negation*, Chicago, Chicago University Press.
- Kamp, Hans/Reyle, Uwe (1993), *From Discourse to Logic. Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht, Kluwer.
- Karttunen, Lauri (1971), *The Logic of English Predicate Complement Constructions*, Bloomington, Indiana University Linguistics Club.
- Kratzer, Angelika (1989), *An Investigation of the Lumps of Thought*, *Linguistics and Philosophy* 12:5, 607–653.
- Kratzer, Angelika (2013), *Modals and Conditionals*, Oxford, Oxford University Press.
- Lobo, Maria (2013), *Subordinação adverbial*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática da Língua Portuguesa*, vol. 2, cap. 38, Lisboa, Fundação Calouste Gulbenkian, 1981–2057.
- Marques, Rui (1995), *Sobre os valores dos modos conjuntivo e indicativo em português*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Marques, Rui (2009), *On the selection of mood in complement clauses*, in: Lotte Hogeweg/Helen de Hoop/Andrej Malchukov (orgs.), *Cross-linguistic Semantics of Tense, Aspect and Modality*, Amsterdam/Philadelphia, Benjamins, 179–204.
- Marques, Rui (2010), *Sobre a semântica dos tempos do conjuntivo*, in: *XXV Encontro da Associação Portuguesa de Linguística, Textos Seleccionados*, Porto, APL, 549–565.
- Marques, Rui (2013), *Modo*, in: Eduardo B. Paiva Raposo et al. (orgs.), *Gramática da Língua Portuguesa*, vol. 1, cap. 19, Lisboa, Fundação Calouste Gulbenkian, 671–693.
- Marques, Rui (2014), *Référence au réel et concordance des temps en portugais européen*, *Travaux linguistiques du Cerlico* 27, 31–46, <http://www.pur-editions.fr/detail.php?idOuv=3659> (06.01.2016).
- Marques, Rui/Alves, Ana Teresa (2014), *Sobre as variações de modo nas frases subordinadas temporais com «antes»*, in: *XXIX Encontro da Associação Portuguesa de Linguística, Textos Seleccionados*, Porto, APL, 333–343.
- Mateus, Maria Helena Mira, et al. (orgs.) (2003), *Gramática da Língua Portuguesa*, 5.ª ed. revista e aumentada, Lisboa, Caminho.
- Oliveira, Fátima (2003), *Modalidade e modo*, in: Maria Helena Mira Mateus et al. (orgs.), *Gramática da Língua Portuguesa*, 5.ª ed. revista e aumentada, Lisboa, Caminho, 243–272.
- Palmer, F. R. (1986), *Mood and Modality*, Cambridge, Cambridge University Press.
- Panzeri, Francesca (2003), *In the (indicative or subjunctive) mood*, in: Matthias Weisgerber (ed.), *Proceedings of the Conference «SuB7 – Sinn und Bedeutung»*, Arbeitspapier 114, Konstanz, Universität Konstanz, 215–227, <http://ling.uni-konstanz.de/pages/conferences/sub7/> (06.01.2016).
- Peres, João de Andrade (1993), *Towards an Integrated View of the Expression of Time in Portuguese*, Lisboa, Faculdade de Letras da Universidade de Lisboa.
- Picallo, Carme (1984), *The Infl node and the Null Subject Parameter*, *Linguistic Inquiry* 15, 75–102.
- Raposo, Eduardo Paiva (1987), *Case Theory and Infl-to-Comp: The inflected infinitive in European Portuguese*, *Linguistic Inquiry* 18:1, 85–109.
- Salmon, Wesley C. (1998), *Causality and Explanation*, Oxford, Oxford University Press.
- Santos, Maria Joana de Almeida Vieira dos (2000), *Os usos do conjuntivo em língua portuguesa: uma proposta de análise sintáctica e semântico-pragmática*, tese de Doutoramento, Universidade de Coimbra [publicado em 2003, Lisboa, Fundação Calouste Gulbenkian/Fundação para a Ciência e a Tecnologia].

- Searle, John R. (1969), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge, Cambridge University Press.
- Soares Barbosa, Jerónimo (1822, 1881), *Grammatica Philosophica da Lingua Portugueza ou Principios da Grammatica Geral Aplicados á Nossa Linguagem*, Lisboa, Academia Real das Sciencias.
- Stalnaker, Robert (1979), *Assertion*, in: Peter Cole, *Syntax and Semantics*, vol. 9, New York, Academic Press, 315–332.
- Vogel, Rozemarijn (1997), *Aspects of Tense*, tese de Doutoramento, The Hague, Holland Academic Graphics.
- Werner, Thomas A. (2006), *Future and non-future modal sentences*, *Natural Language Semantics* 14:3, 235–255.

João Veloso

24 O sistema vocálico e a redução e neutralização das vogais átonas

Abstract: Neste capítulo, propõe-se que o sistema vocálico do português inclua lexicalmente quatro graus de abertura e duas vogais centrais não baixas, embora se admita que nem todas as realizações fonéticas das vogais resultantes de tal proposta estejam obrigatoriamente representadas no léxico como tais. É ainda proposta uma descrição do processo de redução átona inspirada no modelo da fonologia dos elementos que assume que a redução átona do português resulta da sobreposição de pelo menos dois mecanismos fonológicos contraditórios entre si, eventualmente com origens históricas diferentes: reforço dos traços de labialidade vs. perda total dos traços de palatalidade. Finalmente, é ainda adiantada a hipótese de a redução do vocalismo átono do português estar, no estágio contemporâneo da história da língua, a sofrer um processo de perda de produtividade, deixando aparentemente de se aplicar a um grande número de neologismos admitidos em português.

Keywords: vocalismo, vogal não marcada, redução átona, nasalização, ditongação

1 Introdução e questões de partida

1.1 Considerações introdutórias

As descrições fonológicas do português costumam dividir o sistema vocálico desta língua em dois subsistemas principais: um sistema tónico, no qual são foneticamente encontradas todas as vogais de (1) exceto a central alta [i],¹ e um sistema átono, como

1 [e] ocorre em sílaba tónica somente nas palavras funcionais 'p[e]ra e 'c[e]da. Foneticamente, e em alguns dialetos, [e] acentuado pode ainda ser encontrado antes de consoante nasal ('b[e]nho, 'c[e]mara) ou palatal (es'p[e]lho, 'l[e]nha), o que é habitualmente interpretado como o resultado da aplicação de regras contextuais a um /á/ ou /é/ subjacentes (cf., entre outros, Mateus/Andrade 2000, 19).

Agradecimentos: Agradeço à Prof^a Doutora Ana Maria Martins a leitura atenta e crítica de uma primeira versão deste texto, bem como as inúmeras sugestões de reformulação que me fez chegar e que, a meu ver, o melhoraram substancialmente. Não obstante, todos os erros de análise e de interpretação são, obviamente, da minha única e inteira responsabilidade.

o exemplificado em (2) (cf., entre outros, J. M. Barbosa ²1983, 51ss.; 1994, 167ss.; Mateus/Andrade 2000, 17, 19; Mateus et al. 2003, 991–992).²

- (1) Inventário completo das vogais tónicas e átonas (pré- e pós-acentuadas, finais e não finais) foneticamente atestadas em português europeu

	Anteriores	Centrais	Recuadas
Altas	[i]	[i]	[u]
(Fechadas)	<i>milha</i> : ['míɫɐ]	<i>que</i> : [ki]	<i>muro</i> : ['muru]
Médias-altas	[e]		[o]
(Semifechadas)	<i>saber</i> : [sɐ'ber]		<i>povo</i> : ['povu]
Médias-baixas	[ɛ]	[ɛ]	[ɔ]
(Semiabertas)	<i>terra</i> : ['tɛrɐ]	<i>cada</i> : ['kɛdɐ]	<i>sorte</i> : ['sɔrtɨ]
Baixas		[a]	
(Abertas)		<i>fado</i> : ['fadu]	

(Exemplos: Palavras em que a vogal em causa ocorre em posição tónica, na variedade-padrão, com exceção de [i], não atestada em posição tónica.)

O sistema átono é ainda dividido, nos mesmos autores acima referidos, em três subconjuntos mais restritos: (i) vocalismo átono pré-acentuado; (ii) vocalismo átono pós-acentuado não final; (iii) vocalismo átono pós-acentuado final.

Esta divisão, seguindo a proposta de J. M. Barbosa (²1983, 135ss., 105ss.), é a que se apresenta em (2).

- (2) Divisão do vocalismo átono do português europeu contemporâneo em três subsistemas: vocalismo átono pré-acentuado; vocalismo átono pós-acentuado não final; vocalismo átono pós-acentuado final (*apud* J. M. Barbosa ²1983)

² No quadro de (1), assumimos a tradicional divisão das vogais em três categorias quanto à posição longitudinal do corpo da língua: anteriores, centrais e recuadas.

Ao longo de todo o capítulo, as generalizações referentes ao português terão por base, fundamentalmente, dados e premissas relativos à norma padrão da variedade europeia da língua.

Vocalismo átono pré-acentuado	Vocalismo átono pós-acentuado não final	Vocalismo átono pós-acentuado final
[i]: <i>f[i]l'trar</i>	[u]: <i>'mác[u]la</i>	Posição final absoluta:
[e]: <i>[e]uro'peu</i>	[e]: <i>fa'zĩ[e]mos</i>	[u]: <i>'magr[u]</i>
[ɛ]: <i>[ɛ]colo'gista</i>	[i]: <i>dí'ssés[s[i]mos</i>	[e]: <i>'fac[e]</i>
[a]: <i>infl[a]'ção</i>	[i]: <i>'últ[i]mo</i>	[i]: <i>'príncip[i]</i>
[e]: <i>[e]rr[e]s[e]'dor</i>		[i]: <i>'táx[i]</i>
[ɔ]: <i>ec[ɔ]no'mia</i>		Sílaba final travada por consoante soante:
[o]: <i>aç[o]i'tar</i>		[i]: <i>'fác[i]l</i>
[u]: <i>[u]tili'zar</i>		[a]: <i>a'çúc[a]r</i> (*)
[i]: <i>r[i]c[i]'ber</i>		[ɛ]: <i>re'pórt[ɛ]r</i>
		[ɔ]: <i>'plânct[ɔ]n</i>
		[u]: <i>'côns[u]l</i>

(*) – Exemplo de J. M. Barbosa (²1983, 118) (coexiste com *a'çúc[e]r*).

Um quadro como o apresentado em (2) parte de uma perspetiva «maximalista» sobre o(s) inventário(s) vocálico(s) do português. Esta enumeração corresponderá, assim, a um *sobreinventário*, do qual outras descrições fonológicas da língua geralmente excluem determinados itens, com base em argumentos de natureza eminentemente fonológica.

Com efeito, um exame mais criterioso aos dados de (2), como o que pretendemos desenvolver neste capítulo, poderia conduzir-nos a inventários mais simplificados e mais económicos, como os encontrados em Mateus/Andrade (2000, 19) e Mateus et al. (2003, 991–992), entre outros, que reduzem as vogais átonas do português, nestes mesmos três contextos acentuais, ao subconjunto formado por [e i u], em posição final, e a [e i u i], em posição não final.³

Convém não ignorar, efetivamente, que a atestação das seguintes vogais em posição não acentuada, embora não contrarie a informação transposta para (2), é, no entanto, passível de explicações particulares. Trata-se de realizações de certa forma excepcionais que devem ser obrigatoriamente interpretadas como o resultado da aplicação de processos fonológicos muito particulares (atuando sobre segmentos cuja especificação teórica difere muito da encontrada ao nível fonético), irregulares ou lexicalmente marcados. Vejamos:

- a realização de [e] e [o] em posição átona pré-tónica é dificilmente encontrada fora do contexto de ditongo decrescente fonético; este padrão distribucional particular, contextualmente muito restritivo, obriga-nos a aceitar estas vo-

³ Estas observações circunscrevem-se, exclusivamente, ao português europeu. Para uma discussão do mesmo tópico relativamente ao português do Brasil, cf., entre outros, Bisol (2010).

gais, nesta posição, como o produto específico das regras pós-lexicais de formação de ditongos em português e não propriamente como vogais lexicalmente especificadas com este grau de abertura em níveis de representação fonológica mais teóricos e abstratos;

- a ocorrência de vogais abertas ou semiabertas em posição não acentuada corresponde, com elevada frequência, ao resultado das «exceções regulares» que inibem a redução átona (conforme as expostas, por exemplo, em Mateus et al. 2003, 1013ss. – cf., neste capítulo, os dados reunidos em (13)): produção de semiabertas em sílaba final fechada por consoante soante – *re'pórt[ɛ]r*, *'plânc[tɔ]n*; neutralização das oposições de abertura em posição inicial absoluta – {[ɛ]/[i]}*colo'gista*, {[u]/[o]/[ɔ]}*bri'gado*; exceções lexicalmente marcadas – *infl[a]'ção*, *ec[ɔ]no'mia*; etc.;
- sobreinventariações como a encontrada em (2) tornam ainda invisíveis certos aspetos muito relevantes para a descrição fonológica da língua como, entre outros: (i) a correspondência (morfofonológica) regular e sistemática entre vogais subjacentes e vogais de superfície (não estabelecendo uma relação inequívoca e explícita, por exemplo, entre [i] e [e] correspondentes a duas realizações possíveis da vogal temática /E/ em formas verbais como *de'v[e]mos* e *dev[i]'ríamos*); ou (ii) o caráter muito excepcional da terminação nominal átona [i] (em palavras como *táx[i]*, *júr[i]*, *biquín[i]*, *mín[i]*, etc.).⁴

Poderemos, assim, extrair uma primeira generalização de partida: não ignorando a possibilidade de, no nível de superfície, encontrarmos vogais que não respeitem rigidamente esta restrição – em muitos casos, devido a variáveis que podemos categorizar e explicitar de forma minimamente sistemática (cf. novamente (12)) –, é possível afirmar que em português europeu contemporâneo prevalecem, em posição átona, as vogais que ocupam posições mais elevadas e mais centralizadas/recuadas do espectro vocálico.

A posição final átona absoluta (vogal em posição final absoluta, eventualmente seguida do marcador de plural /S/) caracteriza-se mesmo por ser a mais restritiva de todas, admitindo, em português, um número muito escasso de vogais. Nesta variedade da língua, esta posição exhibe somente as vogais [i], [e] e [u], sendo praticamente inexistente ou muito excepcional a ocorrência, nesta posição, de vogais que delas difiram (como [i] ou, mais excepcionalmente ainda, [ɛ], conforme antes referido).

⁴ Por deterem, na verdade, uma terminação muito excepcional e marcada em português, estas palavras são frequentemente sujeitas a um fenómeno de *regularização* que consiste em substituir, em certas variedades dialetais/socioletais, o [i] final por [ɛ], por exemplo, em produções como *táx[ɛ]* e *biquín[ɛ]*. De sobreinventários como o de (2) não constam porém, em geral, referências às raras palavras terminadas em [ɛ] átono final, diretamente importadas do latim clássico, e sujeitas ao mesmo fenómeno de regularização: *inclusiv{[ɛ]/[i]}*, *exclusiv{[ɛ]/[i]}* (porventura também: *maxim{[ɛ]/[i]}*).

1.2 Questões de partida

Sistematizações do vocalismo português como as que resumimos na secção anterior, seguindo de perto algumas das principais descrições fonológicas do português, dão-nos uma primeira ideia aproximada dos padrões de distribuição mais frequentes, ao nível de superfície, das vogais da norma padrão desta língua.

Na apresentação mais desenvolvida do vocalismo do português que pretendemos empreender ao longo deste capítulo, esta divisão entre um inventário tónico e um ou vários inventários átonos das vogais da língua servir-nos-á enquanto ponto de partida para uma discussão mais ampla, em que pretendemos aprofundar e problematizar a discussão em torno de duas dimensões particulares e específicas, por sua vez desdobradas em algumas subquestões de âmbito mais restrito:

- 1) o contraste entre o inventário fonético e o inventário fonológico do vocalismo português, questão clássica no estudo do sistema vocálico do português (em grande parte relacionada com o processo de redução átona, que trataremos à parte). Algumas interrogações mais circunscritas, que serão objeto de reflexão mais desenvolvida em secções individualizadas do texto, são por sua vez suscitadas por esta questão de âmbito mais geral:⁵
 - a. qual o estatuto fonológico das vogais centrais não baixas do português?
 - b. qual o estatuto fonológico dos quatro graus de abertura vocálica do português?
- 2) *o processo de redução átona*, para o qual proporemos uma explicação assente numa interação entre processos, baseada no modelo da fonologia dos elementos (dando continuidade às análises propostas em Veloso 2012; 2013).⁶

⁵ Ainda no âmbito da discussão sobre as diferenças entre um inventário *fonético* e um inventário *fonológico* das vogais do português, poderíamos incluir a problematização do estatuto fonológico das vogais foneticamente nasais e das vogais assilábicas (*semivogais*), bem como o das estruturas mais complexas em que estas últimas se combinam (ditongos). No presente estudo, porém, circunscrevemos a nossa atenção às vogais foneticamente realizadas como orais e silábicas.

⁶ Ao longo do texto, servir-nos-emos da designação «fonologia dos elementos» para nos referirmos às abordagens teóricas e descritivas que, apoiando-se na «fonologia das partículas» de Schane (1984), concebem os segmentos fonológicos como unidades complexas e potencialmente decomponíveis em unidades primitivas (indecomponíveis) que não diferem ontologicamente dos segmentos: os *elementos/partículas* (cf., para uma apresentação deste quadro teórico e conceptual, estudos como Schane 1984; Kaye/Lowenstamm/Vergnaud 1985; Van Der Hulst 1989; Brandão de Carvalho 1993; Boltanski 1999; Angoujard 2003; 2006; Durand 2005; Brandão de Carvalho/Nguyen/Wauquier 2010; ou Backley 2011, entre outros). De acordo com esta proposta, cada segmento fonológico resulta da combinação de um elemento dominante com um ou mais elementos em posição dominada, criando-se uma relação hierárquica que em fonologia das dependências e em fonologia do governo se descreve como {Cabeça. Operador} (Scheer 1998, 141–142; Boltanski 1999, 79; Angoujard 2003, 176ss.; 2006, 36ss.). No caso das vogais – que correspondem aos segmentos mais estudados pela fonologia dos elementos –, elas correspondem universalmente à combinação de um elemento de abertura/sonoridade {A} com um ou ambos de dois elementos de tonalidade: {I} (palatalidade) e {U} (labialidade) (Schane 1984, 131;

Cada um destes temas será discutido em separado num subcapítulo diferenciado do texto.

2 As vogais do português europeu contemporâneo: inventariação fonética e inventariação fonológica

Tal como afirmámos anteriormente, as vogais encontradas em (1) e (2) correspondem aos segmentos vocálicos que é possível encontrar, *ao nível das realizações fonéticas* e em diversas posições prosódicas, na maior parte dos dialetos do português (nomeadamente, na norma padrão da língua).

Tais inventários correspondem, por conseguinte, a *inventários fonéticos*. Conforme posto em destaque, entre outros, por Mateus/Andrade (2000, 19ss.) e Mateus et al. (2003, 990ss.), torna-se necessário contrastar estes inventários, baseados na pura observação empírica de dados, com uma observação rigorosa do comportamento linguístico das vogais que eles contemplam, a fim de determinarmos quais as eventuais representações teóricas correspondentes a estas mesmas vogais. Por outras palavras: a partir dos inventários fonéticos de (1) e (2), teremos de reunir dados e argumentos que nos conduzam à proposta de um *inventário fonológico* das vogais teóricas que lhes subjazem. Trata-se de uma questão que tem sido debatida em anteriores descrições fonológicas do português; na discussão deste tema, isolaremos, no presente estudo, as questões mais específicas identificadas em cada uma das subsecções seguintes do texto.

2.1 O estatuto fonológico das vogais centrais não baixas em português europeu contemporâneo

Entre as vogais foneticamente atestadas no português, encontramos as duas vogais centrais [e] e [i] – cf. (1) e (2) –, que a maior parte das anteriores descrições fonológicas da língua explicitamente circunscrevem ao domínio restrito das *vogais puramente fonéticas*, obtidas por derivação e correspondendo, ao nível subjacente, a vogais como /a/, /e/ ou /ɛ/ (cf. e.g. Mateus/Andrade 2000, 18, 20, 30; Mateus et al. 2003, 991–992, 1001, 1009).

Esta explicação encontra uma forte sustentação em todos os casos de alternância morfológica em que uma vogal *tónica* diferente de [e] ou [i] alterna, em superfície, com estas mesmas vogais, como ocorre nos exemplos de (3). Tais exemplos demons-

Brandão de Carvalho/Nguyen/Wauquier 2010, 87). Anteriores aplicações deste modelo à descrição das vogais do português podem ser encontradas, por exemplo, em Brandão de Carvalho (1993; 2011) ou Veloso (2012; 2013).

tram que [e] e [i], nos pontos segmentais identificados – cf. representações ortográficas sublinhadas, em (3) –, correspondem a vogais com uma especificação fonológica diferente da sua especificação fonética: a oscilação fonética da mesma vogal ocupando um mesmo ponto exato da cadeia morfológica (em geral, a última vogal da raiz derivacional) entre uma realização tónica com determinado grau de abertura e anterioridade/recuo e uma realização átona em que estas dimensões diferem permite postular a existência de uma mesma vogal subjacente partilhada pelas duas realizações fonéticas, responsabilizando a atribuição de acento por tal oscilação (ausente, naturalmente, da especificação lexical das raízes morfológicas). De igual modo, [u] pode ser identificado como a realização fonética de uma vogal diferente de /u/, quando alterne morfofonologicamente com vogais cuja especificação fonética não coincida com a de [u], tal como contemplado também nos exemplos de (3).

Assumindo o contexto tónico como aquele que melhor preserva os traços fonológicos das vogais, admitimos, de acordo com descrições anteriores do português como Mateus (1975), que a especificação fonológica subjacente da vogal comum às duas realizações fonéticas é a que é preservada pela vogal em contexto acentuado. Este ponto de vista equivale a aceitar que, nos casos de que aqui nos ocupamos, a especificação fonética das vogais átonas resulta claramente da aplicação das regras de acentuação e redução átona (estas últimas serão apresentadas na secção 3 deste capítulo).

- (3) Alternâncias morfofonológicas pondo em evidência a existência de vogais diferentes de [e] e [i] (e [u]) no nível subjacente

Formas com realização fonética da UVRD em posição tónica	Formas com realização fonética da UVRD em posição átona	Vogal subjacente proposta (para a UVRD)	DERIVAÇÃO ≡ Tónica → Átona
<i>ca<u>sa</u></i> ['kaze]	<i>ca<u>se</u>iro</i> [ke'zejru]	/a/	/a/ → [e]
<i>pe<u>dr</u>a</i> ['pedre]	<i>pe<u>d</u>reira</i> [pi'drejre]	/ɛ/	/ɛ/ → [i]
<i>me<u>d</u>o</i> ['medu]	<i>me<u>d</u>ricas</i> [mi'drike]	/e/	/e/ → [i]
<i>mo<u>r</u>te</i> ['mɔrti]	<i>mo<u>r</u>tal</i> [mur'tat]	/ɔ/	/ɔ/ → [u]
<i>no<u>v</u>o</i> ['novu]	<i>no<u>v</u>idade</i> [nuvi'dadi]	/o/	/o/ → [u]

UVRD: Última Vogal da Raiz Derivacional

Em alguns trabalhos anteriores (Veloso 2007; 2010a; 2012; 2013), porém, defendemos que [e] e [i], em casos em que não alternam morfofonologicamente com vogais como [a], [ɛ] ou [e] (ou seja, em casos não diretamente comparáveis com os exemplos de (3)), correspondam, mesmo sacrificando a economia do inventário segmental fonológico da língua, às vogais subjacentes /e/ e /i/.⁷

⁷ Aplica-se a esta proposta o mesmo *rationale* que mostra que a vogal fonética [u] tanto pode ser o resultado de superfície de uma vogal fonologicamente especificada pelos mesmos traços – cf. *puro*/

Estes casos – em que, segundo a proposta apresentada, [e] e [i] correspondem a /e/ e /i/ fonológicos – são exemplificados de forma muito particular pelas terminações nominais de nomes como *casa* e *base* e pela vogal [i] encontrada em clíticos como *que*, *dee*, *mee*, *lhe*, *see*, etc. A falta de realizações destes morfemas em que uma vogal tónica, diferente de [e] ou [i], alterne morfofonologicamente com [e] ou [i] átonos é o principal argumento para não encontrarmos evidência forte que suporte a aceitação – nestes casos – de uma vogal como /a/, /ɛ/ ou /e/ na subjacência de [e] ou [i].⁸

Aceitando a existência de /e/ e /i/ como segmentos *fonológicos* do português, nos termos que acabamos de sintetizar, nada obsta, ainda, a que passemos a considerar, conforme defendido em Veloso (2010a; 2012; 2013), /i/ como a vogal (*fonológica*) não marcada do português. Os dois principais argumentos em que fundamentamos, nesses trabalhos, a defesa desta posição são os seguintes: (i) [i] corresponde à vogal *epentética* nesta variedade da língua; (ii) /i/ corresponde à vogal *vazia*, desprovida de qualquer representação fonológica (nos termos da fonologia dos elementos, ela é representável como: /i/={ } – cf. Veloso 2012; 2013).⁹

/pureza([u]=/u/) –, como o resultado de uma alteração da especificação fonológica da vogal subjacente por via da aplicação das regras de redução átona – conforme se torna patente em pares como *podre/ podridão*, em que [u] (átono, na palavra sufixada)=/o/, ou *nobre/nobreza*, em que [u] (átono, na palavra sufixada)=/ɔ/.

