Sixth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2019)

Minneapolis, Minnesota, USA
7 June 2019

ISBN: 978-1-5108-8768-8

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.

Printed by Curran Associates, Inc. (2019)

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

Conference Program

Friday, June 7, 2019

9:15–9:30 ***Opening***

9:30–10:00 *A Report on the Third VarDial Evaluation Campaign*
Marcos Zampieri, Shervin Malmasi, Yves Scherrer, Tanja Samardzic, Francis Tyers, Miikka Silfverberg, Natalia Klyueva, Tung-Le Pan, Chu-Ren Huang, Radu Tudor Ionescu, Andrei M. Butnaru and Tommi Jauhiainen

10:00–10:30 *Improving Cuneiform Language Identification with BERT*
Gabriel Bernier-Colborne, Cyril Goutte and Serge Leger

10:30–11:00 ***Coffee break***

11:00–11:30 *Joint Approach to Deromanization of Code-mixed Texts*
Rashed Rubby Riyadh and Grzegorz Kondrak

11:30–12:00 *Char-RNN for Word Stress Detection in East Slavic Languages*
Ekaterina Chernyak, Maria Ponomareva and Kirill Milintsevich

12:00–12:30 *Modeling Global Syntactic Variation in English Using Dialect Classification*
Jonathan Dunn

12:30–14:00 ***Lunch***

14:00–15:00 ***Invited talk — David Yarowsky (Johns Hopkins University): Massively Multilingual Translingual Knowledge Transfer***

15:00–15:30 *Language Discrimination and Transfer Learning for Similar Languages: Experiments with Feature Combinations and Adaptation*
Nianheng Wu, Eric DeMattos, Kwok Him So, Pin-zhen Chen and Çağrı Çöltekin

15:30–16:00 ***Coffee break***

16:00–17:00 ***Poster Session***

Friday, June 7, 2019 (continued)

Variation between Different Discourse Types: Literate vs. Oral
Katrin Ortmann and Stefanie Dipper

Neural Machine Translation between Myanmar (Burmese) and Rakhine (Arakanese)
Thazin Myint Oo, Ye Kyaw Thu and Khin Mar Soe

Language and Dialect Identification of Cuneiform Texts
Tommi Jauhiainen, Heidi Jauhiainen, Tero Alstola and Krister Lindén

Leveraging Pretrained Word Embeddings for Part-of-Speech Tagging of Code Switching Data
Fahad AlGhamdi and Mona Diab

Toward a deep dialectological representation of Indo-Aryan
Chundra Cathcart

Naive Bayes and BiLSTM Ensemble for Discriminating between Mainland and Taiwan Variation of Mandarin Chinese
Li Yang and Yang Xiang

BAM: A combination of deep and shallow models for German Dialect Identification.
Andrei M. Butnaru

The R2I_LIS Team Proposes Majority Vote for VarDial's MRC Task
Adrian-Gabriel Chifu

Initial Experiments In Cross-Lingual Morphological Analysis Using Morpheme Segmentation
Vladislav Mikhailov, Lorenzo Tosi, Anastasia Khorosheva and Oleg Serikov

Neural and Linear Pipeline Approaches to Cross-lingual Morphological Analysis
Çağrı Çöltekin and Jeremy Barnes

Ensemble Methods to Distinguish Mainland and Taiwan Chinese
Hai Hu, Wen Li, He Zhou, Zuoyu Tian, Yiwen Zhang and Liang Zou

SC-UPB at the VarDial 2019 Evaluation Campaign: Moldavian vs. Romanian Cross-Dialect Topic Identification
Cristian Onose, Dumitru-Clementin Cercel and Stefan Trausan-Matu

Friday, June 7, 2019 (continued)

Discriminating between Mandarin Chinese and Swiss-German varieties using adaptive language models
Tommi Jauhiainen, Krister Lindén and Heidi Jauhiainen

Investigating Machine Learning Methods for Language and Dialect Identification of Cuneiform Texts
Ehsan Doostmohammadi and Minoo Nassajian

TwistBytes - Identification of Cuneiform Languages and German Dialects at VarDial 2019
Fernando Benites, Pius von Däniken and Mark Cieliebak

DTeam @ VarDial 2019: Ensemble based on skip-gram and triplet loss neural networks for Moldavian vs. Romanian cross-dialect topic identification
Diana Tudoreanu

Experiments in Cuneiform Language Identification
Gustavo Henrique Paetzold and Marcos Zampieri

17:00–17:30 *Comparing Pipelined and Integrated Approaches to Dialectal Arabic Neural Machine Translation*
Pamela Shapiro and Kevin Duh

17:30–18:00 *Cross-lingual Annotation Projection Is Effective for Neural Part-of-Speech Tagging*
Matthias Huck, Diana Dutka and Alexander Fraser

18:00–18:15 ***Closing Remarks***