

EMNLP 2018

Second Workshop on Universal Dependencies (UDW 2018)

Proceedings of the Workshop

November 1, 2018
Brussels, Belgium

Sponsored by:



©2018 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-948087-78-0

Preface

These proceedings include the program and papers that are presented at the second workshop on Universal Dependencies, held in conjunction with EMNLP in Brussels (Belgium) on November 1, 2018.

Universal Dependencies (UD) is a framework for cross-linguistically consistent treebank annotation that has so far been applied to over 70 languages (<http://universaldependencies.org/>). The framework is aiming to capture similarities as well as idiosyncrasies among typologically different languages (e.g., morphologically rich languages, pro-drop languages, and languages featuring clitic doubling). The goal in developing UD was not only to support comparative evaluation and cross-lingual learning but also to facilitate multilingual natural language processing and enable comparative linguistic studies.

After a successful first UD workshop at NoDaLiDa in Gothenburg last year, we decided to continue to bring together researchers working on UD, to reflect on the theory and practice of UD, its use in research and development, and its future goals and challenges.

We received 39 submissions of which 26 were accepted. Submissions covered several topics: some papers describe treebank conversion or creation, while others target specific linguistic constructions and which analysis to adopt, sometimes with critiques of the choices made in UD; some papers exploit UD resources for cross-linguistic and psycholinguistic analysis, or for parsing, and others discuss the relation of UD to different frameworks.

We are honored to have two invited speakers: Barbara Plank (Computer Science Department, IT University of Copenhagen), with a talk on “Learning χ^2 – Natural Language Processing Across Languages and Domains”, and Dag Haug (Department of Philosophy, Classics, History of Arts and Ideas, University of Oslo), speaking about “Glue semantics for UD”. Our invited speakers target different aspects of UD in their work: Barbara Plank’s talk is an instance of how UD facilitates cross-lingual learning and transfer for NLP components, whereas Dag Haug will address how UD and semantic formalisms can intersect.

We are grateful to the program committee, who worked hard and on a tight schedule to review the submissions and provided authors with valuable feedback. We thank Google, Inc. for its sponsorship which made it possible to feature two invited talks. We also want to thank Jan Hajic for giving us the impetus to put together and submit a workshop proposal to the ACL workshops, Sampo Pyysalo for his invaluable help with the website and prompt reactions as always, and Joakim Nivre for his constant support and helpful suggestions on the workshop organization.

We wish all participants a productive workshop!

Marie-Catherine de Marneffe, Teresa Lynn and Sebastian Schuster

Workshop Co-Chairs:

Marie-Catherine de Marneffe, The Ohio State University, USA
Teresa Lynn, Dublin City University, Ireland
Sebastian Schuster, Stanford University, USA

Organizers:

Joakim Nivre, Uppsala University, Sweden
Filip Ginter, University of Turku, Finland
Yoav Goldberg, Bar Ilan University, Israel
Jan Hajic, Charles University in Prague, Czech Republic
Sampo Pyysalo, University of Cambridge, UK
Reut Tsarfaty, Open University of Israel, Israel
Francis Tyers, Higher School of Economics, Moscow, Russia
Dan Zeman, Charles University in Prague, Czech Republic

Program Committee:

Željko Agić, IT University of Copenhagen, Denmark
Marie Candito, Université Paris Diderot, France
Giuseppe Celano, University of Leipzig, Germany
Çağrı Çöltekin, Tübingen, Germany
Miryam de Lhoneux, Uppsala University, Sweden
Tim Dozat, Stanford University, USA
Kaja Dobrovoljc, University of Ljubljana, Slovenia
Jennifer Foster, Dublin City University, Ireland
Kim Gerdes, Sorbonne nouvelle Paris 3, France
Koldo Gojenola, Euskal Herriko Unibertsitatea, Spain
Sylvain Kahane, Université Paris Ouest - Nanterre, France
Natalia Kotsyba, Polish Academy of Sciences, Poland
John Lee, City University of Hong Kong, Hong Kong
Alessandro Lenci, University of Pisa, Italy
Christopher D. Manning, Stanford University, USA
Héctor Martínez Alonso INRIA - Paris 7, France
Ryan McDonald, Google, UK
Simonetta Montemagni, CNR, Italy
Lilja Ovreliid, University of Oslo, Norway
Martin Popel, Charles University, Czech Republic
Peng Qi, Stanford University, USA
Siva Reddy, Stanford University, USA
Rudolf Rosa, Charles University in Prague, Czech Republic
Petya Osenova, Bulgarian Academy of Sciences, Bulgaria
Tanja Samardžić, University of Zurich, Switzerland
Nathan Schneider, Georgetown University, USA
Djamé Seddah, INRIA/ Université Paris 4 La Sorbonne, France
Maria Simi, Università di Pisa, Italy
Zdeněk Žabokrtský, Charles University in Prague, Czech Republic
Amir Zeldes, Georgetown University, USA

