

UDW 2025

**Eighth Workshop on Universal Dependencies  
(UDW, SyntaxFest 2025)**

**Proceedings**

August 27, 2025

The UDW organizers gratefully acknowledge the support from the following sponsors.

VITASIS



Ljubljana Tourism



Mestna občina  
Ljubljana



Flanders  
State of the Art



**CJVT** Centre for  
Language Resources  
and Technologies



**AI4DH** CENTRE OF EXCELLENCE IN AI  
FOR DIGITAL HUMANITIES

**Organized by**



**As part of SyntaxFest 2025**



©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
317 Sidney Baker St. S  
Suite 400 - 134  
Kerrville, TX 78028  
USA  
Tel: +1-855-225-1962  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 979-8-89176-292-3

## Introduction

Universal Dependencies (UD) is a framework for cross-linguistically consistent treebank annotation that has so far been applied to over 150 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 brings 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, 8th Workshop on Universal Dependencies (UDW 2025) is held as a part of SyntaxFest 2025 in Ljubljana, Slovenia, which brought together five related but independent events:

- 18th International Conference on Parsing Technologies (IWPT 2025)
- 8th Universal Dependencies Workshop (UDW 2025)
- 8th International Conference on Dependency Linguistics (DepLing 2025)
- 23rd Workshop on Treebanks and Linguistic Theories (TLT 2025)
- 3rd Workshop on Quantitative Syntax (QUASY 2025)

In addition, a pre-conference workshop organized by the COST Action CA21167 – Universality, Diversity and Idiosyncrasy in Language Technology (UniDive) was held prior to the main event, with dedicated sessions on the 1st UniDive Shared Task on Morphosyntactic Parsing and the 2nd Workshop on Universal Dependencies for Turkic Languages.

SyntaxFest 2025 continues the tradition of SyntaxFest 2019 (Paris, France), SyntaxFest 2021 (Sofia, Bulgaria), and GURT/SyntaxFest 2023 (Washington DC, USA) in bringing together multiple events that share a common interest in using corpora and treebanks for empirically validating syntactic theories, studying syntax from quantitative and theoretical points of view, and training machine learning models for natural language processing. Much of this research is increasingly multilingual and cross-lingual and requires continued systematic analysis from various theoretical, applied, and practical perspectives. By co-locating these workshops under a shared umbrella, SyntaxFest fosters dialogue between overlapping research communities and supports innovation at the intersection of linguistics and language technology. As in previous editions, all five workshops at SyntaxFest 2025 shared a common submission and reviewing process, with a unified timeline, identical submission formats, and a shared program committee. During submission, authors could indicate one or more preferred venues, but the final assignment of papers was determined by the collective program chairs, composed of the individual workshop chairs, based on thematic alignment. All accepted submissions were peer-reviewed by at least three reviewers from the shared program committee.

In total, SyntaxFest 2025 received 94 submissions, of which 73 (78%) were accepted for presentation. The final program included a total of 47 long papers, 21 short papers, and 5 non-archival contributions, distributed across the five workshops: 5 papers were presented at IWPT (2 long, 3 short); 20 at UDW (14 long, 5 short, 1 non-archival); 16 at DepLing (12 long, 2 short, 2 non-archival); 18 at TLT (10 long, 7 short, 1 non-archival); and 14 at QUASY (9 long, 4 short, 1 non-archival).

Our sincere thanks go to everyone who made this event possible. We thank all authors for their submissions and the reviewers for their time and thoughtful feedback, which contributed to a diverse and high-quality program. Special thanks go to the local organizing team at the University of Ljubljana and the Slovene Language Technologies Society for hosting the event, and to the sponsors for their generous support. Finally, we gratefully acknowledge ACL SIGPARSE for endorsing the event and the ACL Anthology for publishing the proceedings.

