

SIGMORPHON 2019

**The 16th SIGMORPHON Workshop
on Computational Research in Phonetics
Phonology, and Morphology**

Proceedings of the Workshop

August 2, 2019
Florence, Italy

©2019 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 978-1-950737-36-9

Preface

Welcome to the 16th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, to be held on August 2, 2019 in Florence, Italy. 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 20 submissions, and after a competitive reviewing process, we accepted 12. 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 fourth iteration of the SIGMORPHON Shared Task in Morphological Inflection, previously co-located with CoNLL. This year's task encouraged submissions in two important inflectional sub-tasks: cross-lingual inflection generation, and contextual morphological tagging.

23 teams participated in the task, with 14 teams submitting system papers describing their work, one of which is non-archival.

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 task. We are looking forward to a workshop covering a wide range of topics, and we hope for lively discussions.

Garrett Nicolai
Ryan Cotterell

Organizers:

Garrett Nicolai, Johns Hopkins University
Ryan Cotterell, University of Cambridge

Program Committee:

Noor Abo Mokh, Indiana University
Jane Chandlee, Haverford College
Çağrı Çöltekin, University of Tübingen
Daniel Dakota, Indiana University
Ewan Dunbar, Université Paris Diderot
Jason Eisner, Johns Hopkins University
Micha Elsner, The Ohio State University
Kyle Gorman, CUNY / Google AI
Nizar Habash, NYU Abu Dhabi
Mans Hulden, University of Colorado
Adam Jardine, Rutgers University
Gaja Jarosz, University of Massachusetts Amherst
Christo Kirov, Google AI
Greg Kobele, Universität Leipzig
Greg Kondrak, University of Alberta
Sandra Kübler, Indiana University
Andrew Lamont, University of Massachusetts Amherst
Karen Livescu, TTI Chicago
Arya D. McCarthy, Johns Hopkins University
Kevin McMullin, University of Ottawa
Kemal Oflazer, CMU Qatar
Jeff Parker, Brigham Young University
Gerald Penn, University of Toronto
Jelena Prokic, Ludwig-Maximilians-University Munich
Mohamad Salameh, Diffbot
Miikka Silfverberg, University of Colorado
Kairit Sirts, University of Tartu
Kenneth Steimel, Indiana University
Francis Tyers, Indiana University
Anssi Yli-Jyrä, University of Helsinki
Kristine Yu, University of Massachusetts Amherst

Shared Task Organizers:

Arya D. McCarthy, Johns Hopkins University
Ekaterina Vylomova, University of Melbourne
Shijie Wu, Johns Hopkins University
Chaitanya Malaviya, Allen Institute for Artificial Intelligence
Lawrence Wolf-Sonkin, Google AI
Garrett Nicolai, Johns Hopkins University
Christo Kirov, Google AI
Miikka Silfverberg, University of Helsinki
Sebastian J. Mielke, Johns Hopkins University
Svetlana Toldova, National Research University Higher School of Economics
Olga Lyashevskaya, National Research University Higher School of Economics
Karina Mishchenkova
Elena Klyachko
Jeffrey Heinz, Stony Brook University
Ryan Cotterell, University of Cambridge
Mans Hulden , University of Colorado

Invited Speakers:

Sharon Goldwater, University of Edinburgh
Janet Pierrehumbert, Oxford University
Adina Williams, Facebook AI
Géraldine Walther, University of Zurich
Benoît Sagot, Inria