8 A respeito deste argumento, poderíamos, naturalmente, contrapor o seguinte questionamento crítico: o postulado de formas subjacentes como /e/ e /i/ aqui proposto, com base nos fundamentos expostos, assenta fortemente num tipo de evidência que pode ser avaliado com algum ceticismo – a *evidência negativa*. Dado que, para formas como as apresentadas no texto, não existem, no português contemporâneo, realizações de [e] e [i] que alternem morfofonologicamente com vogais com outros graus de abertura, assumimos que a única especificação possível para tais vogais, no nível subjacente, fosse, respetivamente, /e/ e /i/.

Num plano argumentativo puramente racional e conjetural, não é inteiramente legítimo, na verdade, excluir que só devido a um acaso histórico formas tónicas dos mesmos morfemas em que estas vogais ocorrem não tenham emergido na língua em posição tónica e com outros graus de abertura. Reconhecemos que, nesse plano de raciocínio maximamente abstrato, tais formas não deixam de corresponder, de algum modo, a um «formato disponível» nos recursos fonológicos potenciais do português. Num plano mais empírico, porém, o que é factualmente atestável é que tais formas não emergiram, pelo que aqui insistimos na proposta apresentada em trabalhos anteriores (Veloso 2007; 2010a; 2012; 2013).

9 Uma notação alternativa seria /i/={@, @}, já que, em fonologia dos elementos, @ simboliza um vazio de elemento.

2.2 O estatuto fonológico da distinção entre quatro graus de abertura vocálica em português

De acordo com os inventários de (1) e (2), o português pertenceria ao grupo das línguas que distinguem, *pelo menos no nível fonético*, entre as vogais altas ([i], [ĩ] e [u]) e a vogal baixa ([a]), dois graus de abertura intermédios: as médias-altas (ou semi-fechadas) [e] e [o] e as médias-baixas (ou semiabertas) [ɛ] e [ɔ].¹⁰

Nem todas as línguas românicas ostentam oposições deste tipo. Línguas como o italiano e o espanhol, por exemplo, admitem unicamente, quer na série das vogais anteriores, quer na das recuadas, uma vogal intermédia, /E/ ou /O/, foneticamente realizável, respetivamente, como [ɛ]/[e] e [ɔ]/[o] (Lausberg 1981, 59–62, 110, 116ss.; Alarcos Llorach ⁴1991, 146ss.; Dardano 2005, 27; Kenstowicz 2010). A oscilação entre uma vogal semi-fechada ([e]/[o]) e uma vogal semiaberta ([ɛ]/[ɔ]), nestas línguas, depende, na maior parte dos casos, de fatores de natureza contextual, prosódica ou dialetal (cf. e.g. Alarcos Llorach ⁴1991, 149; Dardano 2005, 27; Kenstowicz 2010).

Em línguas como o português e o francês (cf., para a história do vocalismo francês, Anglade 1958, 7–32), divergentemente, encontramos, nos graus intermédios de abertura e ao nível das realizações fonéticas, a já referida oposição entre vogais semi-fechadas e vogais semiabertas. Esta oposição manifesta-se plenamente em posição tónica, no caso do português – cf. (1).

Numa primeira análise, poderíamos atribuir a esta oposição um caráter lexical. Ou seja: é possível encontrar pares léxicos do português, não relacionados entre si por qualquer tipo de alternância morfofonológica, cujos termos se distinguem entre si unicamente pelo grau de abertura (semiaberto≠semi-fechado) da vogal, normalmente em posição tónica. Reunimos em (4) alguns exemplos desta condição, face aos quais nos é permitido aceitar esta oposição já ao nível das oposições fonológicas, lexicais, distintivas e subjacentes.

- (4) Pares léxicos do português baseados na oposição {semiaberto≠semi-fechado}

/E/	/O/
<i>sede</i> ['sedɨ] ~ <i>sede</i> ['sedɨ]	<i>molho</i> ['mɔʎu] ~ <i>molho</i> ['moʎu]
<i>pega</i> ['peɣɛ] ~ <i>pega</i> ['peɣe]	<i>forma</i> ['fɔrmɛ] ~ <i>forma</i> ['forme]
<i>besta</i> ['bestɛ] ~ <i>besta</i> ['beste]	<i>cor</i> [kɔr] ~ <i>cor</i> [kor]

¹⁰ Esta distinção de dois graus de abertura intermédios faria do português uma das línguas românicas em que melhor se preservariam as oposições vocálicas do próprio latim, dado que as semiabertas [ɛ] e [ɔ] descendem das vogais breves latinas ĕ e ō e as semi-fechadas [e] e [o] descendem das vogais longas latinas ē e ō e das vogais breves ĭ e ū (Nunes ³1956, 40ss.; Lausberg 1981, 129; Huber 1986, 53ss.; Castro 1991, 113–114; Calabrese 2005).

Em termos autosegmentais, autores como Wetzels (1992) e Bisol (2010) descrevem este tipo de oposições recorrendo a explicações que pressupõem a natureza escalar (gradual) da abertura vocálica. É em propostas dessa natureza que aqui nos baseamos para propor a classificação encontrada em (5) (que diverge, no escalonamento crescente de [aberto], das duas propostas acima citadas e que assume a gradação de quatro traços unários de abertura vocálica).

- (5) A descrição dos quatro graus de abertura vocálica do português de acordo com um modelo escalar de traços unários

[aberto1]	[i, ī, u]
[aberto2]	[e, o]
[aberto3]	[ɛ, e, ɔ]
[aberto4]	[a]

De acordo com as propostas da fonologia dos elementos,¹¹ a diferença entre vogais semiabertas e semifechadas radicaria numa combinação, em graus diferentes, dos elementos de sonoridade ({A}) e tonalidade (palatalidade ou labialidade: {I} e {U}), e invertendo, na estrutura interna de cada segmento, as relações de dominância entre Cabeça e Operador: nas vogais semiabertas ([aberto3]), o elemento de sonoridade {A} ocorre como Cabeça do segmento; nas vogais semifechadas ([aberto2]), essa posição é ocupada pelo elemento de tonalidade {I} ou {U}. É esta a explicação que encontramos, para as vogais do francês e do português, por exemplo, em Angoujard (2003; 2006) e Brandão de Carvalho/Nguyen/Wauquier (2010) e em Veloso (2012; 2013), respetivamente. É em propostas como as que acabamos de referir que nos apoiamos para, consequentemente, explicar a diferença entre [aberto2] e [aberto3], num modelo baseado em elementos, nos termos encontrados em (6).

- (6) A distinção entre vogais [aberto2] e [aberto3] do português num modelo de elementos

	Anteriores	Recuadas
Vogais semifechadas (Cabeça: Tonalidade; Operador: Sonoridade)	[e]={I, A}	[o]={U, A}
Vogais semiabertas (Cabeça: Sonoridade; Operador: Tonalidade)	[ɛ]={A, I}	[ɔ]={A, U}

11 Cf. nota 6.

A questão principal de que nos ocuparemos neste momento não se prende tanto com a descrição estrutural da oposição em si mesma, mas com a discussão do estatuto fonológico que lhe cabe.

Interessa-nos, por outras palavras, discutir se a oposição entre estes dois graus de abertura vocálica:

- (i) está presente já no próprio léxico, correspondendo, como tal, a uma verdadeira oposição *fonológica* que legitime a aceitação de /e/, /ɛ/, /o/ e /ɔ/ como quatro segmentos distintos e distintivos em português;
- (ii) ou se, em contrapartida, as especificações dos graus de abertura destas vogais são atribuídas em níveis posteriores da derivação, circunscrevendo as oposições [e]~[ɛ] e [o]~[ɔ] ao nível meramente fonético. Caso aceitemos uma explicação como esta, o inventário segmental teórico das vogais do português ficaria reduzido a apenas duas vogais intermédias (/E/ e /O/), resultando daqui um inventário segmental mais económico (consequentemente, o português seria retirado do conjunto de línguas com quatro graus de abertura vocálica *fonológica*).

A questão, que tem sido discutida por alguns estudos anteriores (Wetzels 1992; 1995; 2011; Bisol 2010; Brandão de Carvalho 2011), parece-nos motivada pela observação de alguns dados pertinentes, bem como pelas reflexões levadas a cabo por alguns destes estudos, conforme pretendemos explorar de seguida.

Como podemos observar nos exemplos de (4), dispomos de um argumento principal que nos permite aceitar, em português, a existência de uma oposição *fonológica, subjacente*, entre vogais [aberto2] e [aberto3]. Esse argumento, claramente ilustrado nos exemplos dados, reside na função distintiva que tal oposição detém em pares léxicos como *sede* ['sedɨ] ~ *sede* ['sedɨ], *molho* ['moʎu] ~ *molho* ['moʎu], etc.

Pares desta natureza, porém, não são muito abundantes em português, e outros argumentos nos poderiam levar a colocar a hipótese de que estivéssemos perante uma oposição pós-lexical:

1. Muitos casos de alternância fonética [aberto2]/[aberto3] são *gramaticalmente motivados*, não estando, portanto, presentes no léxico:
 - a) casos de alternância morfofonológica em que a abertura vocálica é atribuída à última vogal do radical derivacional /E/ ou /O/ em função do marcador temático com que se combina para dar origem: (i) a formas verbais da primeira conjugação (marcador temático /a/, [aberto4], foneticamente realizado ou omissivo), com última vogal do radical derivacional foneticamente semiaberta (*eu gov[ɛ]rno*, *eu [ɛ]rro*, *eu despr[ɛ]zo*; *eu ch[ɔ]ro*, *eu alm[ɔ]ço*, *eu j[ɔ]go*); ou (ii) a formas nominais (marcador temático /u/, [aberto1]), com última vogal do radical derivacional foneticamente semifechada (*o gov[e]rno*, *o [ɛ]rro*, *o despr[e]zo*; *o ch[o]ro*, *o alm[o]ço*, *o j[o]go*). Estes casos poderiam ser explicados como casos particulares de *harmonização vocálica*, em que o grau de abertura da última vogal do radical derivacional se aproxima do do

marcador temático, ficando apenas a um grau de distância deste, de acordo com um esquema como (7).

- (7) Harmonização vocálica entre UVRD e marcador temático verbal/nominal
 /E V O/_{UVRD} & MT__[aberto4]_(Verbo_1ªConjug; MT=/a//) → [E V ɔ]_{UVRD}__[aberto3] (gov[ε]mo; ch[ɔ]ro)
 /E V O/_{UVRD} & MT__[aberto1]_(Nome; MT=/u/) → [e V o]_{UVRD}__[aberto2] (gov[e]mo; ch[o]ro)

(&=combinação de raiz derivacional com marcador temático teórico, foneticamente expresso ou omitido)

- b) Na flexão nominal do português, encontramos um número considerável de pares gramaticais com última vogal do radical flexional /O/ que se realiza como [o] no masculino singular e como [ɔ] nas restantes formas, conforme evidenciado pelos exemplos de (8), também explicáveis como casos de *harmonização vocálica*.¹²

- (8) Alternância morfofonológica entre última vogal do radical flexional [o] no masculino singular e última vogal do radical flexional [ɔ] nas restantes formas¹³

n[o]vo; n[ɔ]va, n[ɔ]vos, n[ɔ]vas
 s[o]gro; s[ɔ]gra, s[ɔ]gros, s[ɔ]gras

2. Noutros casos, a alternância [aberto2]/[aberto3] é simplesmente *neutralizada* em posição átona, como se torna evidente nos casos em que um mesmo radical flexional ou derivacional pode ostentar, em posição tónica, uma oscilação entre [ε]/[e] ou [ɔ]/[o] (fruto, normalmente, das alternâncias morfofonológicas apresentadas no §1 *supra*), completamente subsumida, em contexto átono, por [i] ou [u], conforme ilustrado pelos exemplos de (9). Este argumento serve precisamente para Brandão de Carvalho (2011) propor que, em português, as oposições [ε]/[e] ou [ɔ]/[o] correspondam, no nível subjacente, a duas vogais fonológicas apenas (/E/ e /O/): visto que o *output* da redução vocálica que neutraliza esses dois pares em posição átona é o mesmo para as duas vogais de cada par ([ε]/[e] neutralizados em [i]; [ɔ]/[o] neutralizados em [u]), o autor admite que, em subjacência, corresponda a cada par uma só vogal teórica, à qual se aplica a regra de neutralização (Brandão de Carvalho 2011, 59, 61ss.).

¹² Junto de diversos autores, este fenómeno é designado por *metafonia*, termo que engloba também as assimilações discutidas no ponto anterior (cf. e.g. J. M. Barbosa 1994, 178ss.).

¹³ Estes casos correspondem a um subgrupo de palavras bem delimitado na língua, sujeitas a limitações gramaticais e históricas distintas das que abrangem os casos discutidos anteriormente e expostas a fenómenos de variação diatópica que não tornam a descrição aqui proposta diretamente aplicável a todos os dialetos da língua.

- (9) Neutralização átona das oposições fonéticas tónicas [ɛ]/[e] e [ɔ]/[o]

*go*ʋ[ɛ]ʀmo_{Verbo} ~ *go*ʋ[e]ʀmo_{Nome} vs. *gov*[i]ʀma'çãõ
con's[ɔ]lo_{Verbo} ~ *con*'s[o]lo_{Nome} vs. *cons*[u]la'çãõ

3. Finalmente, um terceiro argumento que poderia ser invocado na defesa de que, a nível subjacente, não existe uma oposição entre vogais [aberto2] e vogais [aberto3], remetendo tal distinção para níveis mais próximos das estruturas de superfície e das realizações fonéticas, encontra-se no fenómeno de *abaixamento dactílico*, posto em destaque por autores como Wetzels (1992; 1995; 2011), que focaliza a sua atenção nas variedades brasileiras do português. Este fenómeno consiste no abaixamento que se dá na última vogal do radical derivacional, de [aberto2] para [aberto3], quando certos radicais derivacionais, realizados com última vogal do radical derivacional semifechada em formas paroxítonas, dão origem a derivados proparoxítonos em que a mesma última vogal do radical derivacional sofre abaixamento. Em posição proparoxítona dar-se-ia, então, um outro caso de neutralização da oposição entre os dois graus de abertura vocálica. O fenómeno encontra-se exemplificado em (10) e é apontado por Wetzels (1992; 1995; 2011) como uma evidência de que, no nível subjacente, só as vogais semiabertas têm cabimento. Wetzels (1992, 32–33) tenta reforçar este argumento com um outro: a suposta excepcionalidade de proparoxítonos não sufixados com vogal semifechada em posição tónica; tal excepcionalidade não parece suficientemente comprovada, porém, quando considerada a relativa frequência de palavras como 's[o]frego, 'tr[o]pego, es't[o]mago, 'p[e]ssego, 'n[e]spera, etc.).

- (10) Exemplos de *abaixamento dactílico*, segundo Wetzels (1992; 1995; 2011)

Formas paroxítonas com vogal [aberto2] em UVRD		Formas proparoxítonas com vogal [aberto3] em UVRD
/E/= [e] <i>esque</i> '[e]to <i>fo</i> 'n[e]ma	→	/E/= [ɛ] <i>esque</i> '[ɛ]tico <i>fo</i> 'n[ɛ]mico
/O/= [o] <i>ge</i> 'h[o]ma		/O/= [ɔ] <i>ge</i> 'h[ɔ]mico

Como avaliar evidências como as reunidas em (7), (8), (9) e (10) à luz da questão de sabermos quais as vogais admissíveis no inventário fonológico vocálico do português, nomeadamente no que diz respeito à questão particular da especificação lexical dos graus de abertura [aberto2] e [aberto3]?

Pensamos que o mesmo argumento em que, em estudos anteriores, nos apoiámos para defender a existência das vogais fonológicas /e/ e /i/ em português pode aplicar-se também neste caso, conforme passamos a explicar.

Para todos os casos em que dados da sincronia nos mostrem alternância morfológica entre [aberto2] e [aberto3] – ou seja, relativamente aos casos em que um mesmo radical flexional ou derivacional apresente, num dado ponto da cadeia segmental e morfológica, uma vogal com um grau de abertura claramente relacionável com propriedades ou processos gramaticais (cf. exemplos em (7), (8) e (9) *supra*) –, julgamos adequada a postulação de uma mesma vogal subjacente às duas vogais que alternam na forma de superfície, sem nos comprometermos com a especificação de um desses dois graus de abertura como aquele que estaria presente já a nível das representações lexicais. Para tais casos, seria suficiente postular a existência de uma vogal subjacente como /E/ ou /O/, deixando a especificação do seu grau de abertura fonético para etapas posteriores da derivação (em concordância, neste aspeto e recordando estudos anteriores referentes ao português, com a proposta encontrada em Brandão de Carvalho 2011, 59, 61ss.).

No entanto, esta explicação não nos parece aplicável aos casos – que, pouco numerosos ou não, são atestados em português – em que [aberto2] ou [aberto3] não resultam de um processo de alternância morfológica e, por isso, não se integram imediatamente em processos de neutralização como o apresentado em (9). Referimo-nos explicitamente aos casos em que a oposição [aberto2]/[aberto3] assegura oposições lexicais distintivas como as exemplificadas em (4).

Para estes casos, propomos que a especificação de abertura ocorra no próprio léxico, opondo, em subjacência, quatro vogais fonológicas: /e/, /ɛ/, /o/ e /ɔ/. Reiteramos que estas não são as únicas vogais que poderão ser encontradas, a nível teórico, na subjacência de todas as realizações de [e], [ɛ], [o] e [ɔ]: mantemos que, nos casos em que a especificação do grau de abertura seja claramente atribuível a processos fonológicos ou morfológicos mais próximos das últimas etapas da derivação, uma vogal fonológica não inteiramente especificada quanto ao grau de abertura (/E/, /O/) será suficiente para integrar a representação subjacente de tais vogais.

3 A redução do vocalismo átono em português

Deixamos para a secção final deste capítulo uma reflexão mais focalizada sobre um dos processos fonológicos mais idiossincráticos do português, graças ao qual podemos explicar, entre outras características desta língua, a coexistência de um inventário vocálico tónico e de um (ou vários) inventário(s) vocálico(s) átono(s) (cf. secções 1 e 2 deste capítulo): a *redução do vocalismo átono*, também designada como *elevação e centralização/recuo das vogais átonas*.

Em termos puramente descritivos, trata-se do fenómeno que faz com que uma mesma vogal subjacente possa ocorrer com diferentes graus de abertura em função da posição prosódica que ocupa em cada palavra ou forma de palavra (neutralizando um número considerável de oposições atestadas em posição tónica): [aberto4], [aberto3] ou [aberto2] em posição tónica; [aberto1] (sem palatalidade), preferencialmente, em

posição átona. A par da diminuição do grau de abertura das vogais átonas, este processo é ainda responsável por uma centralização ou recuo das mesmas.

Os exemplos de (3) ilustram, uma vez mais, este processo em português.

Conforme defendemos anteriormente (Veloso 2013), a redução do vocalismo átono do português é, simultaneamente, um processo *fonético* – pois confina o espaço de realização articulatória, em posição átona, a um número mais reduzido de configurações articulatórias do que as encontradas em posição tónica – e um processo *fonológico*, precisamente por neutralizar, em posição átona, um certo número de contrastes atestados em posição tónica.¹⁴

Nas suas manifestações fonéticas específicas, a redução do vocalismo átono é bastante característica do português: por exemplo, está na origem de fenómenos não só como a redução/neutralização acima referida, mas também de outros processos pós-lexicais, com relevo para uma tendência frequente de apagamento fonético das vogais átonas (este apagamento, por sua vez, é responsável, entre outros efeitos, pelo afastamento – muito típico do português coloquial – em relação ao chamado «silabismo máximo», de acordo com a terminologia de Emiliano 2009). Noutras línguas, porém, fenómenos de neutralização semelhantes, igualmente associados à perda de acento, são também encontrados (cf., para uma revisão da literatura e de alguns dados contemplando essa diversidade linguística, Veloso 2013, 660ss.). Com base, justamente, na comparação do funcionamento dos processos de redução átona atestados em várias línguas do mundo, autores como Donegan (1973), Crosswhite (2004) e Harris (2005) propõem que os processos de redução vocálica átona sejam tipificados de acordo com as seguintes dicotomias:

- *coloração vs. descoloração* (Donegan 1973): a coloração consiste num reforço dos traços de tonalidade (palatalidade ou labialidade) das vogais átonas; a descoloração, pelo contrário, consiste na perda de tais traços, que pode ou não ser acompanhada de um reforço da abertura/sonoridade das vogais átonas;
- *reforço de contraste vs. redução da sonoridade* (Crosswhite 2004): o reforço de contraste elimina da posição átona as vogais não coincidentes com os vértices do triângulo vocálico ([i a u]); a redução de sonoridade diminui o grau de abertura das vogais átonas, confinando-as preferencialmente a posições centrais e mais elevadas do triângulo vocálico;
- *redução centrífuga vs. redução centrípeta* (Harris 2005): na redução centrífuga, as vogais átonas tendem a aproximar-se de [i a u], ao passo que na redução

14 Segundo Padgett/Tabain (2005, 14), entre outros, a redução fonética consistiria sobretudo numa compressão do espaço articulatório das vogais às porções mais centrais do espectro vocálico. Uma vez que, em português, da Redução do Vocalismo Átono nem sempre resultam vogais centrais (/o/ e /ɔ/ «reduzem-se», por exemplo, em [u]), este processo seria mais apropriadamente descrito em português como um fenómeno eminentemente *fonológico*, para o qual poderíamos preferir designações como *neutralização das oposições vocálicas em posição átona*, muito mais fiel, em nosso entender, ao comportamento dos dados linguísticos observados (cf. Veloso 2013).

centrípeta as vogais átonas se afastam dos vértices do triângulo vocálico para se aproximarem (tal como na redução de sonoridade, de acordo com a tipologia de Crosswhite 2004) de zonas centrais do triângulo vocálico.

A observação de processos de redução vocálica numa quantidade razoável de línguas do mundo mostra que muitas destas línguas se apresentam como muito regulares e coerentes na correspondência dos seus processos de redução átona com tipologias como as que acabamos de apresentar. Um exemplo, verificado numa variedade diferente do português, encontra-se no fenómeno que as descrições fonológicas do português do Brasil designam comumente como *harmonização vocálica*. Esta harmonização consiste, em muitos dialetos do português do Brasil, em uma vogal não alta adquirir o mesmo grau de altura da vogal tónica [aberto1], desde que respeitadas certas condições, entre as quais (cf. e.g. Bisol 2013): a vogal que sofre elevação ocorre na sílaba imediatamente anterior à sílaba tónica; entre a vogal tónica e a vogal que sofre elevação não existe mais do que um grau de diferença quanto à abertura vocálica; certas estruturas morfológicas, como diminutivos e superlativos, são imunes ao processo.¹⁵ Este fenómeno corresponde a uma instância de redução vocálica (não associada à perda de acento, ao contrário do que sucede com a redução do vocalismo átono do português) que se identifica somente com um dos três tipos de redução de Donegan (1973), Crosswhite (2004) ou Harris (2005). Vejam-se os exemplos de (11), em que a elevação da átona pré-tónica ([o]→[u]; [e]→[i]) configura sempre casos de *coloração*, *reforço de contraste* e *redução centrífuga*, de acordo com as referidas propostas.

- (11) A harmonização vocálica do português do Brasil como aplicação de processos sistemáticos de *coloração*, *reforço de contraste* e *redução centrífuga*

g[o]verno → g[u]verno

m[e]nino → m[i]nino

([aberto 2] → [aberto1], independentemente dos traços da palatalidade ou labialidade)

Regularidades deste tipo encontram-se em muitas outras línguas, como já foi afirmado.

A redução do vocalismo átono do português, porém, e conforme posto em destaque por Brandão de Carvalho (2011) e Veloso (2013), apresenta-se-nos como um processo híbrido à luz destas divisões, uma vez que nele parecem confluir, de certa forma contraditoriamente, comportamentos *assimétricos* relativamente aos elementos de tonalidade {I} e {U}. As vogais anteriores, com efeito, sofrem uma perda total de

¹⁵ No português medieval e clássico, ter-se-ia produzido um fenómeno semelhante à harmonização vocálica do moderno português do Brasil, conforme posto em destaque por autores como, entre outros, Carvalho (²1984, 95), Maia (1986, 530–533) e Marquilhas (2003, 10–11).

palatalidade (em reduções como $m[e]do \rightarrow m[i]dricas$ e $p[\varepsilon]dra \rightarrow p[i]dreira$ – cf. (3)), cabendo assim nos casos de *descoloração* (Donegan 1973), *redução de sonoridade* (Crosswhite 2004) ou *centripetação* (Harris 2005). Pelo contrário, as vogais recuadas, que apresentam um reforço total de labialidade/recuo (em «reduções» como $m[\text{ɔ}]rte \rightarrow m[u]rtal$ e $n[o]vo \rightarrow n[u]vidade$, recorrendo novamente aos exemplos de (3)), correspondem a casos de *coloração* (Donegan 1973), *reforço de contraste* (Crosswhite 2004) ou *centrifugação* (Harris 2005).

Descrito à luz da fonologia dos elementos, o processo de redução do vocalismo átono do português pode ser formalizado nos termos propostos em (12), que retomam Veloso (2013).

