# Table of Contents

*Findings of the VarDial Evaluation Campaign 2017*
Marcos Zampieri, Shervin Malmasi, Nikola Ljubešić, Preslav Nakov, Ahmed Ali, Jörg Tiedemann, Yves Scherrer and Noëmi Aepli .......................................................... 1

*Dialectometric analysis of language variation in Twitter*
Gonzalo Donoso and David Sanchez .................................................. 16

*Computational analysis of Gondi dialects*
Taraka Rama, Çağrı Çöltekin and Pavel Sofroniev ............................... 26

*Investigating Diatopic Variation in a Historical Corpus*
Stefanie Dipper and Sandra Waldenberger ........................................ 36

*Author Profiling at PAN: from Age and Gender Identification to Language Variety Identification (invited talk)*
Paolo Rosso ....................................................................................... 46

*The similarity and Mutual Intelligibility between Amharic and Tigrigna Varieties*
Tekabe Legesse Feleke ........................................................................ 47

*Why Catalan-Spanish Neural Machine Translation? Analysis, comparison and combination with standard Rule and Phrase-based technologies*
Marta R. Costa-jussà ........................................................................... 55

*Kurdish Interdialect Machine Translation*
Hossein Hassani ................................................................................ 63

*Twitter Language Identification Of Similar Languages And Dialects Without Ground Truth*
Jennifer Williams and Charlie Dagli .................................................. 73

*Multi-source morphosyntactic tagging for spoken Rusyn*
Yves Scherrer and Achim Rabus ......................................................... 84

*Identifying dialects with textual and acoustic cues*
Abualsoud Hanani, Aziz Qaroush and Stephen Taylor ......................... 93

*Evaluating HeLI with Non-Linear Mappings*
Tommi Jauhiainen, Krister Lindén and Heidi Jauhiainen ....................... 102

*A Perplexity-Based Method for Similar Languages Discrimination*
Pablo Gamallo, Jose Ramom Pichel and Iñaki Alegria .............................. 109

*Improving the Character Ngram Model for the DSL Task with BM25 Weighting and Less Frequently Used Feature Sets*
Yves Bestgen .................................................................................... 115

*Discriminating between Similar Languages with Word-level Convolutional Neural Networks*
Marcelo Criscuolo and Sandra Maria Aluisio ......................................... 124

*Cross-lingual dependency parsing for closely related languages - Helsinki’s submission to VarDial 2017*
Jörg Tiedemann .................................................................................. 131
Discriminating between Similar Languages Using a Combination of Typed and Untyped Character N-grams and Words
Helena Gomez, Ilia Markov, Jorge Baptista, Grigori Sidorov and David Pinto .......................... 137

Tübingen system in VarDial 2017 shared task: experiments with language identification and cross-lingual parsing
Çağrı Çöltekin and Taraka Rama ................................................................. 146

When Sparse Traditional Models Outperform Dense Neural Networks: the Curious Case of Discriminating between Similar Languages
Maria Medvedeva, Martin Kroon and Barbara Plank .......................................... 156

German Dialect Identification in Interview Transcriptions
Shervin Malmasi and Marcos Zampieri ..................................................... 164

CLUZH at VarDial GDI 2017: Testing a Variety of Machine Learning Tools for the Classification of Swiss German Dialects
Simon Clematide and Peter Makarov .................................................... 170

Arabic Dialect Identification Using iVectors and ASR Transcripts
Shervin Malmasi and Marcos Zampieri .................................................... 178

Discriminating between Similar Languages using Weighted Subword Features
Adrien Barbaresi ......................................................................................... 184

Exploring Lexical and Syntactic Features for Language Variety Identification
Chris van der Lee and Antal van den Bosch ........................................... 190

Learning to Identify Arabic and German Dialects using Multiple Kernels
Radu Tudor Ionescu and Andrei Butnaru .................................................. 200

Slavic Forest, Norwegian Wood
Rudolf Rosa, Daniel Zeman, David Mareček and Zdeněk Žabokrtský ............................ 210