

SIGMORPHON 2020

**The 17th SIGMORPHON Workshop
on Computational Research in Phonetics
Phonology, and Morphology**

Proceedings of the Workshop

July 10, 2020



©2020 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-952148-19-4

Preface

Welcome to the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, to be held on July 10, 2020 as part of a virtual ACL. The workshop aims to bring together researchers interested in applying computational techniques to problems in morphology, phonology, and phonetics. Our program this year highlights the ongoing and important interaction between work in computational linguistics and work in theoretical linguistics. This year, work in both theoretical phonology and computational morphology were strongly represented in the workshop submissions. We received 14 submissions, and after a competitive reviewing process, we accepted 8. The workshop is privileged to present four invited talks this year, all from very respected members of the SIGMORPHON community.

This year also marks the fifth iteration of the SIGMORPHON Shared Task. Unlike previous years, this year, we hosted three distinct tasks:

Task 0: SIGMORPHON's fifth installment of its inflection generation shared task focuses on languages that are typologically diverse from languages in our previous tasks. Many of these languages are extremely low-resource. In this edition, we are specifically interested in inflection generation systems' ability to generalize to new languages, including languages that are typologically distinct. For example, if you have a neural network architecture that works well for a sample of Indo-European languages, should you expect the same architecture to also work well for Tupi–Guarani languages (where nouns are "declined" for tense)?

Task 1: This new task, the first of its kind at SIGMORPHON, focuses on grapheme-to-phoneme conversion. This technology is a key component of speech recognition and synthesis engines, but much of the existing published research is either limited to a small number of closely related languages/scripts, or uses proprietary data sets, limiting replicability. The training and development data consists of words and corresponding IPA pronunciations extracted from Wiktionary, a free online encyclopedia, in 15 languages and scripts. 9 teams submitted a total of 23 different systems.

Task 2: Task 2 fills the gap between recent SIGMORPHON shared tasks on morphological inflection learned from limited training data and completely unsupervised morphological generation by proposing the task of unsupervised morphological paradigm completion. The goal is to generate complete inflection tables exclusively from raw text and a lemma list for a known part of speech. 3 teams submitted a total of 7 different systems to tackle this new task.

We are grateful to the program committee for their careful and thoughtful reviews of the papers submitted this year. Likewise, we are thankful to the shared task organizers for their hard work in preparing the shared tasks. We are looking forward to a workshop covering a wide range of topics, and we hope for lively discussions.

Garrett Nicolai
Kyle Gorman
Ryan Cotterell

Organizers:

Garrett Nicolai (University of British Columbia, Canada)
Kyle Gorman (The Graduate Center, City University of New York, USA)
Ryan Cotterell (ETH Zürich, Switzerland)

Program Committee:

Çağrı Çöltekin (University of Tübingen, Germany)
Daniel Dakota (Indiana University, USA)
Ewan Dunbar (Université Paris Diderot, France)
Micha Elsner (The Ohio State University, USA)
Jeffrey Heinz (Stony Brook University, USA)
Mans Hulden (University of Colorado, USA)
Adam Jardine (Rutgers University, USA)
Christo Kirov (Google AI, USA)
Greg Kondrak (University of Alberta, Canada)
Sandra Kübler (Indiana University, USA)
Andrew Lamont (University of Massachusetts Amherst, USA)
Fred Mailhot (Dialpad, Inc., Canada)
Arya D. McCarthy (Johns Hopkins University, USA)
Kemal Oflazer (CMU Qatar, Qatar)
Jeff Parker (Brigham Young University, USA)
Gerald Penn (University of Toronto, Canada)
Jelena Prokic (Ludwig Maximilian University of Munich, Germany)
Mohamad Salameh (Huawei, Canada)
Miikka Silfverberg (University of British Columbia, Canada)
Kairit Sirts (University of Tartu, Estonia)
Kenneth Steimel (Indiana University, USA)
Francis Tyers (Indiana University, USA)
Ivan Vulić (University of Cambridge, United Kingdom)
Ekaterina Vylomova (University of Melbourne, Australia)
Anssi Yli-Jyrä (University of Helsinki, Finland)
Kristine Yu (University of Massachusetts Amherst, USA)

