



LREC 2026

**The Ninth Workshop on Universal Dependencies
(UDW 2026)**

Workshop Proceedings

Editors
Çağrı Çöltekin and Kaja Dobrovoljc

May 16, 2026

©ELRA Language Resources Association (ELRA), 2026

These proceedings are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-62-3

Preface

These proceedings include the program and papers that are presented at the ninth workshop on Universal Dependencies, held in conjunction with LREC on May 16, 2026.

Universal Dependencies (UD) is a framework for cross-linguistically consistent treebank annotation that has so far been applied to over 180 languages. With the goal of enabling both comparative linguistic studies and multilingual natural language processing, the framework aims to capture similarities as well as idiosyncrasies among different languages. Since the first UD workshop in 2017 in Gothenburg, the UD workshops 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. This year’s workshop, The Ninth Workshop on Universal Dependencies (UDW 2026), is held in conjunction with LREC 2026 in Mallorca, Spain.

We received 42 submissions of which 29 regular and three non-archival submissions were accepted. As in previous instances of the UD workshop, accepted submissions cover a broad range of papers related to UD, including introduction of new treebanks, discussion of specific linguistic constructions, the applications of UD for certain languages and language families, and the use cases of UD for linguistic research and natural language processing.

We are honored to have two invited speakers at UDW 2026: Marie-Catherine de Marneffe (FNRS – UCLouvain) with a talk on “Diversity in NLP: why, what, where and how”, and Stephen Mayhew (Duolingo) with a talk on “Universal NER: Standing on the Shoulders of Giants”. We would also like to thank the UDW 2026 Program Committee for their timely and meticulous reviews.

We are looking forward to a productive and engaging workshop!

Çağrı Çöltekin, Kaja Dobrovoljc and Joakim Nivre

Organizing Committee

Çağrı Çöltekin, Tübingen University
Kaja Dobrovoljc, University of Ljubljana & Jozef Stefan Institute
Joakim Nivre, Uppsala University

Program Committee

Noor Abo Mokh, Indiana University
Jane Adkins, Dublin City University
Leonel de Alencar, UFC
Giuseppe Attardi, Università di Pisa
David Beck, University of Alberta
Aleksandrs Berdicevskis, University of Gothenburg
Ann Bies, Linguistic Data Consortium, Univ. Of Pennsylvania
Bernd Bohnet, Google, London
Gosse Bouma, University of Groningen
Miriam Butt, University of Konstanz
Giuseppe Celano, University of Leipzig
Xinying Chen, University of Ostrava
Daniel Dakota, Indiana University
Stefania Degaetano-Ortlieb, Saarland University
Kilian Evang, Heinrich Heine University Düsseldorf
Pegah Faghiri, CNRS
Marcos Garcia, University of Santiago de Compostela
Kim Gerdes, Université Paris-Saclay
Carlos Gómez-Rodríguez, Universidade da Coruña,
Bruno Guillaume, LORIA
Eva Hajičová, Charles University, Prague
Dag Haug, University of Oslo
Santiago Herrera, Université Paris Nanterre
Maarten Janssen, Charles University Prague
Mayank Jobanputra, Saarland University
Sylvain Kahane, Université Paris Nanterre
Sandra Kübler, Indiana University
Guy Lapalme, Université de Montréal
Teresa Lynn, Dublin City University
Rob Malouf, San Diego State University
Nicolas Mazziotta, Université de Liège
Alexander Mehler, Goethe-Universität Frankfurt am Main
Wolfgang Menzel, Hamburg University
Aleksandra Miletić, CNRS
Simon Mille, ADAPT, Dublin City University
Jiří Mirovský, Charles University, Prague
Simonetta Montemagni, Istituto di Linguistica Computazionale “A. Zampolli” - CNR
Kaili Müürisep, University of Tartu
Anna Nedoluzhko, Charles University Prague
Ruo Chen Niu, Beijing Language and Culture University
Stephan Oepen, University of Oslo