Table of Contents

<i>AX Semantics' Submission to the SIGMORPHON 2019 Shared Task</i> Andreas Madsack and Robert Weißgraeber	1
<i>Cognate Projection for Low-Resource Inflection Generation</i> Bradley Hauer, Amir Ahmad Habibi, Yixing Luan, Rashed Rubby Riyadh and Grzegorz Kondrak	6
<i>Cross-Lingual Lemmatization and Morphology Tagging with Two-Stage Multilingual BERT Fine-Tuning</i> Dan Kondratyuk	12
<i>CBNU System for SIGMORPHON 2019 Shared Task 2: a Pipeline Model</i> Uygun Shadikhodjaev and Jae Sung Lee	19
<i>Morpheus: A Neural Network for Jointly Learning Contextual Lemmatization and Morphological Tagging</i> Eray Yildiz and A. Cüneyd Tantuğ	25
<i>Multi-Team: A Multi-attention, Multi-decoder Approach to Morphological Analysis.</i> Ahmet Üstün, Rob van der Goot, Gosse Bouma and Gertjan van Noord	35
<i>IT-IST at the SIGMORPHON 2019 Shared Task: Sparse Two-headed Models for Inflection</i> Ben Peters and André F. T. Martins	50
<i>CMU-01 at the SIGMORPHON 2019 Shared Task on Crosslinguality and Context in Morphology</i> Aditi Chaudhary, Elizabeth Salesky, Gayatri Bhat, David R. Mortensen, Jaime Carbonell and Yulia Tsvetkov	57
<i>Cross-lingual morphological inflection with explicit alignment</i> Çağrı Çöltekin	71
<i>THOMAS: The Hegemonic OSU Morphological Analyzer using Seq2seq</i> Byung-Doh Oh, Pranav Maneriker and Nanjiang Jiang	80
<i>Sigmorphon 2019 Task 2 system description paper: Morphological analysis in context for many languages, with supervision from only a few</i> Brad Aiken, Jared Kelly, Alexis Palmer, Suleyman Olcay Polat, Taraka Rama and Rodney Nielsen	87
<i>UDPipe at SIGMORPHON 2019: Contextualized Embeddings, Regularization with Morphological Categories, Corpora Merging</i> Milan Straka, Jana Straková and Jan Hajic	95
<i>CUNI-Malta system at SIGMORPHON 2019 Shared Task on Morphological Analysis and Lemmatization in context: Operation-based word formation</i> Ronald Cardenas, Claudia Borg and Daniel Zeman	104
<i>A Little Linguistics Goes a Long Way: Unsupervised Segmentation with Limited Language Specific Guidance</i> Alexander Erdmann, Salam Khalifa, Mai Oudah, Nizar Habash and Houda Bouamor	113
<i>Equiprobable mappings in weighted constraint grammars</i> Arto Anttila, Scott Borgeson and Giorgio Magri	125

<i>Unbounded Stress in Subregular Phonology</i> Yiding Hao and Samuel Andersson	135
<i>Data mining Mandarin tone contour shapes</i> Shuo Zhang	144
<i>Convolutional neural networks for low-resource morpheme segmentation: baseline or state-of-the-art?</i> Alexey Sorokin	154
<i>What do phone embeddings learn about Phonology?</i> Sudheer Kolachina and Lilla Magyar	160
<i>Inverting and Modeling Morphological Inflection</i> Yohei Oseki, Yasutada Sudo, Hiromu Sakai and Alec Marantz	170
<i>Augmenting a German Morphological Database by Data-Intense Methods</i> Petra Steiner	178
<i>Unsupervised Morphological Segmentation for Low-Resource Polysynthetic Languages</i> Ramy Eskander, Judith Klavans and Smaranda Muresan	188
<i>Weakly deterministic transformations are subregular</i> Andrew Lamont, Charlie O’Hara and Caitlin Smith	195
<i>Encoder-decoder models for latent phonological representations of words</i> Cassandra L. Jacobs and Fred Mailhot	205
<i>Action-Sensitive Phonological Dependencies</i> Yiding Hao and Dustin Bowers	217
<i>The SIGMORPHON 2019 Shared Task: Morphological Analysis in Context and Cross-Lingual Transfer for Inflection</i> Arya D. McCarthy, Ekaterina Vylomova, Shijie Wu, Chaitanya Malaviya, Lawrence Wolf-Sonkin, Garrett Nicolai, Miikka Silfverberg, Sebastian J. Mielke, Jeffrey Heinz, Ryan Cotterell and Mans Hulden	228

Conference Program

August 2, 2019

8:30–10:30 Morning Session

8:30–9:30 *Invited Speaker:* Adina Williams

9:30–10:30 *Invited Speaker:* Janet Pierrehumbert

10:30–11:00 Break

11:00–12:30 Shared Task

11:00–11:30 *Shared Task Wrapup*
TBA

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

AX Semantics' Submission to the SIGMORPHON 2019 Shared Task
Andreas Madsack and Robert Weißgraeber