Shared Task Organizers:

Task 0:

Adina Williams (Facebook AI Research NYC, USA)
Antonios Anastasopoulos (Carnegie Mellon University, USA)
Christo Kirov (Google Research NYC, USA)
Ekaterina Vylomova (University of Melbourne, Australia)
Eleanor Chodroff (University of York, UK)
Elizabeth Salesky (Johns Hopkins University, USA)
Garrett Nicolai (University of British Columbia, Canada)
Mans Hulden (University of Colorado Boulder, USA)
Miikka Silfverberg (University of British Columbia, Canada)
Ryan Cotterell (ETH Zürich, Switzerland)
Sabrina Mielke (Johns Hopkins University, USA)
Shijie Wu (Johns Hopkins University, USA)
Andrej Krizhanovsky (Karelian Research Centre, Russia)
Edoardo Ponti (University of Cambridge, UK)
Elena Klyachko (National Research University Higher School of Economics, Russia)
Francis Tyers (Indiana University, US)
Hilaria Cruz (University of Louisville, US)
Ilya Yegorov (Lomonosov Moscow State University, Russia)
Irene Nikkarinen (University of Cambridge, UK)
Jennifer White (University of Cambridge, UK)
Josef Valvoda (University of Cambridge, UK)
Lucas Torroba Hennigen (University of Cambridge, UK)
Natalia Krizhanovsky (Karelian Research Centre, Russia)
Paula Czarnowska (University of Cambridge, UK)
Ran Zmigrod (University of Cambridge, UK)
Rowan Hall Maudslay (University of Cambridge, UK)
Svetlana Toldova (National Research University Higher School of Economics, Russia)
Tiago Pimentel (University of Cambridge, UK)

Task 1:

Kyle Gorman (Graduate Center, City University of New York, USA)
Lucas F. E. Ashby (Graduate Center, City University of New York, USA)
Aaron Goyzueta (Graduate Center, City University of New York, USA)
Arya D. McCarthy (Johns Hopkins University, USA)
Shijie Wu (Johns Hopkins University, USA)
Daniel You (Jericho High School, USA)

Task 2:

Arya McCarthy (Johns Hopkins University, USA)
Katharina Kann (University of Colorado Boulder, USA)
Garrett Nicolai (University of British Columbia, Canada)
Mans Hulden (University of Colorado Boulder, USA)
Chen Xia (Carnegie Mellon University, USA)
Huiming Jin (Carnegie Mellon University, USA)
Liwei Cai (Carnegie Mellon University, USA)
Yihui Peng (Carnegie Mellon University, USA)

Invited Speakers:

Jane Chandlee, (Haverford College, USA)

Bruce Hayes, (University of California, Los Angeles, USA)

Rob Malouf, (San Diego State University, USA)

Clara Vania (New York University, USA)