- (12) A redução do vocalismo átono do português descrita à luz da fonologia dos elementos e confrontada com as tipologias de redução vocálica de Donegan (1973), Crosswhite (2004) e Harris (2005)

Vogal subjacente (preservada em posição tónica)	Realização de superfície em posição átona	Comparação tónica/átona	Donegan (1973)	Crosswhite (2004)	Harris (2005)
/e/: { <u>A</u> , A}	[i]: { <u>@</u> , @}	{ <u>A</u> , A} → { <u>@</u> , @}	Descoloração	Redução de sonoridade	Centripetação
/ɛ/: { <u>A</u> , I}	[i]: { <u>@</u> , @}	{ <u>A</u> , I} → { <u>@</u> , @}	Descoloração	Redução de sonoridade	Centripetação
/a/: { <u>A</u> , A}	[ɐ]: { <u>A</u> , @}	{ <u>A</u> , A} → { <u>A</u> , @}	Descoloração	Redução de sonoridade	Centripetação
/o/: { <u>U</u> , A}	[u]: { <u>U</u> , U}	{ <u>U</u> , A} → { <u>U</u> , U}	Coloração	Reforço de contraste	Centrifugação
/ɔ/: { <u>A</u> , U}	[u]: { <u>U</u> , U}	{ <u>A</u> , U} → { <u>U</u> , U}	Coloração	Reforço de contraste	Centrifugação

OBS.: @=vazio/apagamento de elemento (rasurado na forma subjacente)

Em face destes dados, podemos reafirmar, reiterando propostas anteriores (Veloso 2013), que a redução do vocalismo átono corresponde, em português, não a um único processo fonológico, mas à interação de três processos intrinsecamente distintos:

- perda ou atenuação de {A}: todas as vogais átonas apresentam um decréscimo de pelo menos um grau de abertura relativamente às suas correspondentes tónicas;
- perda total de {I}: as vogais anteriores, na passagem de tónicas a átonas, perdem completamente o elemento {I}, quer em Cabeça, quer em Operador, realizando-se foneticamente como [i] (={});
- reforço total de {U}: as vogais recuadas, na passagem de tónicas a átonas, substituem o elemento {A}, quer em Cabeça, quer em Operador, por {U}, dando origem, invariavelmente, a realizações fonéticas [u] (={}).

O funcionamento diferente dos processos de coloração/descoloração, reforço de contraste/redução de sonoridade e centrifugação/centripetação, na redução do vocalismo átono do português, perante vogais anteriores e perante vogais recuadas poderá ainda relacionar-se com fatores de natureza histórica. Sendo a redução do vocalismo átono um fenómeno de datação algo controversa na história do português (Carvalho ²1984; Martins 1985; 2006; 2007; Teyssier ⁴1990; Castro 1991; Marquilhas 2003; Paiva 2008; 2009; Maia 2013), dispomos de indícios de que ele poderá ter começado a manifestar-se em épocas diferentes e diante de vogais diferentes (variando ao longo da diacronia). Entre os indícios que podemos aceitar para a validação desta suposição contam-se (entre outros, certamente) os seguintes:

- a possível oscilação, num período muito recuado da história da língua, entre [i] e [e], que podemos aceitar com base em indícios gráficos recolhidos em textos do século XIII e seguintes (cf. Martins 1985; Maia 1986, 367, 422);
- a inexistência de referências explícitas, absolutamente claras e inequívocas, à produção de uma vogal central alta fechada como [i] em textos anteriores ao século XIX.¹⁶ Com efeito – e apesar de Marquilhas (2003, 13) encontrar em grafias seiscentistas como «Ceristam» (para *cristão*) ou «foremos» (para *formas*) indícios «indubitáveis» de um «e mudo» já nos dialetos meridionais da língua de então–,¹⁷ só em escritos como os de J. Soares Barbosa (1822) e Viana (1883) tais referências são encontradas – e Lhomond (³1838, v), por sua vez, num manual de pronúncia do francês para falantes do português considera a vogal [ə] do francês «huma das cousas, que mais custão aos Portuguezes», o que não encontraria fácil explicação se, à época, o português já dispusesse de uma vogal central como [i] ou [ə]).

Teremos de admitir, portanto, que a redução do vocalismo átono possa corresponder, efetivamente, à sobreposição de processos distintos na natureza e na cronologia da língua, o que poderia em parte explicar o já mencionado comportamento *assimétrico* de [I] e [U].

Como acima afirmámos, a redução do vocalismo átono do português, nos moldes exemplificados em (3) e acima desenvolvidos, corresponde a uma característica muito idiossincrática do português. Trata-se, com efeito, e conforme dissemos já também, do fenómeno responsável pelo afastamento do «silabismo máximo» (Emiliano 2009) que caracteriza o português coloquial, causando a impressão intuitiva, sublinhada por J. M. Barbosa (²1983) na introdução à sua obra, de que «[...] les Portugais <avalent> la plupart de leurs voyelles, voire de leurs syllabes, et que de ce fait le portugais ne

¹⁶ Teyssier (⁴1990) data o surgimento desta vogal em português de finais do século XVIII: «Seja derivado directamente de [ɛ=], ou tenha passado por uma fase intermediária [i], como em posição final átona, uma coisa pelo menos parece segura: o [ē] pretónico, tão característico da língua contemporânea de Portugal, surgiu no século XVIII, provavelmente depois de 1750» (Teyssier ⁴1990, 62).

¹⁷ Tais testemunhos gráficos poderiam, porém, ser interpretados como o registo escrito de uma vogal diferente de [i], à luz de argumentos como, entre outros, os de Teyssier (⁴1990, 62), precisamente.

ressemble pas du tout à l'espagnol» (Barbosa ²1983, 11). Trata-se, ainda, de um fenómeno que se relaciona de perto com a possibilidade de classificarmos genericamente o português como uma língua de ritmo acentual (ou como uma língua de ritmo predominantemente acentual) (Pamies Bertrán 1999).¹⁸

Sendo um processo de aplicação muito regular em português, a redução do vocalismo átono está ainda sujeita, todavia, a um conjunto relativamente bem delimitado de contextos que a inibem. Por outras palavras: existem alguns contextos fonético-fonológicos em que a redução átona é bloqueada de forma consideravelmente sistemática. Partindo da lista de exceções encontrada em Mateus et al. (2003, 1013–1016), à qual acrescentamos observações como as que antes reunimos em Veloso (2009; 2010b), tais contextos podem ser resumidos como em (13).

(13) Contextos de inibição sistemática de redução do vocalismo átono em português (*apud* Mateus et al. 2003; Veloso 2009; 2010b).

- a. Em início absoluto de palavra: /e ε/ nunca são realizadas como [i]; as vogais recuadas não são obrigatoriamente reduzidas para [u]:

ermida: [e]r'mida, *[i]r'mida;

olhar: {[o]/[ɔ]/[u]}lhar

- b. Em ditongo decrescente:

flautista: fl[aw]'tista, *fl[ew]'tista

coisinha: [koj]'sinha, *[kuj]'sinha

- c. Antes de [ʃ]:

relvado: [reʃ]'vado, *[riʃ]'vado

maldade: [maʃ]'dade, *[mɛʃ]'dade

- d. Em sílaba final de palavra terminada por soante segmental ou por coda ramificada excecional:¹⁹

revólver: re'vólvr[ε]r, *re'vólvr[i]r

abdómen: ab'dóm[ε]n, *ab'dóm[i]n

plâncton: 'plânct[ɔ]n, *'plânct[u]n

córtex: 'córt[ε]x, *'córt[i]x

fórceps: 'fórc[ε]ps, *'fórc[i]ps

18 A categorização definitiva do português como uma língua tipicamente de ritmo acentual – e, mais ainda, a distinção entre as variedades europeias e brasileiras da língua com base na dicotomia ritmo acentual/ritmo silábico – constitui um outro tópico de abundante discussão, que aqui não aprofundaremos. Remetemos, para uma apreciação dos argumentos em que se baseia esta controvérsia, para estudos de referência como Major (1981; 1985), Pamies Bertrán (1999), P. Barbosa (2000) e Frota/Vigário/Martins (2002).

19 J.M. Barbosa (²1983), como vimos em (2), inclui /a/ entre as vogais abrangidas por esta exceção, dando o exemplo de *açúca*[a].

- e. Em palavras formadas com os sufixos *-mente* ou *z-avaliativo*:
rapidamente: r[a]pida'mente, *r[e]pida'mente
cafezinho: ca,f[ɛ]'zinho, *ca,f[i]'zinho
- f. Nos casos de vogais descendentes de hiatos medievais, na sua maioria resultantes da síncope histórica de consoante latina intervocálica:
padeiro: lat. PA[N]ATARIUM > port. med. p[aa]deiro > p[a]'deiro, *p[e]'deiro
caveira: lat. CA[L]AVARIAM > port. med. c[aa]veira > c[a]'veira, *c[e]'veira
- g. Em casos de oscilação/variação em curso:
mestrado: m{[ɛ,i]}s'trado
ingenuidade: ing{[ɛ,i]}nui'dade
- h. Em exceções lexicalmente marcadas:
economia: ec[ɔ]no'mia, *ec[u]no'mia
inflação: infl[a]'ção, *infl[e]'ção
invasão: inv[a]'são, *inv[e]'são

A observação destes contextos excepcionais em que a aplicação da redução do vocalismo átono se apresenta bloqueada conduz-nos a uma observação final, em que nos parece oportuno destacar que este processo (ou interação de processos), muito produtivo e regular em estádios da língua devidamente documentados, estará porventura a assistir, no presente, a um provável retrocesso. Esta é uma hipótese suscitada, relativamente ao português, por trabalhos como Marquilhas (2003, 9ss.).²⁰

Na verdade, comparando a realização *átona* de certas raízes derivacionais integradas no léxico de herança da língua (ou mesmo no léxico culto até meados do século XX) com a das mesmas (ou outras) raízes em palavras que deram entrada no português nas últimas décadas, podemos observar indícios de que a redução do vocalismo átono não será porventura, no estágio contemporâneo da língua, tão produtiva como foi em estádios passados. Os exemplos de (14),²¹ em que contrapomos palavras (por vezes partilhando a mesma raiz derivacional ou outros formativos morfológicos) admitidas no português até à primeira metade do século XX a palavras mais recentes, sugerem que a redução do vocalismo átono seja bastante menos frequente nos neologismos. Esta inferência, que carece de aprofundamento futuro, pode contribuir para a identificação de um indício de que estejamos perante um eventual processo de variação em curso, que acabará eventualmente por conduzir a fonologia do português a uma provável «atenuação» das regras de redução vocálica.

20 Hipótese semelhante é sugerida por Nadeu/Hualde (2013) para o catalão (uma outra língua românica sujeita a regras de redução do vocalismo átono idênticas, em certos aspetos, às do português) a partir da comparação do comportamento de certos constituintes morfológicos átonos em nomes compostos e não compostos.

21 Deixamos deliberadamente de fora desta breve apresentação da questão a discussão do estatuto das unidades acentuais, porventura de natureza diferente, originadas pelas formas resultantes das combinações com e sem redução do vocalismo átono.

Se estudos ulteriores vierem a comprovar estarmos em presença de uma variação em curso, não serão naturalmente de excluir significativos índices de variação motivada por fatores de diferente natureza, nomeadamente dialetal e, mesmo, idioletal.

- (14) Raízes derivacionais e outros morfemas, em posição átona, com e sem RVA: palavras de inserção lexical mais antiga (com RVA) vs. palavras de inserção lexical mais recente (sem RVA) em português europeu

Constituinte morfológico	Palavras introduzidas em português até meados do século XX (<u>COM</u> RVA)	Palavras introduzidas em português após meados do século XX (<u>SEM</u> RVA)
<i>ferr-</i>	<i>f[i]ʹrroso, f[i]ʹrrugem</i>	<i>f[ɛ]rroʹvia, f[ɛ]rrocaʹrril</i>
<i>tele-</i>	<i>t[i]ʹl[i]ʹfone, t[i]ʹl[i]ʹfoʹnia, t[i]ʹl[i]ʹférico, t[i]ʹl[i]ʹgrafia</i>	<i>t[ɛ]ʹl[ɛ]ʹmóvel, t[ɛ]ʹl[ɛ]ʹtransporte</i>
<i>bloc-</i>		<i>bl[ɔ]ʹquista</i>
<i>blog-</i>		<i>bl[ɔ]ʹgueiro, bl[ɔ]g[ɔ]ʹsʹfera</i>
<i>fad-</i>	<i>f[e]ʹdista</i>	
<i>jihad-</i>		<i>jih[a]ʹdista</i>
<i>mono-</i>	<i>m[u]ʹnóculo, m[u]n[u]ʹgrafia</i>	<i>m[ɔ]n[ɔ]ʹnucleʹose</i>
<i>foto-</i>	<i>f[u]t[u]ʹgrafia</i>	<i>f[ɔ]t[ɔ]ʹcópia, f[ɔ]t[ɔ]ʹteraʹpia, f[ɔ]t[ɔ]ʹssíntese</i>

4 Observações finais

Concluindo esta apresentação dos aspetos que nos pareceram mais importantes para a caracterização do vocalismo do português, reforçaremos algumas conclusões anteriormente contempladas.

Mais uma vez, fazemos notar que deste estudo foram excluídas as *vogais nasais* e as *semivogais*, tendo sido tomados em consideração predominantemente dados relativos à norma padrão do português europeu.

Em primeiro lugar, consideramos pertinente continuar a distinguir entre um *inventário vocálico tónico* e um (ou mais) *inventário(s) vocálico(s) átono(s)*. De acordo com a exposição desenvolvida anteriormente, porém, esta distinção terá um caráter mais probabilístico do que estritamente categórico: com efeito, à exceção de [i] e, com algumas reservas particulares, de [e],²² qualquer vogal de (1) se pode encontrar em

²² Cf. nota 1.

posição tónica em português. A ocorrência de vogais não centrais com graus de abertura [aberto4], [aberto3] e, com menor frequência, [aberto2] em posição átona não se encontra totalmente proibida nesta língua, embora se revista de um carácter relativamente excepcional. Podemos, em suma, afirmar que a norma padrão do português admite em contexto acentuado praticamente todas as vogais (1), ao passo que o contexto não acentuado, nesta mesma variedade, privilegia as vogais [e i u] (e [i], sobretudo em posição não final), embora não exclua categoricamente a ocorrência, mais excepcional, de outras vogais também.

Da mesma forma, julgamos pertinente manter uma distinção entre um *inventário fonético* e um *inventário fonológico* das vogais do português. Tal como relativamente à distinção referida no parágrafo anterior, no entanto, esta divisão não é aqui assumida exatamente nos mesmos termos em que é apresentada em alguns estudos anteriores. De acordo com anteriores descrições fonológicas do português, o inventário fonológico das vogais corresponderia sempre a um subconjunto de especificações das especificações identificadas no nível fonético. Neste último, seriam encontradas vogais – como [e] e [i] – *nunca* especificadas como tais no inventário fonológico da língua (e sempre interpretadas, conseqüentemente, como o resultado fonético, obtido por derivação fonológica, de vogais como /a/, /ε/ ou /e/).

A distinção entre um inventário vocálico fonético e um inventário vocálico fonológico, na proposta aqui apresentada, não assenta numa exclusão ou circunscrição categórica de determinados itens num ou noutra, mas, antes, na observação dos dados linguísticos e do comportamento linguístico dos segmentos vocálicos nesses dados. Assim, consideramos qualquer vogal – [e], [i], mas também [u], nos casos em que esta avaliação se aplique – como a contraparte fonética de uma vogal fonologicamente dotada de uma especificação diferente da que é identificada na forma de superfície sempre que, num mesmo morfema, estas vogais alternem com outras vogais foneticamente distintas, como sucede muito particularmente nos casos de redução do vocalismo átono – cf. exemplos de (3): 'c[a]sa~c[e]'sinha, {[a],[e]=/a/}; 'p[ε]dra~p[i]'dreira, {[ε],[i]=/ε/}; 'm[ɔ]rte~m[u]r'tal, {[ɔ],[u]=/ɔ/}; etc. Na falta de alternância morfofonológica deste tipo, não consideramos reunidos os argumentos suficientes para postular, no nível teórico subjacente, vogais com uma especificação fonológica diferente da que se encontra nas realizações fonéticas – assim, as terminações [e] e [i] de *casa* e *base*, por exemplo, corresponderiam aos segmentos teóricos /e/ e /i/, tal como o [u] de *p[u]reza*, que nunca alterna com vogal diferente de [u] ('*p[u]ro*, [u] tónico, vs. *p[u]reza*, [u] átono), corresponde, por isso, a /u/ fonológico). De acordo com este entendimento, e em linha com a argumentação desenvolvida em secções anteriores do texto, /i/ corresponderia não só a um segmento vocálico fonológico do português como, ainda, à sua vogal (fonológica) *não marcada*.

O mesmo se aplica, em nosso entender, à distinção entre vogais semiabertas e semifechadas. Nos casos em que uma alternância morfofonológica as faça alternar, em superfície, entre si ou com uma terceira vogal, designadamente em função de operações morfológicas (como em pares do tipo *gov[ε]rmo*_{Verbo}/*gov[e]rmo*_{Nome} ou

alm[ɔ]ço_{Verbo}/alm[o]ço_{Nome} – cf. mais exemplos em (7) e em (9)), aceitamos a existência de uma vogal subjacente, não especificada quanto ao grau de abertura, comum a todas essas realizações fonéticas. Contudo, nos casos em que estas vogais tenham um valor distintivo lexical – cf. (4) –, admitimos a sua presença, com a distinção [aberto2]/[aberto3] preservada já a nível subjacente, no próprio inventário vocálico fonológico do português.

Em consonância com esta proposta, o inventário vocálico teórico do português apresentar-se-á mais extenso do que é habitualmente assumido nas anteriores descrições fonológicas da língua. Concretamente, propomos, para o português, um inventário vocálico *fonológico* composto pelos segmentos constantes de (15), lexicalmente especificados como tais nos casos a que se apliquem as explicações desenvolvidas mais acima. A incorrespondência entre um inventário vocálico fonético e um inventário vocálico fonológico do português, diante dos argumentos expostos, derivará então, e somente, de uma relação de tipo multívoco entre segmentos fonológicos e realizações fonéticas, identificada e analisada a partir de um exame criterioso das derivações fonológicas que estão na origem de cada forma de superfície. A multivocidade destas relações materializa-se na possibilidade de um mesmo segmento fonológico ser especificável foneticamente através de duas ou mais realizações distintas – e.g. {/e/= [e]V[i]}, em pares como *dēdo~dēdada* – e vice-versa – {[u]=/o/V/u/}, em palavras como *fogueira* (cf. *fogo*) e *furador* (cf. *furo*). Complementarmente, admite-se a hipótese de duas ou mais vogais fonéticas idênticas – e.g. diferentes realizações de [e], [i] ou [u] – poderem corresponder a vogais com especificações fonológicas coincidentes com estas mesmas especificações fonéticas – [e]=/e/; [i]=/i/, [u]=/u/ – ou a vogais com especificações fonéticas diferentes – [e]=/a/; [i]=/e, ε/, [u]=/o, ɔ/. Tentamos formalizar esta rede de relações em (16).

(15) O inventário *fonológico* das vogais do português

/i e ε a ɔ o u e i/

(16) Correspondência multívoca entre vogais fonéticas e vogais fonológicas em português

/i/	/e/	/ε/	/i/	/a/	/e/	/o/	/ɔ/	/u/
[i]	[e]	[ε]		[a]		[o]	[ɔ]	
((i))			[i]		[e]			[u]

Em português, um dos processos responsáveis pela multivocidade ilustrada em (16) é a redução do vocalismo átono, que quisemos aqui integrar numa explicação mais abrangente de todo o sistema vocálico português. Tal explicação integrada assentou, num primeiro momento, numa revisão das principais características e manifestações de superfície quer do processo de redução vocálica em si mesmo nesta língua, quer

das principais exceções determinadas por alguns contextos relativamente bem identificados em estudos anteriores. Em articulação com essa revisão, pretendemos reafirmar uma descrição da redução do vocalismo átono que, baseada na fonologia dos elementos – permitindo-nos identificar um funcionamento claramente assimétrico de {I} e {U} neste processo –, coloca em evidência indícios sugestivos de que a redução do vocalismo átono pode ser de alguma forma «desdobrada» em mais de um processo, eventualmente como consequência da sobreposição de processos que deram entrada na língua em fases distintas da diacronia. Esta proposta de análise foi ainda completada com a sugestão de que a redução do vocalismo átono poderá estar porventura a entrar, no estágio atual do português, numa fase de retrocesso, dados os casos, relativamente numerosos, de neologismos em que o processo não se aplica, mesmo diante de constituintes morfológicos que, em estádios anteriores da língua, foram objeto da sua aplicação.

Naturalmente, várias dimensões pertinentes para uma descrição cabal do sistema vocálico do português não foram objeto de exploração neste estudo, onde procurámos contrapor algumas propostas tradicionais com análises mais recentes de diversos fenómenos e cujas propostas deixamos para discussão e escrutínio futuros.

5 Referências

- Alarcos Llorach, Emilio (⁴1991), *Fonología Española*, Madrid, Gredos.
- Anglade, Joseph (1958), *Grammaire élémentaire de l'ancien français*, Paris, Colin.
- Angoujard, Jean-Pierre (2003), *Phonologie et diachronie*, in: Jean-Pierre Angoujard/Sophie Wauquier-Gravelines (edd.), *Phonologie: Champs et perspectives*, Lyon, ENS Éditions, 173–194.
- Angoujard, Jean-Pierre (2006), *Phonologie déclarative*, Paris, CNRS.
- Backley, Philip (2011), *An Introduction to Element Theory*, Edinburgh, Edinburgh University Press.
- Barbosa, Jorge Morais (²1983), *Études de Phonologie Portugaise*, Évora, Universidade de Évora.
- Barbosa, Jorge Morais (1994), *Introdução ao Estudo da Fonologia e Morfologia do Português*, Coimbra, Almedina.
- Barbosa, Plínio (2000), «Syllable-Timing in Brazilian Portuguese»: *Uma crítica a Roy Major*, DELTA 16:2, 369–402.
- Bisol, Leda (2010), *A simetria no sistema vocálico do português brasileiro*, *Linguística. Revista de Estudos Linguísticos da Universidade do Porto* 5:1, 41–52.
- Bisol, Leda (2013), *Harmonização vocálica: efeito parcial e total*, *Organon* 28:54, 49–61.
- Boltanski, Jean-Élie (1999), *Nouvelles directions en phonologie*, Paris, Presses Universitaires de France.
- Brandão de Carvalho, Joaquim (1993), *De quoi sont faites les voyelles? Phonologie tridimensionnelle des particules et harmonie vocalique*, in: Bernard Laks/Marc Plénat (edd.), *De natura sonorum. Essais de phonologie*, Saint Denis, Presses Universitaires de Vincennes, 65–100.
- Brandão de Carvalho, Joaquim (2011), *Contrastive hierarchies, privative features, and Portuguese vowels*, *Linguística. Revista de Estudos Linguísticos da Universidade do Porto* 6:1, 51–66.
- Brandão de Carvalho, Joaquim/Nguyen, Noël/Wauquier, Sophie (2010), *Comprendre la phonologie*, Paris, Presses Universitaires de France.
- Calabrese, Andrea (2005), *On the Feature [ATR] and the Evolution of the Short High Vowels of Latin into Romance*, *A View from Arjona*. University of Connecticut Working Papers in Linguistics 13, 33–78.

- Carvalho, José G. Herculano de (²1984, ¹1962), *Nota sobre o vocalismo antigo português: valor dos grafemas «e» e «o» em sílaba átona*, in: José G. Herculano de Carvalho (1984), *Estudos Linguísticos*, Coimbra, Coimbra Editora, 2, 75–103.
- Castro, Ivo (1991), *Curso de História da Língua Portuguesa*, Lisboa, Universidade Aberta.
- Crosswhite, Katherine M. (2004), *Vowel reduction*, in: Bruce Hayes et al. (edd.), *Phonetically Based Phonology*, Cambridge, Cambridge University Press, 191–231.
- Dardano, Maurizio (2005), *Nuovo Manualetto di Linguistica Italiana*, Bologna, Zanichelli.
- Donegan [Miller], Patricia (1973), *Bleaching and Coloring*, in: *Papers from the Ninth Regional Meeting. Chicago Linguistic Society*, Chicago, IL, Chicago Linguistic Society, 386–397.
- Durand, Jacques (2005), *Les primitives phonologiques: des traits distinctifs aux éléments*, in: Noël Nguyen/Sophie Wauquier-Gravelines/Jacques Durand (edd.), *Phonologie et phonétique. Forme et substance*, Paris, Hermes Science/Lavoisier, 63–93.
- Emiliano, António (2009), *Fonética do Português Europeu. Descrição e Transcrição*, Lisboa, Guimarães.
- Frota, Sónia/Vigário, Marina/Martins, Fernando (2002), *Discriminação entre línguas: evidência para classes rítmicas*, in: *Actas do XVII Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL/Colibri, 189–199.
- Harris, John (2005), *Vowel reduction as information loss*, in: Philip Carr et al. (edd.), *Headhood, Elements, Specification and Contrastivity*, Amsterdam/Philadelphia, Benjamins, 119–132.
- Huber, Joseph (1986, ¹1933), *Gramática do Português Antigo*, Lisboa, Fundação Calouste Gulbenkian [trad.].
- Kaye, Jonathan/Lowenstamm, Jean/Vergnaud, Jean-Roger (1985), *The internal structure of phonological elements: a theory of charm and government*, *Phonology Yearbook* 2, 305–328.
- Kenstowicz, Michael (2010), *Vocale incerta, vocale aperta*, *Studi e Saggi Linguistici* 48, 25–50.
- Lausberg, Heinrich (1981, ¹1963), *Linguística Românica*, Lisboa, Fundação Calouste Gulbenkian [trad.].
- Leão, Duarte Nunes de (1576), *Ortografia da Língua Portuguesa*, ed. crítica de Maria Leonor Carvalhão Buescu, Lisboa, INCM, 1983.
- Lhomond (²1838), *Elementos da gramática franceza*, Lisboa, Typ. de Antonio José da Rocha.
- Maia, Clarinda de Azevedo (1986), *História do Galego-Português. Estado linguístico da Galiza e do Noroeste de Portugal desde o século XIII ao século XVI (Com referência à situação do galego moderno)*, Coimbra, INIC.
- Maia, Clarinda de Azevedo (2013), *O vocalismo átono na história do Português: contributos para a cronologia das mudanças*, in: Rosario Álvarez et al. (edd.), *Ao sabor do texto. Estudos dedicados a Ivo Castro*, Santiago de Compostela, Universidade de Santiago de Compostela, 335–354.
- Major, Roy C. (1981), *Stress-timing in Brazilian Portuguese*, *Journal of Phonetics* 9, 343–351.
- Major, Roy C. (1985), *Stress and Rhythm in Brazilian Portuguese*, *Language* 61:2, 259–282.
- Marquilhas, Rita (2003), *Mudança analógica e elevação das vogais pretônicas*, in: Ivo Castro/Inês Duarte (edd.), *Razão e Emoção: Miscelânea de estudos em homenagem a Maria Helena Mira Mateus*, vol. II, Lisboa, Imprensa Nacional – Casa da Moeda, 7–18.
- Martins, Ana Maria (1985), *Elementos para um Comentário Linguístico do Testamento de Afonso II (1214)*, dissertação inédita (apresentada a Provas de Aptidão Pedagógica e Capacidade Científica), Faculdade de Letras da Universidade de Lisboa.
- Martins, Ana Maria (2006), *Para a história do vocalismo átono português: a propósito do Testamento de D. Afonso II de 1214*, *Revista Portuguesa de Filologia* 35:1, 295–319.
- Martins, Ana Maria (2007), *O primeiro século do português escrito*, in: Ana Boullón Agrelo (ed.), *Na Nosa Lyngoage Galega. A Emerxencia do Galego como Língua Escrita na Idade Media*, Santiago de Compostela, Consello da Cultura Galega/Instituto da Língua Galega, 161–184.
- Mateus, Maria Helena Mira (1975), *Aspectos da Fonologia Portuguesa*, Lisboa, Centro de Estudos Filológicos.

