

TLT16
January 2018

Proceedings of the 16th International Workshop on
Treebanks and Linguistic Theories

January 23–24, 2018
Prague, Czech Republic

ISBN: 978-80-88132-04-2

Edited by Jan Hajič.

These proceedings with papers presented at the 16th International Workshop on Treebanks and Linguistic Theories are included in the ACL Anthology, supported by the Association for Computational Linguistics (ACL) at <http://aclweb.org/anthology>; both individual papers as well as the full proceedings book are available.

© 2017. Copyright of each paper stays with the respective authors (or their employers).
Distributed under a CC-BY 4.0 licence.

TLT16 takes place in Prague, Czech Republic on January 23–24, 2018.

Preface

The Sixteenth International Workshop on Treebanks and Linguistic Theories (TLT16) is being held at Charles University, Czech Republic, 23–24 January 2018, for the second time in Prague, which hosted TLT already in 2006. This year, TLT16 is co-located with the Workshop on Data Provenance and Annotation in Computational Linguistics 2018 (January 22, 2018, at the same place and venue) and immediately followed by the 2nd Workshop on Corpus-based Research in the Humanities, held in Vienna, Austria.

This year, TLT16 received 32 submissions of which 15 have been selected to be presented as oral presentations, and additional seven have been asked to present as posters. We were happy that our invitation to give a plenary talk has been accepted by both Lilja Øvrelid of University of Oslo in Norway (with a talk “Downstream use of syntactic analysis: does representation matter?”) and Marie Candito, of University of Paris Diderot, France (“Annotating and parsing to semantic frames: feedback from the French FrameNet project”). We have also included a panel discussion on the very topic of Treebanks and Linguistic Theories – to discuss opinions of the future of the field and the workshop with the participants.

We are grateful to the members of the program committee, who worked hard to review the submissions and provided authors with valuable feedback. We would also like to thank various sponsors, mainly the Faculty of Mathematics and Physics of Charles University, providing local and logistical and accounting and financial support, and the Centre for Advanced Study (CAS) at the Norwegian Academy of Science and Letters for supporting one of the invited speakers.

Last but not least, we would like to thank all authors for submitting interesting and relevant papers, and we wish all participants a fruitful workshop.

December 2017

Jan Hajič (also in the name of local organizers), Charles University, Prague, Czech Republic
Sandra Kuebler, Indiana University, Bloomington, Indiana, USA
Stephan Oepen, University of Oslo, Norway, and CAS, Norwegian Academy of Science and Letters, Oslo
Markus Dickinson, Indiana University, Bloomington, Indiana, USA

Program Committee Co-chairs

Program Chairs

Jan Hajič	Charles University, Prague
Sandra Kübler	Indiana University Bloomington
Stephan Oepen	Universitetet i Oslo
Markus Dickinson	Indiana University Bloomington

Reviewers

Patricia Amaral	Indiana University Bloomington
Emily M. Bender	University of Washington
Eckhard Bick	University of Southern Denmark
Ann Bies	Linguistic Data Consortium, University of Pennsylvania
Gosse Bouma	Rijksuniversiteit Groningen
Miriam Butt	Konstanz
Silvie Cinková	Charles University, Prague
Marie-Catherine de Marneffe	The Ohio State University
Koenraad De Smedt	University of Bergen
Tomaž Erjavec	Dept. of Knowledge Technologies, Jožef Stefan Institute
Filip Ginter	University of Turku
Memduh Gokirmak	Istanbul Technical University
Eva Hajicova	Charles University, Prague
Dag Haug	University of Oslo
Barbora Hladka	Charles University, Prague
Lori Levin	Carnegie Mellon University
Teresa Lynn	Dublin City University
Adam Meyers	New York University
Jiří Mirovský	Charles University, Prague
Kaili Müürisep	University of Tartu
Joakim Nivre	Uppsala University
Petya Osenova	Sofia University and IICT-BAS
Agnieszka Patejuk	Institute of Computer Science, Polish Academy of Sciences
Tatjana Scheffler	Universität Potsdam
Olga Scriver	Indiana University Bloomington
Djamé Seddah	Alpage/Université Paris la Sorbonne
Michael White	The Ohio State University
Nianwen Xue	Brandeis University
Daniel Zeman	Charles University, Prague
Heike Zinsmeister	University of Hamburg

