

Ninth Workshop on NLP for Similar Languages, Varities and Dialects (VarDial 2022)

Gyeongju, South Korea
16 October 2022

ISBN: 978-1-7138-6199-7

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© (2022) by the Association for Computational Linguistics
All rights reserved.

Copyright of each paper stays with the respective authors (or their employers).

Printed with permission by Curran Associates, Inc. (2023)

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

<i>Findings of the VarDial Evaluation Campaign 2022</i> Noëmi Aepli, Antonios Anastasopoulos, Adrian-Gabriel Chifu, William Domingues, Fahim Faisal, Mihaela Gaman, Radu Tudor Ionescu and Yves Scherrer	1
<i>Social Context and User Profiles of Linguistic Variation on a Micro Scale</i> Olga Kellert and Nicholas Hill Matlis	14
<i>dialectR: Doing Dialectometry in R</i> Ryan Soh-Eun Shim and John Nerbonne	20
<i>Low-Resource Neural Machine Translation: A Case Study of Cantonese</i> Evelyn Kai-Yan Liu	28
<i>Phonetic, Semantic, and Articulatory Features in Assamese-Bengali Cognate Detection</i> Abhijnan Nath, Rahul Ghosh and Nikhil Krishnaswamy	41
<i>Mapping Phonology to Semantics: A Computational Model of Cross-Lingual Spoken-Word Recognition</i> Iuliia Zaitova, Badr Abdullah and Dietrich Klakow	54
<i>Annotating Norwegian language varieties on Twitter for Part-of-speech</i> Petter Mæhlum, Andre Kåsen, Samia Touileb and Jeremy Barnes	64
<i>OcWikiDisc: a Corpus of Wikipedia Talk Pages in Occitan</i> Aleksandra Miletic and Yves Scherrer	70
<i>Is Encoder-Decoder Transformer the Shiny Hammer?</i> Nat Gillin	80
<i>The Curious Case of Logistic Regression for Italian Languages and Dialects Identification</i> Giacomo Camposampiero, Quynh Anh Nguyen and Francesco Di Stefano	86
<i>Neural Networks for Cross-domain Language Identification. Phlyers @ Vardial 2022</i> Andrea Ceolin	99
<i>Transfer Learning Improves French Cross-Domain Dialect Identification: NRC @ VarDial 2022</i> Gabriel Bernier-Colborne, Serge Leger and Cyril Goutte	109
<i>Italian Language and Dialect Identification and Regional French Variety Detection using Adaptive Naive Bayes</i> Tommi Jauhiainen, Heidi Jauhiainen and Krister Lindén	119