- Mateus, Maria Helena/Andrade, Ernesto d' (2000), *The Phonology of Portuguese*, Oxford, Oxford University Press.
- Mateus, Maria Helena Mira, et al. (2003), *Gramática da Língua Portuguesa*, Lisboa, Caminho.
- Nadeu, Marianna/Hualde, José Ignacio (2013), *Phonologization of unstressed vowel reduction patterns in Catalan compounds*, comunicação apresentada ao PaPI 2013, *Phonetics and Phonology in Iberia*, Book of abstracts, Lisboa, Universidade de Lisboa, 56–57.
- Nunes, José Joaquim (1956), *Compêndio de Gramática Histórica Portuguesa. Fonética e Morfologia*, Coimbra, Clássica Editora.
- Padgett, Jay/Tabain, Marija (2005), *Adaptive Dispersion Theory and Phonological Vowel Reduction in Russian*, *Phonetica* 62, 14–54.
- Paiva, Maria Helena (2008), *A descrição do vocalismo átono quinhentista: linhas e entrelinhas nos textos metalinguísticos coevos*, *Linguística. Revista de Estudos Linguísticos da Universidade do Porto* 3:1, 197–221.
- Paiva, Maria Helena (2009), *Variação e mudança no vocalismo átono quinhentista: práticas escriturais e juízos normativos*, *Linguística. Revista de Estudos Linguísticos da Universidade do Porto* 4:1, 85–110.
- Pamies Bertrán, Antonio (1999), *Prosodic Typology: On the Dichotomy between Stress-Timed and Syllable-Timed Languages*, *Language Design* 2, 103–130.
- Schane, Sanford A. (1984), *The Fundamentals of Particle Phonology*, *Phonology Yearbook* 1, 129–155.
- Scheer, Tobias (1998), *La structure interne des consonnes*, in: Patrick Sauzet (ed.), *Langues et Grammaire (II–III): Phonologie*, Saint Denis, Université de Paris 8, 141–172.
- Soares Barbosa, Jerónimo (1822), *Grammatica Philosophica da Língua Portuguesa ou Principios da Grammatica Geral Aplicados á Nossa Linguagem*, Lisboa, Academia Real das Sciencias.
- Teyssier, Paul (1990), *História da Língua Portuguesa*, Lisboa, Sá da Costa [trad.].
- Van Der Hulst, Harry (1989), *Atoms of Segmental Structure: Components, Gestures and Dependency*, *Phonology* 6:2, 253–284.
- Veloso, João (2007), *Schwa in European Portuguese: The Phonological Status of [i]*, in: Olivier Crouzet/Jean-Pierre Angoujard (edd.), *Actes des JEL'2007. Schwa(s). 5èmes Journées d'Études Linguistiques*, Nantes, Université de Nantes, 55–60.
- Veloso, João (2009), *Découpage de continuums phonétiques en mots: Critères formels vs. indices substantiels*, in: Olivier Crouzet/Alí Tifrit/Jean-Pierre Angoujard (edd.), *Actes des JEL'2009. 6èmes Journées d'Études Linguistiques*, Nantes, Université de Nantes, 85–90.
- Veloso, João (2010a), *Central, epenthetic, unmarked vowels and schwas: A brief outline of some essential differences*, *Linguística. Revista de Estudos Linguísticos da Universidade do Porto* 5:1, 193–213.
- Veloso, João (2010b), *Rimes /VGNS/ en position finale de mot en portugais: Une contrainte «sensible au mot»*, in: Maria Iliescu/Heidi M. Siller-Runggaldier/Paul Danler (edd.), *Actes du XXV^e Congrès International de Linguistique et de Philologie Romanes*, vol. 2, Berlin/Boston, De Gruyter, 231–240.
- Veloso, João (2012), *Vogais centrais do português europeu contemporâneo: Uma proposta de análise à luz da fonologia dos elementos*, *Letras de Hoje* 47:3, 234–243.
- Veloso, João (2013), *Redução do vocalismo átono do português europeu contemporâneo: Assimetria dos elementos de tonalidade e interação entre diversos tipos de redução vocálica*, in: Fátima Silva/Isabel Falé/Isabel Pereira (edd.), *Textos Seleccionados do XXVIII Encontro Nacional da Associação Portuguesa de Linguística*, Coimbra, Associação Portuguesa de Linguística, 655–672.
- Viana, Aniceto R. Gonçalves (1883), *Essai de phonétique et de phonologie de la langue portugaise d'après le dialecte actuel de Lisbonne*, *Romania* 12, 29–98.

Wetzels, W. Leo (1992), *Mid Vowel Neutralization in Brazilian Portuguese*, *Cadernos de Estudos Lingüísticos* 23, 19–55.

Wetzels, W. Leo (1995), *Mid-Vowel Alternations in the Brazilian Portuguese Verb*, *Phonology* 12:2, 281–304.

Wetzels, W. Leo (2011), *The Representation of Vowel Height and Vowel Height Neutralization in Brazilian Portuguese (Southern Dialects)*, in: John A. Goldsmith/Elizabeth V. Hume/W. Leo Wetzels (edd.), *Tones and Features*, Berlin/Boston, De Gruyter, 331–360.

Maria João Freitas

25 A sílaba na gramática do adulto e na aquisição de língua materna

Abstract: O objetivo central do presente capítulo é o de efetuar uma apresentação do funcionamento da unidade fonológica *sílaba* no português europeu padrão. Os dados da aquisição de língua materna são aqui usados como forma de testar as análises propostas para a gramática do adulto, funcionando como base empírica adicional para a reflexão sobre a natureza do conhecimento fonológico. São fornecidas inicialmente informações sobre o modelo teórico adotado e sobre os princípios universais que regem a organização dos segmentos dentro desta unidade prosódica. Posteriormente, é feita uma descrição geral dos padrões silábicos do sistema linguístico em foco e apresentada a sua ordem de aquisição. Finalmente, são discutidas algumas estruturas problemáticas do ponto de vista da sua estruturação silábica: os grupos consonânticos problemáticos; as sequências consonânticas de tipo *sC*; o estatuto das consoantes em final de sílaba; as consoantes velares labializadas; os ditongos crescentes. Nesta discussão, são evocados os dados da aquisição, descrevendo-se sumariamente o comportamento verbal das crianças portuguesas face às estruturas-alvo problemáticas sob avaliação.

Keywords: sílaba, fonologia, aquisição e desenvolvimento linguísticos, constituintes silábicos, interface sílaba/segmentos

1 Introdução

Contrariamente a outros constituintes fonológicos usados atualmente para descrever a estrutura sonora das línguas, a sílaba é uma unidade com história na tradição gramatical portuguesa, sendo este conceito usado na maior parte das gramáticas do português ao longo dos séculos. Veja-se o exemplo de João de Barros, que, na sua *Grammatica da Lingua Portuguesa* (1540), nos relata «O proveito de saber muitas sílabas (...). E o trabalho que se nestas levar será [de] gram proveito para os mininos, ca lhe fáz a língua tam solta e costumáda a ésta generalidade de sílabas que se nam empeça em a pronunçiaçám das dições» (Barros 1540/1971, 250). Logo no início da obra, é apresentada a seguinte definição do conceito: «Sílaba é ùa das quatro partes da nossa gramática que corresponde à Prosódia, que quer dizer acento e canto. A qual sílaba é ajuntamento de ùa vogal e duas e às vezes três consoantes que juntamente fazem ùa só voz» (Barros 1540/1971, 3–4). A identificação de padrões silábicos surge em algumas gramáticas do português. A título ilustrativo, Jerónimo Soares Barbosa (Soares Barbosa 1822) apresenta uma tipologia de classificação das sílabas quanto às possíveis combi-

natórias de *vozes* (i.e. vogais) e *consonancias* (i.e. consoantes); no entanto, é Viana (1883) quem primeiro disponibiliza uma descrição detalhada dos vários formatos silábicos possíveis no português. Ainda no contexto da tradição gramatical, e esquecendo as variações terminológicas, a referência à sílaba é feita estritamente através das dicotomias *aberta* (terminada em vogal) / *fechada* (terminada em consoante) e *tónica* (portadora de acento de palavra) / *átone* (não portadora de acento de palavra). A unidade é, ainda, usada para classificar as palavras quanto ao número de sílabas (*monossílabas, dissílabas, trissílabas, polissílabas*) (e.g. Cunha/Cintra 1984).

Na linguística do século XX, a sílaba é usada como unidade de descrição da estrutura sonora das línguas, embora o seu uso não seja muito produtivo. Na obra seminal de Jakobson (¹1941), e tendo na base as interações entre a tipologia de línguas e o processo de aquisição, a predição é a de que a sílaba CV¹ é a única possível no início da produção de fala nas crianças, seja qual for a língua em aquisição. No entanto, toda a reflexão sobre as estruturas sonoras, quer em contexto de aquisição quer em contexto de diagnóstico de afasia, é feita com base nos sons da fala e na sua representação fonológica. Numa perspetiva estruturalista de descrição gramatical sincrónica, e na sequência de trabalhos seus anteriores, Câmara Jr. (1970) tece várias reflexões sobre a relação entre sílabas fonéticas e a sua natureza fonológica. No contexto dos estudos sobre o português europeu, uma descrição exaustiva das possíveis estruturas silábicas do português é efetuada em Barbosa (1965), à luz do modelo funcionalista.

No âmbito da fonologia generativa linear, modelo proposto por Chomsky/Halle (1968), a formalização dos contextos nos quais as regras atuam inclui, quando relevante, a fronteira de sílaba (a título ilustrativo, a formalização da regra de velarização de /l/ inclui a notação de fronteira de sílaba na descrição contextual). No entanto, este modelo não considera a sílaba como uma unidade fonológica com estrutura interna. Sendo aquela a perspetiva teórica adotada por Mateus (1975) e Andrade (1977), estes trabalhos não exploram, em consequência, a estrutura da sílaba em português.

É com o advento da fonologia generativa multilinear, nos anos 1970/1980, e a promoção do trabalho sobre a prosódia das línguas que se intensifica a discussão sobre a estrutura interna da sílaba e a reflexão sobre as relações que esta estabelece com outras unidades fonológicas, discussão já iniciada, aliás, em propostas anteriores (Pike/Pike 1947; Anderson 1969; Fudge 1969). Neste contexto, assume-se que as unidades prosódicas se encontram hierarquicamente acima das unidades segmentais. A sílaba passa, assim, a ser definida como um constituinte interno à estrutura prosódica das línguas, num quadro em que a arquitetura do conhecimento fonológico integra:²

1 Os símbolos C, V e G são usados para representar, respetivamente, *consoante*, *vogal* e *semivogal* (ou *glide*). O ponto em CV. CV representa a fronteira de sílaba (como em [pá.tu] *pato*).

2 Para mais informação sobre o funcionamento destes constituintes no português, consulte-se Mateus et al. (2003) e Mateus/Falé/Freitas (2005).

- (i) nos níveis hierarquicamente superiores, o acento de palavra (representado com base no pé métrico), a palavra prosódica, o sintagma fonológico, o sintagma entoacional e o enunciado (Nespor/Vogel 1986);
- (ii) nos níveis hierarquicamente inferiores, os segmentos e os traços distintivos.

Embora nem todos os fonólogos assumam esta definição (Kahn 1976; Clements/Keyser 1983), a sílaba é hoje geralmente entendida como um objeto multidimensional, com uma estrutura interna composta por constituintes hierarquicamente organizados (Selkirk 1982). É responsável pelo primeiro nível de estruturação prosódica das línguas, organizando sequências de segmentos (consoantes, vogais, semivogais) em unidades melódicas intuitivamente identificadas pelos falantes. Esta identificação intuitiva das sílabas regista-se quer em crianças quer em sujeitos não alfabetizados. Veja-se o exemplo em (1), no qual uma criança com 2 anos de idade produz a palavra *chaminé*, identificando as suas fronteiras silábicas através do uso de pausas silenciosas (Freitas 1997):

- (1) Numa situação em que a Laura (2 anos) descreve um desenho:

Mãe: É uma...

Laura: É uma...

Mãe: cha...

Laura: miné

Mãe: Diz lá...

Laura: chaminé

Mãe: Outra vez...

Laura: cha//mi//né³

No exemplo em (2), a criança traduz erradamente uma sequência em inglês para uma outra em português; o paralelismo não é lexical, sendo a tradução feita através da identificação do mesmo número de sílabas nas duas sequências de fala (a inglesa e a portuguesa), que resulta no uso da mesma estrutura rítmica:

- (2) Num contexto em que o Manuel (6 anos) relata a sua aula de inglês à mãe:

Manuel: Mãe, hoje, lá na escola, aprendi a dizer coisas em Inglês!

Mãe: Ah, sim? Então, diz lá o que aprendeste...

Manuel: [wen//tu//tri] <one, two, three>!

Mãe: Muito bem! E o que é que isso quer dizer?

Manuel: [ke//la//du] <calados>!

Dos vários modelos de representação da estrutura interna da sílaba propostos ao longo da história da fonologia,⁴ os mais utilizados têm sido o modelo de *moras*, construído com base no conceito de peso silábico (Hyman 1985), e o modelo de *ataque-rima* (Selkirk 1982). Este último tem sido o mais adotado pelos investigadores

³ A notação // indica a produção de pausa silenciosa.

⁴ Para uma breve síntese, leia-se Blevins (1995).

portugueses e brasileiros que lidam com o conceito de sílaba (Collischonn ¹1996; Mateus/Andrade 2000). Assim, neste capítulo, assumiremos o modelo de *ataque-rima* adotado por Mateus/Andrade (2000) para a descrição do português,⁵ de acordo com o qual a sílaba é uma unidade fonológica com uma estrutura interna descritível em termos de padrões universais regulares, condicionados por princípios de boa formação silábica; os seus constituintes estabelecem relações hierárquicas entre si e com outras unidades fonológicas, sendo domínio de ativação de processos fonológicos.

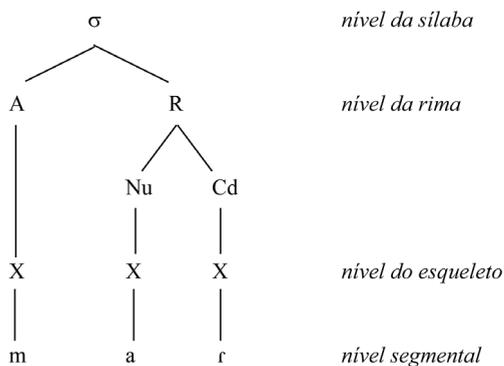
2 A estrutura interna da sílaba

2.1 Os constituintes silábicos

Na sequência do referido anteriormente, trataremos a sílaba como uma unidade com estrutura interna (Selkirk 1982; 1984), assumindo-se que a sílaba (σ) se encontra organizada em constituintes internos hierarquicamente organizados (cf. representação em (3)):

- (i) o *ataque* (A) domina a(s) consoante(s) na margem esquerda da sílaba;
- (ii) a *rima* (R) domina o *núcleo* (Nu), constituído por uma vogal ou por um ditongo, e a *coda* (Cd), que domina a(s) consoante(s) na margem direita da sílaba.

(3) Estrutura da sílaba no modelo de Ataque-Rima



À exceção da rima, todos os outros constituintes são terminais, dominando unidades segmentais. A existência do constituinte não terminal *rima*, que domina o núcleo e a coda, é baseada no funcionamento das línguas do mundo. Se pensarmos nas relações

⁵ Neste capítulo, *português* é usado como sinónimo de *português europeu padrão*. A referência a outras variantes ou variedades será explicitamente feita, sempre que necessário.

de adjacência entre constituintes terminais (*ataque, núcleo, coda*), o ataque e o núcleo apresentam maior autonomia de funcionamento entre si do que o núcleo e a coda. Veja-se o caso da nasalidade na história do português: as consoantes nasais em ataque, em geral, não transferiram a nasalidade para o núcleo (nu.me.ra.re → [nu].[mi].rar), contrariamente às consoantes nasais em coda, que, ao serem suprimidas, deixaram sistematicamente vestígio de si na vogal tautossilábica (can.ta.re → [kẽ].tar), com transferência de informação da coda para o núcleo.

Todos os constituintes terminais estão associados a posições rítmicas, representadas no nível do esqueleto – cf. (3) –, as quais dominam segmentos (consoantes, vogais ou semivogais). Estas posições rítmicas permitem, por exemplo, distinguir vogais breves (uma posição de esqueleto) de vogais longas (duas posições de esqueleto), em línguas como o inglês ou o holandês, em que a duração da vogal implica contraste fonológico.

Todos os constituintes podem assumir os formatos *não ramificado* (uma só posição de esqueleto associada ou a um só segmento ou a uma posição vazia) e *ramificado* (duas posições de esqueleto associadas a dois segmentos). No português, a tipologia de constituintes silábicos é a que se apresenta em (4):

(4) Tipologia de constituintes silábicos em português

		constituintes silábicos		exemplos
ataque	não ramificado	simples		<u>dá</u>
		vazio		<u>_é</u>
	ramificado			<u>cruz</u>
rima	não ramificada	núcleo	não ramificado	<u>pá</u>
			ramificado	<u>pai</u>
	ramificada	núcleo+coda		<u>paz</u>

O ataque não ramificado pode ser simples (associado a uma só consoante, como em dá) ou vazio (sem preenchimento segmental, como em _é); o ataque ramificado domina duas consoantes (como em cruz). A rima pode ser não ramificada, apresentando apenas o núcleo (como em pá), ou ramificada, com núcleo seguido de coda (como em paz). O núcleo, de preenchimento obrigatório, pode ser não ramificado (associado apenas a uma vogal) ou ramificado⁶ (dominando, normalmente, um ditongo). Como veremos na secção 2.1.2.2, a coda em português não ramifica, estando, portanto, associada a apenas um segmento (como em paz).

6 São também usados, na literatura, os termos *ataque complexo* e *núcleo complexo*.

2.1.1 O ataque

Embora seja pouco produtivo em algumas línguas (como acontece nas línguas do grupo Bantu), o ataque não ramificado, em português, pode ser vazio, tanto em posição inicial como em posição medial de palavra:

- (5) Ataque vazio
- | | |
|---------------|---------------|
| <i>_á.gua</i> | <i>lu_ a</i> |
| <i>_i.lha</i> | <i>di_ a</i> |
| <i>_o.vo</i> | <i>te_ or</i> |

Nesta língua, todas as consoantes podem preencher o ataque não ramificado simples, havendo, no entanto, restrições quanto ao formato de palavra, que proíbem [ʎ], [ɲ] e [r] no seu início. Os exemplos em (6), retirados de Mateus et al. (2003, 1039), ilustram a presença das consoantes do português em ataque simples, nas posições inicial e medial de palavra:

- (6) Ataque simples
- | | | | |
|-------------------|---------------------|--------------------------|--------------------|
| a. oclusivas | b. fricativas | c. líquidas ⁷ | d. nasais |
| <i>papa</i> [p-p] | <i>fofa</i> [f-f] | <i>lilás</i> [l-l] | <i>mamar</i> [m-m] |
| <i>baba</i> [b-b] | <i>vive</i> [v-v] | <i>talha</i> [-ʎ] | <i>ninar</i> [n-n] |
| <i>teta</i> [t-t] | <i>senso</i> [s-s] | <i>caro</i> [-r] | <i>senha</i> [-ɲ] |
| <i>dedo</i> [d-d] | <i>zozna</i> [z-z] | <i>carro</i> [-R] | |
| <i>coca</i> [k-k] | <i>chocho</i> [ʃ-ʃ] | <i>raro</i> [R-] | |
| <i>gago</i> [g-g] | <i>jejum</i> [ʒ-ʒ] | | |

Finalmente, os ataques ramificados são constituídos por seqüências de obstruinte seguida de líquida,⁸ observáveis em posições inicial e medial de palavra, embora nem todas as combinatórias destes dois tipos de segmentos estejam presentes no português. Em (7), ilustramos as combinatórias possíveis:

- (7) Ataques ramificados
- | | |
|----------------------|---------------------|
| a. oclusiva+vibrante | b. oclusiva+lateral |
| [pr] <i>primo</i> | [pl] <i>placa</i> |
| [br] <i>bruxa</i> | [bl] <i>blusa</i> |
| [tr] <i>traço</i> | [tl]* <i>atleta</i> |
| [dr] <i>drama</i> | [kl] <i>cloro</i> |
| [kr] <i>cruz</i> | [gl] <i>glosa</i> |
| [gr] <i>grama</i> | |

* Atestado apenas em posição medial

⁷ São usadas as transcrições dos róticos adotadas pela autora. Para mais informação sobre o funcionamento dos róticos em português europeu, consulte-se Veloso (2015).

⁸ Assumimos, aqui, a análise de Mateus/Andrade (2000). Para análise alternativa dos grupos constituídos por *obstruinte+líquida* (*obstruinte+vibrante* como tautossilábicas; *obstruinte+lateral* como heterossilábicas), consulte-se Veloso (2006).

c. fricativa+vibrante	d. fricativa+lateral
[fr] <i>fruta</i>	[fl] <i>fluvial</i>
[vr] <i>livro</i>	

Outras sequências consonânticas são identificáveis em português, em palavras como *observar*, *amnistia*, *oftalmológico*, *obter* ou *admirado*. Dada a sua natureza problemática, estas serão descritas na secção 3 deste capítulo.

2.1.2 A rima

Como referido, a rima pode ter um formato não ramificado, sendo constituída apenas pelo núcleo (ramificado ou não) ou pode ramificar em núcleo e coda, como nos exemplos em (8):

(8) Tipos de rima			
a. rima não ramificada	<i>pá</i>	[pá]	(com núcleo não ramificado)
	<i>pái</i>	[páj]	(com núcleo ramificado)
b. rima ramificada	<i>páz</i>	[páj]	(com núcleo não ramificado e coda)
	<i>país</i>	[páj]	(com núcleo ramificado e coda)

Contrariamente ao que ocorre entre o ataque e o núcleo, são várias as relações entre segmentos com função de núcleo e segmentos com função de coda, o que legitima, como referimos já, o uso do constituinte *rima* na representação da estrutura interna da sílaba. Citámos, a este propósito, o caso da nasalidade na história do português. Um outro exemplo de interação núcleo/coda ocorre no português do Brasil, com a semivocalização da lateral em coda (*sal* [sáw]), processo este que não ocorre em ataque (*lá* [lá]). O mesmo processo ocorre também na formação do plural de palavras terminadas em lateral, em ambas as variantes, brasileira e europeia: *anim*[á]/*anim*[májj] (cf. 76 Morfologia nas interfaces). Esta questão será retomada na secção 3, a propósito da observação dos dados da aquisição do português como língua materna, que permitem questionar a organização dos segmentos no domínio da rima.