Petya Osenova, Bulgarian Academy of Sciences
Prokopis Prokopidis, Institute for Language and Speech Processing/Athena RC
Mathilde Regnault, Universität Stuttgart
Tanja Samardžić, University of Zurich
Giuseppe Samo, Beijing Language and Culture University
Nathan Schneider, Georgetown University
Djamé Seddah, Sorbonne University
Achim Stein, University of Stuttgart
Luka Terčon, Faculty of Arts, University of Ljubljana
Giulia Venturi, Istituto di Linguistica Computazionale “A. Zampolli” - CNR
Veronika Vincze, Hungarian Academy of Sciences
Yaqin Wang, Guangdong University of Foreign Studies
Pan Xiaxing, Huaqiao University
Nianwen Xue, Brandeis University, Waltham
Jianwei Yan, Zhejiang University
Zdeněk Žabokrtský, Charles University
Eva Zehentner, University of Zurich
Amir Zeldes, Georgetown University
Daniel Zeman, Charles University
Heike Zinsmeister, University of Hamburg

Table of Contents

<i>Probing the Dynamics of Syntactic Ability Acquisition Throughout LLM Pretraining</i> Hiroshi Matsuda and Masayuki Asahara	1
<i>Which languages are "hot", and which are "cool"? Using Universal Dependencies for large-scale comparisons of subject expression</i> Natalia Levshina	49
<i>Clefting and beyond in Tigrinya</i> Nazareth Amlesom Kifle and Michael Gasser	54
<i>Coconstructions in Spoken Data: UD Annotation Guidelines and First Results</i> Ludovica Pannitto, Kaja Dobrovoljc Zor, Sylvain Kahane, Elena Battaglia, Bruno Guillaume, Caterina Mauri and Eleonora Zucchini	60
<i>Verifying the Menzerath-Altmann law in the verbal domain in 180 languages</i> Pegah Faghiri, Kim Gerdes and Sylvain Kahane	76
<i>Comparing Dependency Distances of Esperanto and Other Languages in a Multi-Lingual Parallel Corpus</i> Masanori Oya	91
<i>Negation of Turkic non-verbal clauses: Analysis and Universal Dependencies Implementation</i> Nikolett Mus, Furkan Akkurt, Bermet Chontaeva, Soudabeh Eslami, Sardana Ivanova, Çağrı Çöltekin, Jonathan N. Washington, Gulnura Dzhumaliev and Aida Kasieva	97
<i>Towards Universal Dependencies for L2 Learners of Modern Greek: Annotation and Challenges</i> Christina Klironomou, Thelka Pasparaki, Arianna Masciolini, Alexandros Tantos, Despoina Ourania Touriki, Konstantinos Tsiotskas and Eleni Tsourilla	103
<i>A Comparative Linguistic Analysis of Ottoman and Modern Turkish through UD Treebanks</i> Enes Yilandiloğlu	109
<i>Word segmentation for UD: a comparison of isiZulu and Sepedi</i> Laurette Marais and Laurette Pretorius	116
<i>CoBra: A Compound Branching Resource for Nominal Triconstituent Compounds in English and German</i> Carmen Schacht, Isabell Landwehr, Diana Davidson, Konrad Grabowski, Magdalena Meiser and Sophia Wiedmann	128
<i>SE Constructions Revisited: Focus on Treebanks for Romance Languages</i> Verginica Barbu Mititelu, Elena Irimia, Adriana S. Pagano, Roxana Ciolanescu and Ioana Buhnica	142
<i>Say "No" to Missing Polarity: A Negation Enrichment of Porttinari UD Treebank</i> Isaac Souza de Miranda Junior, Oto Araújo Vale and Marie-Catherine de Marneffe ...	153
<i>The Grammar Does the Work: Functional vs. Lexical Dependency Length Minimization Across the UD Languages</i> Kim Gerdes	163

