

COLING 2020

**Fourth Workshop on Universal Dependencies (UDW 2020)**

**Proceedings of the Workshop**

December 13, 2020  
Barcelona, Spain (Online)

**Sponsored by:**



Copyright of each paper stays with the respective authors (or their employers).

ISBN 978-1-952148-48-4

## Preface

These proceedings include the program and papers that are presented at the fourth workshop on Universal Dependencies, held in conjunction with COLING online on December 13, 2020.

Universal Dependencies (UD) is a framework for cross-linguistically consistent treebank annotation that has so far been applied to over 90 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 three successful editions of the workshop, 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 33 submissions of which 24 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 historical analysis, or for parsing, and some develop tools using UD.

We are honored to have an invited speaker: Martha Palmer (Department of Linguistics, University of Colorado at Boulder), with a talk on “Transcending Dependencies” which talks about contextual interpretation in dialogues and the role of Abstract Meaning Representation in this context.

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. and the National Science Foundation (NSF) for grants which allowed us to cover registration fees of some of the participants.

We wish all participants a productive workshop!

Marie-Catherine de Marneffe, Miryam de Lhoneux, Joakim Nivre and Sebastian Schuster



**Workshop Co-Chairs:**

Marie-Catherine de Marneffe, The Ohio State University, USA  
Miryam de Lhoneux, University of Copenhagen, Denmark  
Joakim Nivre, Uppsala University, Sweden  
Sebastian Schuster, Stanford University, USA

**Program Committee:**

Željko Agić, IT University of Copenhagen, Denmark  
Emily Bender, University of Washington, USA  
Marcel Bollman, University of Copenhagen, Denmark  
Gosse Bouma, University of Groningen, The Netherlands  
Marie Candito, Université Paris Diderot, France  
Giuseppe Celano, University of Leipzig, Germany  
Çağrı Çöltekin, Tübingen, Germany  
Valeria dePaiva, Samsung Research, USA  
Timothy Dozat, Google, USA  
Kira Droganova, Charles University Prague, Czech Republic  
Richard Futrell, University of California Irvine, USA  
Kim Gerdes, Sorbonne University, France  
Carlos Gómez-Rodríguez, University of Coruña, Spain  
Michael Hahn, Stanford University, USA  
Jan Hajic, Charles University Prague, Czech Republic  
Johannes Heinecke, Orange Labs  
Sylvain Kahane, Université Paris Ouest - Nanterre, France  
Arne Köhn, University of Hamburg, Germany  
Natalia Kotsyba, Polish Academy of Sciences, Poland  
John Lee, City University of Hong Kong, Hong Kong  
Teresa Lynn, Dublin City University, Ireland  
Arya McCarthy, John Hopkins University, USA  
Stephan Oepen, University of Oslo, Norway  
Lilja Øvrelid, University of Oslo, Norway  
Guy Perrier, University of Lorraine, France  
Tommi Pirinen, University of Hamburg, Germany  
Martin Popel, Charles University, Czech Republic  
Sampo Pyysalo, University of Cambridge, UK  
Peng Qi, Stanford University, USA  
Alexandre Rademaker, IBM Research, Brazil  
Rudolf Rosa, Charles University in Prague, Czech Republic  
Tanja Samardžić, University of Zurich, Switzerland  
Nathan Schneider, Georgetown University, USA  
Francis Tyers, Indiana University, USA  
Zdeněk Žabokrtský, Charles University in Prague, Czech Republic  
Amir Zeldes, Georgetown University, USA  
Daniel Zeman, Charles University Prague, Czech Republic