2.1.2.1 O núcleo

De acordo com os princípios universais que regem a boa formação silábica (cf. secção 2.2), o núcleo é de preenchimento segmental obrigatório, definindo a identidade da sílaba (na secção 3, referiremos estruturas cujas natureza problemática levou alguns fonólogos a proporem a existência de um núcleo vazio). Como já mencionado, este constituinte pode assumir os formatos não ramificado e ramificado. As vogais (orais e nasais) que podem ocorrer em núcleo não ramificado estão ilustradas em (9):

(9) Vogais em núcleo não ramificado

[ɛ]	[e]	[i]	[a]	[ɐ]	[i]	[ɔ]	[o]	[u]	[ẽ]	[ē]	[ĩ]	[õ]	[ũ]
<u>sé</u>	<u>vê</u>	<u>vi</u>	<u>paz</u>	<u>da</u>	<u>de</u>	<u>só</u>	<u>dor</u>	<u>luz</u>	<u>lã</u>	<u>lente</u>	<u>sim</u>	<u>som</u>	<u>um</u>

De acordo com Mateus/Andrade (2000), todos os ditongos decrescentes são considerados núcleos ramificados em português, ocupando duas posições de esqueleto.⁹ Em (10), listam-se os ditongos decrescentes do português nas posições tónica, pós-tónica e pré-tónica (exemplos retirados de Mateus et al. 2003, 1044):

(10) Ditongos decrescentes

	a. tónicos	b. pré-tónicos	c. pós-tónicos
[ɛj]	<u>quei<u>xa</u></u>	<u>quei<u>xume</u></u>	<u>fá<u>ceis</u></u>
[ɛj]	<u>pap<u>éis</u></u>		
[aj]	<u>bai<u>rr</u>o</u>	<u>bai<u>rr</u>ista</u>	
[ɔj]	<u>r<u>ói</u></u>		
[oj]	<u>boi<u>ada</u></u>	<u>boi<u>ada</u></u>	
[uj]	<u>azi<u>uis</u></u>	<u>cui<u>dado</u></u>	
[iw]	<u>vi<u>u</u></u>		
[ew]	<u>de<u>us</u></u>	<u>ende<u>us</u>ar</u>	
[ɛw]	<u>v<u>eu</u></u>		
[aw]	<u>pau<u>ta</u></u>	<u>pau<u>ta</u>re</u>	
[ẽĩ]	<u>m<u>ãe</u></u>		<u>ba<u>tem</u></u>
[ẽw̃]	<u>m<u>ã</u>o</u>		<u>fa<u>lam</u></u>
[õĩ]	<u>p<u>õe</u></u>		

2.1.2.2 A coda

A coda não é de preenchimento obrigatório, impondo, em várias línguas do mundo, fortes restrições ao tipo de segmentos possíveis nesta posição. Várias línguas exibem codas ramificadas: vejam-se os casos das sequências consonânticas finais das palavras *past*, no inglês, e *feest*, no holandês. No entanto, no caso do português, este constituinte é não ramificado, dominando apenas uma consoante.¹⁰ No português, como em várias outras línguas, existe um contraste forte entre o inventário segmental possível em ataque (como vimos acima, se consideradas as posições inicial e medial

⁹ No entanto, Bisol (1999) propõe que os ditongos que exibem alternância VG/V, chamados ditongos leves, sejam considerados núcleos não ramificados (c[aj]xa/c[a]xa), dominando uma só posição de esqueleto, podendo esta estar associada a um ou dois segmentos (só a vogal; a vogal e a semivogal).

¹⁰ Palavras como *pers.pe.ti.va* e *sols.ti.cio* exibem, na primeira sílaba, duas consoantes à direita da vogal, ilustrando a presença de codas ramificadas no português. No entanto, estas estruturas são raras na língua, sendo, portanto, consideradas exceção ao padrão de coda não ramificada, típico do português.

de palavra, todas as consoantes do português são possíveis nesta posição) e o inventário que ocorre em coda, sendo este limitado a quatro segmentos fonéticos [ʃ, ʒ, r, ʎ]:

(11) Codas		
a. fricativas	b. lateral	c. vibrante
[ʃ] <i>feʃta</i>	[ʎ] <i>saʎto</i>	[r] <i>moʁno</i>
[ʒ] <i>muʒgo</i>		

Dado que as produções [ʃ]/[ʒ] decorrem da ativação do processo de assimilação do vozeamento, no âmbito do qual a fricativa assume o vozeamento da consoante adjacente à direita, estas são consideradas variantes alofônicas de uma única fricativa fonológica em coda, representada, de forma abstrata, através de /s/ ou de /S/, conforme os autores. Neste sentido, apesar de quatro segmentos fonéticos serem identificáveis no nível fonético, na representação fonológica, na verdade, apenas três segmentos fonológicos podem ocupar a posição de coda, a saber, /s, l, r/.¹¹ No caso das laterais, as formas [l] e [ʎ] são consideradas variantes alofônicas de /l/, uma vez que [l] ocorre sistematicamente em ataque e [ʎ], em coda.

A escassez de consoantes em coda no português levaram Morales-Front/Holt (1997) a discutir a tendência do português para a nuclearização, listando, para tal, argumentos que decorrem de factos diacrónicos e de factos sincrónicos. Com base em dados da aquisição de língua materna, retomaremos este assunto na secção 3, a propósito da natureza silábica das consoantes em final de sílaba (codas ou membros de núcleos ramificados).

A combinação dos vários tipos de constituintes leva à produção de um conjunto reduzido de padrões silábicos, sendo o mais frequente no português o que incorpora apenas constituintes não ramificados, a saber, o padrão CV, com valores a rondar os 50%, variando ligeiramente consoante os estudos consultados. Na tabela abaixo, identificamos os vários tipos silábicos listados na literatura e os valores de frequência por estudo citado:

¹¹ A coda nasal não foi incluída neste inventário de segmentos por se assumir a análise de Mateus/Andrade (2000) para o português europeu, segundo a qual a nasalidade no núcleo decorre de um autossegmento nasal associado ao nó rima, espreado-se no seu domínio aquando da produção. Outras representações da nasalidade na rima têm sido propostas na literatura, nomeadamente o arquifonema nasal (/N/) e a consoante /n/ (Barbosa 1965; Mateus 1975; Andrade 1977; para o português do Brasil, Wetzels 1997; Bisol 2005).

(12) Padrões silábicos (%)

	Andrade/Viana (1994)	Vigário/Falé (1994)	Vigário/Martins/Frota (2006)
CV	59,49	52,8	46,36
V	9,93	7,4	15,83
CVC	14,91	14,3 (Coda=r, s)	11,01
CVGN		3,2	5,62
CVN		6,0	5,37
VC	2,95	2,0 (C=s)	3,03
CVG	5,56	2,3	2,66
VN		2,0	2,64
CCV	4,79	1,9 (CC=pr)	2,18
VG	0,54		1,51
CVGC	0,71		1,21
Outros	<1%	<1%	<1%

2.2 A aquisição da estrutura silábica

Desde Jakobson (¹1941) que se assume que a estrutura silábica CV é a primeira a estar disponível no sistema linguístico da criança. Os vários estudos desenvolvidos para várias línguas do mundo demonstraram que assim é (Bernhardt/Stemberger 1998; Fikkert 2005). A presença precoce de sílabas do tipo CV nos primeiros enunciados infantis, a sua alta frequência nas línguas do mundo e o facto de haver línguas que usam quase exclusivamente este formato levam os linguistas a considerá-la o formato silábico universal. No entanto, tem sido registado um contraste entre línguas que apenas permitem o formato CV nas primeiras produções, como é o caso do inglês e do holandês, e línguas que permitem CV e V nas primeiras produções. Note-se que, no caso das duas primeiras línguas, as crianças produzem inicialmente uma consoante na posição de ataque vazio (V → CV), o que mostra a indisponibilidade do padrão V no estágio inicial de aquisição. Vejam-se os exemplos em (13), retirados de Fikkert (1994) e de Freitas (1997), respetivamente:

(13) Primeiros padrões silábicos

a. Holandês	<i>poes</i>	/pu:s/	→	[pu:]	(Jarmo, 1;5)
	<i>auto</i>	/o:to:/	→	[to:to:]	(Tom, 1;2)
b. Português	<i>quer</i>	/kɛr/	→	[kɛ]	(João, 1;0)
	<i>água</i>	/ag ^w e/	→	[abe]	(João, 0;11)

A exclusividade de CV e de V nas primeiras produções corresponde à presença apenas de constituintes não ramificados. O primeiro constituinte a ramificar na aquisição do português é a rima, com a produção de uma consoante em coda, o [ʃ]. Só muito mais tarde as crianças produzirão as codas [r] e [ʎ], sendo referido na literatura que esta última é a estrutura de aquisição mais tardia (Mendes et al. 2009; Amorim 2014).

Embora os ditongos de tipo VG possam estar presentes desde muito cedo, a sua estabilização é tardia, com oscilações frequentes entre VG e V – e.g. *já* [dáj] (João 1;7) e *pau* [pá] (João 1;6). O desaparecimento das oscilações VG/V é interpretado como o momento de disponibilização de núcleos ramificados no sistema da criança. Tem sido descrito que a estabilização dos ditongos ocorre paralelamente à estabilização das codas líquidas, o que levou a propor, na sequência de Fikkert (1994), que estas consoantes seriam processadas pelas crianças portuguesas como membros de núcleos ramificados e não como codas.¹² Este tópico carece, no entanto, de fundamentação empírica adicional, a partir de amostras de sujeitos mais alargadas.

Finalmente, como em várias línguas do mundo, os ataques ramificados são de aquisição tardia. Tal como a coda /l/, a sua aquisição pode não estar terminada à entrada na escola. A síntese relativa à aquisição da estrutura silábica em (14) é o produto das descrições disponíveis na literatura sobre a estabilização da estrutura silábica no português europeu (as duas colunas nos estádios 3 e 4 representam os dois percursos atestados até ao momento).

(14) Ordem de aquisição dos constituintes silábicos

Estádio 1	ataque não ramificado + rima não ramificada: CV / V	
Estádio 2	ataque não ramificado + rima ramificada: (C)VC _{fricativa}	
Estádio 3	núcleo ramificado: (C)VG / (C)VC _{líquida}	ataque ramificado: CCV
Estádio 4	ataque ramificado: CCV	núcleo ramificado: (C)VC _{líquida}

Como referido na introdução a este capítulo, um dos aspetos centrais nos modelos fonológicos não lineares é a possibilidade de representação das relações entre diferentes tipos de constituintes fonológicos. No caso dos dados da aquisição relativos à relação entre segmentos e sílabas, estudos para várias línguas têm mostrado correlações fortes entre produção segmental e disponibilização de constituintes fonológicos no sistema linguístico infantil (entre outros, Fikkert 1994; Freitas 1997; Bernhardt/Stemberger 1998; Amorim 2014). Por outras palavras, um mesmo segmento pode ser produzido pela criança em ataque não ramificado, presente no estágio 1, mas ainda não ser produzido noutras posições silábicas, pelo facto de os constituintes ainda não estarem presentes no sistema da criança. Veja-se o caso do Luís (1;9–2;2), que produz

¹² Para mais informação, consulte-se a secção 3 deste capítulo.

[r] em ataque não ramificado (*amarelo* [mɛrɛ'w], 1;9) mas ainda não o produz nem em coda nem em ataque ramificado (*barco* [báku], 1;9, e *frutas* [fúte]], 2;2).¹³

A relação entre desenvolvimento silábico e aquisição do inventário de segmentos é crucial para o diagnóstico e intervenção clínicas. No caso do diagnóstico, a maior parte dos instrumentos de avaliação não distingue os diferentes contextos silábicos, avaliando preferencialmente o segmento em ataque não ramificado, o que pode levar a taxas de sucesso superiores às reais. No caso da planificação da intervenção, uma criança com todos os [r] em ataque não ramificado mas sem [r] em coda e em ataque ramificado precisa de ser estimulada com alvos dos segundo e terceiro tipos (*porta*; *praia*) e não com alvos do primeiro tipo, já adquirido. É, portanto, crucial que os terapeutas da fala considerem, na sua prática clínica, a relação entre constituintes silábicos e segmentos, sob pena de a intervenção ser menos eficaz.

Em contexto escolar, embora a maior parte dos trabalhos relativos ao português não se centre especificamente no impacto da estrutura silábica sobre os desempenhos ortográficos infantis, vários estudos têm referido que estruturas silábicas complexas motivam a produção de taxas mais elevadas de erros ortográficos (Miranda/Matzenauer 2010; Santos 2013). Neste sentido, a formação inicial de professores de 1.º Ciclo deve incorporar, entre outros aspetos fonológicos, uma reflexão sistemática sobre o efeito da estrutura silábica nos desempenhos ortográficos infantis, no sentido de promover uma intervenção didática eficaz na superação dos problemas ortográficos de base silábica apresentados pelas crianças portuguesas. Vejam-se os exemplos em (15), retirados de Santos (2013):

(15) Erros ortográficos produzidos por crianças portuguesas (Santos 2013)

1º ano de escolaridade	<i>bruxa</i>	*<berucha>	(criança 1)
	<i>globo</i>	*<gulobu>	(criança 6)
	<i>pneu</i>	*<paneo>	(criança 2)
	<i>afta</i>	*<afata>	(criança 22)
4º ano de escolaridade	<i>atlântico</i>	*<atalântico>	(criança 20)
	<i>astrónomo</i>	*<asterónumo>	(criança 11)
	<i>Neptuno</i>	*<nepetuno>	(criança 27)
	<i>pictograma</i>	*<picatograma>	(criança 10)

Embora a epêntese vocálica não seja a única estratégia identificada nos erros ortográficos de estruturas problemáticas (neste caso, ataques ramificados e grupos consonânticos problemáticos), elas espelham um comportamento verbal oral muito usado pelas crianças portuguesas na aquisição destas estruturas (Freitas 1997; Santos 2013). Vejam-se os exemplos em (16):

¹³ As produções infantis neste parágrafo foram retiradas de Freitas (1997).

(16) Produções orais de crianças no 1.º ano de escolaridade (Santos 2013)

<i>globo</i>	→	[gilóbu]	(criança 5)
<i>pneu</i>	→	[pinéw]	(criança 14)
		[punéw]	(criança 22)

A produtividade da epêntese vocálica nas crianças portuguesas, por oposição a crianças falantes de outras línguas, pode decorrer das diferenças entre representações fonológicas correspondentes a estruturas silábicas simples e produção fonética de palavras com taxas elevadas de supressão vocálica – e.g. *telefone* [tlfõn], *desprevenir* [dʃprvnr].¹⁴ Esta assimetria pode estar na base do comportamento epentético das crianças portuguesas. Note-se que, na aquisição dos ataques ramificados, para além da redução do grupo consonântico em ataque ramificado – $C_1C_2 \rightarrow C_1\emptyset$, como em *preta* [péta] (Inês, 1;10) –, muito frequente na aquisição das línguas do mundo, as crianças portuguesas usam produtivamente a epêntese vocálica – e.g. *bruxa* [birúʃe] (Luís, 2;11) –, pouco frequente na aquisição de outros sistemas linguísticos (Bernhardt/Stemberger 1998).¹⁵

2.3 Princípios de organização dos segmentos na sílaba

O trabalho sobre tipologia de línguas desenvolvido ao longo do século XX permitiu observar que nem todas as sequências de segmentos são possíveis no domínio de uma sílaba. As relações de vizinhança entre os segmentos na sílaba são regidas por princípios universais de boa formação silábica, de que daremos conta nesta secção.

Todas as estruturas silábicas apresentadas em 2.1 (exemplos (5) a (11)) são consideradas regulares por obedecerem aos princípios universais que regem a estrutura interna da unidade em foco. De todas, a mais frequente é a sílaba com o formato CV, que está presente em todas as línguas do mundo e que condiciona o formato das primeiras produções infantis – veja-se o exemplo da produção [péte] para *preta* (Inês, 1;10). Por estas razões, CV é o formato silábico normalmente designado como universal. Com base neste facto, o Princípio do Ataque Máximo assume que «o preenchimento dos Ataques é preferível ao preenchimento das codas». De acordo com este princípio, a sequência VCV é preferencialmente silabificada como V.CV, sendo C um ataque, e não como *VC.V, estrutura na qual C seria uma coda. Deste modo, numa palavra como *ovo*, a silabificação prevista pelo Princípio do Ataque Máximo é [ó.vu] e não *[óv.u].

Os segmentos dentro da sílaba estão organizados em função de padrões melódicos que facilitam a identificação intuitiva das fronteiras silábicas. Esta organização

¹⁴ Exemplos retirados de Mateus/Andrade (2000, 44).

¹⁵ As produções infantis neste parágrafo foram retiradas de Freitas (1997).

decorre do grau intrínseco de sonoridade dos segmentos, organizados do seguinte modo na escala crescente de sonoridade abaixo apresentada:

(17) Escala de Sonoridade

grau mínimo oclusivas < fricativas < nasais < líquidas < semivogais < vogais grau máximo

O Princípio de Sonoridade trabalha com base na escala de sonoridade e dá conta da organização melódica dos segmentos dentro da sílaba, postulando que «a sonoridade dos segmentos cresce da margem esquerda até ao núcleo e decresce do núcleo até à margem direita» (Selkirk 1984). A título ilustrativo, na sílaba inicial [plaʃ] de *plástico*, o som [p], oclusiva, tem grau de sonoridade inferior ao [l], líquida, que por sua vez tem grau de sonoridade inferior a [a]; aqui termina a subida de sonoridade da margem esquerda da sílaba até ao seu núcleo; o som [ʃ], fricativa, tem grau de sonoridade inferior a [a], registando-se uma descida de grau de sonoridade do núcleo até à margem direita da sílaba.

As línguas gostam de contrastes máximos de sonoridade entre os segmentos adjacentes dentro da sílaba. O contraste na estrutura $C_{\text{oclusiva}}V$ (como em *pá*), em que cada segmento provém de pontos opostos na escala de sonoridade, dá origem ao padrão mais natural nas línguas do mundo (Jakobson ¹1941). Por outro lado, a maior frequência de sequências do tipo *oclusiva+líquida* em ataque ramificado (como em *gra.vo*), quando comparadas com as sequências do tipo *fricativa+líquida* (como em *li.vro*), decorre desta preferência por contrastes de sonoridade elevados dentro da sílaba. Este aspeto é captado pelo Princípio de Dissemelhança,¹⁶ de acordo com o qual «quanto maior for a distância de sonoridade entre dois segmentos adjacentes dentro da sílaba, maior é a sua naturalidade» (Harris 1983; Selkirk 1984).

Por fim, refira-se o Princípio da Binaridade Máxima dos Constituintes, ativo na representação de vários constituintes linguísticos em diferentes módulos gramaticais, segundo o qual «os constituintes linguísticos são maximamente binários» (Kayne 1984). O modelo de *ataque-rima* (cf. representação em (3)) é construído tendo na base este princípio.

Os vários princípios universais de boa formação silábica enunciados acima serão usados na secção que se segue para refletir sobre algumas estruturas silábicas problemáticas no português, tanto em termos de identificação de fronteiras silábicas como de papéis silábicos.

¹⁶ Em alguns textos, é possível encontrar as designações alternativas Princípio de Dissimilaridade ou Condição de Dissemelhança.

3 Estruturas problemáticas e contributos dos dados da aquisição

Nesta secção, debruçar-nos-emos sobre algumas estruturas silábicas problemáticas em português, que têm suscitado, em alguns casos, diferentes propostas de análise por parte dos fonólogos. Apresentaremos a(s) análise(s) proposta(s) e adicionaremos informação sobre o que sabemos relativamente à sua aquisição em contexto de língua materna. Esta escolha decorre do facto de os dados da aquisição serem terreno adequado à testagem das análises construídas para a gramática do adulto, fornecendo evidência empírica para a reflexão sobre a natureza do conhecimento linguístico (Chomsky 1986).

3.1 Grupos consonânticos problemáticos

Os grupos consonânticos listados em (7), na secção 2, são considerados ataques ramificados, constituídos por sequências de uma obstruinte (oclusiva ou fricativa) seguida de uma líquida (vibrante ou lateral) e ocorrem na periferia esquerda da sílaba, respeitando, assim, o Princípio da Binaridade Máxima dos Constituintes. O grau de sonoridade das duas consoantes é distinto e crescente, como previsto, respetivamente, pelo Princípio de Dissemelhança e pelo Princípio de Sonoridade. Contrariamente a estas sequências, também designadas como *grupos consonânticos próprios*, as sequências listadas em (18), os chamados *grupos consonânticos problemáticos*, violam os segundo e terceiro princípios acima listados (Princípio de Dissemelhança e Princípio de Sonoridade), pelo que não são considerados ataques ramificados (Mateus/Andrade 2000):

(18) Grupos consonânticos problemáticos

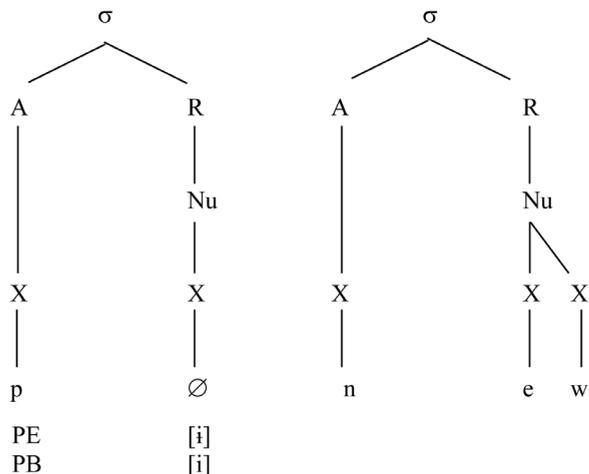
- | | |
|-----------------------|-----------------|
| a. oclusiva+oclusiva | <i>optar</i> |
| b. oclusiva+nasal | <i>pneu</i> |
| c. oclusiva+fricativa | <i>absurdo</i> |
| d. fricativa+oclusiva | <i>afta</i> |
| e. nasal+nasal | <i>amnistia</i> |

Como referido em Mateus/Andrade (2000), todos estes grupos problemáticos permitem a epêntese vocálica no português europeu padrão [pinéw], sendo a epêntese obrigatória no português do Brasil (*p[i]neu*). Esta epêntese vocálica não acontece com os ataques ramificados nas produções dos adultos (**p[i]rato* no português europeu; **p[i]rato* no português do Brasil), o que argumenta a favor da natureza silabicamente distinta dos dois tipos de grupos consonânticos.¹⁷ Mateus/Andrade (2000) propõem, assim, que as duas consoantes dos grupos problemáticos sejam consideradas como

¹⁷ Para mais argumentos, consulte-se Mateus/Andrade (2000, 44–45).

dois ataques simples heterossilábicos, sendo a primeira ataque de uma sílaba com núcleo vazio e a segunda ataque de uma sílaba com núcleo preenchido:

(19) Natureza silábica dos grupos consonânticos problemáticos



A natureza problemática destes grupos consonânticos, espelhada numa análise que assume a existência de núcleos vazios, não canónica no âmbito do modelo de ataque-rima, é confirmada pelos dados da aquisição. Santos (2013) explora a hipótese formulada em Freitas (1997) segundo a qual os ataques ramificados seriam adquiridos antes dos grupos consonânticos problemáticos. Os dados da oralidade confirmaram a hipótese, sendo a ordem atestada a seguinte: *ataques ramificados* >> *grupos consonânticos problemáticos*. Os dados de escrita avaliados espelham igualmente este facto, com mais problemas ortográficos no registo dos grupos consonânticos problemáticos do que no dos ataques ramificados – *pneu* *<paneio> (criança 2, 1.º ano); *Neptuno* *<nepetuno> (criança 27, 4.º ano) –, sendo clara a preferência pela epêntese vocálica entre as duas consoantes do grupo.

3.2 Grupos sC

Um outro tipo de grupo consonântico que viola princípios de boa formação silábica consiste na sequência de fricativa seguida de uma ou duas consoantes em início de palavra; estes são designados, na literatura, como grupos sC:¹⁸

¹⁸ Para mais informação sobre esta estrutura, consulte-se Andrade/Rodrigues (1999), Mateus/Andrade (2000), Veloso (2002) e Freitas/Rodrigues (2003).

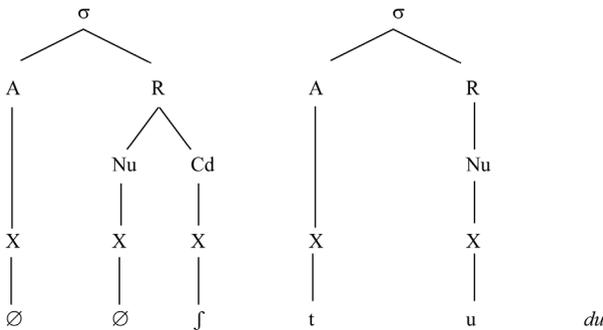
(20) Grupos sC

a. sequências sC	b. sequências sCC
[ft] <i>est<u>u</u>do</i>	[ftr] <i>est<u>ra</u>da</i>
[fk] <i>es<u>co</u>la</i>	[fkr] <i>es<u>cr</u>avo</i>
[ff] <i>es<u>f</u>inge</i>	[ffr] <i>es<u>f</u>rega</i>
[ʒg] <i>es<u>g</u>ana</i>	[ʒgr] <i>es<u>gr</u>ima</i>
[ʒm] <i>es<u>m</u>aga</i>	[ʒdr] <i>es<u>d</u>rxula</i>

A hipótese de estes grupos constituírem ataques ramificados é questionável em função da violação de princípios universais: (i) a consoante na margem esquerda é uma fricativa, seguida de uma oclusiva ou de uma fricativa, o que viola o Princípio de Sonoridade e o Princípio de Dissemelhança; (ii) nos casos de (20b), identificamos três consoantes, o que viola o Princípio da Binaridade Máxima dos Constituintes.

