

ACL 2010

**SIGMORPHON 2010**

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

**Proceedings of the Workshop**

15 July 2010  
Uppsala University  
Uppsala, Sweden

Production and Manufacturing by  
*Taberg Media Group AB*  
*Box 94, 562 02 Taberg*  
*Sweden*

©2010 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](mailto:acl@aclweb.org)

ISBN 978-1-932432-76-3 / 1-932432-76-0

## Introduction

We are delighted to present the Proceedings of the Eleventh Meeting of the ACL Special Interest Group on Computational Morphology and Phonology (SIGMORPHON), to be held on July 15, 2010 at Uppsala University, Uppsala, Sweden.

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.

We received a large number of submissions, on the full range of sub-areas, and accepted around forty 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 are especially grateful to those reviewers who stepped in at the last minute when the number of submissions became clear.

We hope that you enjoy the workshop and these proceedings.

Jeffrey Heinz  
Lynne Cahill  
Richard Wicentowski



**Organizers:**

Jeffrey Heinz, University of Delaware  
Lynne Cahill, University of Brighton  
Richard Wicentowski, Swarthmore College

**Program Committee Members:**

Adam Albright, Massachusetts Institute of Technology  
Jason Eisner, Johns Hopkins University  
Mark Ellison, University of Western Australia  
Sharon Goldwater, University of Edinburgh  
Grzegorz Kondrak, University of Alberta  
Kimmo Koskenniemi, University of Helsinki  
Karen Livescu, Toyota Technological Institute at Chicago/University of Chicago  
Mike Maxwell, University of Maryland  
Jason Riggle, University of Chicago  
Shuly Wintner, University of Haifa

**External Reviewers:**

Timothy Bunnell, A.I. DuPont Hospital/University of Delaware  
Roger Evans, University of Brighton  
Dafydd Gibbon, University of Bielefeld  
Jon Herring, British Library  
William Idsardi, University of Maryland  
Greg Kobele, University of Chicago  
Katya Pertsova, University of North Carolina



## Table of Contents

<i>Instance-Based Acquisition of Vowel Harmony</i> Fred Mailhot .....	1
<i>Verifying Vowel Harmony Typologies</i> Sara Finley .....	9
<i>Complexity of the Acquisition of Phonotactics in Optimality Theory</i> Giorgio Magri .....	19
<i>Maximum Likelihood Estimation of Feature-Based Distributions</i> Jeffrey Heinz and Cesar Koirala .....	28
<i>A Method for Compiling Two-Level Rules with Multiple Contexts</i> Kimmo Koskeniemi and Miikka Silfverberg .....	38
<i>Exploring Dialect Phonetic Variation Using PARAFAC</i> Jelena Prokic and Tim Van de Cruys .....	46
<i>Quantitative Evaluation of Competing Syllable Parses</i> Jason A. Shaw and Adamantios I. Gafos .....	54
<i>Toward a Totally Unsupervised, Language-Independent Method for the Syllabification of Written Texts</i> Thomas Mayer .....	63
<i>Comparing Canonicalizations of Historical German Text</i> Bryan Jurish .....	72
<i>Semi-Supervised Learning of Concatenative Morphology</i> Oskar Kohonen, Sami Virpioja and Krista Lagus .....	78
<i>Morpho Challenge 2005-2010: Evaluations and Results</i> Mikko Kurimo, Sami Virpioja, Ville Turunen and Krista Lagus .....	87





# Conference Program

## Thursday, 15 July 2010

- 9:00–9:30     *Instance-Based Acquisition of Vowel Harmony*  
Fred Mailhot
- 9:30–10:00    *Verifying Vowel Harmony Typologies*  
Sara Finley
- 10:00–10:30   *Complexity of the Acquisition of Phonotactics in Optimality Theory*  
Giorgio Magri
- 10:30–11:00   Morning Break
- 11:00–11:30   *Maximum Likelihood Estimation of Feature-Based Distributions*  
Jeffrey Heinz and Cesar Koirala
- 11:30–12:00   *A Method for Compiling Two-Level Rules with Multiple Contexts*  
Kimmo Koskenniemi and Miikka Silfverberg
- 12:00–12:30   *Exploring Dialect Phonetic Variation Using PARAFAC*  
Jelena Prokic and Tim Van de Cruys
- 12:30–14:00   Lunch
- 14:00–14:30   *Quantitative Evaluation of Competing Syllable Parses*  
Jason A. Shaw and Adamantios I. Gafos
- 14:30–15:00   *Toward a Totally Unsupervised, Language-Independent Method for the Syllabification of Written Texts*  
Thomas Mayer
- Comparing Canonicalizations of Historical German Text*  
Bryan Jurish
- 15:30–16:00   Afternoon Break
- 16:00–16:30   *Semi-Supervised Learning of Concatenative Morphology*  
Oskar Kohonen, Sami Virpioja and Krista Lagus
- 16:30–17:00   *Morpho Challenge 2005-2010: Evaluations and Results*  
Mikko Kurimo, Sami Virpioja, Ville Turunen and Krista Lagus

