

ACL 2014

**2014 Joint Meeting of SIGMORPHON and SIGFSM
MORPHFSM 2014**

Proceedings of the Workshop

June 27, 2014
Baltimore, Maryland, USA

©2014 The Association for Computational Linguistics

ISBN 978-1-941643-13-6

Preface

These proceedings contain the contributions presented at the MORPHFSM workshop held on June 27, 2014 in conjunction with ACL 2014 in Baltimore, Maryland, USA. The workshop was a joint meeting of two special interest groups of the ACL:

- SIGMORPHON — the special interest group in computational morphology, phonology and phonetics, and
- SIGFSM — the special interest group on finite-state methods.

It was the thirteenth meeting of SIGMORPHON and an off-year event for SIGFSM. The full-day workshop consisted of an invited presentation by JASON EISNER, contributed presentations, and a special panel session on open problems.

The workshop covered a wide range of topics from theoretical to applied morphology and finite-state technology in natural language processing. This volume contains the 7 regular and 1 panel paper that were presented at the workshop. In total, 12 papers (10 regular and 2 panel papers) were submitted to a doubly blind refereeing process, in which each paper was reviewed by 3 program committee members. The overall acceptance rate was 67%. The program committee was composed of internationally leading researchers and practitioners selected from academia, research labs, and companies.

The organizing committee would like to thank the program committee for their hard work and valuable feedback, the invited speaker JASON EISNER for his innovative and inspiring keynote, our panelists for their interesting discussion and expertise, the local organizers for their tireless efforts, the ACL administration for their support, and last but not least the authors for their contributions.

ÖZLEM ÇETİNOĞLU
JEFFREY HEINZ
ANDREAS MALETTI
JASON RIGGLE

Organizers:

Özlem Çetinoğlu, IMS, University of Stuttgart (Germany)
Jeffrey Heinz, University of Delaware, Delaware (USA)
Andreas Maletti, IMS, University of Stuttgart (Germany)
Jason Riggle, University of Chicago, Illinois (USA)

Program Committee:

Adam Albright, MIT (USA)
Lynne Cahill, University of Sussex (UK)
Francisco Casacuberta, Instituto Tecnológico de Informática (Spain)
Jason Eisner, Johns Hopkins University (USA)
Roger Evans, University of Brighton (UK)
Sharon Goldwater, University of Edinburgh (UK)
Colin de la Higuera, University of Nantes (France)
Mans Hulden, University of Helsinki (Finland)
André Kempe, Nuance Communications (Germany)
Grzegorz Kondrak, University of Alberta (Canada)
Kimmo Koskenniemi, University of Helsinki (Finland)
Marco Kuhlmann, Linköping University (Sweden)
Karen Livescu, Toyota Technological Institute at U. Chicago (USA)
Giorgio Magri, IJN, ENS (France)
Mark-Jan Nederhof, University of St. Andrews (UK)
Kemal Oflazer, Carnegie Mellon University (Qatar)
Katya Pertsova, University of North Carolina (USA)
Max Silberstein, Université de Franche-Comté (France)
Richard Sproat, Google (USA)
Shuly Wintner, University of Haifa (Israel)
Anssi Yli-Jyrä, University of Helsinki (Finland)
Kristine Yu, University of Massachusetts, Amherst (USA)

Invited Speaker:

Jason Eisner, Johns Hopkins University (USA)

Panelists:

Jason Eisner, Johns Hopkins University (USA)
Mans Hulden, University of Helsinki (Finland)
Grzegorz Kondrak, University of Alberta (Canada)
Jason Riggle, University of Chicago, Illinois (USA)

Table of Contents

<i>Revisiting Word Neighborhoods for Speech Recognition</i> Preethi Jyothi and Karen Livescu	1
<i>The Error-driven Ranking Model of the Acquisition of Phonotactics: How to Keep the Faithfulness Constraints at Bay</i> Giorgio Magri	10
<i>Comparing Models of Phonotactics for Word Segmentation</i> Natalie Schrimpf and Gaja Jarosz	19
<i>Generalizing Inflection Tables into Paradigms with Finite State Operations</i> Mans Hulden	29
<i>Automatic Conversion of Dialectal Tamil Text to Standard Written Tamil Text using FSTs</i> Marimuthu K and Sobha Lalitha Devi	37
<i>Rule Based Morphological Analyzer of Kazakh Language</i> Gulshat Kessikbayeva and Ilyas Cicekli	46
<i>Rules, Analogy, and Social Factors Codetermine Past-tense Formation Patterns in English</i> Peter Racz, Clayton Beckner, Jennifer B. Hay and Janet B. Pierrehumbert	55
<i>10 Open Questions in Computational Morphology</i> Grzegorz Kondrak	64

Conference Program

Keynote session

9:00–9:15 Welcome by Jason Riggle

9:15–10:15 Keynote by Jason Eisner

10:30–11:00 Morning break

Morning session

11:00–11:30 *Revisiting Word Neighborhoods for Speech Recognition*
Preethi Jyothi and Karen Livescu

11:30–12:00 *The Error-driven Ranking Model of the Acquisition of Phonotactics: How to Keep the Faithfulness Constraints at Bay*
Giorgio Magri

12:00–12:30 *Comparing Models of Phonotactics for Word Segmentation*
Natalie Schrimpf and Gaja Jarosz

12:30–14:00 Lunch break

Early afternoon session

14:00–14:30 *Generalizing Inflection Tables into Paradigms with Finite State Operations*
Mans Hulden

14:30–15:00 *Automatic Conversion of Dialectal Tamil Text to Standard Written Tamil Text using FSTs*
Marimuthu K and Sobha Lalitha Devi

15:00–15:30 *Rule Based Morphological Analyzer of Kazakh Language*
Gulshat Kessikbayeva and Ilyas Cicekli

15:30–16:00 Afternoon break

No Day Set (continued)

Afternoon session

16:00–16:30 *Rules, Analogy, and Social Factors Codetermine Past-tense Formation Patterns in English*
Peter Racz, Clayton Beckner, Jennifer B. Hay and Janet B. Pierrehumbert

Panel discussion

16:30–16:40 *10 Open Questions in Computational Morphology*
Grzegorz Kondrak

16:40–17:30 Panel discussion by Jason Eisner, Mans Hulden, Grzegorz Kondrak, Jason Riggle