A fricativa dos grupos sC em início de palavra é sempre [ʃ] ou [ʒ], como acontece nos exemplos das codas fricativas em (11a), *fe[ʃ]ta* e *mu[ʒ]go*. Esta variação alofónica decorre, como então referimos, da assimilação do vozeamento presente na consoante adjacente à direita. Nos casos dos grupos sC, o mesmo se verifica: identificamos um [ʃ] inicial sempre que a consoante à direita é não vozeada e um [ʒ] sempre que a consoante à direita é vozeada. Este argumento leva a que se considere que a fricativa e a(s) consoante(s) que se lhe segue(m) sejam heterossilábicas: a fricativa inicial é coda de uma sílaba inicial com núcleo vazio; a(s) consoante(s) seguinte(s) constituem o ataque (simples ou ramificado) da sílaba adjacente à direita (Andrade/Rodrigues 1999; Mateus/Andrade 2000):

(21) Natureza silábica dos grupos sC



Esta análise permite dar conta da esporádica presença de [i] inicial em algumas produções artificiais no português europeu padrão ([i]scola),¹⁹ hospedando o núcleo vazio a vogal epentética produzida no início de palavra. Os dados da aquisição do português fornecem evidência empírica a favor da análise representada em (21):

- (i) a produção da fricativa inicial surge quando estão a ser adquiridas as codas fricativas (Freitas/Rodrigues 2003);
- (ii) num primeiro estágio da aquisição destes grupos, as crianças começam por apagar a fricativa inicial, produzindo apenas a obstruinte ($C_1C_2 \rightarrow \emptyset C_2$), como em *escova* [øtúke] (Marta, 1;2); se se tratasse de um ataque ramificado, esperar-se-ia que as crianças comessem por preservar a fricativa, a consoante mais à esquerda, apagando a obstruinte, à direita; é o que acontece nas sequências de *obstruinte+líquida* em ataque ramificado, e.g. *preta* [pøéte] (Inês, 1;10), em que a obstruinte é preservada e a líquida é suprimida ($C_1C_2 \rightarrow C_1\emptyset$);
- (iii) no segundo estágio, regista-se a produção de vogal inicial (normalmente, [i]), com ou sem fricativa, e.g. *escreve* [iké'v] (Marta, 1;11), *estrelinha* [iʃtírĩɲe] (Marta, 2;0).

Os dados acima descritos podem ser interpretados como decorrentes do facto de a fricativa estar a ser processada pelas crianças como coda e não como C_1 de um ataque ramificado. A produção de [i], vogal preenchedora de categorias prosódicas vazias no português europeu padrão, pode ser interpretada como decorrente da necessidade de preenchimento do núcleo vazio em fases iniciais da aquisição.

Os grupos *sC* em início de palavra são analisados de formas distintas em diferentes línguas. Fikkert/Freitas (2004) compararam a aquisição desta estrutura por crianças holandesas e por crianças portuguesas: no holandês, o grupo *sC* é adquirido tardiamente, sendo das últimas estruturas silábicas a estabilizar, já depois das codas fricativas e dos ataques ramificados; contrariamente, no português, a aquisição de *sC* é precoce, ocorrendo aquando da aquisição das codas fricativas e antes da aquisição dos ataques ramificados. Estes resultados são consistentes com as diferentes análises adotadas nas duas línguas: no holandês, a fricativa do grupo *sC* é considerada extrassilábica, sendo estas estruturas excepcionais e, como tal, adquiridas muito tardiamente; no português, a fricativa é uma coda, estrutura que emerge cedo nesta língua, normalmente antes dos 2 anos.

¹⁹ Com base em Rodrigues (2004).

3.3 Líquidas em final de sílaba

Afirmámos, na secção 2.1.2.2, que o constituinte coda impõe fortes restrições ao tipo de unidades segmentais que domina. No português, apenas /s, l, r/ assumem este papel silábico. No entanto, a sua distribuição é distinta no português:

- (i) /s/ coocorre com núcleos ramificados (VGC_{fricativa}, como em *fausto*), o que não acontece com as líquidas, que apenas coocorrem com núcleos não ramificados (VC_{líquida}, como em *parto* e *palco*);
- (ii) em final de palavra, /l, r/ finais permitem inserção vocálica (*mar[i]*; *anel[i]*), o que não acontece com /s/ (**bolos[i]*).

Alguns autores têm refletido sobre a natureza das líquidas em final de sílaba e sobre as diferenças entre /l/ e /r/. No caso particular do /l/, Andrade (1998) relembra as semelhanças fonéticas entre laterais e semivogais, mostrando interferência da lateral velarizada na produção da vogal tautossilábica, adjacente à esquerda. Esta interferência tem impacto fonológico no processo de redução vocálica, bloqueando-o: uma vogal átona com /l/ à direita não eleva, à imagem do que acontece com uma vogal com semivogal à direita – *m[aɫ]dade* e não **m[ɐɫ]dade*; *c[aj]xinha* e não **c[ɛj]xinha*. Tal poderia levar-nos a pensar que a lateral e a semivogal têm um estatuto silábico semelhante, sendo ambas membros de um núcleo ramificado.

Em termos de análises silábicas presentes na literatura, Miguel (1993) propõe a representação de /l/ em final de sílaba no domínio de um núcleo ramificado. A velarização da lateral no português europeu (*sa[ɫ]*) e a sua semivocalização no português do Brasil (*sa[w]*) têm sido interpretadas como diferentes manifestações da tendência do português para a nuclearização (Girelli 1988; Morales-Front/Holt 1997). Note-se, como referido acima, que o caso da nasalidade em coda no latim e a sua transferência para a vogal tautossilábica tem também sido usado como argumento para a nuclearização (*CANTARE* → *c[ẽ]tar*). Um outro argumento no mesmo sentido é o do plural das palavras terminadas em lateral, dado que, com a adjunção do marcador de número /+s/, uma coda fricativa, a lateral final do radical semivocaliza, integrando o núcleo (*anima[ɫ]*, *anima[j]s*).

Alguns dos comportamentos registados para /l/ não se verificam com /r/ em final de sílaba. A título ilustrativo, /r/, contrariamente a /l/, não bloqueia a redução vocálica (*p[ɐr]tida*; *m[aɫ]dade*). Por outro lado, no português do Brasil (Bisol 2005; Mateus/Andrade 2000), enquanto /l/ se aproxima das vogais do núcleo, semivocalizando, o /r/ pode até fricativizar, tornando-se, portanto, mais próximo das codas fricativas (a produção como fricativa glotal – e.g. *par* como [páh] – pode também indiciar a tendência para a sua supressão).

Nos dados da aquisição do português, existe um forte contraste entre a aquisição das fricativas e a das líquidas em final de sílaba (*fricativas* >> *líquidas*), sendo que o intervalo de tempo entre a emergência das primeiras e a das segundas pode ser superior a 1 ano (as primeiras codas fricativas podem emergir por volta dos 2 anos,

normalmente em final de palavra – Freitas 1997). Os dados de Mendes et al. (2009) apresentam, como faixas etárias para a aquisição destas consoantes, as seguintes:

(22) Aquisição das consoantes em final de sílaba

estrutura	faixa etária
a. fricativa /s/ em final de sílaba	3;6–3;12
b. vibrante /r/ em final de sílaba	4;6–4;12
c. lateral /l/ em final de sílaba	5;0–5;6

O intervalo temporal entre a aquisição das codas fricativas e a das líquidas em final de sílaba é inexplicável se considerarmos que, numa dada faixa etária, por exemplo, aos 4 anos, a criança (1) já tem o constituinte *coda* disponível e (2) já produz os segmentos /l/ e /r/ em ataque simples (o constituinte e os segmentos já estão, portanto, disponíveis no sistema) mas (3) ainda não produz estes segmentos em coda (Freitas 1997; Mendes et al. 2009; Amorim 2014). Com base neste perfil, a questão que se coloca é a de saber se as crianças portuguesas estão ou não a processar as líquidas em final de sílaba como codas ou como membros de um núcleo ramificado. Os dados de aquisição do português europeu revelam um paralelismo entre a estabilização dos núcleos ramificados e a das líquidas em final de sílaba (Freitas 1997). Na sequência de Fikkert (1994), o comportamento das crianças portuguesas foi interpretado como decorrendo do processamento das líquidas em final de sílaba como membros de um núcleo ramificado e não como codas.

Os dados acima descritos podem ser interpretados como fornecendo evidência empírica para a proposta de representação de /l/ em final de sílaba no núcleo. Ficam, no entanto, por explicar as assimetrias entre /r/ e /l/, tanto na gramática do adulto como nos dados da aquisição.

3.4 Velares labializadas e ditongos crescentes

Andrade/Viana (1994), na sequência de Bisol (1989), observam a diferença entre os dois paradigmas listados em (23), correspondentes a aparentes ditongos crescentes, em estruturas CGV:

(23) Aparentes sequências CGV

- a. *q*[w]arto **q*[u]arto
q[w]adrado **q*[u]adrado
g[w]arda **g*[u]arda
ág[w]a **ág*[u]a

b.	<i>t[w]alha</i>	<i>t[u]alha</i>
	<i>c[w]elho</i>	<i>c[u]elho</i>
	<i>p[j]ano</i>	<i>p[i]ano</i>
	<i>qu[j]abo</i>	<i>qu[i]abo</i>

Sumariamente, podemos dizer que as estruturas em (23a) se distinguem das de (23b) por:

- (i) não permitirem o hiato: a alternância GV/VV apenas é atestada nos casos em (23b);
- (ii) em (23a), apenas as consoantes velares /k, g/ ocorrerem à esquerda da aparente semivogal e esta apenas apresentar o formato [w];
- (iii) em (23b), todas as consoantes serem possíveis à esquerda de semivogal e esta assumir os formatos [w] ou [j].

Estes factos permitem argumentar a favor da natureza distinta das duas estruturas: em (23a), os autores assumem tratar-se de consoantes velares labializadas (/k^w/ e /g^w/), segmentos complexos e, como tal, representados num ataque não ramificado; em (23b), é proposta a presença de um hiato lexical, que legitima produções do tipo VV (*p[i]ano*; *c[u]elho*), que podem alternar com GV (*p[j]ano*; *c[w]elho*) nos enunciados de fala. Neste último caso, a semivocalização da vogal alta decorre da tendência das línguas para a resolução de hiatos e é aceite tradicionalmente como o formato fonético mais frequente nos enunciados orais.

A questão que se coloca na aquisição é a de saber se estruturas aparentemente semelhantes (*q[w]arto* / *c[w]elho*) são ou não adquiridas do mesmo modo pelas crianças portuguesas ou se estas são precocemente sensíveis aos factos distribucionais acima listados.

Registam-se, em (24) e em (25), alguns exemplos de produções das crianças portuguesas observadas em Freitas (1997):

(24) Produções de alvos do tipo *q[w]arto*

<i>água</i>	[ábe] / [áwe]	(Inês, 1;9)
<i>água</i>	[ágo]	(Marta, 1;7)
<i>quarto</i>	[kátu]	(Inês, 1;8)
<i>quarto</i>	[átu]	(João, 2;1)

(25) Produções de alvos do tipo *c[w]elho*

<i>cuecas</i>	[kié'keʃ]	(Marta, 1;6)
<i>o coelho</i>	[ubukéj]	(Marta, 1;6)
<i>viola</i>	[fi'we]	(Luís, 2;5)
<i>coelho</i>	[kwélu]	(Laura, 2;3)
<i>joelho</i>	[zuelu]	(Laura, 2;3)

Os dados acima listados ilustram as estratégias mais usadas em cada um dos dois paradigmas sob observação. Note-se desde já que, em (24), e contrariamente a (25), não se regista a alternância GV/VV.

As estratégias mais comuns em (24) são a supressão de [w] ou a supressão de C ou, ainda, a de ambas (CGV → C∅V; CGV → ∅GV; CGV → ∅∅V). Note-se, ainda, o uso da coalescência – *água* [ágo] (Marta, 1;7) e *guardados* [godáduj] (Laura, 2;6), com projeção do arredondamento do [w] na vogal tautossilábica. Este facto é usado por Clements (1991) como argumento a favor da presença de segmentos complexos (neste caso, /k^w/ e /g^w/), tal como proposto por Bisol (1989) e por Andrade/Viana (1994) para o paradigma em (23a). O comportamento das crianças portuguesas, ilustrado em (24), argumenta, assim, a favor da presença de consoantes velares labializadas (/k^w/ e /g^w/) no português.

Em (25), existe uma tendência inicial para a preservação de VV, contrariamente ao que se regista na fala dos adultos, por vezes até com produção de material segmental epentético a preencher a posição de ataque vazio – *coelho* [ubukéj] (Marta, 1;6). A alternância VV/GV é possível (vejam-se as produções da Laura para *coelho* e *joelho*, na mesma sessão), ocorrendo esporadicamente a substituição da semivogal alvo – *cecas* [kié'kej] (Marta, 1;6).

Os dados acima sumariados argumentam a favor da análise proposta para o português por Bisol (1989) e por Andrade/Viana (1994), mostrando que as crianças são precocemente sensíveis às propriedades fonológicas da língua materna (cf. contraste entre os paradigmas em (23a) e em (23b)).

4 Considerações finais

Neste capítulo, começámos por fazer um breve apontamento sobre o uso do conceito de sílaba tanto nas gramáticas do português como nos modelos fonológicos disponibilizados ao longo do século XX. Detivemo-nos na apresentação do modelo de *ataque-rima* (Selkirk 1982; 1984), adotado por Mateus/Andrade (2000) para o estudo da sílaba no português europeu. Procedemos, em seguida, à descrição da estrutura silábica deste sistema linguístico com base nas propostas destes autores, após o que fornecemos informação sobre o desenvolvimento silábico infantil nas crianças portuguesas. Foram, posteriormente, apresentados alguns princípios universais de boa formação silábica, com o objetivo de discutir estruturas silábicas problemáticas: (i) os grupos consonânticos problemáticos; (ii) as sequências consonânticas de tipo sC; (iii) o estatuto das consoantes em final de sílaba; (iv) as consoantes velares labializadas e os ditongos crescentes. Paralelamente, e em cada estrutura problemática tratada, foram fornecidos dados da sua aquisição, no sentido de testar as análises fonológicas propostas para a gramática do adulto.

5 Referências

- Amorim, Clara (2014), *Padrão de aquisição de contrastes do PE: a interação entre traços, segmentos e sílabas*, tese de Doutorado, Universidade do Porto.
- Anderson, John (1969), *Syllabic or non-syllabic phonology*, *Journal of Linguistics* 5, 136–143.
- Andrade, Amália (1998), *Varição fonética do /l/ em ataque silábico em Português Europeu*, in: Antónia Mota/Rita Marquilha (edd.), *Actas do XIII Encontro Nacional da APL*, Lisboa, Colibri/APL, 55–76.
- Andrade, Ernesto d' (1977), *Aspects de la Phonologie (Générative) du Portugais*, Lisboa, INIC.
- Andrade, Ernesto d'/Rodrigues, Celeste (1999), *Das escolas e das culturas: história de uma sequência consonântica*, in: Ana Lopes/Cristina Martins (edd.), *Actas do XIV Encontro da Associação Portuguesa de Linguística*, Lisboa, APL/Colibri, 117–133.
- Andrade, Ernesto d'/Viana, Céu (1994), *Sinérese, diérese e estrutura silábica*, in: *Actas do IX Encontro Nacional da Associação Portuguesa de Linguística*, Lisboa, APL, 31–42.
- Barbosa, José Morais (1965), *Études de phonologie portugaise*, Lisboa, Junta de Investigações do Ultramar.
- Barros, João de (1971, ¹1540), *Grammatica da Língua Portuguesa com os Mandamentos da Santa Madre Igreja*, Lisboa, Faculdade de Letras, Universidade de Lisboa.
- Bernhardt, Barbara/Stemberger, Joseph (1998), *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*, San Diego, Academic Press.
- Bisol, Leda (1989), *O ditongo na perspectiva da fonologia atual*, *DELTA* 5:2, 168–185.
- Bisol, Leda (1999), *A sílaba e os seus constituintes*, in: Maria Helena Moura Neves (ed.), *Gramática do Português Falado*, vol. 7, São Paulo, Humanitas, 701–742.
- Bisol, Leda (2005, ¹1996), *Introdução a estudos de fonologia do português brasileiro*, Porto Alegre, EDIPUCRS.
- Blevins, Juliette (1995), *The syllable in phonological theory*, in: John Goldsmith (ed.), *The Handbook of Phonological Theory*, Cambridge MA, Blackwell, 206–244.
- Câmara Jr., Joaquim (1970), *Estrutura da Língua Portuguesa*, Rio de Janeiro, Editora Vozes.
- Chomsky, Noam (1986), *Knowledge of Language. Its Nature, Origin and Use*, New York, Praeger.
- Chomsky, Noam/Halle, Morris (1968), *The Sound Pattern of English*, New York, Harper & Row.
- Clements, Nick (1991), *Place of articulation in consonants and vowels: a unified theory*, *Working Papers of the Cornell Phonetics Laboratory* 5, 77–123.
- Clements, Nick/Keyser, Jay (1983), *CV Phonology: A Generative Theory of the Syllable*, Cambridge MA, MIT Press.
- Collischonn, Gisela (2005, ¹1996), *A sílaba em Português*, in: Leda Bisol (ed.), *Introdução a Estudos de Fonologia do Português Brasileiro*, Porto Alegre, EDIPUCRS, 101–134.
- Cunha, Celso/Cintra, Luís F. Lindley (1984), *Gramática da Língua Portuguesa*, Coimbra, Almedina.
- Fikkert, Paula (1994), *On the Acquisition of Prosodic Structure*, Leiden, HIL.
- Fikkert, Paula (2005), *Getting sound structures in mind. Acquisition bridging linguistics and psychology*, in: Anne Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*, Mahwah, NJ, Erlbaum, 43–56.
- Fikkert, Paula/Freitas, Maria João (2004), *The role of language-specific phonotactics in the acquisition of onset clusters*, in: Leonie Cornips/Jenny Doetjes (edd.), *Linguistics in the Netherlands*, Amsterdam/Philadelphia, Benjamins, 58–68.
- Freitas, Maria João (1997), *Aquisição da Estrutura Silábica do Português Europeu*, tese de Doutorado, Universidade de Lisboa.
- Freitas, Maria João/Rodrigues, Celeste (2003), *On the nature of sC-clusters in European Portuguese*, *Journal of Portuguese Linguistics* 2, 55–86.

- Fudge, Erick (1969), *Syllables*, *Journal of Linguistics* 5, 253–286.
- Girelli, Carl (1988), *Brazilian Portuguese Syllable Structure*, tese de Doutorado, University of Connecticut.
- Harris, James W. (1983), *Syllable Structure and Stress in Spanish: A Nonlinear Analysis*, Cambridge MA, MIT Press.
- Hyman, Larry (1985), *A Theory of Phonological Weight*, Dordrecht, Foris.
- Jakobson, Roman (1968, ¹1941), *Child Language, Aphasia and Phonological Universals*, The Hague, Mouton.
- Kahn, Daniel (1976), *Syllable-based generalizations in English phonology*, New York, Garland Press.
- Kayne, Richard (1984), *Connectedness and Binary Branching*, Dordrecht, Foris.
- Mateus, Maria Helena Mira (1975), *Aspectos da Fonologia Portuguesa*, Lisboa, INIC.
- Mateus, Maria Helena/Andrade, Ernesto d' (2000), *The Phonology of Portuguese*, Oxford, Oxford University Press.
- Mateus, Maria Helena Mira/Falé, Isabel/Freitas, Maria João (2005), *Fonética e Fonologia do Português*, Lisboa, Universidade Aberta.
- Mateus, Maria Helena Mira, et al. (2003), *Gramática da Língua Portuguesa*, Lisboa, Caminho.
- Mendes, Ana, et al. (2009), *Teste Fonético Fonológico – Avaliação da Linguagem Pré-Escolar (TFF-ALPE)*, Aveiro, Universidade de Aveiro.
- Miguel, Maria Augusta Cavaco (1993), *Os Padrões das Alternâncias Vocálicas e da Vogal Zero na Fonologia Portuguesa*, tese de Doutorado, Universidade dos Açores.
- Miranda, Ana Ruth/Matzenauer, Carmen (2010), *Aquisição da fala e da escrita: relações com a fonologia*, *Cadernos de Educação* 35, 359–405.
- Morales-Front, Alfonso/Holt, Eric (1997), *The interplay of phonology, morphology, and faithfulness in Portuguese pluralization*, in: Fernando Martínez-Gil/Alfonso Morales-Front (edd.), *Issues in the Phonology and Morphology of the Major Iberian Languages*, Washington, D.C./Georgetown, University Press, 393–437.
- Nespor, Marina/Vogel, Irene (1986), *Prosodic Phonology*, Dordrecht, Foris.
- Pike, Kenneth/Pike, Eunice (1947), *Immediate constituents of Mazatec syllables*, *International Journal of American Linguistics* 13, 78–91.
- Rodrigues, Celeste (2004), *Lisboa e Braga: Fonologia e Variação*, Lisboa, Fundação Calouste Gulbenkian.
- Santos, Rita (2013), *Aquisição de Grupos Consonânticos e seu Impacto nos Desempenhos Escritos do 1.º Ciclo do Ensino Básico*, dissertação de Mestrado, Faculdade de Letras da Universidade de Lisboa.
- Selkirk, Elisabeth (1982), *Syllables*, in: Harry van der Hulst/Norval Smith (edd.), *The Structure of Phonological Representations*, Dordrecht, Foris, 337–383.
- Selkirk, Elisabeth (1984), *On the Major Class Features and Syllable Theory*, in: Mark Aronoff/Richard Oehrle (edd.), *Language and Sound Structure*, Cambridge MA, MIT Press, 107–136.
- Soares Barbosa, Jerónimo (1822), *Grammatica Philosophica da Língua Portuguesa ou Princípios da Grammatica Geral Aplicados á Nossa Linguagem*, Lisboa, Academia Real das Ciências.
- Veloso, João (2002), *Do fricative+plosive onsets exist word-initially in European Portuguese? Gathering preliminary data and suggestive evidence from a small group of children in the phonological explicit task of syllable segmentation and examining some theoretical implications for the relationship between phonology and psycholinguistics*, in: Krasimira Petrova/Andreana Eftimova/Radko Shopov (edd.), *Litora Psycholinguística*, Sófia, Sema RSH, 49–56.
- Veloso, João (2006), *Reavaliando o estatuto silábico das sequências obstruinte+lateral em português europeu*, *DELTA* 22:1, 127–158.
- Veloso, João (2015), *The English R coming! The never ending story of Portuguese rhotics*, *Oslo Studies in Language* 7:1, 323–336.

- Viana, Aniceto Gonçalves (1883), *Essai de phonétique et de phonologie de la langue portugaise d'après le dialecte actuel de Lisbonne*, Romania 12, 29–98.
- Vigário, Marina/Falé, Isabel (1994), *A sílaba do português fundamental: uma descrição e algumas considerações de ordem teórica*, in: *Actas do IX Encontro da APL*, Lisboa, Colibri/APL, 465–478.
- Vigário, Marina/Martins, Fernando/Frota, Sónia (2006), *Frequências no Português Europeu: a ferramenta FreP*, in: Inês Duarte/Isabel Leiria (edd.), *Actas do XX Encontro da Associação Portuguesa de Linguística*, Lisboa, APL, 897–908.
- Wetzels, Leo (1997), *The lexical representation of nasality in Brazilian Portuguese*, Probus 9, 203–232.