Local Organizing Committee

Jan Hajič (chair)	Charles University, Prague
Kateřina Bryanová	Charles University, Prague
Zdeňka Urešová	Charles University, Prague
Eduard Bejček (publication chair)	Charles University, Prague

Table of Contents

Invited Speakers	
Annotating and parsing to semantic frames: feedback from the French FrameNet project	v
<i>Marie Candito</i>	
Downstream use of syntactic analysis: does representation matter?	vi
<i>Lilja Øvrelid</i>	
<hr/>	
Tuesday, 23rd January	
Distributional regularities of verbs and verbal adjectives: Treebank evidence and broader implications	1
<i>Daniël de Kok, Patricia Fischer, Corina Dima and Erhard Hinrichs</i>	
UD Annotatrix: An annotation tool for Universal Dependencies	10
<i>Francis M. Tyers, Maria Sheyanova and Jonathan Washington</i>	
The Treebanked Conspiracy. Actors and Actions in Bellum Catilinae	18
<i>Marco Passarotti and Berta Gonzalez Saavedra</i>	
Universal Dependencies-based syntactic features in detecting human translation varieties	27
<i>Maria Kunilovskaya and Andrey Kutuzov</i>	
Graph Convolutional Networks for Named Entity Recognition	37
<i>Alberto Cetoli, Stefano Bragaglia, Andrew O’Harney and Marc Sloan</i>	
Extensions to the GrETEL Treebank Query Application	46
<i>Jan Odijk, Martijn van der Klis and Sheean Spoel</i>	
The Relation of Form and Function in Linguistic Theory and in a Multilayer Treebank	56
<i>Eduard Bejček, Eva Hajičová, Marie Mikulová and Jarmila Panevová</i>	
Literal readings of multiword expressions: as scarce as hen’s teeth	64
<i>Agata Savary and Silvio Ricardo Cordeiro</i>	
Querying Multi-word Expressions Annotation with CQL	73
<i>Natalia Klyueva, Anna Vernerova and Behrang Qasemizadeh</i>	
REALEC learner treebank: annotation principles and evaluation of automatic parsing	80
<i>Olga Lyashevskaya and Irina Panteleeva</i>	
A semiautomatic lemmatisation procedure for treebanks. Old English strong and weak verbs	88
<i>Marta Tío Sáenz and Darío Metola Rodríguez</i>	
Data point selection for genre-aware parsing	95
<i>Ines Rehbein and Felix Bildhauer</i>	
Error Analysis of Cross-lingual Tagging and Parsing	106
<i>Rudolf Rosa and Zdeněk Žabokrtský</i>	
<hr/>	
Wednesday, 24th January	
A Telugu treebank based on a grammar book	119
<i>Taraka Rama and Sowmya Vajjala</i>	
Recent Developments within BulTreeBank	129
<i>Petya Osenova and Kiril Simov</i>	

Towards a dependency-annotated treebank for Bambara	138
<i>Ekaterina Aplonova and Francis M. Tyers</i>	
Merging the Trees – Building a Morphological Treebank for German from Two Resources	146
<i>Petra Steiner</i>	
What I think when I think about treebanks	161
<i>Anders Søgaard</i>	
Syntactic Semantic Correspondence in Dependency Grammar	167
<i>Catalina Maranduc, Catalin Mititelu and Victoria Bobicev</i>	
Multi-word annotation in syntactic treebanks – Propositions for Universal Dependencies	181
<i>Sylvain Kahane, Marine Courtin and Kim Gerdes</i>	
A Universal Dependencies Treebank for Marathi	190
<i>Vinit Ravishankar</i>	
Dangerous Relations in Dependency Treebanks	201
<i>Chiara Alzetta, Felice Dell’Orletta, Simonetta Montemagni and Giulia Venturi</i>	