

SIGTYP 2026

**The 8th Workshop on Research in Computational Linguistic
Typology and Multilingual NLP**

Proceedings of the Workshop

March 29, 2026

©2026 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

ISBN 979-8-89176-374-6

Introduction

We are very pleased to welcome you to the 8th edition of the workshop for typology-related research and its integration into multilingual Natural Language Processing (SIGTYP 2026). The workshop is co-located with the 19th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026), which takes place in Rabat, Morocco.

Encouraged by the 2019 – 2025 workshops, the aim of the eighth edition of SIGTYP workshop is to act as a platform and a forum for the exchange of information between typology-related research, multilingual NLP, and other research areas that can lead to the development of truly multilingual NLP methods. The workshop is specifically aimed at raising awareness of linguistic typology and its potential in supporting and widening the global reach of multilingual NLP, as well as at introducing computational approaches to linguistic typology. It fosters research and discussion on open problems, not only within the active community working on cross- and multilingual NLP but also inviting input from leading researchers in linguistic typology.

The workshop provides focused discussions on a range of topics, including the following:

1. Integration of typological features in language transfer and joint multilingual learning. In addition to established techniques such as “selective sharing”, are there alternative ways to encoding heterogeneous external knowledge in machine learning algorithms?
2. Development of unified taxonomy and resources. Building universal databases and models to facilitate understanding and processing of diverse languages.
3. Automatic inference of typological features. The pros and cons of existing techniques (e.g. heuristics derived from morphosyntactic annotation, propagation from features of other languages, supervised Bayesian and neural models) and discussion on emerging ones.
4. Typology and interpretability. The use of typological knowledge for interpretation of hidden representations of multilingual neural models, multilingual data generation and selection, and typological annotation of texts.
5. Improvement and completion of typological databases. Combining linguistic knowledge and automatic data-driven methods towards the joint goal of improving the knowledge on cross-linguistic variation and universals.
6. Linguistic diversity and universals. Challenges of cross-lingual annotation. Which linguistic phenomena or categories should be considered universal? How should they be annotated?

In addition, this year we have introduced a novel theme: using LLMs for typological studies. Can LLMs be utilised to formulate or prove typological hypotheses? Are they capable of making useful cross-linguistic generalisations?

The final program of SIGTYP contains 2 keynote talks, 6 archival papers, 2 EACL Findings and 2 extended abstracts, intertwined with linguistic trivia elements. This workshop would not have been possible without the contribution of its program committee, to whom we would like to express our gratitude! We also thank Terry Regier and Jennifer Culbertson for kindly accepting our invitation to be keynote speakers. Finally, we are also very thankful to Vilém Zouhar who offered his help in running the in-person part of the workshop.

Wishing you a wonderful time at SIGTYP 2026!

Ekaterina Vylomova, Priya Rani, Andrei Shcherbakov on behalf of the SIGTYP 2026 Program Co-Chairs

Organizing Committee

Program Chairs

Priya Rani, University of Galway
Michael Hahn, Saarland University
Alexey Sorokin, Lomonosov Moscow State University
Vilém Zouhar, ETH Zurich
Oleg Serikov, King Abdullah University of Science and Technology
Andrei Shcherbakov, CTI
Ryan Cotterell, ETH Zurich
Ekaterina Vylomova, The University of Melbourne

Local Chair

Vilém Zouhar, ETH Zurich

Publication Chairs

Ekaterina Vylomova, The University of Melbourne
Andrei Shcherbakov, CTI

SIGTYP Officers

President: Ekaterina Vylomova, The University of Melbourne
Secretary: Ryan Cotterell, ETH Zurich

Program Committee

Program Committee

Giuseppe G. A. Celano, Universität Leipzig
Elisabetta Jezek, University of Pavia
Esther Ploeger, Aalborg University
Leonie Weissweiler, Uppsala University
Barend Beekhuizen, University of Toronto
Jannic Alexander Cutura, DSTI School of Engineering
Ioana-Madalina Silai, Paris Nanterre University
Priya Rani, University of Galway
Michael Hahn, Saarland University
Alexey Sorokin, Lomonosov Moscow State University
Andrei Shcherbakov, CTI
Ekaterina Vylomova, The University of Melbourne

Invited Speakers

Terry Regier, University of California, Berkeley
Jennifer Culbertson, University of Edinburgh

