

# **18th SIGMORPHON Workshop on Computational Research in Phonetics Phonology, and Morphology (SIGMORPHON 2021)**

Held online due to COVID-19

Bangkok, Thailand  
5 August 2021

ISBN: 978-1-7138-3385-7

**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© (2021) by the Association for Computational Linguistics and  
the Asian Federation of Natural Language Processing  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

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>Towards Detection and Remediation of Phonemic Confusion</i> Francois Roewer-Despres, Arnold Yeung and Ilan Kogan .....	1
<i>Recursive prosody is not finite-state</i> Hossep Dolatian, Aniello De Santo and Thomas Graf .....	11
<i>The Match-Extend serialization algorithm in Multiprecedence</i> Maxime Papillon .....	23
<i>Incorporating tone in the calculation of phonotactic probability</i> James Kirby .....	32
<i>MorphyNet: a Large Multilingual Database of Derivational and Inflectional Morphology</i> Khuyagbaatar Batsuren, Gábor Bella and fausto giunchiglia .....	39
<i>A Study of Morphological Robustness of Neural Machine Translation</i> Sai Muralidhar Jayanthi and Adithya Pratapa .....	49
<i>Sample-efficient Linguistic Generalizations through Program Synthesis: Experiments with Phonology Problems</i> Saujas Vaduguru, Aalok Sathe, Monojit Choudhury and Dipti Sharma .....	60
<i>Findings of the SIGMORPHON 2021 Shared Task on Unsupervised Morphological Paradigm Clustering</i> Adam Wiemerslage, Arya D. McCarthy, Alexander Erdmann, Garrett Nicolai, Manex Agirrezabal, Miikka Silfverberg, Mans Hulden and Katharina Kann .....	72
<i>Adaptor Grammars for Unsupervised Paradigm Clustering</i> Kate McCurdy, Sharon Goldwater and Adam Lopez .....	82
<i>Orthographic vs. Semantic Representations for Unsupervised Morphological Paradigm Clustering</i> E. Margaret Perkoff, Josh Daniels and Alexis Palmer .....	90
<i>Unsupervised Paradigm Clustering Using Transformation Rules</i> Changbing Yang, Garrett Nicolai and Miikka Silfverberg .....	98
<i>Paradigm Clustering with Weighted Edit Distance</i> Andrew Gerlach, Adam Wiemerslage and Katharina Kann .....	107
<i>Results of the Second SIGMORPHON Shared Task on Multilingual Grapheme-to-Phoneme Conversion</i> Lucas F.E. Ashby, Travis M. Bartley, Simon Clemenide, Luca Del Signore, Cameron Gibson, Kyle Gorman, Yeonju Lee-Sikka, Peter Makarov, Aidan Malanoski, Sean Miller, Omar Ortiz, Reuben Raff, Arundhati Sengupta, Bora Seo, Yulia Spektor and Winnie Yan .....	115
<i>Data augmentation for low-resource grapheme-to-phoneme mapping</i> Michael Hammond .....	126
<i>Linguistic Knowledge in Multilingual Grapheme-to-Phoneme Conversion</i> Roger Yu-Hsiang Lo and Garrett Nicolai .....	131
<i>Avengers, Ensemble! Benefits of ensembling in grapheme-to-phoneme prediction</i> Vasundhara Gautam, Wang Yau Li, Zafarullah Mahmood, Frederic Mailhot, Shreekantha Nadig, Riqiang WANG and Nathan Zhang .....	141

<i>CLUZH at SIGMORPHON 2021 Shared Task on Multilingual Grapheme-to-Phoneme Conversion: Variations on a Baseline</i>	
Simon Clematide and Peter Makarov .....	148
<i>What transfers in morphological inflection? Experiments with analogical models</i>	
Micha Elsner .....	154
<i>Simple induction of (deterministic) probabilistic finite-state automata for phonotactics by stochastic gradient descent</i>	
Huteng Dai and Richard Futrell .....	167
<i>Recognizing Reduplicated Forms: Finite-State Buffered Machines</i>	
Yang Wang .....	177
<i>An FST morphological analyzer for the Gitksan language</i>	
Clarissa Forbes, Garrett Nicolai and Miikka Silfverberg .....	188
<i>Comparative Error Analysis in Neural and Finite-state Models for Unsupervised Character-level Transduction</i>	
Maria Ryskina, Eduard Hovy, Taylor Berg-Kirkpatrick and Matthew R. Gormley .....	198
<i>Finite-state Model of Shupamem Reduplication</i>	
Magdalena Markowska, Jeffrey Heinz and Owen Rambow .....	212
<i>Improved pronunciation prediction accuracy using morphology</i>	
Dravyansh Sharma, Saumya Sahai, Neha Chaudhari and Antoine Bruguier .....	222