Twelfth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2025)

Abu Dhabi, United Arab Emirates
19 January 2025

ISBN: 979-8-3313-1374-6

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2025) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2026)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Findings of the VarDial Evaluation Campaign 2025: The NorSID Shared Task on Norwegian Slot, Intent and Dialect Identification
Yves Scherrer, Rob van der Goot and Petter Mæhlum 1

Information Theory and Linguistic Variation: A Study of Brazilian and European Portuguese
Diego Alves 9

Leveraging Open-Source Large Language Models for Native Language Identification
Yee Man Ng and Ilia Markov 20

Adapting Whisper for Regional Dialects: Enhancing Public Services for Vulnerable Populations in the United Kingdom
Melissa Torgbi, Andrew Clayman, Jordan J. Speight and Harish Tayyar Madabushi 29

Large Language Models as a Normalizer for Transliteration and Dialectal Translation
Md Mahfuz Ibn Alam and Antonios Anastasopoulos 39

Testing the Boundaries of LLMs: Dialectal and Language-Variety Tasks
Fahim Faisal and Antonios Anastasopoulos 68

Text Generation Models for Luxembourgish with Limited Data: A Balanced Multilingual Strategy
Alistair Plum, Tharindu Ranasinghe and Christoph Purschke 93

Retrieval of Parallelizable Texts Across Church Slavic Variants
Piroska Lendvai, Uwe Reichel, Anna Jouravel, Achim Rabus and Elena Renje 105

Neural Text Normalization for Luxembourgish Using Real-Life Variation Data
Anne-Marie Lutgen, Alistair Plum, Christoph Purschke and Barbara Plank 115

Improving Dialectal Slot and Intent Detection with Auxiliary Tasks: A Multi-Dialectal Bavarian Case Study
Xaver Maria Krückl, Verena Blaschke and Barbara Plank 128

Regional Distribution of the /el/-/æl/ Merger in Australian English
Steven Coats, Chloé Diskin-Holdaway and Debbie Loakes 147

Learning Cross-Dialectal Morphophonology with Syllable Structure Constraints
Salam Khalifa, Abdelrahim Qaddoumi, Jordan Kodner and Owen Rambow 157

Common Ground, Diverse Roots: The Difficulty of Classifying Common Examples in Spanish Varieties
Javier A. Lopetegui, Arij Riabi and Djamé Seddah 168

Add Noise, Tasks, or Layers? MaiNLP at the VarDial 2025 Shared Task on Norwegian Dialectal Slot and Intent Detection
Verena Blaschke, Felicia Körner and Barbara Plank 182

LTG at VarDial 2025 NorSID: More and Better Training Data for Slot and Intent Detection
Marthe Midtgaard, Petter Mæhlum and Yves Scherrer 200

HiTZ at VarDial 2025 NorSID: Overcoming Data Scarcity with Language Transfer and Automatic Data Annotation

Jaione Bengoetxea, Mikel Zubillaga, Ekhi Azurmendi, Maite Heredia, Julen Etxaniz, Markel Ferro and Jeremy Barnes 209

CUFE@VarDial 2025 NorSID: Multilingual BERT for Norwegian Dialect Identification and Intent Detection

Michael Ibrahim 220