

NAACL-HLT 2012

SIGMORPHON2012

**Twelfth Meeting of the ACL Special Interest Group on
Computational Morphology and Phonology**

Proceedings of the Workshop

June 7, 2012
Montréal, Canada

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2012 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

ISBN13: 978-1-937284-20-6.
ISBN10: 1-937284-20-4.

Introduction

We are delighted to present the Proceedings of the Twelfth Meeting of the ACL Special Interest Group on Computational Morphology and Phonology (SIGMORPHON), to be held on June 7, 2012 at Le Centre Sheraton Montréal, Montréal, Canada. The purpose of SIGMORPHON is to foster computational research on the phonological, morphological, and phonetic properties of human language. All three of these sub-areas deal largely with the local structure of words and so share many technical methods. Furthermore, computational work that models empirical data must often draw on at least two of these areas, with explicit consideration of the morphology-phonology or phonology-phonetics interface. In recent years the remit has expanded to also include research on orthographic issues such as transliteration and variable spelling.

We received a good number of submissions, on the full range of sub-areas, and accepted around sixty percent. This has enabled us to provide what we hope you will agree is a high quality program.

We are grateful to the program committee for their careful and thoughtful reviews and discussions of the papers submitted this year. We hope that you enjoy the workshop and these proceedings.

Lynne Cahill
Adam Albright

Organizers:

Lynne Cahill, University of Brighton (UK)
Adam Albright, MIT (USA)

Program Committee:

Jason Eisner, Johns Hopkins University (USA)
Mark Ellison, University of Western Australia (Australia)
Roger Evans, University of Brighton (UK)
Sharon Goldwater, University of Edinburgh (UK)
Jeffrey Heinz, University of Delaware (USA)
Jon Herring, British Library (UK)
William Idsardi, University of Maryland (USA)
Gaja Jarosz, Yale University (USA)
Greg Kobele, University of Chicago (USA)
Grzegorz Kondrak, University of Alberta (Canada)
Kimmo Koskenniemi, University of Helsinki (Finland)
Mikko Kurimo, TKK (Finland)
Karen Livescu, Toyota Technological Institute at University of Chicago (USA)
Giorgio Magri, IJN, ENS (France)
Katya Pertsova, University of North Carolina (USA)
Jason Riggle, University of Chicago (USA)
Shuly Wintner, University of Haifa (Israel)

Table of Contents

<i>A Morphological Analyzer for Egyptian Arabic</i> Nizar Habash, Ramy Eskander and Abdelati Hawwari	1
<i>Hindi Derivational Morphological Analyzer</i> Nikhil Kanuparthi, Abhilash Inumella and Dipti Misra Sharma	10
<i>Phrase-Based Approach for Adaptive Tokenization</i> Jianqiang Ma and Dale Gerdemann	17
<i>A Regularized Compression Method to Unsupervised Word Segmentation</i> Ruey-Cheng Chen, Chiung-Min Tsai and Jieh Hsiang	26
<i>A rule-based approach to unknown word recognition in Arabic</i> Lynne Cahill	35
<i>Bounded copying is subsequential: Implications for metathesis and reduplication</i> Jane Chandlee and Jeffrey Heinz	42
<i>An approximation approach to the problem of the acquisition of phonotactics in Optimality Theory</i> Giorgio Magri	52
<i>Learning probabilities over underlying representations</i> Joe Pater, Robert Staubs, Karen Jesney and Brian Smith	62
<i>Linguistic categorization and complexity</i> Katya Pertsova	72

Conference Program

Thursday June 7, 2012

- 9:00 Welcome
- 9:15–9:45 *A Morphological Analyzer for Egyptian Arabic*
Nizar Habash, Ramy Eskander and Abdelati Hawwari
- 9:45–10:15 *Hindi Derivational Morphological Analyzer*
Nikhil Kanuparthi, Abhilash Inumella and Dipti Misra Sharma
- 10:15–10:30 Discussion
- 10:30–11:00 Coffee break
- 11:00–11:30 *Phrase-Based Approach for Adaptive Tokenization*
Jianqiang Ma and Dale Gerdemann
- 11:30–12:00 *A Regularized Compression Method to Unsupervised Word Segmentation*
Ruey-Cheng Chen, Chiung-Min Tsai and Jieh Hsiang
- 12:00–12:30 *A rule-based approach to unknown word recognition in Arabic*
Lynne Cahill
- 12:30–12:45 Discussion
- 12:45–2:00 Lunch
- 2:00–2:30 *Bounded copying is subsequential: Implications for metathesis and reduplication*
Jane Chandlee and Jeffrey Heinz
- 2:30–3:00 *An approximation approach to the problem of the acquisition of phonotactics in Optimality Theory*
Giorgio Magri
- 3:00–3:30 *Learning probabilities over underlying representations*
Joe Pater, Robert Staubs, Karen Jesney and Brian Smith

Thursday June 7, 2012 (continued)

3:30-4:00 Coffee break

4:00-4:30 *Linguistic categorization and complexity*
Katya Pertsova

4:30-4:45 Discussion

4:45-5:15 Business meeting (all are welcome to attend)