

ACL 2007



ACL 2007

Computing and Historical Phonology

Proceedings of the Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology

**June 28, 2007
Prague, Czech Republic**



Production and Manufacturing by
Omnipress
2600 Anderson Street
Madison, WI 53704
USA

©2007 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

Preface

Welcome to the ACL Workshop on Computing and Historical Phonology, the 9th Meeting of ACL Special Interest Group for Computational Morphology and Phonology, a meeting held in conjunction with the 45th Meeting of the ACL in Prague. An introductory article explains our motivation for holding the workshop, which attracted 16 submissions, all but one of which is included in this volume of proceedings. We are gratified not only by the level of interest, but also by the quality of submissions we received.

We hoped to attract interest not only in the computational linguistics community *sensu stricto* but also in the broader linguistics community, and in the group of geneticists who have begun to apply phylogenetic analysis to linguistic data. As the reader may verify in these proceedings, we were not disappointed in this hope. Perhaps it is worth adding that, while we are in principle interested in further meetings of this sort, there are at the time of this writing no concrete plans for follow-ups.

Our thanks are largely given in the acknowledgments section of the introductory article, but let's add thanks here to Simone Teufel, the workshop chair of the conference, who ushered us through the various steps from the proposal through the production of this publication, and also to the committee she chaired.

John Nerbonne, T. Mark Ellison and Grzegorz Kondrak (Chairs)

Organizers

Chairs:

John Nerbonne, University of Groningen
Grzegorz Kondrak, University of Alberta
T. Mark Ellison, University of Western Australia

Program Committee:

Chris Brew, Ohio State University
Pierre Darlu, Paris
Michael Dunn, Max Planck, Nijmegen
Sheila Embleton, York University, Toronto
Hans Goebel, Salzburg
Russell Gray, Auckland
Sheldon Harrison, Western Australia
Wilbert Heeringa, Groningen
Brian Joseph, Ohio State University
Brett Kessler, Washington University of St. Louis
Simon Kirby, Edinburgh
Bill Kretzschmar, Georgia
Franz Manni, Paris
Hermann Moisl, Newcastle
David Nash, Australian National University, Canberra
Michael Oakes, Sunderland
Jon Patrick, Sydney
Gerald Penn, Toronto
Janet Pierrehumbert, Northwestern
Thomas Pilz, Duisburg
Joe Salmons, Wisconsin
Tandy Warnow, University of Texas

Keynote Speaker:

Brett Kessler, Washington University of St. Louis

Other Reviewers:

Hans J. Holm, Hannover
Sebastian Kürschner, University of Groningen

Table of Contents

<i>Computing and Historical Phonology</i>	
John Nerbonne, T. Mark Ellison and Grzegorz Kondrak	1
<i>Word Similarity Metrics and Multilateral Comparison</i>	
Brett Kessler	6
<i>Bayesian Identification of Cognates and Correspondences</i>	
T. Mark Ellison	15
<i>Testing Cladistics on Dialect Networks and Phyla (Gallo-Romance Vowels, Southern Italo-Romance Di- asystems and Mayan Languages)</i>	
Antonella Gaillard-Corvaglia, Jean-Léo Léonard and Pierre Darlu	23
<i>The Relative Divergence of Dutch Dialect Pronunciations from their Common Source: An Exploratory Study</i>	
Wilbert Heeringa and Brian Joseph	31
<i>Can Corpus Based Measures be Used for Comparative Study of Languages?</i>	
Anil Kumar Singh and Harshit Surana	40
<i>Inducing Sound Segment Differences Using Pair Hidden Markov Models</i>	
Martijn Wieling, Therese Leinonen and John Nerbonne	48
<i>Phonological Reconstruction of a Dead Language Using the Gradual Learning Algorithm</i>	
Eric Smith	57
<i>Evolution, Optimization, and Language Change: The Case of Bengali Verb Inflections</i>	
Monojit Choudhury, Vaibhav Jalan, Sudeshna Sarkar and Anupam Basu	65
<i>On the Geolinguistic Change in Northern France between 1300 and 1900: A Dialectometrical Inquiry</i>	
Hans Goebel	75
<i>Visualizing the Evaluation of Distance Measures</i>	
Thomas Pilz, Axel Philipsenburg and Wolfram Luther	84
<i>Data Nonlinearity in Exploratory Multivariate Analysis of Language Corpora</i>	
Hermann Moisl	93
<i>Emergence of Community Structures in Vowel Inventories: An Analysis Based on Complex Networks</i>	
Animesh Mukherjee, Monojit Choudhury, Anupam Basu and Niloy Ganguly	101
<i>Cognate Identification and Alignment Using Practical Orthographies</i>	
Michael Cysouw and Hagen Jung	109
<i>ParaMor: Minimally Supervised Induction of Paradigm Structure and Morphological Analysis</i>	
Christian Monson, Jaime Carbonell, Alon Lavie and Lori Levin	117