Kenji Sagae, Stephan Oepen (IWPT 2025 Chairs)  
Gosse Bouma, Çağrı Çöltekin (UDW 2025 Chairs)

Eva Hajičová, Sylvain Kahane (DepLing 2025 Chairs)  
Heike Zinsmeister, Sarah Jablotschkin, Sandra Kübler (TLT 2025 Chairs)  
Xinying Chen, Yaqin Wang (QUASY 2025 Chairs)  
Kaja Dobrovoljc (SyntaxFest 2025 Organization Chair)

Ljubljana, August 2025

# Organizing Committee

## TLT Chairs

Heike Zinsmeister, University of Hamburg  
Sarah Jablotschkin, University of Hamburg  
Sandra Kübler, Indiana University

## DepLing Chairs

Eva Hajičová, Charles University, Prague  
Sylvain Kahane, Université Paris Nanterre

## UDW Chairs

Gosse Bouma, University of Groningen  
Çağrı Çöltekin, University of Tübingen

## IWPT Chairs

Kenji Sagae, University of California, Davis  
Stephan Oepen, University of Oslo

## QUASY Chairs

Xinying Chen, University of Ostrava  
Yaqin Wang, Guangdong University of Foreign Studies

## Publication Chair

Sarah Jablotschkin, University of Hamburg

## Local SyntaxFest 2025 Organizing Committee

Kaja Dobrovoljc, University of Ljubljana, SDJT  
Špela Arhar Holdt, University of Ljubljana  
Luka Terčon, University of Ljubljana  
Marko Robnik-Šikonja, University of Ljubljana  
Matej Klemen, University of Ljubljana  
Sara Kos, University of Ljubljana  
Timotej Knez, University of Ljubljana, SDJT  
Tinca Lukan, University of Ljubljana

## Special Thanks for designing the SyntaxFest 2025 logo to

Kim Gerdes, Université Paris-Saclay

# Program Committee

## Shared Program Committee

V.S.D.S.Mahesh Akavarapu, Eberhard-Karls-Universität Tübingen  
Leonel Figueiredo de Alencar, Federal University of Ceará (UFC)  
Patricia Amaral, Indiana University  
Giuseppe Attardi, University of Pisa  
John Bauer, Stanford University  
David Beck, University of Alberta  
Laura Becker, Albert-Ludwigs-Universität Freiburg  
Aleksandrs Berdicevskis, Gothenburg University  
Ann Bies, University of Pennsylvania  
Igor Boguslavsky, Universidad Politécnica de Madrid  
Bernd Bohnet, Google  
Cristina Bosco, University of Turin  
Gosse Bouma, University of Groningen  
Miriam Butt, Universität Konstanz  
G. A. Celano, Universität Leipzig  
Heng Chen, Guangdong University of Foreign Studies  
Xinying Chen, University of Ostrava  
Jinho D. Choi, Emory University  
Çağrı Çöltekin, University of Tuebingen  
Daniel Dakota, Leidos  
Stefania Degaetano-Ortlieb, Universität des Saarlandes  
Kaja Dobrovoljc, University of Ljubljana  
Jakub Dotlacil, Utrecht University  
Gülşen Eryiğit, Istanbul Technical University  
Kilian Evang, Heinrich Heine University Düsseldorf  
Pegah Faghiri, CNRS  
Ramon Ferrer-i-Cancho, Universidad Politécnica de Cataluna  
Marcos Garcia, Universidade de Santiago de Compostela  
Kim Gerdes, Université Paris-Saclay  
Loïc Grobol, Université Paris Nanterre  
Bruno Guillaume, INRIA  
Carlos Gómez-Rodríguez, Universidade da Coruña  
Eva Hajicova, Charles University  
Dag Trygve Truslew Haug, University of Oslo  
Santiago Herrera, University of Paris Nanterre  
Richard Hudson, University College London  
Maarten Janssen, Charles University Prague  
Jingyang Jiang, Zhejiang University  
Mayank Jobanputra, Universität des Saarlandes  
Sylvain Kahane, Université Paris Nanterre  
Václava Kettnerová, Charles University Prague  
Sandra Kübler, Indiana University  
Guy Lapalme, University of Montreal  
François Lareau, Université de Montréal  
Miryam de Lhoneux, KU Leuven  
Zoey Liu, University of Florida