<i>Greenberg's Universal 45 in Universal Dependencies: Gender Distinctions and Annotation Challenges</i>	
Antoni Brosa-Rodriguez and M. Dolores Jimenez Lopez	174
<i>A Proposal for a More Universal Annotation of Relative Clauses in Universal Dependencies</i>	
Santiago Herrera and Sylvain Kahane	183
<i>MesoTree: Annotated Linguistic Resources for Quantitative Comparative Linguistic Analysis and NLP in Mesoamerica</i>	
Robert Pugh, Francis Tyers and Robert Henderson	197
<i>Nonprototypical Predication and Nonpredicational Clauses in Universal Dependencies</i>	
Joakim Nivre, William Croft and Andre Coneglian	208
<i>Complex Predicates in Universal Dependencies</i>	
Andre Coneglian, Joakim Nivre and William Croft	218
<i>Bringing Information Structure to Universal Dependencies</i>	
Nikolett Mus, Andrew Dyer, Claudia Corbetta and Sylvain Kahane	229
<i>Introducing Universal Dependencies for Sardinian: the UD ContSar Treebank</i>	
Nicoletta Puddu, Manuela Sanguinetti and Luigi Talamo	242
<i>Exploiting Parallel Aligned Treebanks</i>	
Maarten Janssen	252
<i>Cross-Dialectal Transfer for Low-Resource Arabic: The Tunisian Arabic Dependency Treebank</i>	
Amal Aissaoui	258
<i>From Treebank Metadata to Sentence-Level Genre in Universal Dependencies: A Reproducible, Versioned Resource</i>	
Egon Stemle	268
<i>Towards a Universal Dependency Corpus for Old Saxon (Old Low German)</i>	
Christian Chiarcos and Janine Siewert	277
<i>Gathering valency frames for annotation and batch corrections</i>	
Mathilde Regnault	289
<i>Syntax is the Key to Semantics: Combining Universal Dependencies and Abstract Meaning Representation</i>	
Johannes Heinecke	295
<i>Extending Retag to Conversion Error Detection: A Case Study on SynTagRus Morphology</i>	
Andrei Movsesian and Daniil Timchenko	305
<i>Speech Act Constructions in Universal Dependencies</i>	
Joakim Nivre, William Croft and Andre Coneglian	315

Workshop Program

Saturday May 16, 2026

09:00 Session 1

09:00 *Opening*

09:05 *Keynote 1: Diversity in NLP: why, what, where and how*
Marie-Catherine de Marneffe

10:00 *Probing the Dynamics of Syntactic Ability Acquisition Throughout LLM Pre-training*
Hiroshi Matsuda and Masayuki Asahara

10:15 *Which languages are "hot", and which are "cool"? Using Universal Dependencies for large-scale comparisons of subject expression*
Natalia Levshina

10:30 Coffee break

11:00 Session 2 (poster)

Clefting and beyond in Tigrinya
Nazareth Amlesom Kifle and Michael Gasser

Coconstructions in Spoken Data: UD Annotation Guidelines and First Results
Ludovica Pannitto, Kaja Dobrovoljc Zor, Sylvain Kahane, Elena Battaglia, Bruno Guillaume, Caterina Mauri and Eleonora Zucchini

Verifying the Menzerath-Altmann law in the verbal domain in 180 languages
Pegah Faghiri, Kim Gerdes and Sylvain Kahane

Comparing Dependency Distances of Esperanto and Other Languages in a Multi-Lingual Parallel Corpus
Masanori Oya

Negation of Turkic non-verbal clauses: Analysis and Universal Dependencies Implementation
Nikolett Mus, Furkan Akkurt, Bermet Chontaeva, Soudabeh Eslami, Sardana Ivanova, Çağrı Çöltekin, Jonathan N. Washington, Gulnura Dzhumaliev and Aida Kasieva

Saturday May 16, 2026 (continued)

Towards Universal Dependencies for L2 Learners of Modern Greek: Annotation and Challenges

Christina Klironomou, Thelka Pasparaki, Arianna Masciolini, Alexandros Tantos, Despoina Ourania Touriki, Konstantinos Tsiotskas and Eleni Tsourilla

A Comparative Linguistic Analysis of Ottoman and Modern Turkish through UD Treebanks

Enes Yilandiloğlu

Word segmentation for UD: a comparison of isiZulu and Sepedi

Laurette Marais and Laurette Pretorius

CoBra: A Compound Branching Resource for Nominal Triconstituent Compounds in English and German

Carmen Schacht, Isabell Landwehr, Diana Davidson, Konrad Grabowski, Magdalena Meiser and Sophia Wiedmann

SE Constructions Revisited: Focus on Treebanks for Romance Languages

Verginica Barbu Mititelu, Elena Irimia, Adriana S. Pagano, Roxana Ciolanescu and Ioana Buhnila