Table of Contents

<i>SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection</i>	
Ekaterina Vylomova, Jennifer White, Elizabeth Salesky, Sabrina J. Mielke, Shijie Wu, Edoardo Maria Ponti, Rowan Hall Maudslay, Ran Zmigrod, Josef Valvoda, Svetlana Toldova, Francis Tyers, Elena Klyachko, Ilya Yegorov, Natalia Krizhanovsky, Paula Czarnowska, Irene Nikkarinen, Andrew Krizhanovsky, Tiago Pimentel, Lucas Torroba Hennigen, Christo Kirov, Garrett Nicolai, Adina Williams, Antonios Anastasopoulos, Hilaria Cruz, Eleanor Chodroff, Ryan Cotterell, Miikka Silfverberg and Mans Hulden	1
<i>The SIGMORPHON 2020 Shared Task on Multilingual Grapheme-to-Phoneme Conversion</i>	
Kyle Gorman, Lucas F.E. Ashby, Aaron Goyzueta, Arya McCarthy, Shijie Wu and Daniel You .	40
<i>The SIGMORPHON 2020 Shared Task on Unsupervised Morphological Paradigm Completion</i>	
Katharina Kann, Arya D. McCarthy, Garrett Nicolai and Mans Hulden	51
<i>One-Size-Fits-All Multilingual Models</i>	
Ben Peters and André F. T. Martins	63
<i>Ensemble Self-Training for Low-Resource Languages: Grapheme-to-Phoneme Conversion and Morphological Inflection</i>	
Xiang Yu, Ngoc Thang Vu and Jonas Kuhn	70
<i>The CMU-LTI submission to the SIGMORPHON 2020 Shared Task 0: Language-Specific Cross-Lingual Transfer</i>	
Nikitha Murikinati and Antonios Anastasopoulos	79
<i>Grapheme-to-Phoneme Conversion with a Multilingual Transformer Model</i>	
Omnia ElSaadany and Benjamin Suter	85
<i>The NYU-CUBoulder Systems for SIGMORPHON 2020 Task 0 and Task 2</i>	
Assaf Singer and Katharina Kann	90
<i>The IMS-CUBoulder System for the SIGMORPHON 2020 Shared Task on Unsupervised Morphological Paradigm Completion</i>	
Manuel Mager and Katharina Kann	99
<i>SIGMORPHON 2020 Task 0 System Description: ETH Zürich Team</i>	
Martina Forster and Clara Meister	106
<i>KU-CST at the SIGMORPHON 2020 Task 2 on Unsupervised Morphological Paradigm Completion</i>	
Manex Agirrezabal and Jürgen Wedekind	111
<i>Low-Resource G2P and P2G Conversion with Synthetic Training Data</i>	
Bradley Hauer, Amir Ahmad Habibi, Yixing Luan, Arnob Mallik and Grzegorz Kondrak	117
<i>Frustratingly Easy Multilingual Grapheme-to-Phoneme Conversion</i>	
Nikhil Prabhu and Katharina Kann	123
<i>Exploring Neural Architectures And Techniques For Typologically Diverse Morphological Inflection</i>	
Pratik Jayarao, Siddhanth Pillay, Pranav Thombre and Aditi Chaudhary	128

<i>University of Illinois Submission to the SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection</i>	
Marc Canby, Aidana Karipbayeva, Bryan Lunt, Sahand Mozaffari, Charlotte Yoder and Julia Hockenmaier	137
<i>One Model to Pronounce Them All: Multilingual Grapheme-to-Phoneme Conversion With a Transformer Ensemble</i>	
Kaili Vesik, Muhammad Abdul-Mageed and Miikka Silfverberg	146
<i>Leveraging Principal Parts for Morphological Inflection</i>	
Ling Liu and Mans Hulden	153
<i>Linguist vs. Machine: Rapid Development of Finite-State Morphological Grammars</i>	
Sarah Beemer, Zak Boston, April Bukoski, Daniel Chen, Princess Dickens, Andrew Gerlach, Torin Hopkins, Parth Anand Jawale, Chris Koski, Akanksha Malhotra, Piyush Mishra, Saliha Muradoglu, Lan Sang, Tyler Short, Sagarika Shreevastava, Elizabeth Spaulding, Testumichi Umada, Beilei Xiang, Changbing Yang and Mans Hulden	162
<i>CLUZH at SIGMORPHON 2020 Shared Task on Multilingual Grapheme-to-Phoneme Conversion</i>	
Peter Makarov and Simon Clematide	171
<i>The UniMelb Submission to the SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection</i>	
Andreas Scherbakov	177
<i>Data Augmentation for Transformer-based G2P</i>	
Zach Ryan and Mans Hulden	184
<i>Transliteration for Cross-Lingual Morphological Inflection</i>	
Nikitha Murikinati, Antonios Anastasopoulos and Graham Neubig	189
<i>Evaluating Neural Morphological Taggers for Sanskrit</i>	
Ashim Gupta, Amrith Krishna, Pawan Goyal and Oliver Hellwig	198
<i>Getting the #life out of living: How Adequate Are Word-Pieces for Modelling Complex Morphology?</i>	
Stav Klein and Reut Tsarfaty	204
<i>Induced Inflection-Set Keyword Search in Speech</i>	
Oliver Adams, Matthew Wiesner, Jan Trmal, Garrett Nicolai and David Yarowsky	210
<i>Representation Learning for Discovering Phonemic Tone Contours</i>	
Bai Li, Jing Yi Xie and Frank Rudzicz	217
<i>Joint learning of constraint weights and gradient inputs in Gradient Symbolic Computation with constrained optimization</i>	
Max Nelson	224
<i>In search of isoglosses: continuous and discrete language embeddings in Slavic historical phonology</i>	
Chundra Cathcart and Florian Wandl	233
<i>Multi-Tiered Strictly Local Functions</i>	
Phillip Burness and Kevin McMullin	245

