

Normas de transcrição de corpus oral adoptadas pelo grupo Anagrama (CLUL)
 Sandra Antunes, Rita Veloso, Amália Mendes, Maria Fernanda Bacelar do Nascimento

Fenómeno	Definição/exemplos	Etiqueta
<p>pausa sintáctica longa; final de enunciado assertivo</p>	<p>Representa o fim de enunciado declarativo ou qualquer outro que não interrogativo, interrompido ou deixado em aberto. É sempre precedido por espaço.</p> <p>XYZ: estou a gostar // estou a gostar //</p> <p>Nas intervenções em que há só elementos extralinguísticos, as mesmas não terminam com nenhum símbolo. Contudo, se os mesmos estiverem inseridos no discurso, ficam entre barras simples.</p> <p>XYZ: hhh</p> <p>YWK: acho que é do / hhh / Marcelo Rebelo de Sousa //</p>	<p>//</p>
<p>interrogação</p>	<p>É sempre precedido por espaço.</p> <p>YWK: então mas que cadeiras é que tens ?</p>	<p>?</p>
<p>pausa sintáctica breve; separação de hesitações e elementos extralinguísticos (riso, tosse, etc.)</p>	<p>Representa a fronteira de uma unidade tonal. No início da fala de um informante, significa que esse acto de fala não deve ser contabilizado como um turno dialógico novo mas sim como a continuação do turno dialógico anterior do mesmo informante</p>	<p>/</p>

	<p>(independentemente de a fala anterior terminar com +, /, // ou sem símbolo).</p> <p>XYZ: no outro dia teve um ataque de / hipoglicemia na aula //</p> <p>YWK: ah / bom //</p> <p>XYZ: / e / olha / por acaso foi interessante //</p> <p>XYZ: &eh / tenho ciências da administração / também //</p> <p>XYZ: hhh / a sério //</p>	
outras pausas ou interrupções silenciosas	Representa um enunciado que o falante deixou por completar, por ser evidente a sua conclusão (como os provérbios)	...
pausas preenchidas, hesitações	São sempre seguidas do símbolo /.	<p>&ah, &eh,</p> <p>&hum</p> <p>(de acordo com a pronúncia)</p>
enunciado interrompido/abandonado	<p>XYZ: aquilo tem uma +</p> <p>YWK: construção em altura ?</p> <p>XYZ: sim //</p>	+
sobreposição de falas	<p>Vai ser marcada no Exmaralda. A anotação dependerá do formato em que se exporta.</p> <p>Ex: O output em .txt marca as sobreposições com []; o export para</p>	

	.chat marca automaticamente as sobreposições de acordo com as normas do chat, etc.	
palavra incompreensível		xxx
sequência de palavras incompreensíveis		yyyy
palavras incompletas	Transcreve-se a palavra truncada como o falante a pronuncia. A mesma é precedida de &.	&anteont &vislum
palavras mal pronunciadas	Se um informante pronunciar mal uma palavra e imediatamente a emendar, as duas formas são mantidas na transcrição (ex: <i>lugal / lugar, sau / céu</i>); se, por outro lado, o informante pronunciar mal uma palavra e não a emendar (continuando com o seu discurso), é transcrita somente a forma correcta, acrescentando-se uma nota no respectivo campo do cabeçalho a indicar que essa palavra foi mal pronunciada (ex: <i>the informant XYZ produces “banrdarilheiro” instead of bandarilheiro</i>).	
formas reduzidas	Embora o falante pronuncie uma forma reduzida (<i>tá, tou, tive</i>), as formas são transcritas na totalidade (<i>estou, está, estive</i>).	
formas truncadas	As formas truncadas são mantidas na transcrição (ex: <i>prof</i>).	

repetições de palavras	XYZ: parece que [/] que [/] que ele foi lá // Se, no entanto, a repetição incidir sobre mais do que uma palavra, estas ficarão entre parêntesis angulares (< >) XYZ: parece <que ele> [/] que ele foi lá	[/]
reformulação do discurso com a mesma ideia	XYZ: um sítio que [//] onde eu fui // Se, no entanto, a reformulação incidir sobre mais do que uma palavra, estas ficarão entre parêntesis angulares (< >) XYZ: <nós fomos> [//] eu fui lá ver como é que ele estava //	[//]
reformulação completa do discurso (diferente de discurso interrompido/abandonado voluntariamente (+))	XYZ: o problema é que nós não temos [///] o senhorio não quer vender a casa //	[///]
variantes de uma palavra	As variantes de uma palavra são transcritas desde que se encontrem registadas nos dicionários de português (ex: <i>mostrar/amostrar, louçalloiça, cacatua/catatua</i>) ou no vocabulário do Rebelo Gonçalves.	
topónimos, antropónimos, nomes de filmes, livros, jornais, programas de televisão e de rádio, bandas, músicas, entidades e edifícios legais		maiúsculas

(segue-se a tradição ortográfica)		
estrangeirismos	São transcritos segundo a ortografia da sua língua original sempre que forem produzidos de forma aproximada à da sua língua de origem (ex: <i>cachet, check-in, stress, voyeurs, workshop</i>); se as palavras estrangeiras forem pronunciadas “à portuguesa”, então são transcritas de acordo com as entradas nos dicionários de português (ex: <i>pochettezinha, shortezinho, stressado, videozinho</i>).	
acrónimos e siglas	Os acrónimos e as siglas são transcritos com letra maiúscula e sem pontos (ex: <i>TAP</i> e não <i>T.A.P.</i>); contudo, se o acrónimo já tiver uma entrada no dicionário, então é transcrito em minúsculas (ex: <i>sida, radar</i>).	
numerais e letras	Todos os numerais (incluindo os que integram datas ou percentagens) são transcritos por extenso; todas as letras são transcritas por extenso (ex: <i>pê, erre, xis</i>).	
interjeições	As interjeições são transcritas de acordo com a norma lexicográfica.	
elementos extralinguísticos (choro, riso, tosse)		hhh
onomatopeias e elementos paralinguísticos	As onomatopeias e os elementos paralinguísticos não atestados nos dicionários são transcritos de modo que representem, da forma mais	

	fiel possível, o som pronunciado (ex: <i>pac-pac-pac</i> , <i>pffff</i> , <i>tanana</i> , <i>tatata</i> , <i>uuu</i>) de acordo com o alfabeto português.	
elementos não standard ou inventados	Todos os elementos não standard ou inventados pelo informante são transcritos da forma mais fiel possível e introduzidos num glossário, que acompanha o <i>corpus</i> .	
marcação do tempo de gravação transcrito (início, fim, interrupções na transcrição; inserção de música (no caso dos media), etc....)	a indicar no metadata	
identificação dos informantes	Três letras maiúsculas (se possível, que dêem indicação do nome dos falantes) seguidas de dois pontos.	XYZ:
elementos retirados dos textos (censurados)	Introdução de <i>beep</i> no ficheiro de som. nome de pessoa = NP nome geográfico = NG nome de empresa = NE nome de barco = NB marca comercial = MC	
discurso directo	XYZ: e ele disse // “ então / não vais lá a casa ? ” (as aspas são seguidas e precedidas de espaço. Atenção: quando se fecha as aspas antes de espaço, estas ficam sempre a abrir. O melhor é inserir o espaço depois de as fechar).	“ ”

campo “observações” dos metadata	Assinalar situações não contempladas pela norma proposta ou casos de dúvida.	
-----------------------------------------	------------------------------------------------------------------------------	--