Keynote Talk
A Keynote talk by Terry Regier

Terry Regier

University of California, Berkeley

2026-03-29 16:00:00 – Room: SALLE LE LIXUS

Bio: Terry Regier is a cognitive scientist and linguist whose research investigates language, meaning, and cognition. He is best known for his work exploring word meanings across languages, examining how word meanings reflect and sometimes shape thought and perception. Regier’s lab integrates computational modeling, cross-linguistic data, and behavioral experiments to study universals and variation in semantic domains such as color, kinship, number, and spatial relations. His research contributes to understanding how languages evolve, and how language and cognition are influenced by cultural diversity.

Keynote Talk
A Keynote talk by Jennifer Culbertson

Jennifer Culbertson
University of Edinburgh
2026-03-29 09:30:00 – Room: SALLE LE LIXUS

Bio: Jennifer Culbertson is Professor in the Department of Linguistics and English Language at the University of Edinburgh. She is a founding member of the Centre for Language Evolution, with her research focusing on how typological universals are shaped by properties of human cognition. She is best known for her work investigating universals of word order and morphological categories using the experimental method of Artificial Language Learning.

Table of Contents

<i>Automatic Grammatical Case Prediction for Template Filling in Case-Marking Languages: Implementation and Evaluation for Finnish</i>	
Johannes Laurmaa	1
<i>A Valency Lexicon for Universal Dependencies</i>	
Petr Kocharov and Lilit Kharatyan	12
<i>Evaluating the Interplay of Information Status and Information Content in a Multilingual Parallel Corpus</i>	
Julius Steuer, Toshiki Nakai, Andrew Thomas Dyer, Luigi Talamo and Annemarie Verkerk ...	18
<i>What does Surprisal have to do with Information Status?</i>	
Andrew Thomas Dyer	26
<i>Beyond Multilinguality: Typological Limitations in Multilingual Models for Meitei Language</i>	
Badal Nyalang	32
<i>A RAG Approach for Typological Database Completion</i>	
Jonathan Hus and Antonios Anastasopoulos	39

Program

Sunday, March 29, 2026

09:00 - 09:10 *Opening Remarks*

09:10 - 09:30 *Session 1*

Automatic Grammatical Case Prediction for Template Filling in Case-Marking Languages: Implementation and Evaluation for Finnish
Johannes Laurmaa

09:30 - 10:30 *Keynote talk by Jennifer Culbertson*

10:30 - 11:00 *Coffee/Tea Break*

11:00 - 11:50 *Session 2*

Beyond Multilinguality: Typological Limitations in Multilingual Models for Meitei Language
Badal Nyalang

A Valency Lexicon for Universal Dependencies
Petr Kocharov and Lilit Kharatyan

Modality Matching Matters: Calibrating Language Distances for Cross-Lingual Transfer in URIEL+
York Hay Ng, Aditya Khan, Xiang Lu, Matteo Salloum, Michael Zhou, Phuong H. Hoang, Seza Dogruoz and En-Shiun Annie Lee

11:50 - 12:30 *Session 3: Findings*

ParsTranslit: Truly Versatile Tajik-Farsi Transliteration
Rayyan Merchant and Kevin Tang

SrcMix: Mixing of Related Source Languages Benefits Extremely Low-resource Machine Translation
Sanjeev Kumar, Preethi Jyothi and Pushpak Bhattacharyya

12:30 - 14:00 *Lunch*

Sunday, March 29, 2026 (continued)

14:00 - 14:30 *Linguistic Trivia*

14:30 - 15:30 *Session 4*

Evaluating the Interplay of Information Status and Information Content in a Multilingual Parallel Corpus

Julius Steuer, Toshiki Nakai, Andrew Thomas Dyer, Luigi Talamo and Annemarie Verkerk

What does Surprisal have to do with Information Status?

Andrew Thomas Dyer

A RAG Approach for Typological Database Completion

Jonathan Hus and Antonios Anastasopoulos

Translation as Efficient Meaning Compression: An Information Bottleneck Analysis of Spatial Terms

Antoine Taroni, Ludovic Moncla and Frederique Laforest

15:30 - 16:00 *Coffee/Tea Break*

16:00 - 17:00 *Keynote talk by Terry Regier*

17:00 - 17:15 *Closing*