**Invited Speakers:**

Martha Palmer, University of Colorado at Boulder, USA



## Table of Contents

<i>A small Universal Dependencies treebank for Hittite</i> Erik Andersen and Benjamin Rozonoyer .....	1
<i>Parsing in the absence of related languages: Evaluating low-resource dependency parsers on Tagalog</i> Angelina Aquino and Franz de Leon .....	8
<i>A Universal Dependencies conversion pipeline for a Penn-format constituency treebank</i> Pórunn Arnardóttir, Hinrik Hafsteinsson, Einar Freyr Sigurðsson, Kristín Bjarnadóttir, Anton Karl Ingason, Hildur Jónsdóttir and Steinþór Steingrímsson .....	16
<i>Corpus evidence for word order freezing in Russian and German</i> Aleksandrs Berdicevskis and Alexander Piperski .....	26
<i>Subjecthood and annotation: The cases of French and Wolof</i> Olivier Bondéelle and Sylvain Kahane .....	34
<i>Verification, reproduction and replication of NLP experiments: A case study on parsing Universal Dependencies</i> Çağrı Çöltekin .....	46
<i>From LFG To UD: A combined approach</i> Cheikh M. Bamba Dione .....	57
<i>Identifying and handling cross-treebank inconsistencies in UD: A pilot study</i> Tillmann Dönicke, Xiang Yu and Jonas Kuhn .....	67
<i>Composing byte-pair encodings for morphological sequence classification</i> Adam Ek and Jean-Philippe Bernardy .....	76
<i>Configurable dependency tree extraction from CCG derivations</i> Kilian Evang .....	87
<i>Unifying the treatment of preposition-determiner contractions in German Universal Dependencies treebanks</i> Stefan Grünewald and Annemarie Friedrich .....	94
<i>Annotation issues in Universal Dependencies for Korean and Japanese</i> Ji Yoon Han, Tae Hwan Oh, LEE JIN and Hansaem Kim .....	99
<i>Exploring diachronic syntactic shifts with dependency length: The case of scientific English</i> Tom S Juzek, Marie-Pauline Krielke and Elke Teich .....	109
<i>UDon2: A library for manipulating Universal Dependencies trees</i> Dmytro Kalpakchi and Johan Boye .....	120
<i>Annotating MWEs in the Irish UD treebank</i> Sarah McGuinness, Jason Phelan, Abigail Walsh and Teresa Lynn .....	126
<i>I've got a construction looks funny – representing and recovering non-standard constructions in UD</i> Josef Ruppenhofer and Ines Rehbein .....	140
<i>Universal Dependencies for Manx Gaelic</i> Kevin Scannell .....	152

<i>Variation in Universal Dependencies annotation: A token-based typological case study on adpossession constructions</i>	
Kaius Sinnemäki and Viljami Haakana . . . . .	158
<i>PALMYRA 2.0: A configurable multilingual platform independent tool for morphology and syntax annotation</i>	
Dima Taji and Nizar Habash . . . . .	168
<i>Universal Dependencies for Albanian</i>	
Marsida Toska, Joakim Nivre and Daniel Zeman . . . . .	178
<i>First Steps towards Universal Dependencies for Laz</i>	
Utku Türk, kaan bayar, Ayşegül Dilara Özercan, Gökem Yiğit Öztürk and Şaziye Betül Özateş	188
<i>Dependency annotation of noun incorporation in polysynthetic languages</i>	
Francis Tyers and Karina Mishchenkova . . . . .	194
<i>Universal Dependency treebank for Xibe</i>	
He Zhou, Juyeon Chung, Sandra Kübler and Francis Tyers . . . . .	204



# Workshop Program

**Sunday, December 13, 2020**

**14:00–14:10 Introduction**

**14:10–15:00 Q&A Session 1**

*Dependency annotation of noun incorporation in polysynthetic languages*

Francis Tyers and Karina Mishchenkova

*I've got a construction looks funny – representing and recovering non-standard constructions in UD*

Josef Ruppenhofer and Ines Rehbein

*Annotating MWEs in the Irish UD treebank*

Sarah McGuinness, Jason Phelan, Abigail Walsh and Teresa Lynn

*Subjecthood and annotation: The cases of French and Wolof*

Olivier Bondéelle and Sylvain Kahane

*Annotation issues in Universal Dependencies for Korean and Japanese*

Ji Yoon Han, Tae Hwan Oh, LEE JIN and Hansaem Kim

**15:00–15:10 Break**

**15:10–16:00 Q&A Session 2**

*Variation in Universal Dependencies annotation: A token-based typological case study on adpossession constructions*

Kaius Sinnemäki and Viljami Haakana

*Corpus evidence for word order freezing in Russian and German*

Aleksandrs Berdicevskis and Alexander Piperski

*Parsing in the absence of related languages: Evaluating low-resource dependency parsers on Tagalog*

Angelina Aquino and Franz de Leon

**Sunday, December 13, 2020 (continued)**

*Verification, reproduction and replication of NLP experiments: A case study on parsing Universal Dependencies*

