

**Proceedings of the Eleventh International Workshop
on Treebanks and Linguistic Theories (TLT11)**

30 November—1 December 2012

Lisbon, Portugal

Editors:

Iris Hendrickx

Sandra Kübler

Kiril Simov

Title: Proceedings of the Eleventh International Workshop
on Treebanks and Linguistic Theories (TLT11)

Editors: Iris Hendrickx, Sandra Kübler,
Kiril Simov

Cover photograph: Pedro Salitre

ISBN 978-989-689-274-6

Depósito legal n.º 351 424/12

Publisher: Edições Colibri, Lisboa
www.edi-colibri.pt

Sponsor:



Faculdade de Letras
da Universidade de Lisboa

Lisbon, November 2012

Preface

The 11th International Workshop on Treebanks and Linguistic Theories is held in Lisbon, Portugal. When the first TLT took place in Sozopol, Bulgaria, it was not clear whether there would be a second workshop. Now, we can look back on more than 10 years of successful workshops and of research on treebanks in a linguistic context.

There are several directions in which research on treebanks has made considerable progress:

- Treebanks have evolved from a necessary resource for NLP applications to a field of research on their own. They are used for parser training as well as for linguistic investigations. Additionally, annotation issues often also become research topics.
- The field also has evolved from a situation where treebanks were available for only a handful of the major languages to one where more and more treebanks for lesser-studied languages become available. This TLT is a good indicator for this development: it features papers on Ancient Greek, Basque, Czech, Bangla, Bulgarian, Danish, Dutch, English, French, German, Hindi, Italian, Norwegian, Persian, Portuguese, Swedish, Telugu, and Urdu.
- Treebanks have also broadened in the spectrum of linguistic phenomena that are tackled: While the first treebanks were restricted to syntactic information, today there is a wide range of other linguistic phenomena that are annotated on top of the syntactic annotations, or parallel to them. This year's TLT features talks concerning the annotation of semantics, coreference, named entities, and discourse structure.
- Another development worth noticing is the emergence of parallel treebanks, which will have an influence on machine translation as well as on linguistic investigations, as illustrated in three contributions to TLT 11.

TLT 11 features a well-rounded program that provides contributions to all these areas. This year, we had 32 submissions, out of which 19 were accepted, either as oral presentation or as poster.

TLT aims at being a forum for all researchers and students working in the area of treebanking. To complete the picture, Mark Steedman and Nianwen Xue accepted the invitation to be keynote speakers at the workshop. We hope that you will enjoy the workshop and the proceedings.

Iris Hendrickx, Sandra Kübler, Kiril Simov

Workshop Organization

Program Chairs

Iris Hendrickx, University of Lisbon, Portugal
Sandra Kübler, Indiana University, USA
Kiril Simov, Bulgarian Academy of Sciences, Bulgaria

Program Committee

Eckhard Bick, University of Southern Denmark, Denmark
Johan Bos, University of Amsterdam, The Netherlands
Gosse Bouma, University of Groningen, The Netherlands
António Branco, University of Lisbon, Portugal
Ernestina Carrilho, University of Lisbon, Portugal
Koenraad De Smedt, Bergen University, Norway
Markus Dickinson, Indiana University, USA
Stefanie Dipper, Bochum University, Germany
Dan Flickinger, Stanford University, USA
Anette Frank, Heidelberg University, Germany
Eva Hajičová, Charles University, Czech Republic
Erhard Hinrichs, University of Tübingen, Germany
Julia Hockenmaier, University of Illinois, USA
Valia Kordoni, Saarland University, Germany
Nuno Mamede, IST / INESC-ID, Portugal
Amália Mendes, University of Lisbon, Portugal
Detmar Meurers, University of Tübingen, Germany
Yusuke Miyao, University of Tokyo, Japan
Kaili Müürisep, Tartu University, Estonia
Kemal Oflazer, Carnegie Mellon University, Qatar
Sebastian Padó, Heidelberg University, Germany
Marco Passarotti, Catholic University of the Sacred Heart, Italy
Petya Osenova, Sofia University, Bulgaria
Adam Przepiórkowski, Polish Academy of Sciences, Poland
Victoria Rosén, Bergen University, Norway
Caroline Sporleder, Saarland University, Germany
Manfred Stede, University of Potsdam, Germany
Gertjan van Noord, University of Groningen, The Netherlands
Martin Volk, University of Zurich, Switzerland
Heike Zinsmeister, Konstanz University, Germany