Conference Program

July 10, 2020

08:30–10:30 Morning Session

08:30–09:30 *Invited Talk: On Understanding Character-level Models for Representing Morphology*
Clara Vania, NYU

09:30–10:30 *Invited Talk: Recursive Schemes for Phonological Analysis*
Jane Chandlee, Haverford College

10:30–10:45 Break

10:45–12:30 Shared Task

10:45–11:00 *SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection*

Ekaterina Vylomova, Jennifer White, Elizabeth Salesky, Sabrina J. Mielke, Shijie Wu, Edoardo Maria Ponti, Rowan Hall Maudslay, Ran Zmigrod, Josef Valvoda, Svetlana Toldova, Francis Tyers, Elena Klyachko, Ilya Yegorov, Natalia Krizhanovsky, Paula Czarnowska, Irene Nikkarinen, Andrew Krizhanovsky, Tiago Pimentel, Lucas Torroba Hennigen, Christo Kirov, Garrett Nicolai, Adina Williams, Antonios Anastasopoulos, Hilaria Cruz, Eleanor Chodroff, Ryan Cotterell, Miikka Silfverberg and Mans Hulden

11:00–11:15 *The SIGMORPHON 2020 Shared Task on Multilingual Grapheme-to-Phoneme Conversion*

Kyle Gorman, Lucas F.E. Ashby, Aaron Goyzueta, Arya McCarthy, Shijie Wu and Daniel You

11:15–11:30 *The SIGMORPHON 2020 Shared Task on Unsupervised Morphological Paradigm Completion*

Katharina Kann, Arya D. McCarthy, Garrett Nicolai and Mans Hulden

11:30–12:30 *Shared Task Poster Session*
Multiple

One-Size-Fits-All Multilingual Models
Ben Peters and André F. T. Martins

Ensemble Self-Training for Low-Resource Languages: Grapheme-to-Phoneme Conversion and Morphological Inflection
Xiang Yu, Ngoc Thang Vu and Jonas Kuhn

The CMU-LTI submission to the SIGMORPHON 2020 Shared Task 0: Language-Specific Cross-Lingual Transfer
Nikitha Murikinati and Antonios Anastasopoulos

July 10, 2020 (continued)

Grapheme-to-Phoneme Conversion with a Multilingual Transformer Model

Omnia ElSaadany and Benjamin Suter

The NYU-CUBoulder Systems for SIGMORPHON 2020 Task 0 and Task 2

Assaf Singer and Katharina Kann

The IMS-CUBoulder System for the SIGMORPHON 2020 Shared Task on Unsupervised Morphological Paradigm Completion