Çağrı Çöltekin

*Composing byte-pair encodings for morphological sequence classification*

Adam Ek and Jean-Philippe Bernardy

**16:00–17:15 Poster Session**

*A small Universal Dependencies treebank for Hittite*

Erik Andersen and Benjamin Rozonoyer

*Universal Dependencies for Albanian*

Marsida Toska, Joakim Nivre and Daniel Zeman

*Universal Dependencies for Manx Gaelic*

Kevin Scannell

*Universal Dependency treebank for Xibe*

He Zhou, Juyeon Chung, Sandra Kübler and Francis Tyers

*First Steps towards Universal Dependencies for Laz*

Utku Türk, kaan bayar, Ayşegül Dilara Özercan, Görkem Yiğit Öztürk and Şaziye Betül Özateş

*From LFG to UD: A combined approach*

Cheikh M. Bamba Dione

*UDon2: A library for manipulating Universal Dependencies trees*

Dmytro Kalpakchi and Johan Boye

*PALMYRA 2.0: A configurable multilingual platform independent tool for morphology and syntax annotation*

Dima Taji and Nizar Habash

*Unifying the treatment of preposition-determiner contractions in German Universal Dependencies treebanks*

Stefan Grünewald and Annemarie Friedrich

**Sunday, December 13, 2020(continued)**

**17:15–18:00    Keynote**

*Transcending dependencies*

Martha Palmer, University of Colorado Boulder, USA

**Break**

**18:10–19:00    Q&A Session 3**

*Exploring diachronic syntactic shifts with dependency length: The case of scientific English*

Tom S Juzek, Marie-Pauline Krielke and Elke Teich

*Identifying and handling cross-treebank inconsistencies in UD: A pilot study*

Tillmann Dönicke, Xiang Yu and Jonas Kuhn

*Profiling-UD: A tool for linguistic profiling of texts* (cross-submission)

Dominique Brunato, Andrea Cimino, Felice Dell’Orletta, Simonetta Montemagni and Giulia Venturi

*Configurable dependency tree extraction from CCG derivations*

Kilian Evang

*A Universal Dependencies conversion pipeline for a Penn-format constituency tree-bank*

Þórunn Arnardóttir, Hinrik Hafsteinsson, Einar Freyr Sigurðsson, Kristín Bjarnadóttir, Anton Karl Ingason, Hildur Jónsdóttir and Steinþór Steingrímsson



## **Invited Talk: Martha Palmer, University of Colorado Boulder**

### **Transcending Dependencies**

This talk will discuss some of the challenges arising from the Blocks World scenario in the DARPA Communicating with Computers program. The actions are very simple and concrete, such as “Add a block to the tower.” However, even in this restricted world, getting the appropriate contextual interpretation of a sentence can be challenging, especially with respect to spatial relations and implicit information. The talk will review the progress we have made so far on collecting useful data that comprises complete 2 person dialogues discussing block structure constructions, and our attempts to achieve the goal of contextual interpretation in the processing of these dialogues. A main focus will be the ways in which we are expanding AMR annotation to encompass spatial relations and the recovery of implicit arguments. Both expansions play into the task of maintaining a discourse structure and producing the predicate logic sentence representations needed by the down-stream planner. The talk will conclude with our current AMR parsing results, our attempts to pass them along to the planner, and our future goals.

### **Bio**

Martha Palmer is the Helen & Hubert Croft Endowed Professor of Engineering in the Computer Science Department, and an Arts & Sciences Professor of Distinction in the Linguistics Department, at the University of Colorado, with a split appointment. She is also an Institute of Cognitive Science Faculty Fellow, a co-Director of CLEAR, an ACL Fellow, and a AAI Fellow. She was the Director of the 2011 Linguistics Institute in Boulder, CO. Her research is focused on capturing elements of the meanings of words that can comprise automatic representations of complex sentences and documents in English, Chinese, Arabic, Hindi, and Urdu, funded by DARPA, DTRA and NSF. A more recent focus is the application of these methods to biomedical journal articles and clinical notes, funded by NIH. She co-edits LiLT, Linguistic Issues in Language Technology, and has been a co-editor of the Journal of Natural Language Engineering and on the CLJ Editorial Board. She is a past President of ACL, past Chair of SIGLEX, was the Founding Chair of SIGHAN, and has over 300 peer-reviewed publications.