

# **5th Workshop on Research in Computational Typology and Multilingual NLP (SIGTYP 2023)**

Dubrovnik, Croatia  
6 May 2023

ISBN: 978-1-7138-7387-7

**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© (2023) by the Association for Computational Linguistics  
All rights reserved.

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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>You Can Have Your Data and Balance It Too: Towards Balanced and Efficient Multilingual Models</i> Tomasz Limisiewicz, Dan Malkin and Gabriel Stanovsky .....	1
<i>Multilingual End-to-end Dependency Parsing with Linguistic Typology knowledge</i> Chinmay Choudhary and Colm O’riordan .....	12
<i>Identifying the Correlation Between Language Distance and Cross-Lingual Transfer in a Multilingual Representation Space</i> Fred Philippy, Siwen Guo and Shohreh Haddadan .....	22
<i>Using Modern Languages to Parse Ancient Ones: a Test on Old English</i> Luca Brigada Villa and Martina Giarda .....	30
<i>The Denglisch Corpus of German-English Code-Switching</i> Doreen Osmelak and Shuly Wintner .....	42
<i>Trimming Phonetic Alignments Improves the Inference of Sound Correspondence Patterns from Multilingual Wordlists</i> Frederic Blum and Johann-Mattis List .....	52
<i>A Crosslinguistic Database for Combinatorial and Semantic Properties of Attitude Predicates</i> Deniz Özyıldız, Ciyang Qing, Floris Roelofsen, Maribel Romero and Wataru Uegaki .....	65
<i>Corpus-based Syntactic Typological Methods for Dependency Parsing Improvement</i> Diego Alves, Božo Bekavac, Daniel Zeman and Marko Tadić .....	76
<i>Cross-lingual Transfer Learning with Persian</i> Sepideh Mollanorozy, Marc Tanti and Malvina Nissim .....	89
<i>Information-Theoretic Characterization of Vowel Harmony: A Cross-Linguistic Study on Word Lists</i> Julius Steuer, Johann-Mattis List, Badr M. Abdullah and Dietrich Klakow .....	96
<i>Revisiting Dependency Length and Intervener Complexity Minimisation on a Parallel Corpus in 35 Languages</i> Andrew Thomas Dyer .....	110
<i>Does Topological Ordering of Morphological Segments Reduce Morphological Modeling Complexity? A Preliminary Study on 13 Languages</i> Andreas Shcherbakov and Ekaterina Vylomova .....	120
<i>Findings of the SIGTYP 2023 Shared task on Cognate and Derivative Detection For Low-Resourced Languages</i> Priya Rani, Koustava Goswami, Adrian Doyle, Theodorus Franssen, Bernardo Stearns and John P. McCrae .....	126
<i>ÚFAL Submission for SIGTYP Supervised Cognate Detection Task</i> Tomasz Limisiewicz .....	132
<i>CoToHiLi at SIGTYP 2023: Ensemble Models for Cognate and Derivative Words Detection</i> Liviu P. Dinu, Ioan-Bogdan Iordache and Ana Sabina Uban .....	137
<i>Multilingual BERT has an Accent: Evaluating English Influences on Fluency in Multilingual Models</i> Isabel Papadimitriou, Kezia Lopez and Dan Jurafsky .....	143

<i>Grambank’s Typological Advances Support Computational Research on Diverse Languages</i>	
Hannah J. Haynie, Damián Blasi, Hedvig Skirgård, Simon J. Greenhill, Quentin D. Atkinson and Russell D. Gray .....	147
<i>Language-Agnostic Measures Discriminate Inflection and Derivation</i>	
Coleman Haley, Edoardo M. Ponti and Sharon Goldwater .....	150
<i>Gradual Language Model Adaptation Using Fine-Grained Typology</i>	
Marcell Richard Fekete and Johannes Bjerva .....	153
<i>On the Nature of Discrete Speech Representations in Multilingual Self-supervised Models</i>	
Badr M. Abdullah, Mohammed Maqsood Shaik and Dietrich Klakow .....	159