ACL 2025

10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025)

Co-located with the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)

July 31, 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 978-1-959429-57-9

Preface

This volume contains the papers presented at Slavic NLP 2025: the 10^{th} Workshop on Natural Language Processing (NLP) for Slavic Languages. The workshop is organized by ACL SIGSLAV, the Special Interest Group of the Association for Computational Linguistics on NLP for Slavic Languages.

The Slavic NLP (formerly BSNLP) workshops have served as a key venue for over fifteen years, with a mission to advance the state of NLP for Slavic languages—languages that are spoken by more than 400 million people globally and represent an important part of the linguistic and cultural fabric of Eurasia.

The 2025 edition of the workshop continues the proud tradition established by the earlier BSNLP workshops, which were held in conjunction with the following venues:

- ACL 2007 Conference in Prague, Czech Republic.
- IIS 2009: Intelligent Information Systems, in Kraków, Poland.
- TSD 2011: 14th International Conference on Text, Speech and Dialogue in Plzeň, Czech Republic.
- ACL 2013 Conference in Sofia, Bulgaria.
- RANLP 2015 Conference in Hissar, Bulgaria.
- EACL 2017 Conference in Valencia, Spain.
- ACL 2019 Conference in Florence, Italy.
- EACL 2021 Conference in Kyiv, Ukraine.
- EACL 2023 Conference in Dubrovnik, Croatia.

Despite the importance and rich linguistic heritage of Slavic languages, the development of NLP tools and resources still lags behind those available for high-resource languages such as English. Many Slavic languages, particularly those spoken by smaller communities or lacking official EU status, remain underrepresented in both datasets and commercial NLP solutions.

Slavic languages present both practical and theoretical challenges: rich inflectional morphology, aspectual systems, and relatively free word order are among the many features that complicate parsing, translation, and generation. Yet, these features also create opportunities for research that is both linguistically informed and technically innovative.

Slavic NLP 2025 continues to serve as a forum for uniting researchers from academia and industry with a shared interest in advancing NLP for these languages. A recurring theme throughout this year's proceedings is the convergence of state-of-the-art methods –including large language models (LLMs), prompt-based learning, and multilingual transformers– with linguistically motivated problems, ranging from diachronic analysis to humor and persuasion detection.

This edition of the workshop features a diverse and ambitious research program. The accepted papers span a number of different Slavic languages and dialects, including Belarusian, Church Slavonic, Croatian, Czech, Macedonian, Polish, Russian, Slovak, Slovenian, and Rusyn. The topics address a rich spectrum of applications:

- Low-resource adaptation and language-specific modeling, including efforts to adapt definition modeling for Belarusian, build foundational models for Macedonian, and construct efficient summarization tools for Slovak.

- Diachronic and sociolinguistic analysis, through embeddings and curated datasets for Croatian news and Church Slavonic, as well as border effects in dialect variation for Rusyn.
- Bias and ethics in language technologies, such as examining gender representation in Czech and Slovenian LLM outputs and benchmarking LLM safety for Polish.
- Persuasion, propaganda, and disinformation detection, a central theme in this year's Shared Task and related contributions, with a variety of generative, multitask, and ensemble approaches.
- Speech and syntax-focused work, including filled pause detection across South and West Slavic languages and high-efficiency transformer-based speech models.

A highlight of Slavic NLP 2025 is the Shared Task on the Detection and Classification of Persuasion Techniques in Parliamentary Debates and Social Media, attracting a record number of participants. This task aligns directly with the global relevance of NLP for combating manipulation and fostering informed public discourse. Several papers in this volume present novel methods –ranging from multitask debiasing to explanation-based data augmentation– for tackling this complex challenge.

Altogether, this volume features papers selected through rigorous peer review. The contributions represent a mix of theoretical insights, resource development, and practical systems, offering a snapshot of current research and emerging directions in Slavic NLP.

We thank all authors for their excellent submissions, the reviewers for their thoughtful and constructive feedback, and the participants of the Shared Task for their enthusiasm and innovation. We also express our gratitude to the organizing committee and program chairs for their commitment to ensuring the continued success of Slavic NLP.

We hope that this collection will be a valuable resource for researchers and developers, and that it will continue to foster collaboration across the many communities interested in the Slavic languages and their computational study.

The SlavNLP Organizers: Jakub Piskorski, Preslav Nakov, Nikola Ljubešić, Pavel Přibáň, Roman Yangarber

Program Committee

Chairs

Nikola Ljubešić, Jožef Stefan Institute
Michal Marcinczuk, Samurai Labs
Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence
Jakub Piskorski, Polish Academy of Sciences
Pavel Přibáň, University of West Bohemia, Faculty of Applied Sciences
Roman Yangarber, University of Helsinki

Program Committee

Zeljko Agic, Unity Technologies
Ekaterina Artemova, Toloka.AI
Dimitar Dimitrov, University of Sofia St. Kliment Ohridski"
Filip Dobranić, Institute of Contemporary History
Marina Ernst, University of Koblenz
Radovan Garabik, L. Stur Institute of Linguistics, Slovak Academy of Sciences
Jacek Haneczok, Erste Group IT
Milos Jakubicek, Lexical Computing
Mikhail Kopotev, University of Helsinki
Ivan Koychev, Sofia University St. Kliment Ohridski"
Vladislav Kubon, Charles University
Gaurav Kumar, University of California San Diego
Wojciech Kusa, NASK National Research Institute
Arkadiusz Modzelewski, Polish-Japanese Academy of Information Technology
Ivo Moravski, Sofia University
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences
Petya Osenova, Sofia University St. Kl. Ohridskiand IICT-BAS
Alexander Panchenko, Skolkovo Institue of Science and Technology
Lidia Pivovarova, University of Helsinki
Senja Pollak, Jožef Stefan Institute
Marko Robnik-Sikonja, University of Ljubljana, Faculty of Computer and Information Science
Alexandr Rosen, Charles University, Prague
Agata Savary, Paris-Saclay University
Serge Sharoff, University of Leeds
Marko Tadić, University of Zagreb, Faculty of Humanities and Social Sciences
Marcin Woliński, Institute of Computer Science, Polish Academy of Sciences
Daniel Zeman, Charles University, Faculty of Mathematics and Physics

Table of Contents