Lista de autores

Raquel Amaro

Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas,
Centro de Linguística da Universidade Nova de Lisboa,
Avenida de Berna, 26-C, 1069-061 Lisboa, Portugal
e
Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600-214 Lisboa, Portugal
raquel.amaro@fcsh.unl.pt

Hugo C. Cardoso

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600-214 Lisboa, Portugal
hugoccardoso@gmail.com

Ernestina Carrilho

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600-214 Lisboa, Portugal
ernestina.carrilho@campus.ul.pt

Madalena Colaço

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600-214 Lisboa, Portugal
mmcolaco@gmail.com

Armanda Costa

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600-214 Lisboa, Portugal
armandacosta@letras.ulisboa.pt

João Costa

Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas,
Centro de Linguística da Universidade Nova de Lisboa,
Avenida de Berna, 26-C, 1069-061 Lisboa, Portugal
jcosta@fcsh.unl.pt

Inês Duarte

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600-214 Lisboa, Portugal
iduarte@letras.ulisboa.pt

Alexandra Fiéis

Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas,
Centro de Linguística da Universidade Nova de Lisboa,
Avenida de Berna, 26-C, 1069–061 Lisboa, Portugal
afieis@fcsh.unl.pt

Maria João Freitas

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
joaofreitas@letras.ulisboa.pt

Anabela Gonçalves

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
a.goncalves@letras.ulisboa.pt

Tjerk Hagemeyer

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
t.hagemeyer@letras.ulisboa.pt

Iris Hendrickx

Radboud University, Faculty of Arts, Centre for Language Studies,
Erasmusplein 1, 6525 HT Nijmegen, The Netherlands
i.hendrickx@let.ru.nl

Maria Lobo

Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas,
Centro de Linguística da Universidade Nova de Lisboa,
Avenida de Berna, 26-C, 1069–061 Lisboa, Portugal
maria.lobo@fcsh.unl.pt

Rui Marques

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
rmarques@letras.ulisboa.pt

Rita Marquilhas

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
rmarquilhas@letras.ulisboa.pt

Ana Maria Martins

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
anamartins@letras.ulisboa.pt

Ana Isabel Mata

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
aim@letras.ulisboa.pt

Amália Mendes

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
amalia.mendes@clul.ul.pt

Sara Mendes

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
sara.mendes@clul.ul.pt

Telmo Móia

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
tmoia@netcabo.pt

Helena Moniz

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
e

Universidade de Lisboa, Instituto Superior Técnico, INESC-ID,
Rua Alves Redol, 9, 1000–029 Lisboa, Portugal
helena.moniz@inesc-id.pt

Maria Antónia Mota

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
mar.mota@clul.ul.pt

Sandra Pereira

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600–214 Lisboa, Portugal
spereira@clul.ul.pt

Celeste Rodrigues

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600-214 Lisboa, Portugal
celesterodrigues@campus.ul.pt

Ana Lúcia Santos

Universidade de Lisboa, Faculdade de Letras, Centro de Linguística,
Alameda da Universidade, 1600-214 Lisboa, Portugal
als@letras.ulisboa.pt

João Paulo Silvestre

King's College London, Department of Spanish, Portuguese & Latin American Studies,
Camões Centre for Portuguese Language and Culture,
22 Kingsway, London WC2B 6LE, United Kingdom
joao.silvestre@kcl.ac.uk

João Veloso

Universidade do Porto, Faculdade de Letras,
Centro de Linguística da Universidade do Porto,
Via Panorâmica, s/n, 4150-564 Porto, Portugal
jveloso@letras.up.pt

Índice remissivo

- a gente* 14, 20, 242, 423, 498, 561n, 573, 574n, 575n, 576–577
- abaixamento, cf. *baixa* / *abaixamento*
- aberta 103–105, 636, 637, 639, 640, 642, 643n, 644–652, 657, 658, 664, cf. também *grau de abertura*
- acento 663
- de palavra 101, 108, 158, 164, 166–168, 173, 285n, 286–287, 302, 664, 665
 - nuclear 122, 126, 138, 139–141
 - prosódico contrastivo 384, 392
 - prosódico neutro 383, 392
 - secundário 173
 - tonal 119–121, 125, 127, 130, 133, 134, 148
 - vocálico 101n, 642, 650, 651
- Acordo Ortográfico 212
- adição de segmentos, cf. *inserção*
- adjunto / adjunção 56, 240, 309, 311, 320, 321, 324–326, 347, 355, 356, 358, 469, 530, cf. também *tempo*
- frase adjunta 453, 458–460, 470, 473, 474, 477, 478
- adverbial 315, 317, 328, 372, 461, 462, 604
- orações adverbiais 21, 434n, 444, 445, 481, 487–491, 498, 571
 - resposta ~ 581, 585, 586, 588, 590, 591, 592, 596–598
 - subordinação / subordinada ~ 317, 332, 361, 434, 445, 612, 613, 626
- advérbio 61, 159, 189, 189n, 243, 317, 327, 330, 385, 405–407, 409, 412, 422, 423, 440, 441, 444, 445, 445n, 461, 462, 494, 509–511, 539, 542, 586, 587, 590–593, 596–598, 601, 604
- afirmação 403, 581–607, cf. também *respostas*
- concordante / discordante 583–586, 588–607
 - enfática 406n, 410n, 586, 601–607
- afixo 61, 156n, 166, 169, 170, 170n, 213, 529, cf. também *prefixo*, *sufixo*
- africada, cf. *consoante(s)*
- Aktionsart* 190, 190n, 308, 318, 324, 325
- alomorfia 158, 161–163, 168
- alta / altura 10, 11, 103–107, 163, 636, 637, 644, 651, 653
- Alta Guiné, cf. *crioulo(s)*
- alternância / alternativo/a 182, 296, 346, 348, 358, 397, 414, 438, 504, 515, 519, 545, 549, 554, 587, 597–599, 601, 624, 631, 632
- alternância consonântica 15, 160
 - alternância de transitividade 437, 377
 - alternância posicional 375, 394, 394n, 397, 406
 - alternância pronominal 359, 361, 363, 432, 436, 559, 560, 563–566, 570
 - alternância VG/V 670n
 - alternância GV/VV 683, 684
 - alternância vocálica 15, 16, 158, 160, 641, 642–644, 646, 647, 649, 657
- ambiguidade 308, 309–311, 345, 347, 348, 352, 355, 358, 361, 376, 377n, 432, 454, 476, 485, 574, 575
- estrutural 336, 343, 345, 347, 349, 350, 352, 355–358, 406, 459, 469n, 476, 551
 - lexical 25, 178–180, 182, 309, 406
 - local 343, 346
 - temporária 344, 347
- anáfora 330, 331, 338, 446, 447, 465n
- anafórico/a 309, 310, 311, 330, 331, 344, 548, 576
- advérbio ~ 243
 - cadeia ~ 346, 358–363
 - dependências anafóricas 308, 312, 318, 330–331
 - locativo ~ 14
 - pronome ~ 330, 331, 358–363, 446–449, 465n, 571
 - retoma ~ 313, 351
 - sujeito ~ 351, 571
 - tempo ~ 322, 628
- Angola, cf. *português angolano* / *Angola*
- angolar 45, 47
- animacidade, cf. *traço (semântico: animacidade / +humano / definitude)*
- anotação (linguística / de corpora) 49, 129, 146, 147, 220, 224, 226, 228, 230, 232–242, 244, 251, 252, 258, 268, 269, 271–275, 574n
- antecedente 173, 388, 446, 455n, 456n, 466, 470, 471, 478, 507, 531n, 536, 548, 549, 564, 565, 565n, 570–572, 574–576, 613n
- crioulo 87, 88
 - da relativa 55–58, 356–358, 508

- (no sistema responsivo) 586, 589, 590, 593–596, 598, 600
- anterior / anteriorização 5n, 99, 102, 102n, 103, 113, 167, 169, 284, 289, 637, 642, 644, 645, 651–653
- anterioridade / sobreposição / posterioridade 172, 322, 542, 543, 629–631
- antroponímia 210, 218
- antropônimo 214–217
- apagamento 56, 59, 83, 92, 143, 145, 163, 358, 412, 512, 582
- (de segmentos) 7, 9, 10, 51, 106, 139, 160–162, 650, 652, 655, 675, 681, 684
- aquisição 2, 52, 85, 88, 92, 225, 227, 237, 246, 286–289, 292, 296, 297, 301, 339, 340, 401, 403, 410, 411, 431–433, 439, 439n, 440, 444–447, 449, 453, 460, 472, 476, 478, 558, 559, 566–568, 570, 571, 582, 663, 669, 672–675, 677, 678, 680–684, cf. também *desenvolvimento linguístico*
- de língua materna / L1 1–3, 9, 10, 161, 227, 237, 340, 351, 411, 431, 453, 455, 477, 478, 481n, 495, 558, 663, 664, 671
- de língua não-materna / L2 43, 52, 61, 85, 88, 92, 227, 237, 238, 247, 298
- argumental 60, 185, 189, 195, 318, cf. *sujeito*
- estrutura / grelha ~ 53, 55, 58, 59, 159, 182, 183, 189, 190, 194, 195, 301, 344, 348, 360, 465
- monoargumental 371, 376, 378, 380, 382, 467
- não ~, cf. *sujeito*
- quase ~, cf. *sujeito*
- argumento 54, 75, 78, 182–184, 189–191, 192n, 194, 195, 308, 312, 313, 318, 327, 360, 457, 464, 471, 506, 513, 531, 541, 614
- cognato 559
- dativo 54, 55, 457, 471
- externo 352, 457, 458, 468, 497, 621
- interno 54, 55, 352, 458, 471, 475, 497, 498, 512, 513, 547, 621
- arredondada, cf. *labial / labialização*
- artigo (definido / indefinido) 22, 52, 88, 217, 239, 240, 314, 355, 463
- artigo definido + possessivo 19, 21
- Ásia 1, 2, 68–72, 78–81, 85, 92
- asiático/a 68–73, 75, 77, 79–81, 84, 85, 89, 92, 582
- aspeto 158, 159, 172, 173, 294, 318, 320–322, 407, 423, 425, 468, 487, 527, 530, 534, 543, 545, 549, 549n
- asserção 320, 376, 384, 407n, 599, 600, 610, 619–621
- assimilação 8, 26, 87, 671, 679
- ataque (silábico) 6, 9, 10, 99, 110–113, 167n, 285, 287, 288, 302, 665–680, 682–684
- atético, cf. *telicidade / tético*
- atitudes proposicionais, cf. *proposicional*
- ato de fala 125, 147, 308, 310, 323, 329, 594n, 619, 620
- /b/ e /v/, cf. *consoante(s)*
- baixa / abaixamento 10, 101, 105, 163, 636, 637, 640, 641, 644, 648
- bantu, cf. *línguas bantu*
- Brasil, cf. *português brasileiro / do Brasil*
- c-comando 446, 512, 513
- assimétrico 515, 519
- Cabo Verde, cf. *crioulo(s), português cabo-verdiano / Cabo Verde*
- cadeia anafórica, cf. *anafórico/a*
- Casamansa, cf. *crioulo(s)*
- castelhano 5n, 6n, 8, 14, 200, 201, 209, cf. também *espanhol*
- catalão 8, 12–14, 21, 24n, 25, 118, 285, 401–403, 417, 434, 582n, 589n, 590n, 602, 603, 655n
- causativa 21, 524–526, 529–532, 535–537, 545–554, 621, 624, cf. também *verbo causativo*
- central / centralização 8, 9, 100, 100n, 101, 102n, 107, 143, 163, 164, 636, 637, 639–641, 649–651, 657
- clítico 14, 18–20, 48, 55, 59, 61–63, 76, 108, 239, 291, 294, 295, 301, 340, 382, 386, 394, 401–426, 431–450, 457, 523–527, 527n, 532, 536, 537, 544, 546, 550, 605, 607, 643
- acusativo 59, 60, 294, 432, 435, 436, 457, 526, 535, 547
- colocação dos (pronomes) clíticos 2, 3, 20, 61–63, 389, 393, 401–426, 431, 440, 441, 444–446, 524, 532, 532n, 606, cf. também *ênclise, interpolação, mesóclise, próclise*
- dativo 13, 14, 18, 58, 59, 457, 535, 547
- inerente 59–61, 536, 536n, 546
- reflexo 59, 76, 432, 436, 536, 546
- se 92, 442, 527n

- subida de ~ 19, 20, 404, 404n, 416, 419, 523, 524, 526, 527, 529, 530, 532–539, 544–546, 550
- coda 4, 6, 9, 10, 101, 104, 105, 106n, 109n, 161, 162, 164, 164n, 167, 167n, 286–288, 302, 603, 654, 666, 667, 669–675, 679–682
- codificação de caracteres 258, 264, 265, 266, 271
- codificação lexical 180, 182, 186, 191, 193, 196, 197
- cognição 181, 278, 279, 336, 337, 341, 345
- cognitivo/a 2, 181, 225, 278, 279, 284, 299, 336–343, 345, 346, 350, 351, 354, 355, 358, 375, 376
- colocação dos pronomes clíticos / átonos, cf. *clítico*
- colocações 205, 207, 211, 230, 232, 243
- compilação de corpora, cf. *corpus / corpora*
- completiva 18, 21, 301, 317, 330, 331, 347, 348, 403, 441, 443–445, 453, 453n, 456, 457, 461–466, 470, 474, 475, 477, 478, 523, 527, 531, 538n, 539, 562, 564, 570, 614–624, 626, 628, 631
- complexo verbal 467, 487–491, 498
- compreensão 131, 132, 137, 150, 208, 280, 290–292, 296, 298, 299, 338–340, 342, 343, 355, 358, 363, 431, 433, 437, 439, 446–449, 570
- concordância 2, 19, 21, 24, 51, 52, 56, 87, 165, 166, 168, 169, 169n, 171, 293, 294, 330, 354, 355, 434, 453n, 481, 485, 486, 492, 502, 503, 505, 507, 510, 512, 513, 515, 521, 558, 562, 563, 568, 574n, 575n, 589, 591, 595, 616, 627–629
- com o primeiro termo coordenado 502, 504, 506, 508, 510–513, 515, 516, 518, 519
- de número 48, 50, 88, 354, 453n, 516
- negativa 23, 329, 330
- nominal 50, 90, 91, 242, 301, 354, 355, 502, 513, 514, 521
- parcial 502–512, 514–516, 518–521
- plena 502–514, 516, 517, 519, 520
- sujeito–verbo 50, 88, 166, 169, 242, 354, 356–358, 467, 481, 483n, 485, 486, 492, 502, 504–506, 508, 509, 511, 515, 521
- concordância (de corpus) 230–232, 234, 235, 237, 241, 256
- concordanciador 230, 231, 243
- conector 13, 16, 61, 243, 310, 324, 325, 332, 431, 440, 445, 446, 481, 492–495, 497, 497n, 498, 611, 612, 632
- conexão interproposicional, cf. *proposicional*
- confirmação / inversão 581–583, 585, 586, 595, 600
- conhecimento lexical 178, 180, 186, 191, 197, 445, 446
- conjuntivo 2, 18, 21, 157, 160, 163, 172, 327, 445n, 494, 530, 570, 587, 610–633
- consoante(s) 4–7, 9–11, 82, 99–101, 107, 109, 110, 113, 160, 162–165, 167, 167n, 206, 283–285, 636, 638, 639, 655, 663–668, 670–673, 677–679, 682–684, cf. também *alternância consonântica*
- africada palatal /tʃ/ 4–6, 87, 110, 111
- africadas predorsodentais 4, 5
- /b/ e /v/ 11, 111, 112
- fricativas 4, 5, 5n, 87, 100, 101, 105, 106, 106n, 111–113, 129, 284, 286, 287, 289, 668, 669, 671, 673, 676–682
- laterais 164, 164n, 287n, 668, 669, 671, 677, 681, 682
- líquidas 87, 286, 287, 289, 668, 668n, 673, 676, 677, 680–682
- nasais 10, 100n, 163, 167n, 168, 289, 636, 667, 668, 671, 676, 677, 681
- obstruinte + líquida 668, 677, 680
- oclusivas 5, 112n, 135, 284, 289, 668, 676, 677, 679
- palatais 4–7, 10, 11, 100, 110, 206, 603, 636
- sibilantes 4, 5, 6, 90, 603
- vibrantes 4, 6, 6n, 87, 113, 288, 668, 688n, 669, 671, 677, 682
- constituente silábico 9, 288, 663, 666, 667, 673, 674
- contacto (de línguas / linguístico) 2, 29, 43, 44, 48, 50, 51, 53, 54, 55, 57, 58, 60, 61, 63, 64, 68, 69, 71, 72, 75, 77–80, 82, 84–88, 91, 92, 209, 213, 219, 410, 411, 418, 582
- contexto discursivo 310, 314, 322, 354, 620
- contexto verídico / de veridicidade 618, 619, 625
- coordenação 2, 301, 331, 332, 395, 416, 445, 502–521, 582n, 593, 595, 596, 598
- aditiva 503, 504
- corpus / corpora 2, 49, 51, 53, 62, 63, 70, 79, 116–119, 119n, 125n, 129, 133n, 138, 142n, 144, 145, 147–150, 169, 169n, 200, 201, 207, 210, 211, 212, 214, 220, 224–247, 256, 258, 259, 269, 271–274, 310, 323, 353, 389, 409n, 420n, 425, 472–474, 477n,

- 482, 486, 490, 520, 535n, 538, 552, 565n, 573n, 574, 574n, 577
- comparáveis 227, 236
 - compilação de ~ 224-226, 228, 230, 231, 233, 235, 237, 238
 - de fala 116-118, 129, 147-149, 228, 233-234, 242
 - de referência 227, 232
 - diacrônico / histórico 210, 214, 234, 235, 243, 244, 245, 252, 258, 259, 271, 273
 - linguística de corpus 2, 200, 203, 224, 225, 226, 271, 272, 389
 - paralelos 227, 237, 246
 - correferência 331, 336, 351, 358-363, 447, 449, 470, 478, 570, 573
 - correferente 432, 495, 570-575, 577
 - crioulo(s) 2, 43-47, 50, 52, 53n, 57, 58, 60-64, 68, 71, 72, 78, 79, 81, 84-90, 92
 - da Alta Guiné 43-45, 50, 51, 57
 - da Guiné Bissau / ~ guineense / Kriol 44-46
 - de Cabo Verde / ~ cabo-verdiano / Kabuverdianu 44-47, 53n, 58, 62, 560n, 561n
 - de Casamansa 44, 45
 - de S. Tomé / santome / forro 45-47, 53, 54, 57, 59
 - do Golfo da Guiné 43-45, 50, 51, 57
 - do Príncipe / lung'ie 45, 47
 - fa d'ambô 45
- definitude, cf. *traço (semântico: animacidade) / +humano / definitude)*
- dêítico 51, 311, 315, 322, 423-425, 462, 542
- locativo 12, 14, 407, 411, 412, 425
- dependência anafórica, cf. *anafórico/a*
- descrição lexical 200, 206, 207, 214
- desenvolvimento linguístico 282, 298, 299, 300, 301, 663, cf. também *aquisição*
- atípico 278, 282-290, 339, 340
 - típico 284-290, 292, 294-296, 298, 339
- derivação / derivado (morfologia) 78, 81, 156, 167, 167n, 170, 209, 210, 212, 213, 215-218, 605, 606, 648
- determinação 308, 312, 313-315, 517
- determinante 22, 315, 354, 406, 413, 433, 517-519
- diacronia, cf. *mudança (linguística) / diacronia*
- diacrônico/a 1, 3, 9, 20, 168, 204, 206, 208-210, 212-215, 227, 234, 235, 242, 243, 245, 253, 386n, 401, 404, 414, 416, 423, 426, 559, 594, 671
- dialetal 11, 14, 20, 111, 206, 235, 245, 246, 401, 423, 425, 492, 538, 552, 644, 656, 666
- português (europeu) ~ 1-3, 6, 8, 9, 11, 13, 20-22, 24, 26, 61, 98-100, 104-106, 109, 226, 410, 423, 481-498, 560n, 572-577, 636, 641, 647
 - variação ~ , cf. *variação linguística*
- dialeto(s) 11, 13, 21, 24n, 98, 99n, 100, 103n, 104, 106, 108-110, 112, 227, 549, 560n, 636, 647n
- do português do Brasil 651
 - do português (europeu) 6, 8, 9, 13, 20, 22, 24, 26, 61, 80, 98, 105, 410, 423, 641
 - (portugueses) centro-meridionais 5, 9, 14, 21, 99n, 100, 100n, 103, 107, 109n, 481, 485, 653
 - (portugueses) setentrionais 5, 10, 11, 16, 21, 100, 100n, 104-108, 110-113, 130, 133, 483, 484
- dialetologia 483
- dialetológico/a 98, 99
- dicionários bilingues 201-205, 208-210
- dissimilação 7, 11, 26, 105, 165, 168
- ditongaço 7, 8, 10, 11, 12, 636
- ditongo 7, 10-12, 49n, 87, 101n, 163-165, 167-169, 431, 638, 639, 640n, 654, 663, 666, 667, 670, 670n, 673, 682-684
- nasal 7, 11, 12, 167, 168
- duplo objecto 54, 63
- edição acadêmica digital 252, 259, 263, 271-274
- edição crítica 253, 258, 263, 266, 268, 270
- elementos, cf. *fonologia*
- elevação (de vogais) 4, 7-9, 104-106, 163, 164, 649, 651, cf. também *verbo*
- empréstimo (lexical) 15, 26, 27, 29, 68, 71-74, 81, 82, 88, 90, 91, 213, 214, 217, 258n
- ênclise 19, 61, 63, 389, 392, 393, 401-406, 408-414, 416-421, 433, 440-446, 532, 606
- enclítico/a 60, 61, 389, 403-405, 414, 416n, 418, 431, 443, 449
- entoação 116-118, 131, 150, 600, 601, 603-605, 607
- entoacional 107-109, 116-121, 124, 129-131, 134-136, 138, 139, 141, 148, 150, 229, 285, 355, 358, 665

- entrada lexical 157, 159, 166, 170, 171, 191, 541
 – rica 157
 epêntese, cf. *inserção*
 epônimo 216, 217
 espanhol 21–25, 118, 205, 233, 235, 274, 355,
 373, 386, 401–404, 417, 418, 434, 460,
 514, 520, 550, 551n, 582n, 589n, 590n, 602,
 603, 619, 644, cf. também *castelhano*
 especificação lexical 178, 186–192, 198, 642,
 648
 estatístico/a 46, 109, 203, 217, 224, 238
 estrutura argumental, cf. *argumental*
 etimologia 81, 178, 204, 206, 208, 209, 213,
 215, 217
 etimológico/a 74, 80, 81, 203, 206, 208–210,
 212–216, 220, 257, 258
 etiqueta 202, 226, 228, 239, 241, 242,
 266–270, 272, 274, 411
 evidencialidade 323, 376, 377n, 378, 379, 412
 exclamativas 371, 373, 407, 409, 440, 455
 expansão (linguística) 5, 6, 11, 18, 22, 68–70,
 72, 79, 113, 416, 417
 expletivo 241, 242, 309, 329, 491, 495,
 559–563, 566, 567, 572, 574n, 577

 fa d'ambô, cf. *crioulo(s)*
 fazer-Infinitivo / fazer-Inf, cf. *infinitivo*
 flexão / flexionado / flexional 12, 15, 18, 20,
 21, 50, 51, 156, 157–163, 165–167, 169–172,
 201, 212, 226, 232, 239, 242, 281, 293–295,
 314, 340, 351, 356–359, 408, 409, 453–498,
 523, 531, 531n, 534, 537, 540n, 545, 547n,
 549–552, 560, 561, 563, 567, 568, 594, 610,
 612, 627, 628, 647, 649
 fechada 637, 639, 644–646, 648, 653, 657,
 664, cf. também *grau de abertura*
 filologia 252–255, 256, 257, 258, 259, 260, 262,
 275
 finito/a 344, 402–404, 417–419, 492, 494
 – domínio ~ 20, 481, 491, 492, 494, 495, 498
 – frase ~ 19, 389, 403, 404, 408, 410, 414, 416,
 417, 420, 458, 466, 612
 – subordinada ~ 20, 62, 403, 405, 440,
 443–445, 466, 474
 finitude 481, 494
 focalização 371, 373, 374, 378, 384–397, 406,
 407, 411–413, 576, 577
 foco contrastivo 371, 383, 384, 387–389, 393–
 397, 397n, 406, 411, 412, 592, 593
 foco informacional 371, 373, 374, 378,
 380–384, 387, 396, 397
 fonologia 1–12, 18, 48, 113, 118, 156, 158–161,
 163–174, 262, 278, 281, 282–290, 299, 301,
 302, 369, 640, 640n, 655, 663, 664, 665
 – dos elementos 636, 640, 640n, 643, 643n,
 645, 651, 652, 659
 formação de palavras 170, 173, 210, 339
 forro, cf. *crioulo(s)*
 francês 8, 12, 13, 14, 15, 16, 21, 24, 24n, 25,
 27, 218, 233, 294, 353–355, 402, 403,
 433, 434, 437, 439, 524, 526, 528–531,
 533–535, 546, 550, 558, 559, 566, 584,
 644, 645, 653
 fricativa, cf. *consoante(s)*

 galego 8–11, 14, 16, 21, 23, 24, 25, 214, 401,
 403n, 420, 453, 481n, 483, 483n, 484, 550,
 581, 582, 586, 602, 603
 gênero (morfologia / sintaxe) 15, 55, 56, 87, 88,
 165, 166, 171, 205, 218, 239, 351, 353–355,
 361, 502, 503, 505, 513, 514n, 515, 516, 519,
 520, 561
 geolinguístico/a 1, 6, 8, 21, 29, 99, 130
 – variação ~ 98, 114, 117, 130–133
 gerúndio 18, 19, 21, 402, 468, 481–485,
 487–490, 492, 492n, 494, 549
 – flexionado 2, 3, 21, 242, 481–487, 489–498,
 549
 glide / semivogal 7, 10, 100, 101, 107–109, 163,
 164, 168, 664, 670, 681, 683, 684, cf. tam-
 bém *alternância VG/V*, *alternância GV/VV*
 gramática histórica 4–29
 grau de abertura 636, 639, 640, 642, 643n,
 644–652, 657, 658
 grelha argumental, cf. *argumental*

 harmonização (vocalica) 7–10, 163, 646, 647,
 651, 651n
 haver / ter, cf. *verbo*
 hiato 7, 9, 10, 12, 14, 108, 164, 167, 655, 683
 história da língua 3, 29, 206, 243, 252, 423,
 636, 653
 homonímia 207, 208, 309, 454
 homônimo/a 158, 208, 209, 210, 347, 454

 ilha sintática 295, 434–437, 439
 indicativo 12, 16, 100, 157–160, 163, 166, 172,
 239, 322, 323, 327, 347, 376, 445n, 494,