Invited Speakers:

Barbara Plank, IT University of Copenhagen, Denmark
Dag Haug, University of Oslo, Norway

Table of Contents

<i>Assessing the Impact of Incremental Error Detection and Correction. A Case Study on the Italian Universal Dependency Treebank</i>	
Chiara Alzetta, Felice Dell’Orletta, Simonetta Montemagni, Maria Simi and Giulia Venturi	1
<i>Using Universal Dependencies in cross-linguistic complexity research</i>	
Aleksandrs Berdicevskis, Çağrı Çöltekin, Katharina Ehret, Kilu von Prince, Daniel Ross, Bill Thompson, Chunxiao Yan, Vera Demberg, Gary Lupyán, Taraka Rama and Christian Bentz	8
<i>Expletives in Universal Dependency Treebanks</i>	
Gosse Bouma, Jan Hajic, Dag Haug, Joakim Nivre, Per Erik Solberg and Lilja Øvrelid	18
<i>Challenges in Converting the Index Thomisticus Treebank into Universal Dependencies</i>	
Flavio Massimiliano Cecchini, Marco Passarotti, Paola Marongiu and Daniel Zeman	27
<i>Er ... well, it matters, right? On the role of data representations in spoken language dependency parsing</i>	
Kaja Dobrovoljc and Matej Martinc	37
<i>Mind the Gap: Data Enrichment in Dependency Parsing of Elliptical Constructions</i>	
Kira Droganova, Filip Ginter, Jenna Kanerva and Daniel Zeman	47
<i>Integration complexity and the order of cosisters</i>	
William Dyer	55
<i>SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD</i>	
Kim Gerdes, Bruno Guillaume, Sylvain Kahane and Guy Perrier	66
<i>Coordinate Structures in Universal Dependencies for Head-final Languages</i>	
Hiroshi Kanayama, Na-Rae Han, Masayuki Asahara, Jena D. Hwang, Yusuke Miyao, Jinho D. Choi and Yuji Matsumoto	75
<i>Investigating NP-Chunking with Universal Dependencies for English</i>	
Ophélie Lacroix	85
<i>Marrying Universal Dependencies and Universal Morphology</i>	
Arya D. McCarthy, Miikka Silfverberg, Ryan Cotterell, Mans Hulden and David Yarowsky	91
<i>Enhancing Universal Dependency Treebanks: A Case Study</i>	
Joakim Nivre, Paola Marongiu, Filip Ginter, Jenna Kanerva, Simonetta Montemagni, Sebastian Schuster and Maria Simi	102
<i>Enhancing Universal Dependencies for Korean</i>	
Youngbin Noh, Jiyoung Han, Tae Hwan Oh and Hansaem Kim	108
<i>UD-Japanese BCCWJ: Universal Dependencies Annotation for the Balanced Corpus of Contemporary Written Japanese</i>	
Mai Omura and Masayuki Asahara	117
<i>The First Komi-Zyrian Universal Dependencies Treebanks</i>	
Niko Partanen, Rogier Blokland, KyungTae Lim, Thierry Poibeau and Michael Rießler	126
<i>The Hebrew Universal Dependency Treebank: Past Present and Future</i>	
Shoval Sade, Amit Seker and Reut Tsarfaty	133