Manuel Mager and Katharina Kann

SIGMORPHON 2020 Task 0 System Description: ETH Zürich Team

Martina Forster and Clara Meister

KU-CST at the SIGMORPHON 2020 Task 2 on Unsupervised Morphological Paradigm Completion

Manex Agirrezabal and Jürgen Wedekind

Low-Resource G2P and P2G Conversion with Synthetic Training Data

Bradley Hauer, Amir Ahmad Habibi, Yixing Luan, Arnob Mallik and Grzegorz Kondrak

Frustratingly Easy Multilingual Grapheme-to-Phoneme Conversion

Nikhil Prabhu and Katharina Kann

Exploring Neural Architectures And Techniques For Typologically Diverse Morphological Inflection

Pratik Jayarao, Siddhanth Pillay, Pranav Thombre and Aditi Chaudhary

University of Illinois Submission to the SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection

Marc Canby, Aidana Karipbayeva, Bryan Lunt, Sahand Mozaffari, Charlotte Yoder and Julia Hockenmaier

One Model to Pronounce Them All: Multilingual Grapheme-to-Phoneme Conversion With a Transformer Ensemble

Kaili Vesik, Muhammad Abdul-Mageed and Miikka Silfverberg

Leveraging Principal Parts for Morphological Inflection

Ling Liu and Mans Hulden

Linguist vs. Machine: Rapid Development of Finite-State Morphological Grammars

Sarah Beemer, Zak Boston, April Bukoski, Daniel Chen, Princess Dickens, Andrew Gerlach, Torin Hopkins, Parth Anand Jawale, Chris Koski, Akanksha Malhotra, Piyush Mishra, Saliha Muradoglu, Lan Sang, Tyler Short, Sagarika Shreevastava, Elizabeth Spaulding, Testumichi Umada, Beilei Xiang, Changbing Yang and Mans Hulden

July 10, 2020 (continued)

CLUZH at SIGMORPHON 2020 Shared Task on Multilingual Grapheme-to-Phoneme Conversion

Peter Makarov and Simon Clematide

The UniMelb Submission to the SIGMORPHON 2020 Shared Task 0: Typologically Diverse Morphological Inflection

Andreas Scherbakov

Data Augmentation for Transformer-based G2P

Zach Ryan and Mans Hulden

12:30–14:00 Lunch

14:00–15:06 Paper Session

14:00–14:10 *Transliteration for Cross-Lingual Morphological Inflection*

Nikitha Murikinati, Antonios Anastasopoulos and Graham Neubig

14:11–14:21 *Evaluating Neural Morphological Taggers for Sanskrit*

Ashim Gupta, Amrith Krishna, Pawan Goyal and Oliver Hellwig

14:22–14:32 *Getting the ##life out of living: How Adequate Are Word-Pieces for Modelling Complex Morphology?*

Stav Klein and Reut Tsarfaty

14:33–14:43 *Induced Inflection-Set Keyword Search in Speech*

Oliver Adams, Matthew Wiesner, Jan Trmal, Garrett Nicolai and David Yarowsky

14:44–14:54 *Representation Learning for Discovering Phonemic Tone Contours*

Bai Li, Jing Yi Xie and Frank Rudzicz

14:55–15:05 *Joint learning of constraint weights and gradient inputs in Gradient Symbolic Computation with constrained optimization*

Max Nelson

July 10, 2020 (continued)

15:06–15:30 Best Paper Session

15:06–15:18 *In search of isoglosses: continuous and discrete language embeddings in Slavic historical phonology*

Chundra Cathcart and Florian Wandl

15:18–15:30 *Multi-Tiered Strictly Local Functions*

Phillip Burness and Kevin McMullin

15:30–16:00 Break

16:00–18:00 Afternoon Session

16:00–17:00 *Invited Talk: Inflectional data science and human/computer-aided linguistic analysis*

Robert Malouf, San Diego State University

Invited Talk: Modeling failure in morphophonological learning

Bruce Hayes, UCLA