Cognate Projection for Low-Resource Inflection Generation
Bradley Hauer, Amir Ahmad Habibi, Yixing Luan, Rashed Rubby Riyadh and Grzegorz Kondrak

Cross-Lingual Lemmatization and Morphology Tagging with Two-Stage Multilingual BERT Fine-Tuning
Dan Kondratyuk

CBNU System for SIGMORPHON 2019 Shared Task 2: a Pipeline Model
Uygun Shadikhodjaev and Jae Sung Lee

Morpheus: A Neural Network for Jointly Learning Contextual Lemmatization and Morphological Tagging
Eray Yildiz and A. Cüneyd Tantuğ

August 2, 2019 (continued)

Pushing the Limits of Low-Resource Morphological Inflection

Antonios Anastasopoulos and Graham Neubig

Multi-Team: A Multi-attention, Multi-decoder Approach to Morphological Analysis.

Ahmet Üstün, Rob van der Goot, Gosse Bouma and Gertjan van Noord

IT-IST at the SIGMORPHON 2019 Shared Task: Sparse Two-headed Models for Inflection

Ben Peters and André F. T. Martins

CMU-01 at the SIGMORPHON 2019 Shared Task on Crosslinguality and Context in Morphology

Aditi Chaudhary, Elizabeth Salesky, Gayatri Bhat, David R. Mortensen, Jaime Carbonell and Yulia Tsvetkov

Cross-lingual morphological inflection with explicit alignment

Çağrı Çöltekin

THOMAS: The Hegemonic OSU Morphological Analyzer using Seq2seq

Byung-Doh Oh, Pranav Maneriker and Nanjiang Jiang

Sigmorphon 2019 Task 2 system description paper: Morphological analysis in context for many languages, with supervision from only a few

Brad Aiken, Jared Kelly, Alexis Palmer, Suleyman Olcay Polat, Taraka Rama and Rodney Nielsen

UDPipe at SIGMORPHON 2019: Contextualized Embeddings, Regularization with Morphological Categories, Corpora Merging

Milan Straka, Jana Straková and Jan Hajic

CUNI-Malta system at SIGMORPHON 2019 Shared Task on Morphological Analysis and Lemmatization in context: Operation-based word formation

Ronald Cardenas, Claudia Borg and Daniel Zeman

August 2, 2019 (continued)

12:30–14:00 Lunch

14:00–14:30 Best Paper Session

14:00–14:15 *A Little Linguistics Goes a Long Way: Unsupervised Segmentation with Limited Language Specific Guidance*

Alexander Erdmann, Salam Khalifa, Mai Oudah, Nizar Habash and Houda Bouamor

14:15–14:30 *Equiprobable mappings in weighted constraint grammars*

Arto Anttila, Scott Borgeson and Giorgio Magri

14:30–15:30 Poster Session

Unbounded Stress in Subregular Phonology

Yiding Hao and Samuel Andersson

Data mining Mandarin tone contour shapes

Shuo Zhang

Convolutional neural networks for low-resource morpheme segmentation: baseline or state-of-the-art?

Alexey Sorokin

What do phone embeddings learn about Phonology?

Sudheer Kolachina and Lilla Magyar

Inverting and Modeling Morphological Inflection

Yohei Oseki, Yasutada Sudo, Hiromu Sakai and Alec Marantz

Augmenting a German Morphological Database by Data-Intense Methods

Petra Steiner

Unsupervised Morphological Segmentation for Low-Resource Polysynthetic Languages

Ramy Eskander, Judith Klavans and Smaranda Muresan

August 2, 2019 (continued)

Weakly deterministic transformations are subregular

Andrew Lamont, Charlie O'Hara and Caitlin Smith

Encoder-decoder models for latent phonological representations of words

Cassandra L. Jacobs and Fred Mailhot

Action-Sensitive Phonological Dependencies

Yiding Hao and Dustin Bowers

15:30–16:00 Break

16:00–18:00 Afternoon Session

16:00–17:00 *Invited Speakers* : Géraldine Walther and Benoît Sagot

17:00–18:00 *Invited Speaker* : Sharon Goldwater