- 530, 564, 570, 571, 587, 594, 602n, 606,
610–623, 625, 626, 627, 628, 629
- infinitivo 15, 18–21, 402, 403, 408, 409, 409n,
416, 416n, 419, 481, 484, 487, 488, 492n,
494, 523–528, 530–553, 563, 588, 590,
604, 611
- fazer-Infinitivo / fazer-Inf 523, 525, 532,
534–537, 545–547, 549–554
- flexionado 2, 3, 18, 20, 21, 157, 314, 340, 408,
409, 453–478, 481n, 483, 484, 491, 494,
531, 537, 549, 550
- pessoal 460, 473, 476, 477, 478
- preposicionado 453, 455, 468, 472, 474, 475,
477, 478, 549n, 552
- inserção 10, 53, 58, 59, 139, 162, 164n, 681,
684
- de glide 100, 101, 107–109, 164, 168
- de vogal epentética 9, 161, 162, 643, 674,
675, 677, 678, 680, 684
- de vogal paragógica 9, 109n, 140
- de vogal protética 9
- instrumentos de avaliação linguística 278, 281,
282, 285, 289–291, 294, 295, 298–302,
674
- interface entre módulos gramaticais 2, 156–174,
179, 182, 288, 290, 296, 396
- interpolação 19, 242, 401, 404, 414–415,
420–426
- interpretação 23, 51, 118, 119, 132, 166, 174,
180, 201, 205, 217, 243, 308–314, 319–321,
330, 331, 336, 340–363, 373, 375, 376,
379–381, 387, 388, 391, 395, 404, 411–
413, 431, 438, 448, 449, 453, 455, 459,
465, 466, 469, 470, 472, 475–478, 495,
506, 507, 511, 513–520, 543, 546, 559, 560,
564–566, 570–572, 575, 587, 589, 590–
594, 600, 611, 613, 616, 618, 619, 622,
623, 625, 626, 629, 630, 632
- interrogativo/a 2, 12–14, 82, 116, 117, 121,
122, 124–127, 131–133, 147, 150, 229,
290–293, 295–297, 329, 343, 373, 380,
407, 409, 431, 440, 441, 562, 565, 581,
582, 585–596, 598–601, 605, 607, 610,
620, cf. também *QU-* (*interrogativa /
palavra / sintagma* ~)
- inversão do sujeito / sujeito invertido (VSO /
VOS) 20, 21, 351, 371, 373–385, 390, 391,
394, 396, 397, 406, 411, 413, 496, 496n,
497, 561, 562, 570, 571, 577
- inversão nominal negativa com *algum* 19, 22
- italiano 12, 13, 14, 21, 24, 24n, 70, 118, 233,
359, 373, 386, 386n, 402, 403, 433, 434,
435, 447, 524–534, 539, 546, 548, 550,
560n, 561n, 567, 569, 602, 603, 644
- itens de polaridade, cf. *polaridade*
- kabuverdianu, cf. *crioulo(s)*
- kriol, cf. *crioulo(s)*
- kriyol, cf. *crioulo(s)*
- L2, cf. *língua segunda / L2*
- labial / labialização 10, 101n, 104, 105, 112,
164n, 286, 636, 640n, 645, 650–652, 663,
682–684
- lateral, cf. *consoante(s)*
- lema 200–202, 206–209, 211, 212, 228, 232,
237, 238, 269
- lematização 201, 207, 226, 232, 238, 239, 273,
274
- lexema(s) 75, 77, 80, 83, 90, 156–174, 200, 201,
207–210
- estrutura interna dos ~ 157, 170
- léxico 1, 3, 11, 26–29, 43, 45, 80, 156–159, 165,
167, 170, 171, 174, 178–187, 191, 193, 194,
196, 197, 200, 201, 205, 207, 210, 214, 220,
242–244, 255, 256, 283, 302, 344, 352,
636, 644, 646, 649, 655
- computacional 180, 181, 184, 185, 192
- generativo 178, 181–183, 190n, 191–193, 195
- lexicográfica (técnica ~ / terminologia ~) 200,
202, 206, 211, 220
- língua segunda / L2 42, 44–49, 52, 61, 68, 70,
71n, 82, 84, 86, 89, 91, 92, 146, 204, 205,
209, 210, 227, 237, 238, 242, 244, 246,
247, 298, 401
- linguagem de marcação 252, 266, 267
- línguas bantu 43, 45–47, 50–52, 52n, 54, 56–
64, 668
- línguas crioulas, cf. *crioulo(s)*
- línguas pidgin, cf. *pidgin(s)*
- línguas românicas, cf. *românico/a*
- linguística clínica 2, 278, 280–282, 290, 291,
297, 298, 300, 302, 340
- linguística de corpus, cf. *corpus / corpora*
- linguística histórica 200, 214, 253, 582
- líquida, cf. *consoante(s)*
- locativo anafórico, cf. *anafórico/a*
- lung'ie / crioulo do Príncipe, cf. *crioulo(s)*

- marcação de texto 263, 266, 267, 268, 271, 272, 274
 marcadores de ênfase 406, 407, 412
 marcadores de foco 385, 391, 395, 406, 497, 412, 591, 593
 matriz de traços 156, 157, 166
 mesóclise 402n, 418–420, 420n, 441, 532n, 606
 metadados 229, 231, 235, 237, 268, 270, 274
 metátese 7, 10, 423
 minimizador 23, 25, 379, cf. também *itens de polaridade negativa*
 Moçambique, cf. *português moçambicano / Moçambique*
 modalidade 21, 159, 172, 173, 308, 312, 325–327, 587, 633
 modelização 178, 180, 186, 191, 196, 197, 345
 modelos de processamento 336, 345, 348, 349, cf. *processamento da linguagem / das línguas naturais*
 modo 16, 158, 159, 172, 173, 326, 327, 473, 494, 530, 567, 610–633, cf. também *indicativo*
 – conjuntivo 610–633, cf. também *conjuntivo*
 monotongação 7, 10, 11, 87, 101, 206
 morfema 14, 15, 75, 100n, 294, 441, 453, 454, 481, 483, 563, 627, 643, 643n, 656, 657
 morfologia (nominal / verbal) 11–17, 50, 60, 156–174, 178, 204, 242, 293–295, 299, 352, 354, 355, 402, 420, 473, 481, 482, 485, 486, 492, 495, 529, 560–563, 568, 571, 574, 577, cf. também *alternância vocálica*, *alternância consonântica*
 movimento (sintático) 292, 295–297, 346, 528, 530, 533–535, 538, 545, 547
 – movimento simultâneo 511–513
 mudança (linguística) / diacronia 1, 3–29, 43, 47, 49, 50, 56, 63, 64, 78, 98, 99, 109, 110, 112, 158, 178, 206, 208–210, 214, 215, 325, 362, 389, 401, 411, 414–426, 559, 566, 577, 653, 659
 mundos possíveis 610, 622–627, 631–633

 nasal, cf. *consoante(s)*, *vogal*, *ditongo*
 nasalização 100n, 168, 636
 nativização 44, 46, 47, 63
 negação 19, 20, 22, 23, 25, 52, 187, 308, 309, 312, 327–330, 378, 379, 403, 409, 421–423, 431, 440, 441, 444, 445, 539, 544, 545, 552, 581–586, 588, 589, 591, 593–601, 605, 607, 615, 619, 625, cf. também *polaridade*
 – concordante / discordante 583, 586, 588, 589, 591, 592, 595–599, 601
 – enfática 600n
 neutralização 2, 5, 8, 9, 99, 104, 636, 639, 647, 648, 649, 650
 núcleo (silábico) 9, 11, 101, 101n, 105, 106, 109n, 116, 164, 167, 287, 302, 666, 667, 669, 671, 673, 676, 678–682
 – núcleo vazio 9, 669, 678–680
 número (morfologia) 12, 14, 48, 50–56, 87, 88, 157, 158, 160, 161, 163, 165–168, 172, 239, 351–357, 453, 453n, 454, 460, 470, 473, 476, 478, 502, 503, 505, 514n, 516, 527n, 561, 586, 681

 objeto anteposto (OSV / OVS) 20, 361, 371, 372, 373, 385–396
 objeto nulo 295, 431, 433–439, 449
 obstruinte, cf. *consoante(s)*
 oclusiva, cf. *consoante(s)*
 omissão de pronomes 294, 295, 431, 433–437, 439, 442–444, 560
 onomástica 200, 210, 214–219
 ordem de constituintes 2, 20, 21, 288, 371–397, 406, 411, 413, 535, 546, cf. também *alternância posicional*
 ordem de palavras 20, 52, 76, 83, 91, 92, 349, 351, 352, 353, 354, 355, 423, 456–458, 463, 464, 481, 495–498, 502–504, 512, 514, 519, 520, 549, 550, cf. também *alternância posicional*
 ordem OSV / OVS, cf. *objeto anteposto (OSV / OVS)*
 ordem SVO 19, 75, 76, 342, 347, 351, 352, 361, 363, 371, 372, 374, 377, 381, 383, 385, 396, 397
 ordem SVO / SOV no português medieval 19, 20, 42, 421
 ordem VSO / VOS / OVS, cf. *inversão do sujeito / sujeito invertido (VSO / VOS)*

 P_toBI 118–120, 129, 148, 149
 Pacífico 2, 68
 padrão (dialeto ~ / língua ~ / português europeu ~ / variedade ~) 1, 2, 5, 11, 14–16, 79, 89, 98–104, 106–109, 111, 113, 114, 116, 117, 130, 132, 133, 164, 168, 226, 227,

- 404n, 410, 417, 419, 453–455, 460, 477, 482, 487, 488, 491, 494, 495, 498, 537, 538, 552, 558, 560, 637, 640, 641, 656, 657, 663, 666n, 680
- padrões de concordância 2, 169n, 502, 504, 511, 521, cf. também *concordância*
- padrões de resposta 581, 582, 582n, 585, 587, 588, 589, 593, 594, 559, 600
- palatal, palatalização 4–6, 7, 9, 10, 11, 100–103, 107, 108, 110, 113, 603, 636, 640n, 645, 649, 650, 651, 652
- palavra abstrata 156, 157
- palavra morfossintática 156, 165
- palavra prosódica 108, 133, 285, 431, 665
- papiamentu 44
- paradigmas flexionais 15, 158–161, 163, 166, cf. *flexão* / *flexionado* / *flexional*
- paragoge, cf. *inserção*
- parâmetro do sujeito nulo 2, 558, 566, 567, 568n, 570, 572, cf. também *pro*
- particípio passado 12, 15, 19, 21, 24, 208, 209, 212, 247, 402, 434, 481, 544, 588, 604
- passivas 18, 20, 54, 76, 159, 301, 343, 362, 363, 377n, 382, 527, 527n, 528, 533, 536, 544, 546
- passivas dativas 54
- perceptiva (estrutura ~) 21, 551n, cf. também *verbo perceptivo*
- Perl* 264, 269, 270
- Perturbação Específica do Desenvolvimento da Linguagem (PEDL) 281, 286–288, 291, 292, 294, 297, 300
- peessoa (morfologia) 12–14, 16, 50, 56, 58–60, 88, 100n, 157, 158, 163, 166, 168, 172, 239, 294, 323, 352, 432, 433, 435, 436, 453, 453n, 454, 456, 460, 464, 465, 470, 471, 473, 475, 476, 478, 481–483, 483n, 485–487, 491, 492, 495, 496, 498, 502, 503, 505, 527n, 551, 561, 561n, 564, 565, 569, 573–577, 586, 594, 602n
- pidgin(s) 68, 70, 71, 74, 77, 78, 79, 86
- polaridade 22, 26, 403, 408, 581, 583, 589, 590, 593–595, 599, 600, 602, 610
- itens bipolares 23, 25
- itens de ~ 26
- itens de ~ negativa 22, 23, 25, 610, cf. também *minimizador*
- itens de ~ positiva 23, 25, 610
- posição final dos marcadores de ~ 599, 600, 607
- polissemia 208, 243
- polissêmico/a 207–209, 228
- português angolano / Angola 48–50, 52–54, 57–59, 61–63, 236, 410
- português brasileiro / do Brasil 1, 2, 3, 8, 9, 15, 20–24, 50, 55, 56, 58, 61, 62, 98, 105, 118, 130, 131, 133, 164n, 169, 169n, 200, 201, 202, 211, 212, 220, 232, 236, 242, 315, 325, 351, 356–358, 361, 362, 401–403, 410, 411, 418–420, 432, 439n, 447, 453, 460, 464n, 472, 476–478, 487, 514, 520, 536, 537, 540n, 553, 561, 565, 572, 576, 577, 581, 586, 596n, 601, 602, 602n, 603, 607, 638n, 648, 651, 651n, 654, 666, 669, 671, 677, 681
- português cabo-verdiano / Cabo Verde 48, 51, 52, 236
- português dialetal, cf. *dialetal*
- português em África / variedades africanas do português 43–64, 131, 133, 169, 236, 237, 242, 410, 537
- português moçambicano / Moçambique 46–48, 50–59, 61, 62, 63, 236, 410, 537
- português na Ásia e no Pacífico 68–91
- português oriental 80–81, 83, 90
- português são-tomense / S. Tomé e Príncipe 47, 48, 49n, 51, 53, 54, 56, 58, 59, 62, 63, 236, 237, 410
- POS 239, 244, 271
- pragmática 2, 16, 98, 291 297–299, 308, 310–312, 321, 323, 329, 331, 332, 358, 372, 540, 587n, 593, 633
- predicação 301, 308, 312, 313, 327, 332, 372, 373, 375, 377n, 413, 468, 491
- predicado complexo 523–526, 529–531, 533–535, 537–539, 541, 545, 549, 550, 552n, 554
- prefixo 52, 54–56, 59, 60, 170, 605, 606, cf. também *afixo*, *sufixo*
- prime* / *priming* / *primada* 350, 357, 358
- pro* 563
- pro-drop 558, 558n, 560n
- processamento automático 116–117, 129, 146–150, 178, 181, 233, 234
- processamento da linguagem / das línguas naturais 178, 179, 181, 194, 230, 282, 283, 332, 336–363
- processos cognitivos 336, 338, 339, 342, 350, cf. também *cognitivo/a*
- processos fonológicos 5–12, 98, 99, 114, 164, 284n, 431, 532, 638, 649, 650, 652, 666

- proclisador 20, 61–63, 389, 407–411, 413, 414, 421, 422, 425, 532n
- próclise 19, 61, 62, 63, 389, 393, 401–414, 416–418, 420–423, 431, 433, 440–446, 449, 450, 532, 533
- próclítico/a 389, 403, 404, 406, 416, 417, 423, 431, 440, 441, 444
- produção escrita 382, 419, 420, 425
- produção espontânea 441, 474, 477, 568–570
- produção oral 98–114, 116–150, 663–684
- produção provocada / induzida 49, 434, 441, 474
- produção sintática (teste de ~) 300–301
- programação 263–266, 271
- pronom(e)s 12–14, 16, 18, 20, 22–24, 26, 55–57, 59–63, 82, 84, 85, 294, 295, 314, 316, 330, 340, 357–359, 361–363, 380, 386, 389, 394, 401–426, 431–450, 456, 460, 464, 465n, 467–469, 473, 476, 498, 502n, 508, 523, 524, 526, 536, 546, 548, 558–567, 570–573, 577, 605, 606, cf. também *alternância pronominal*
- Princípio Evitar Pronome 360, 361
 - pronome acusativo 13, 14, 59, 294, 457n, 468, 473, 526
 - pronome anafórico, cf. *anafórico/a*
 - pronome clítico, cf. *clítico*
 - pronome dativo 13, 14, 18, 53–55, 58, 59, 386, 433, 457, 457n, 473, 526, 535n, 547, 551
 - pronome demonstrativo 12, 13, 52, 314, 330, 498
 - pronome forte 18, 59, 432, 433, 435, 436, 441, 447–449, 469
 - pronome interrogativo 82, 380
 - pronome nominativo 13, 14, 16, 20, 454, 456, 460, 465n, 468, 473, 476
 - pronome nulo 359–363, 469, 565, 571
 - pronome pessoal 13, 52n, 467, 469, 498, 502, 526, 536
 - pronome possessivo 12, 13, 19, 21, 22, 315
 - pronome recíproco 508
 - pronome reflexo 59–61, 546, 548
 - pronome relativo 12, 13, 314, 357
 - pronome resuntivo 57
- proposição 311, 323, 326–328, 331, 332, 613n, 615–625, 632
- proposicional 318, 327, 329, 397
- atitude ~ 329, 610, 621–623, 625
 - conexões interproposicionais 308, 312, 318, 331–333
- prosódia 1, 113, 116–118, 130, 133, 136, 146, 150, 228, 288, 355, 356, 358, 663, 664
- prótese, cf. *inserção*
- QU- (interrogativa / palavra / sintagma ~) 121, 122, 131, 373, 380, 405–407, 581
- quantificação 224, 243, 312, 313, 315–317, 325, 326, 328, 632, 632n
- quantificadores 22, 239, 309, 310, 316, 317, 372, 390, 393, 405–407, 413, 422, 446, 513, 518, 519, 548, 631, 632
- radical 15, 16, 57, 59, 60, 156, 158, 160–173, 213, 646–649, 681
- recuada, recuo 5n, 6, 8, 11, 100, 104, 105, 637, 637n, 639, 642, 644, 645, 649, 650, 652–654
- reduplicação do verbo 581, 586, 598n, 600–607
- relativo/a 12, 13, 19, 20, 22, 52, 55–58, 75, 76, 83, 242, 291, 292, 296, 301, 314, 317, 331, 336, 347–349, 355–358, 382, 393, 409, 434n, 459, 459n, 470, 470n, 473, 474, 477, 478, 508, 613, 614, 630, 631
- relativa cortadora 56, 63
 - relativa de cópia defetiva 56, 58, 63
 - relativa resuntiva 55–58
 - relativização 55, 58, 242
- reestruturação (construção de ~) 18, 416, 523, 525, 525n, 527, 528, 530, 532–546, 553, 554
- reestruturação (linguística) 43, 53, 55, 59
- regras de resolução 502, 503, 504, 511, 519, 520
- relações semânticas 174, 188, 208, 211, 331
- representação do léxico 178, 181, 183, 186, 191, 194, 220
- representatividade 226, 227, 291
- respostas (afirmativas / negativas), cf. *padrões de resposta, afirmação, negação*
- respostas adverbiais 586, 590–592, 596–598
- respostas com *ser* 587, 594–596
- respostas com *sim* 587–594, 596, 599–601
- respostas mínimas (afirmativas / negativas) 586–596
- respostas verbais 586–596
- rима 9, 105, 109n, 287, 302, 665–667, 669, 671n, 673, 676, 678, 684
- românico/a 2, 5, 8, 130, 174, 420, 550
- línguas românicas 1, 3, 8, 10, 11, 13, 19–21, 25–27, 206, 215, 216, 233, 315, 329, 351,

- 353, 363, 371, 375, 386, 387, 397n,
401–403, 426, 460, 505, 523, 530, 540n,
546, 550, 552–554, 581, 582, 586,
601–603, 616, 617, 644n, 655n
- S. Tomé e Príncipe, cf. *português são-tomense* /
S. Tomé e Príncipe
- se impessoal 18, 20, 574n
- se passivo 18, 20, 547
- semântica 158, 159, 165, 167n, 171–174,
308–333
- semântica lexical 3, 13, 22–26, 90, 157,
178–198
- semivogal, cf. *glide* / *semivogal*
- sibilante, cf. *consoante(s)*
- significado em contexto 179, 180, 181, 182, 191,
192
- sílabas 3, 9, 19, 99, 101–107, 109, 116, 119–121,
124, 125, 128, 131, 134, 147, 161, 163–166,
167n, 206, 212, 281, 282, 285–288, 636n,
638, 639, 651, 654, 663–684
- sinóníma 184, 204, 211
- sinónimo/a 184, 188, 205, 211, 311
- sintaxe 1–3, 17–19, 21–23, 25, 44, 48, 58, 61,
63, 76, 156, 159, 160, 165–167, 169–172,
182, 200, 242, 243, 278, 281, 290, 296,
297, 299, 301, 346, 348, 354, 372, 396,
401, 411, 413, 492, 529, 573n, 581, 610,
616, 627, 628
- sistema vocálico 3, 8, 9, 99, 101, 636, 640,
658, 659, cf. também *vocalismo*
- subida de clítico, cf. *clítico*
- subordinação, cf. *adverbial*, *completiva*,
relativo/a
- sufixo 75, 104, 170–172, 206, 217, 317, 482,
655, cf. também *afixo*, *prefixo*
- sujeito 20, 54, 56, 75, 76, 166, 240, 242, 292,
310, 343, 344, 347, 351, 359–363,
372–375, 377, 378, 380–382, 384, 385,
385n, 397, 397n, 412, 413, 421, 432, 440,
441, 444–446, 453n, 454–478, 481, 482,
491, 495–498, 504, 506–513, 521,
526–528, 530, 547–549, 551, 558–562,
564–566, 568, 573–577, 621, 623
- anafórico, cf. *anafórico/a*
- argumental 559
- invertido, cf. *inversão do sujeito* / *sujeito in-*
vertido (VSO / VOS)
- não argumental 559
- nulo 3, 330, 340, 351, 353, 359, 363, 439n,
453, 455, 463, 470–472, 475, 476, 478,
481, 491, 495, 497, 498, 531n, 533, 535,
536, 548, 558–577, cf. também *pro*, *parâ-*
metro do sujeito nulo
- pós-verbal 336, 347, 351, 352, 361, 374, 375,
497, 502, 504, 506, 511–513, 521
- pronominal 360, 361, 363, 374, 558, 559, 563,
564, 565n, 566, 568–577
- quase argumental 559
- supletivismo / supletivo 158, 162, 170
- supressão de segmentos, cf. *apagamento*
- telicidade / tético 183, 309, 319, 320, 324, 325
- temática cf. *vogal temática*
- tempo (morfologia / semântica) 84, 100n, 158,
159, 169, 172, 173, 183, 190, 190n, 240,
241, 293, 294, 308, 309, 311, 312, 315,
318–327, 331–333, 376, 454, 456, 458,
458n, 461–463, 465, 469, 473, 477, 487–
489, 509, 510, 534, 539–543, 545, 571, 573,
587, 594, 610–613, 615, 616, 622, 626–633
- tempo anafórico, cf. *anafórico/a*
- tempo adjunto 324–326
- tempo verbal 321–323
- tempos compostos 19, 23, 24, 24n, 26, 528,
530, 534, 540, 544, 627
- tempo (processamento) 336, 339, 342–345,
347, 349–354, 357, 358, 360–363
- tempo (prosódia) 116, 130, 134, 136, 137, 141–
145, 147
- teoria linguística 278, 279, 280, 281, 282, 291,
292, 293, 294, 295, 296, 298, 300, 302,
341, 342
- tético (juízo ~) / teticidade 371, 373–379, 383,
396, 397
- texto 252–258, 260–262, 275
- topicalização 19, 21, 301, 371, 373, 385, 385–
396, 406, 411, 412, 574n
- tópico 142, 148, 359, 372–375, 377, 380, 381,
382, 385, 386, 389, 392–397, 406, 411,
412, 413, 467, 560, 567
- toponímia 215
- topónimo 77, 83, 214–216, 218
- traço (morfologia) cf. *género*, *número*, *pessoa*
- traço (semântico: animacidade / +humano /
definitude) 23, 25, 25n, 26, 51, 52, 54, 57,
58, 75, 165, 171, 190, 209, 217, 310, 313–
316, 329, 349, 352, 356, 360, 374, 457,

- 464, 565, 565n, 572, 573, cf. também *traços de polaridade*
- traços de polaridade 581, 590, 593–595,
cf. também *polaridade*
- traços distintivos 285, 289, 665
- traços fonológicos 99, 642
- transferência 62, 88, 90
- unidade lexical 111, 178–182, 186, 187, 189,
191, 201, 210, 211, 216–218, 324
- variação linguística 3, 13, 15, 43, 49, 50, 59, 63,
98, 99, 206, 227, 242, 294, 339, 404, 546,
558, 560, 563
- dialetal 1, 3, 5, 6, 9–12, 15, 24n, 100, 105,
340, 351, 419, 495, 498, 558, 572, 647n,
cf. também *dialetal*
- estilística 116, 117, 136–146
- fatores externos de ~ 98, 99, 106, 110
- geolinguística 98, 114, 117, 130–133
- sociolinguística / social 1–3, 6, 8, 9, 11, 12,
15, 29, 98–14, 116, 117, 130, 133–136, 148,
419
- variedades africanas do português, cf. *portu-
guês em África / variedades africanas do
português*
- variedades do português (europeu) 43, 57, 63,
81, 118, 131–133, 164n, 224, 225, 231, 236,
242, 246, 356, 401, 538, 562, 572,
cf. também *dialetal, dialeto(s)*
- verbo, cf. também *alternância de transitividade*
- auxiliar 21, 23, 24, 24n, 26, 61, 63, 239, 327,
523, 525, 525n, 527, 527n, 528, 529, 530,
534, 540, 543, 544, 588, 590, 604
- causativo 18, 20, 21, 63, 457, 458, 467, 474,
475, 477, 524–532, 534, 535, 545, 546,
548–554, 621, 623, 624, cf. também
causativa
- cópula (elipse de ~) 82
- copulativo / cópula 23, 24, 78, 79, 83, 159,
460, 594, 602n
- de controlo 456, 457, 460, 461, 462, 464,
465, 470, 471, 474, 475, 477, 478, 531, 533,
534, 538, 541, 543
- de elevação 20, 456, 466–468, 471, 477, 478,
531, 532, 534, 535, 538, 543, 545
- de movimento 54, 55, 189–191, 195, 527
- existencial 22–24, 75, 79
- *haver / ter* 19, 22–24
- impessoal 241, 463n
- inacusativo 24, 378, 491, 497, 498, 570
- leve 243
- perceptivo 18–21, 467, 468, 474, 475, 477,
478, 525, 526, 530–532, 534, 535, 545,
546, 547n, 549–554, cf. também *perceptiva
(estrutura ~)*
- *ser / estar* 23, 24, 159, 497
- *ser / haver / ter* (como auxiliares) 23, 24
- veridicidade 610, 618, 619, 621
- vibrante, cf. *consoante(s)*
- vocalismo 8, 11, 636, 640, 644, 656
- átono 7, 8, 9, 106, 110, 113, 636–638
- redução do ~ átono 2, 8, 9, 99, 104, 106,
636, 639, 640, 642, 643n, 649–655, 657–
659
- vogal, cf. *aberta, alta / altura, alternância
vocalica, anterior / anteriorização, baixa /
abaixamento, central / centralização,
elevação, fechada, grau de abertura,
inserção, labial / labialização, palatal /
palatalização, recuada / recuo, sistema
vocalico, vocalismo*
- vogal nasal 7, 8, 11, 100, 167, 169, 640n, 656,
669
- vogal temática 12, 15, 160, 163, 166, 168,
639
- vogal vazia 643
- Wordnet* 178, 181, 184, 185, 192–198, 211
- XML 233, 235, 238, 244, 267–270, 272–274