Teresa Lynn, Dublin City University  
Jan Macutek, Slovak Academy of Sciences  
Robert Malouf, San Diego State University  
Marie-Catherine de Marneffe, UCLouvain  
Nicolas Mazziotta, Université de Liège  
Alexander Mehler, Johann Wolfgang Goethe Universität Frankfurt am Main  
Maitrey Mehta, University of Utah  
Wolfgang Menzel, Universität Hamburg  
Marie Mikulová, Charles University  
Aleksandra Miletić, University of Helsinki  
Jasmina Milićević, Dalhousie University  
Simon Mille, Dublin City University  
Yusuke Miyao, The University of Tokyo  
Noor Abo Mokh, Indiana University  
Simonetta Montemagni, Institute for Computational Linguistics “A. Zampolli” (ILC-CNR)  
Jiří Mírovský, Charles University Prague  
Kaili Müürisep, Institute of computer science, University of Tartu  
Anna Nedoluzhko, Charles University Prague  
Ruochen Niu, Beijing Language and Culture University  
Joakim Nivre, Uppsala University  
Stephan Oepen, University of Oslo  
Timothy John Osborne, Zhejiang University  
Petya Osenova, Sofia University “St. Kliment Ohridski”  
Agnieszka Patejuk, Polish Academy of Sciences  
Lucie Poláková, Charles University Prague  
Prokopis Prokopidis, Athena Research Center  
Mathilde Regnault, Universität Stuttgart  
Kateřina Rysová, University of South Bohemia  
Magdaléna Rysová, Charles University Prague  
Tanja Samardžić, University of Zurich  
Giuseppe Samo, Beijing Language and Culture University  
Haruko Sanada, Ritssho University  
Nathan Schneider, Georgetown University  
Djamé Seddah, Sorbonne University  
Anastasia Shimorina, Orange  
Maria Simi, University of Pisa  
Achim Stein, University of Stuttgart  
Daniel G. Swanson, Indiana University  
Luka Terčon, Faculty of Arts, University of Ljubljana  
Giulia Venturi, Institute for Computational Linguistics “A. Zampolli” (ILC-CNR)  
Veronika Vincze, University of Szeged  
Yaqin Wang, Guangdong University of Foreign Studies  
Pan Xiaxing, Huaqiao University  
Chunshan Xu, Anhui Jianzhu University  
Nianwen Xue, Brandeis University  
Jianwei Yan, Zhejiang University  
Zdenek Zabokrtsky, Faculty of Mathematics and Physics, Charles University Prague  
Eva Zehentner, University of Zurich  
Amir Zeldes, Georgetown University  
Daniel Zeman, Charles University Prague  
Šárka Zikánová, Charles University Prague



Heike Zinsmeister, Universität Hamburg

## Keynote

# Typologically informed NLP evaluation

Miryam de Lhoneux  
KU Leuven, Belgium



**Abstract:** NLP has a long history of focusing mainly on English. While increasing efforts are being made towards making language technology more multilingual, English remains the language on which NLP technology is developed first, and applied to other languages next, which inevitably leads to degraded performance compared to English. This talk is about reversing this trend and putting multilinguality at the core of NLP, rather than at the periphery. I describe how typology can inform NLP evaluation, using our recently proposed language sampling framework. A strong limitation of the approach is the state of multilingual datasets, which tend to lack coverage, be machine-translated or have questionable quality. UD is an exception, and I emphasize the role it can play in establishing best practices in multilingual NLP evaluation.

**Bio:** Miryam de Lhoneux is an assistant professor in the department of Computer Science at KU Leuven in Belgium, researching and teaching Natural Language Processing. She heads the LAGoM NLP lab where the focus is on multilingual and interpretable models. Previously, she was a postdoc at the University of Copenhagen in Denmark. She has a PhD from Uppsala University, Sweden, an MSc in Artificial Intelligence from the University of Edinburgh, UK, and a BA and MA in languages and literatures from UCLouvain, Belgium.