Local Committee

Amália Mendes, CLUL, University of Lisbon, Portugal

Iris Hendrickx, CLUL, University of Lisbon, Portugal

Sandra Antunes, CLUL, University of Lisbon, Portugal

Aida Cardoso, CLUL, University of Lisbon, Portugal

Table of Contents

Liesbeth Augustinus, Frank Van Eynde: A Treebank-based Investigation of IPP-triggering Verbs in Dutch	7
Kathrin Beck, Erhard W. Hinrichs. Profiling Feature Selection for Named Entity Classification in the TüBa-D/Z Treebank	13
Riyaz Ahmad Bhat, Dipti Mishra Sharma: Non-Projective Structures in Indian Language Treebanks	25
Riyaz Ahmad Bhat, Sambhav Jain, Dipti Misra Sharma: Experiments on Dependency Parsing of Urdu	31
Sonja Bosch, Key-Sun Choi, Éric de La Clergerie, Alex Chengyu Fang, Gertrud Faass, Kiyong Lee, Antonio Pareja-Lora, Laurent Romary, Andreas Witt, Amir Zeldes, Florian Zipser: Tiger2 as a Standardised Serialisation for ISO 24615 – SynAF	37
Marie Candito, Djamé Seddah: Effectively Long-distance Dependencies in French: Annotation and Parsing Evaluation	61
Rodolfo Delmonte: Logical Form Representation for Linguistic Resources	73
Dan Flickinger, Valia Kordoni, Yi Zhang. DeepBank: A Dynamically Annotated Treebank of the Wall Street Journal	85
Dan Flickinger, Valia Kordoni, Yi Zhang, António Branco, Kiril Simov, Petya Osenova, Catarina Carneiro, Francisco Costa, Sérgio Castro: ParDeepBank: Multiple Parallel Deep Treebanking	97
Masood Ghayoomi, Omid Moradiannasab: The Effect of Fine- and Coarse-grained Treebank Annotation on Parsing: A Comparative Study	109
Iakes Goenaga, Olatz Arregi, Klara Ceberio, Arantza Diaz de Ilarraza, Amane Jimeno: Automatic Coreference Annotation in Basque	115
Pavla Jínová, Jiří Mirovský, Lucie Poláková: Analyzing the Most Common Errors in the Discourse Annotation of the Prague Dependency Treebank	127
Francesco Mambrini, Marco Passarotti: Will a Parser Overtake Achilles? First Experiments on Parsing the Ancient Greek Dependency Treebank	133
Magdalena Plamada, Martin Volk: Using Parallel Treebanks for Machine Translation Evaluation	145
Victoria Rosén, Paul Meurer, Gyri Smørdal Losnegaard, Gunn Inger	

Lyse, Koenraad De Smedt, Martha Thunes, Helge Dyvik: An integrated web-based treebank annotation system	157
Manuela Sanguinetti, Cristina Bosco: Translational Divergences and Their Alignment	169
Djamé Seddah, Benoît Sagot, Marie Candito, Virginie Moulleron, Vanessa Combet: Building a Treebank of Noisy User Generated Content: The French Social Media Bank	181
Arne Skjærholt, Lilja Øvrelid: Impact of Treebank Characteristics on Cross-lingual Parser Adaptation	187
Nitesh Surtani, Soma Paul: Genitives in Hindi Treebank: An Attempt for Automatic Annotation	199