

SIGMORPHON 2018

**The 15th SIGMORPHON Workshop
on Computational Research in Phonetics,
Phonology, and Morphology**

Proceedings of the Workshop

October 31, 2018
Brussels, Belgium

©2018 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 978-1-948087-76-6

Preface

Welcome to the 15th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, to be held on October 31, 2018 in Brussels, Belgium. 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, however, a new focus on low resource approaches to morphology is emerging as a consequence of the SIGMORPHON 2016 and CoNLL shared tasks on morphological reinflection across a wide range of languages.

We received 36 submissions, and after a competitive reviewing process, we accepted 19. Due to time limitations, 6 papers were chosen for oral presentations, the remaining papers are presented as posters. The workshop also includes a joint poster session with CoNLL, on the CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection.

We are grateful to the program committee for their careful and thoughtful reviews of the papers submitted this year. We are looking forward to a workshop covering a wide range of topics, and we hope for lively discussions.

Sandra Kübler
Garrett Nicolai

Organizers:

Sandra Kübler, University of Indiana
Garrett Nicolai, Johns Hopkins University

Program Committee:

Noor Abo Mokh, Indiana University
Adam Albright, MIT
Jane Chandlee, Haverford College
Çağrı Çöltekin, University of Tübingen
Ryan Cotterell, Johns Hopkins University
Daniel Dakota, Indiana University
Thierry Declerck, DFKI
Ewan Dunbar, Laboratoire de Sciences Cognitives et Psycholinguistique, Paris
Jason Eisner, Johns Hopkins University
Micha Elsner, The Ohio State University
Sharon Goldwater, University of Edinburgh
Kyle Gorman, Google
Nizar Habash, NYU Abu Dhabi
Mike Hammond, University of Arizona
Mans Hulden, University of Colorado
Adam Jardine, Rutgers University
Gaja Jarosz, University of Massachusetts Amherst
Christo Kirov, Johns Hopkins University
Greg Kobele, University of Chicago
Grzegorz Kondrak, University of Alberta
Kimmo Koskenniemi, University of Helsinki
Andrew Lamont, University of Massachusetts Amherst
Karen Livescu, TTI Chicago
Arya 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, CMU-Qatar
Miikka Silfverberg, University of Colorado
Andrea Sims, The Ohio State University
Kairit Sirts, Macquarie University
Richard Sproat, Google
Kenneth Steimel, Indiana University
Reut Tsarfaty, The Open University of Israel
Francis Tyers, Higher School of Economics
Richard Wicentowski, Swarthmore University
Anssi Yli-Jyrä, University of Helsinki
Kristine Yu, University of Massachusetts Amherst

Table of Contents

<i>Efficient Computation of Implicational Universals in Constraint-Based Phonology Through the Hyperplane Separation Theorem</i>	
Giorgio Magri	1
<i>Lexical Networks in !Xung</i>	
Syed-Amad Hussain, Micha Elsner and Amanda Miller	11
<i>Acoustic Word Disambiguation with Phonological Features in Danish ASR</i>	
Andreas Søeborg Kirkedal	21
<i>Adaptor Grammars for the Linguist: Word Segmentation Experiments for Very Low-Resource Languages</i>	
Pierre Godard, Laurent Besacier, François Yvon, Martine Adda-Decker, Gilles Adda, H�el�ene Maynard and Annie Riolland	32
<i>String Transduction with Target Language Models and Insertion Handling</i>	
Garrett Nicolai, Saeed Najafi and Grzegorz Kondrak	43
<i>Complementary Strategies for Low Resourced Morphological Modeling</i>	
Alexander Erdmann and Nizar Habash	54
<i>Modeling Reduplication with 2-way Finite-State Transducers</i>	
Hossep Dolatian and Jeffrey Heinz	66
<i>Automatically Tailoring Unsupervised Morphological Segmentation to the Language</i>	
Ramy Eskander, Owen Rambow and Smaranda Muresan	78
<i>A Comparison of Entity Matching Methods between English and Japanese Katakana</i>	
Michiharu Yamashita, Hideki Awashima and Hidekazu Oiwa	84
<i>Seq2Seq Models with Dropout can Learn Generalizable Reduplication</i>	
Brandon Prickett, Aaron Traylor and Joe Pater	93
<i>A Characterwise Windowed Approach to Hebrew Morphological Segmentation</i>	
Amir Zeldes	101
<i>Phonetic Vector Representations for Sound Sequence Alignment</i>	
Pavel Sofroniev and �ađrı �oltekin	111
<i>Sounds Wilde. Phonetically Extended Embeddings for Author-Stylized Poetry Generation</i>	
Aleksy Tikhonov and Ivan Yamshchikov	117
<i>On Hapax Legomena and Morphological Productivity</i>	
Janet Pierrehumbert and Ramon Granell	125
<i>A Morphological Analyzer for Shipibo-Konibo</i>	
Ronald Cardenas and Daniel Zeman	131
<i>An Arabic Morphological Analyzer and Generator with Copious Features</i>	
Dima Taji, Salam Khalifa, Ossama Obeid, Fadhl Eryani and Nizar Habash	140
<i>Sanskrit n-Retroflexion is Input-Output Tier-Based Strictly Local</i>	
Thomas Graf and Connor Mayer	151

