VarDial 2025

VarDial 2025 - The Twelfth Workshop on NLP for Similar Languages, Varieties and Dialects

Proceedings of the Workshop

January 19, 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 979-8-89176-208-4

Preface

These proceedings include the 17 papers presented at the Twelfth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2025), co-located with the 31st International Conference on Computational Linguistics (COLING 2025). VarDial was held in Abu Dhabi, UAE.

Despite the short interval between the 2024 and 2025 editions of VarDial, we are glad to see that VarDial continues to serve the community as the main venue for researchers interested in the computational processing of language variation. The papers accepted this year address a wide range of topics, such as normalization and dialectal translation, native language identification, and slot and intent detection. We also see several papers making use of and evaluating large language models on variety-related tasks. Once again, these proceedings are characterized by great linguistic diversity, with work on regional English dialects, Portuguese, Luxemburgish, Church Slavic, and Arabic, to name just a few.

As in previous editions, VarDial 2025 features an evaluation campaign with the NorSID shared task on slot, intent and dialect identification for Norwegian dialects. Slot and intent detection were already included in a VarDial shared task in 2023, but without including Norwegian data. Likewise, language and dialect identification tasks have been very common at past editions of VarDial, but this is the first dialect identification featuring varieties of Norwegian. This volume includes the system description papers prepared by the four participating teams, as well as a report written by the task organizers summarizing the results and findings of the evaluation campaign.

Finally, we would like to take this opportunity to thank all the shared task organizers and the participants for their hard work. We further thank the VarDial program committee members for being an important part of the workshop's success.

The VarDial workshop organizers:

Yves Scherrer, Tommi Jauhiainen, Nikola Ljubešić, Preslav Nakov, Jörg Tiedemann, and Marcos Zampieri

http://sites.google.com/view/vardial-2025/

Organizing Committee

Organizers:

Yves Scherrer, University of Oslo (Norway)
Tommi Jauhiainen, University of Helsinki (Finland)
Nikola Ljubešić, Jožef Stefan Institute and University of Ljubljana (Slovenia)
Preslav Nakov, Mohamed bin Zayed University of Artificial Intelligence (UAE)
Jörg Tiedemann, University of Helsinki (Finland)
Marcos Zampieri, George Mason University (USA)

Program Committee

Program Committee:

Noëmi Aepli (University of Zurich, Switzerland)
César Aguilar (Universidad Veracruzana, Mexico)
Sina Ahmadi (George Mason University, United States)
Laura Alonso y Alemany (University of Cordoba, Argentina)
Jorge Baptista (University of Algarve and INESC-ID, Portugal)
Delphine Bernhard (University of Strasbourg, France)
Gabriel Bernier-Colborne (National Research Council, Canada)
Verena Blaschke (LMU Munich, Germany)
Francis Bond (Nanyang Technological University, Singapore)
David Chiang (University of Notre Dame, United States)
Steven Coats (University of Oulu, Finland)
Çağrı Çöltekin (University of Tübingen, Germany)
Stefanie Dipper (Ruhr University Bochum, Germany)
Sascha Diwersy (University of Montpellier, France)
Mark Dras (Macquarie University, Australia)
Jonathan Dunn (University of Illinois Urbana-Champaign, United States)
Pablo Gamallo (University of Santiago de Compostela, Spain)
Rob van der Goot (IT University Copenhagen, Denmark)
Cyril Goutte (National Research Council, Canada)
Nizar Habash (New York University Abu Dhabi, UAE)
Radu Ionescu (University of Bucharest, Romania)
Anjali Kantharuban (Carnegie Mellon University, United States)
John McCrae (University of Galway, Ireland)
Surafel Melaku Lakew (FBK , Italy)
Aleksandra Miletić (University of Helsinki, Finland)
Filip Miletić (University of Stuttgart, Germany)
John Nerbonne (University of Groningen, Netherlands and University of Freiburg, Germany)
Ekaterina Lapshinova-Koltunski (University of Hildesheim, Germany)
Lung-Hao Lee (National Yang Ming Chiao Tung University, Taiwan)
Maciej Ogrodniczuk (Institute of Computer Science, Polish Academy of Sciences, Poland)
Petya Osenova (Bulgarian Academy of Sciences, Bulgaria)
Jelena Prokic (Leiden University, Netherlands)
Christoph Purschke (University of Luxembourg, Luxembourg)
Francisco Rangel (Autoritas Consulting, Spain)
Reinhard Rapp (University of Mainz, Germany)
Tanja Samardžić (University of Zurich, Switzerland)
Serge Sharoff (University of Leeds, United Kingdom)
Miikka Silfverberg (University of British Columbia, Canada)
Milena Slavcheva (Bulgarian Academy of Sciences, Bulgaria)
Aarohi Srivastava (University of Notre Dame, United States)
Marco Tadić (University of Zagreb, Croatia)
Joel Tetreault (Dataminr, United States)
Pidong Wang (Google Inc., United States)
Taro Watanabe (Google Inc., Japan)

Table of Contents