<i>Dynamic Correspondences: An Object-Oriented Approach to Tracking Sound Reconstructions</i>	
Tyler Peterson and Gessiane Picanco	126
<i>Creating a Comparative Dictionary of Totonac-Tepéhua</i>	
Grzegorz Kondrak, David Beck and Philip Dilts	134

Workshop Program

Thursday, June 28, 2007

- 8:30–8:45 Opening
Computing and Historical Phonology
John Nerbonne, T. Mark Ellison and Grzegorz Kondrak
- 8:45–9:45 Keynote Speaker
Word Similarity Metrics and Multilateral Comparison
Brett Kessler
- 9:45–10:15 *Bayesian Identification of Cognates and Correspondences*
T. Mark Ellison
- 10:15–10:45 *Testing Cladistics on Dialect Networks and Phyla (Gallo-Romance Vowels, Southern Italo-Romance Diasystems and Mayan Languages)*
Antonella Gaillard-Corvaglia, Jean-Léo Léonard and Pierre Darlu
- 10:45–11:15 Break
- 11:15–11:45 *The Relative Divergence of Dutch Dialect Pronunciations from their Common Source: An Exploratory Study*
Wilbert Heeringa and Brian Joseph
- 11:45–12:15 *Can Corpus Based Measures be Used for Comparative Study of Languages?*
Anil Kumar Singh and Harshit Surana
- 12:15–12:45 *Inducing Sound Segment Differences Using Pair Hidden Markov Models*
Martijn Wieling, Therese Leinonen and John Nerbonne
- 12:45–14:15 Lunch
- 14:15–14:45 *Phonological Reconstruction of a Dead Language Using the Gradual Learning Algorithm*
Eric Smith
- 14:45–15:15 *Evolution, Optimization, and Language Change: The Case of Bengali Verb Inflections*
Monojit Choudhury, Vaibhav Jalan, Sudeshna Sarkar and Anupam Basu

Thursday, June 28, 2007 (continued)

15:15–15:45 Posters

On the Geolinguistic Change in Northern France between 1300 and 1900: A Dialectometrical Inquiry

Hans Goebel

Visualizing the Evaluation of Distance Measures

Thomas Pilz, Axel Philipsenburg and Wolfram Luther

Data Nonlinearity in Exploratory Multivariate Analysis of Language Corpora

Hermann Moisl

Emergence of Community Structures in Vowel Inventories: An Analysis Based on Complex Networks

Animesh Mukherjee, Monojit Choudhury, Anupam Basu and Niloy Ganguly

Cognate Identification and Alignment Using Practical Orthographies

Michael Cysouw and Hagen Jung

ParaMor: Minimally Supervised Induction of Paradigm Structure and Morphological Analysis

Christian Monson, Jaime Carbonell, Alon Lavie and Lori Levin

15:45–16:15 Break

16:15–16:45 *Dynamic Correspondences: An Object-Oriented Approach to Tracking Sound Reconstructions*

Tyler Peterson and Gessiane Picanco

16:45–17:15 *Creating a Comparative Dictionary of Totonac-Tepehua*

Grzegorz Kondrak, David Beck and Philip Dilts

17:15–17:45 Discussion

17:45–18:00 Closing