<i>Multi-source synthetic treebank creation for improved cross-lingual dependency parsing</i> Francis Tyers, Mariya Sheyanova, Aleksandra Martynova, Pavel Stepachev and Konstantin Vinogradskiy	144
<i>Toward Universal Dependencies for Shipibo-Konibo</i> Alonso Vásquez, Renzo Ego Aguirre, Candy Angulo, John Miller, Claudia Villanueva, Željko Agić, Roberto Zariquiey and Arturo Oncevay	151
<i>Transition-based Parsing with Lighter Feed-Forward Networks</i> David Vilares and Carlos Gómez-Rodríguez	162
<i>Extended and Enhanced Polish Dependency Bank in Universal Dependencies Format</i> Alina Wróblewska	173
<i>Approximate Dynamic Oracle for Dependency Parsing with Reinforcement Learning</i> Xiang Yu, Ngoc Thang Vu and Jonas Kuhn	183
<i>The Coptic Universal Dependency Treebank</i> Amir Zeldes and Mitchell Abrams	192

Workshop Program

Thursday, November 1, 2018

9:00–10:30 **Opening, Invited Talk & Oral Presentations 1**

9:00–9:10 *Opening*

9:10–10:00 *Invited Talk: **Glue semantics for UD***
Dag Haug

10:00–10:15 *Using Universal Dependencies in cross-linguistic complexity research*
Aleksandrs Berdicevskis, Çağrı Çöltekin, Katharina Ehret, Kilu von Prince, Daniel Ross, Bill Thompson, Chunxiao Yan, Vera Demberg, Gary Lupyan, Taraka Rama and Christian Bentz

10:15–10:30 *Integration complexity and the order of cosisters*
William Dyer

10:30–11:00 **Coffee Break**

11:00–12:30 **Poster Session**

From LFG to Enhanced Universal Dependencies (in LFG 2018 and LAW-MWE-CxG-2018)
Adam Przepiórkowski and Agnieszka Patejuk

Approximate Dynamic Oracle for Dependency Parsing with Reinforcement Learning
Xiang Yu, Ngoc Thang Vu and Jonas Kuhn

Transition-based Parsing with Lighter Feed-Forward Networks
David Vilares and Carlos Gómez-Rodríguez

UD-Japanese BCCWJ: Universal Dependencies Annotation for the Balanced Corpus of Contemporary Written Japanese
Mai Omura and Masayuki Asahara

Challenges in Converting the Index Thomisticus Treebank into Universal Dependencies
Flavio Massimiliano Cecchini, Marco Passarotti, Paola Marongiu and Daniel Zeman

Thursday, November 1, 2018 (continued)

Investigating NP-Chunking with Universal Dependencies for English

Ophélie Lacroix

Extended and Enhanced Polish Dependency Bank in Universal Dependencies Format

Alina Wróblewska

Mind the Gap: Data Enrichment in Dependency Parsing of Elliptical Constructions

Kira Droганova, Filip Ginter, Jenna Kanerva and Daniel Zeman

The Coptic Universal Dependency Treebank

Amir Zeldes and Mitchell Abrams

Parsing Japanese Tweets into Universal Dependencies (non-archival submission)

Hayate Iso, Kaoru Ito, Hiroyuki Nagai, Taro Okahisa and Eiji Aramaki

Toward Universal Dependencies for Shipibo-Konibo

Alonso Vásquez, Renzo Ego Aguirre, Candy Angulo, John Miller, Claudia Villanueva, Željko Agić, Roberto Zariquiey and Arturo Oncevay

All Roads Lead to UD: Converting Stanford and Penn Parses to English Universal Dependencies with Multilayer Annotations (in LAW-MWE-CxG-2018)

Siyao Peng and Amir Zeldes

The First Komi-Zyrian Universal Dependencies Treebanks

Niko Partanen, Rogier Blokland, KyungTae Lim, Thierry Poibeau and Michael Rießler

The Hebrew Universal Dependency Treebank: Past Present and Future

Shoval Sade, Amit Seker and Reut Tsarfaty

Enhancing Universal Dependencies for Korean

Youngbin Noh, Jiyoung Han, Tae Hwan Oh and Hansaem Kim

Multi-source synthetic treebank creation for improved cross-lingual dependency parsing

Francis Tyers, Mariya Sheyanova, Aleksandra Martynova, Pavel Stepachev and Konstantin Vinogorodskiy

Thursday, November 1, 2018 (continued)