<i>Phonological Features for Morphological Inflection</i>	
Adam Wiemerslage, Miikka Silfverberg and Mans Hulden	161
<i>Extracting Morphophonology from Small Corpora</i>	
Marina Ermolaeva	167

Conference Program

Wednesday, October 31

08:50–09:00 Session Opening: Opening Session

09:00–10:30 Session 1: Phonology

09:00–09:30 *Efficient Computation of Implicational Universals in Constraint-Based Phonology Through the Hyperplane Separation Theorem*

Giorgio Magri

09:30–10:00 *Lexical Networks in !Xung*

Syed-Amad Hussain, Micha Elsner and Amanda Miller

10:00–10:30 *Acoustic Word Disambiguation with Phonological Features in Danish ASR*

Andreas Søeborg Kirkedal

10:30–11:00 *Tea Break 1*

11:00–12:30 Session ST: CoNLL – SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection

12:30–14:00 *Lunch*

Wednesday, October 31

14:00–15:30 Session 2: Low-Resource Morphology

14:00–14:30 *Adaptor Grammars for the Linguist: Word Segmentation Experiments for Very Low-Resource Languages*

Pierre Godard, Laurent Besacier, François Yvon, Martine Adda-Decker, Gilles Adda, H el ene Maynard and Annie Rialland

14:30–15:00 *String Transduction with Target Language Models and Insertion Handling*

Garrett Nicolai, Saeed Najafi and Grzegorz Kondrak

15:00–15:30 *Complementary Strategies for Low Resourced Morphological Modeling*

Alexander Erdmann and Nizar Habash

15:30–16:00 Tea Break 2

16:00–17:30 Session P: Poster Session

Modeling Reduplication with 2-way Finite-State Transducers

Hossep Dolatian and Jeffrey Heinz

Automatically Tailoring Unsupervised Morphological Segmentation to the Language

Ramy Eskander, Owen Rambow and Smaranda Muresan

A Comparison of Entity Matching Methods between English and Japanese Katakana

Michiharu Yamashita, Hideki Awashima and Hidekazu Oiwa

Seq2Seq Models with Dropout can Learn Generalizable Reduplication

Brandon Prickett, Aaron Traylor and Joe Pater

A Characterwise Windowed Approach to Hebrew Morphological Segmentation

Amir Zeldes

Phonetic Vector Representations for Sound Sequence Alignment

Pavel Sofroniev and  ađrı   ltekin

Wednesday, October 31 (continued)

Sounds Wilde. Phonetically Extended Embeddings for Author-Stylized Poetry Generation

Aleksey Tikhonov and Ivan Yamshchikov

On Hapax Legomena and Morphological Productivity

Janet Pierrehumbert and Ramon Granel

A Morphological Analyzer for Shipibo-Konibo

Ronald Cardenas and Daniel Zeman

An Arabic Morphological Analyzer and Generator with Copious Features

Dima Taji, Salam Khalifa, Ossama Obeid, Fadhil Eryani and Nizar Habash

Sanskrit n-Retroflexion is Input-Output Tier-Based Strictly Local

Thomas Graf and Connor Mayer

Phonological Features for Morphological Inflection

Adam Wiemerslage, Miikka Silfverberg and Mans Hulden

Extracting Morphophonology from Small Corpora

Marina Ermolaeva

17:30–17:40 Session Closing: Closing Session