Say "No" to Missing Polarity: A Negation Enrichment of Portinari UD Treebank

Isaac Souza de Miranda Junior, Oto Araújo Vale and Marie-Catherine de Marneffe

12:00 Session 3

The Grammar Does the Work: Functional vs. Lexical Dependency Length Minimization Across the UD Languages

Kim Gerdes

Greenberg's Universal 45 in Universal Dependencies: Gender Distinctions and Annotation Challenges

Antoni Brosa-Rodriguez and M. Dolores Jimenez Lopez

A Proposal for a More Universal Annotation of Relative Clauses in Universal Dependencies

Santiago Herrera and Sylvain Kahane

Saturday May 16, 2026 (continued)

13:00 **Lunch break**

14:00 **Session 4**

14:00 *Keynote 2: Universal NER: Standing on the Shoulders of Giants*
Stephen Mayhew

14:50 *MesoTree: Annotated Linguistic Resources for Quantitative Comparative Linguistic Analysis and NLP in Mesoamerica*
Robert Pugh, Francis Tyers and Robert Henderson

15:10 *Nonprototypical Predication and Nonpredicational Clauses in Universal Dependencies*
Joakim Nivre, William Croft and Andre Coneglian

15:30 *Complex Predicates in Universal Dependencies*
Andre Coneglian, Joakim Nivre and William Croft

15:45 *Bringing Information Structure to Universal Dependencies*
Nikolett Mus, Andrew Dyer, Claudia Corbetta and Sylvain Kahane

16:00 **Coffee break**

16:30 **Session 5 (poster)**

DELTA: A Toolkit for Measuring Linguistic Diversity in Dependency-Parsed Corpora
Louis Estève and Kaja Dobrovoljc

Modelling the Morphology of Verbal Paradigms: A Case Study in the Tokenization of Turkish and Hebrew
Giuseppe Samo and Paola Merlo

Introducing Universal Dependencies for Sardinian: the UD ContSar Treebank
Nicoletta Puddu, Manuela Sanguinetti and Luigi Talamo

Exploiting Parallel Aligned Treebanks
Maarten Janssen

Saturday May 16, 2026 (continued)

Cross-Dialectal Transfer for Low-Resource Arabic: The Tunisian Arabic Dependency Treebank

Amal Aissaoui

From Treebank Metadata to Sentence-Level Genre in Universal Dependencies: A Reproducible, Versioned Resource

Egon Stemle

Towards a Universal Dependency Corpus for Old Saxon (Old Low German)

Christian Chiarcos and Janine Siewert

Gathering valency frames for annotation and batch corrections

Mathilde Regnault

Syntax is the Key to Semantics: Combining Universal Dependencies and Abstract Meaning Representation

Johannes Heinecke

Extending Retag to Conversion Error Detection: A Case Study on SynTagRus Morphology

Andrei Movsesian and Daniil Timchenko

Exploring language relations through syntactic distances and geographic proximity

Juan De Gregorio, Raul Toral and David Sanchez

Speech Act Constructions in Universal Dependencies

Joakim Nivre, William Croft and Andre Coneglian

17:30

Closing: Community discussion

Invited Talks

Universal NER: Standing on the Shoulders of Giants

Stephen Mayhew, Duolingo

The huge success of Universal Dependencies has been an inspiration for similar projects. One such descendant is Universal NER, a large-scale, multilingual, shared-schema named entity recognition project, in the works since 2020. This talk will cover a brief history of the project, the philosophy behind the data, the tagset, and the annotations, as well as details on the data itself, including the latest release, Universal NER v2, presented at LREC 2026.

Diversity in NLP: why, what, where and how

Marie-Catherine de Marneffe, CENTAL - UCLouvain / FNRS

Diversity has become an advocated property of NLP datasets and systems. However, its use remains largely ad hoc, with limited theoretical grounding and definitions and measures that vary across studies. In this talk, I will describe a recent survey with colleagues of 308 diversity-related papers from the ACL Anthology. We propose an NLP-specific framework organized around four perspectives: why diversity matters, what is measured, where it is measured, and how, drawing from insights from other scientific fields where the concept of diversity has been thoroughly conceptualized. I will also look at UD from these perspectives.