# Non-Archival Abstract

## **MultiBLiMP 1.0: A Massively Multilingual Benchmark of Linguistic Minimal Pairs**

Jaap Jumelet<sup>1</sup>, Leonie Weissweiler<sup>2</sup> and Arianna Bisazza<sup>1</sup>

<sup>1</sup>University of Groningen

<sup>2</sup>University of Texas at Austin

We introduce MultiBLiMP 1.0, a massively multilingual benchmark of linguistic minimal pairs, covering 101 languages, 6 linguistic phenomena and containing more than 125,000 minimal pairs. Our minimal pairs are created using a fully automated pipeline, leveraging the large-scale linguistic resources of Universal Dependencies and UniMorph. MultiBLiMP evaluates abilities of LLMs at an unprecedented multilingual scale, and highlights the shortcomings of the current state-of-the-art in modelling low-resource languages.

## Table of Contents

<i>Reference and Modification in Universal Dependencies</i> Joakim Nivre and William Croft .....	1
<i>Annotation of Relative Forms in the Egyptian-UJaen Treebank</i> Roberto A. Diaz Hernandez and Daniel Zeman .....	11
<i>UD Treebanks for Esperanto as a natural language</i> Masanori Oya .....	22
<i>Universal Dependencies for Suansu</i> Jessica K. Ivani and Kira Tulchynska .....	30
<i>Crossing Dialectal Boundaries: Building a Treebank for the Dialect of Lesbos through Knowledge Transfer from Standard Modern Greek</i> Stavros Bompolas, Stella Markantonatou, Angela Ralli and Antonios Anastasopoulos .....	39
<i>UD-English-CHILDES: A Collected Resource of Gold and Silver Universal Dependencies Trees for Child Language Interactions</i> Xiulin Yang, Zhuoxuan Ju, Lanni Bu, Zoey Liu and Nathan Schneider .....	52
<i>A UD Treebank for Bohairic Coptic</i> Amir Zeldes, Nina Speransky, Nicholas E. Wagner and Caroline T. Schroeder .....	59
<i>Negation in Universal Dependencies</i> Jamie Yates Findlay and Dag Trygve Truslew Haug .....	70
<i>TreEn: A Multilingual Treebank Project on Environmental Discourse</i> Adriana Silvina Pagano, Patricia Chiril, Elisa Chierchiello and Cristina Bosco .....	80
<i>Building UD Cairo for Old English in the Classroom</i> Lauren Levine, Junghyun Min and Amir Zeldes .....	97
<i>Universal Dependencies for Sindhi</i> John Bauer, Sakiina Shah, Muhammad Shaheer, Mir Afza Ahmed Talpur, Zubair Sanjrani, Sarwat Qureshi, Shafi Pirzada, Christopher D. Manning and Mutee U Rahman .....	105
<i>Universal Dependencies Treebank for Khoekhoe (KDT)</i> Tulchynska Kira, Job Sylvanus and Witzlack-Makarevich Alena .....	119
<i>Parallel Universal Dependencies Treebanks for Turkic Languages</i> Arofat Akhundjanova, Furkan Akkurt, Bermet Chontaeva, Soudabeh Eslami and Cagri Coltekin	129
<i>Towards better annotation practices for symmetrical voice in Universal Dependencies</i> Andrew Thomas Dyer and Colleen Alena O'Brien .....	137
<i>Extending the Enhanced Universal Dependencies – addressing subjects in pro-drop languages</i> Magali Sanches Duran, Elvis A. de Souza, Maria das Graças Volpe Nunes, Adriana Silvina Pagano and Thiago A. S. Pardo .....	143
<i>Annotating Second Language in Universal Dependencies: a Review of Current Practices and Directions for Harmonized Guidelines</i> Arianna Masciolini, Aleksandrs Berdicevskis, Maria Irena Szawerna and Elena Volodina .....	153

<i>Developing a Universal Dependencies Treebank for Alaskan Gwich'in</i> Matthew Kirk Andrews and Cagri Coltekin .....	164
<i>Quid verbumst? Applying a definition of word to Latin in Universal Dependencies</i> Flavio Massimiliano Cecchini .....	174
<i>ShUD: the First Shanghainese Universal Dependency Treebank</i> Qizhen Yang .....	187