IWCLUL 2025

The 10th International Workshop on Computational Linguistics for Uralic Languages

Proceedings of the Workshop

December 10-12, 2025

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 979-8-89176-360-9

Preface

Welcome to the Proceedings of the 10th International Workshop on Computational Linguistics for Uralic Languages (IWCLUL), a dedicated venue for scholars, practitioners, and researchers working in computational linguistics with a particular emphasis on Uralic languages. This year's workshop continues the IWCLUL tradition of fostering interdisciplinary exchange, shared learning, and a collective dedication to tackling the linguistic, technological, and social issues surrounding Uralic languages in the digital era.

The Uralic language family—stretching across Europe and Asia and including languages as varied as Finnish, Hungarian, and the endangered Udmurt and Khanty—brings with it a distinct set of computational challenges. These languages often feature rich morphology, agglutinative patterns, and unique syntactic and phonological structures, all of which demand specialized methods in computational analysis and language technology. Our workshop aims to highlight these complexities and encourage the creation of innovative tools and approaches that not only facilitate digital use of these languages but also contribute to their long-term vitality.

This year, IWCLUL saw a strong and wide-ranging set of submissions from the international research community, demonstrating the sustained interest in computational work on Uralic languages. The accepted papers span numerous topics and language varieties, offering insights into both well-documented and lesser-resourced Uralic languages. This diversity reflects the ongoing growth of the field and the many directions in which researchers are advancing computational linguistics for Uralic languages.

We are also pleased to celebrate Jack Rueter, who was honored with a lifetime achievement award by ACL SIGUR for his decades-long dedication to Uralic language research, documentation, and technology development. His work has had a lasting impact on the community and serves as an inspiration for future generations of researchers.

We hope that these proceedings inspire continued research and collaboration in computational linguistics for Uralic languages. May the insights, methodologies, and resources shared here contribute to meaningful advances in the field and foster an inclusive future for Uralic languages in the digital landscape.

Sincerely, The IWCLUL 2025 Organizing Committee

Organizing Committee

Organizers

Mika Hämäläinen, Metropolia University of Applied Sciences
Flammie Pirinen, Arctic University of Norway
Lev Kharlashkin, Metropolia University of Applied Sciences
Eiaki V. Morooka, Metropolia University of Applied Sciences
Michael Rießler, University of Eastern Finland
Maarit Koponen, University of Eastern Finland
Ilia Moshnikov, University of Eastern Finland
Diana Kulashekhar, University of Eastern Finland
Mahla Baniasadi, University of Eastern Finland
Nurstrat prova, University of Eastern Finland
Varoon Bakshi, University of Eastern Finland

Program Committee

Laszlo Fejes, Hungarian Research Centre for Linguistics
Gunta Klava, University of Latvia
Sourav Das, Indian Institute of Information Technology Kalyani
Balazs Indig, Eotvos Lorand University
Laleh Davoodi, Abo Akademi University
Flammie A Pirinen, Norgga arktalas universitehta
Trond Trosterud, University of Tromso
Csilla Horvath, University of Szeged
Iaroslav Chelombitko, Neapolis University Paphos
Aleksei Dorkin, University of Tartu
Sebastian Oliver Eck, University of Oxford
Khalid Alnajjar, F-Secure Oyj
Jules Bouton, Universite Paris Cite
Valts Ernstreits, University of Latvia
Michael Riessler, University of Eastern Finland
Timofey Arkhangelskiy, Universitat Hamburg
Samy Ouzerrout, Universite d'Orleans
Jyoti Kunal Shah, Automatic Data Processing
Youngsook Song, Lablup
David Dale, FAIR at Meta
Abhi Desai, New England College

Table of Contents

Program

Wednesday, December 10, 2025

14:00 - 18:00 *Karelian Workshop*

Thursday, December 11, 2025

09:15 - 09:30 *Workshop Opening*

09:30 - 10:30 *Keynote – Josh Wilbur*

10:30 - 11:20 *Lightning Talks*

11:20 - 11:40 *Coffee Break*

11:40 - 12:40 *Oral Session 1*

From NLG Evaluation to Modern Student Assessment in the Era of ChatGPT: The Great Misalignment Problem and Pedagogical Multi-Factor Assessment (P-MFA)
Mika Hämäläinen and Kimmo Leiviskä Leiviskä

Benchmarking Finnish Lemmatizers across Historical and Contemporary Texts
Emily Öhman, Leo Huovinen and Mika Hämäläinen

The world's first South Sámi TTS – a revitalisation effort of an endangered language by reviving a legacy voice
Katri Hiovain-Asikainen, Thomas B. Kjærstad, Maja Lisa Kappfjell and Sjur N. Moshagen

12:40 - 13:40 *Lunch*

13:40 - 15:00 *Oral Session 2*

Can advances in NLP lead to worse results for Uralic languages and how can we fight back? Experiences from the world of automatic spell-checking and correction for Finnish
Flammie A Pirinen

A Hybrid Multilingual Approach to Sentiment Analysis for Uralic and Low-Resource Languages: Combining Extractive and Abstractive Techniques
Mikhail Krasitskii, Olga Kolesnikova, Grigori Sidorov and Alexander Gelbukh

Language technology for the minority Finnic languages
Flammie A Pirinen, Trond Trosterud and Jack Rueter

Thursday, December 11, 2025 (continued)

Kildin Saami-Russian-(English) Parallel Corpus Building
Evan Hansen

15:00 - 15:20 *Coffee Break*

15:20 - 16:40 *Oral Session 3*

SampoNLP: A Self-Referential Toolkit for Morphological Analysis of Subword Tokenizers
Iaroslav Chelombitko, Ekaterina Chelombitko and Aleksey Komissarov

Timur and the Mansi spellchecker
Csilla Horváth

ORACLE: Time-Dependent Recursive Summary Graphs for Foresight on News Data Using LLMs
Lev Kharlashkin, Eiaki V. Morooka, Yehor Tereschenko and Mika Hämäläinen

Creating a multi-layer Treebank for Tundra Nenets
Nikolett Mus, Bruno Guillaume, Sylvain Kahane and Daniel Zeman

Friday, December 12, 2025

10:00 - 10:00 *Day 3 Opening*

10:00 - 11:00 *Oral Session 4*

Benchmarking Large Language Models for Lemmatization and Translation of Finnic Runosongs
Lidia Pivovarova, Kati Kallio, Antti Kanner, Jakob Lindström, Eetu Mäkelä, Liina Saarlo, Kaarel Veskis and Mari Väina

Fine-Tuning Whisper for Kildin Sami
Enzo Gamboni

Digitization Work at the Finno-Ugrian Society: Livonian Case Study
Niko Partanen, Jack Rueter and Valts Ernštreits

11:00 - 12:00 *Lunch*

12:00 - 13:00 *Oral Session 5*

Siberian Ingrian Finnish: FST and IGTs
Ivan Ubaleht

Case–Number Dissociation in Finnish Noun Embeddings:fastText vs. BERT Layer Effects
Alexandre Nikolaev, Yu-Ying Chuang and R. Harald Baayen

Evaluating OpenAI GPT Models for Translation of Endangered UralicLanguages: A Comparison of Reasoning and Non-Reasoning Architectures
Yehor Tereschenko, Mika Hämäläinen and Svitlana Myroniuk

13:00 - 13:20 *Coffee Break*

13:20 - 14:20 *SIGUR Business Meeting*