12:30–14:00 Lunch Break

14:00–15:35 Invited Talk & Oral Presentations 2

14:00–14:50 *Invited Talk: Learning X^2 – Natural Language Processing Across Languages and Domains*
Barbara Plank

14:50–15:05 *Er ... well, it matters, right? On the role of data representations in spoken language dependency parsing*
Kaja Dobrovoljc and Matej Martinc

15:05–15:20 *Assessing the Impact of Incremental Error Detection and Correction. A Case Study on the Italian Universal Dependency Treebank*
Chiara Alzetta, Felice Dell’Orletta, Simonetta Montemagni, Maria Simi and Giulia Venturi

15:20–15:35 *Enhancing Universal Dependency Treebanks: A Case Study*
Joakim Nivre, Paola Marongiu, Filip Ginter, Jenna Kanerva, Simonetta Montemagni, Sebastian Schuster and Maria Simi

15:35–16:00 Coffee Break

16:00–17:30 Oral Presentations 3 & Closing

16:00–16:15 *Marrying Universal Dependencies and Universal Morphology*
Arya D. McCarthy, Miikka Silfverberg, Ryan Cotterell, Mans Hulden and David Yarowsky

16:15–16:30 *Arguments and Adjuncts in Universal Dependencies* (in Coling 2018)
Adam Przepiórkowski and Agnieszka Patejuk

16:30–16:45 *Expletives in Universal Dependency Treebanks*
Gosse Bouma, Jan Hajic, Dag Haug, Joakim Nivre, Per Erik Solberg and Lilja Øvreid

16:45–17:00 *Coordinate Structures in Universal Dependencies for Head-final Languages*
Hiroshi Kanayama, Na-Rae Han, Masayuki Asahara, Jena D. Hwang, Yusuke Miyao, Jinho D. Choi and Yuji Matsumoto

17:00–17:15 *SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD*
Kim Gerdes, Bruno Guillaume, Sylvain Kahane and Guy Perrier

17:15–17:30 Concluding Remarks

Invited Talk: Dag Haug, University of Oslo

Glue semantics for UD

The success of the Universal Dependencies initiative has spurred interest in deriving semantic structures from UD trees. The challenge is to do this while relying as little as possible on language-specific, typically lexical resources that are not available for many of the 60 languages for which there are UD treebanks. In this talk I outline an approach to this problem that builds on techniques developed for LFG + Glue. There are several motivations for this: First, LFG's f-structures track the same aspect of syntactic structure as UD dependency trees. Second, the particular version of dependency grammar that UD embodies has inherited much from LFG via the Stanford Dependencies and the PARC dependencies. Third, unlike many other approaches, LFG + Glue does not assume a one-to-one mapping from syntactic to semantic structures but instead develops a syntax-semantics interface that can map a single syntactic structure to several meaning representations, i.e. the syntax underspecifies the semantics, which is useful when dealing with the lack of information one often encounters in UD trees. In the talk, I will present the theoretical background for UD + Glue and discuss some issues that arose in the development of a proof of concept implementation of the framework.

Bio

Dag Haug is professor of classics and linguistics at the University of Oslo. He has worked extensively in theoretical syntax (mainly Lexical-Functional Grammar) and formal semantics. He has also led various treebanking efforts for ancient languages, which among other things have resulted in the UD treebanks for Ancient Greek, Latin, Old Church Slavonic and Gothic.

Invited Talk: Barbara Plank, IT University of Copenhagen

Learning X^2 – Natural Language Processing Across Languages and Domains

How can we build Natural Language Processing models for new domains and new languages? In this talk I will survey some recent advances to address this challenge, from multi-task learning, data selection, cross-lingual transfer to learning models under distant supervision from disparate sources, and outline open challenges. The talk will focus on two target applications: part-of-speech tagging and dependency parsing.

Bio

Barbara Plank is associate professor at ITU (IT University of Copenhagen), Denmark. She holds a BSc and MSc in Computer Science and received her PhD in Computational Linguistics in 2011. Originally from South Tyrol, Italy, she worked and lived in the Netherlands, Italy and Denmark. Barbara is interested in robust language technology, in particular cross-domain and cross-language learning, learning under annotation bias, and generally, semi-supervised and weakly-supervised machine learning for a broad range of NLP applications, including syntactic parsing, author profiling, opinion mining and information and relation extraction.