We gratefully acknowledge financial support from the U.S. National Science Foundation.

©2009 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

Preface

Welcome to the NAACL HLT 2009 Student Research Workshop and Doctoral Consortium! We are pleased to continue this established tradition at ACL conferences.

The Student Research Workshop and Doctoral Consortium provides a venue for student researchers in Computational Linguistics, Natural Language Processing, and Human Language Technologies to present their work and receive feedback from the community. Unlike regular conference sessions, the workshop welcomes work in progress. In the tradition of previous doctoral consortia, we recruited expert panelists to provide a brief commentary on each oral presentation. We hope that as a result of this workshop, the student participants are able to obtain exposure to the NAACL community and that it helps with their future careers in this field.

This year we received 29 submissions from 11 countries. With a strong program committee of 86 reviewers, from advanced graduate students to senior experts in their field, we were able to keep the individual reviewing load light and assign to each paper a team of three to six reviewers. Most papers were reviewed by two student reviewers and two established researchers.

We would like to thank the reviewers for understanding the spirit of the workshop and giving careful and constructive reviews. We hope their comments will be helpful to all the students who submitted their work.

A grant from the U.S. National Science Foundation enables us to provide financial support to all presenters to assist them in their travel to and attendance of the conference. We gratefully acknowledge this generous contribution.

Finally, we would also like to thank the general chair of NAACL HLT 2009, Mari Ostendorf, the program committee chairs, Michael Collins, Lucy Vanderwende, Doug Oard, and Shri Narayanan, the publicity chairs Matthew Stone, Gokhan Tur, and Diana Inkpen, the publications chairs, Christy Doran and Eric Ringger, the local arrangements chairs, James Martin and Martha Palmer, and the ACL Business Manager, Priscilla Rasmussen, for all the support they have provided in the organization of this workshop.

The faculty advisors and co-chairs of the NAACL HLT 2009 Student Research Workshop and Doctoral Consortium:

Carolyn Penstein Rosé
Anoop Sarkar
Svetlana Stoyanchev
Ulrich Germann
Chirag Shah
Co-chairs:
Ulrich Germann, University of Toronto, Canada
Chirag Shah, University of North Carolina, USA
Svetlana Stoyanchev, Stony Brook University, USA

Faculty Advisors:
Carolyn Penstein Rosé, Carnegie Mellon University, USA
Anoop Sarkar, Simon Fraser University, Canada

Program Committee:

Afra Alishahi       Pierre Isabelle       Joana Paulo Pardal
Nguyen Bach         Howard Johnson       Ted Pedersen
Niranjan Balasubramanian       Pallika Kanani       Gerald Penn
Satnamjee Banerjee       Weimao Ke         Hema Raghavan
Regina Barzilay         Diane Kelly       Ricardo Ribeiro
Shane Bergsma          Kevin Knight       Jason Riesa
Alan Black             Philipp Koehn       Ellen Riloff
Dan Bohus             Greg Kondrak       Mihai Rotaru
Sarah Borys            Roland Kuhn       Alex Rudnicky
Chris Callison-Burch   Shankar Kumar      Frank Rudzicz
Claire Cardie          Giridhar Kumaran     Narges Sharif Razavian
Marine Carpuat         Philippe Langlais    Michel Simard
Colin Cherry           Brian Langner       David Smith
Yejin Choi             Gregor Leusch       Yaxiao Song
Paul Cook              Adam Lopez         Richard Sproat
Steve DeNeefe          Annie Louis        Amanda Stent
Fernando Diaz          Daniel Marcu       Veselin Stoyanov
Gregory Duck           Jonathan May        Joel Tetrauld
Kevin Duh              David McClosky      Dilek Tur
Jason Eisner           David Mimno        Suzan Verberne
Afsaneh Fazly          Saif Mohammad      Stephan Vogel
Jenny Finkel           Antonio Moreno      Qin Iris Wang
Victoria Fossum         Javed Mostafa      Nick Webb
George Foster           Cristina Mota       Ryen White
Yanfen Hao             Smaranda Muresan    Fan Yang
Sanjika Hewavitharana   Ani Nenkova        Su-Youn Yoon
Derrick Higgins         Thuylinh Nguyen    Klaus Zechner
Silja Hildebrand       Oana Nicolov       Xiaodan Zhu
Graeme Hirst           Christopher Parisien
# Table of Contents

**Classifier Combination Techniques Applied to Coreference Resolution**  
Smita Vemulapalli, Xiaqiang Luo, John F. Pitrelli and Imed Zitouni ........................................ 1

**Solving the “Who’s Mark Johnson Puzzle”: Information Extraction Based Cross Document Coreference**  
Jian Huang, Sarah M. Taylor, Jonathan L. Smith, Konstantinos A. Fotiadis and C. Lee Giles . . . 7

**Exploring Topic Continuation Follow-up Questions using Machine Learning**  
Manuel Kirschner and Raffaella Bernardi ................................................................. 13

**Sentence Realisation from Bag of Words with Dependency Constraints**  
Karthik Gali and Sriram Venkatapathy ............................................................................ 19

**Using Language Modeling to Select Useful Annotation Data**  
Dmitriy Dligach and Martha Palmer ............................................................................. 25

**Pronunciation Modeling in Spelling Correction for Writers of English as a Foreign Language**  
Adriane Boyd .................................................................................................................. 31

**Building a Semantic Lexicon of English Nouns via Bootstrapping**  
Ting Qian, Benjamin Van Durme and Lenhart Schubert ................................................. 37

**Multiple Word Alignment with Profile Hidden Markov Models**  
Aditya Bhargava and Grzegorz Kondrak ...................................................................... 43

**Using Emotion to Gain Rapport in a Spoken Dialog System**  
Jaime Acosta .................................................................................................................. 49

**Interactive Annotation Learning with Indirect Feature Voting**  
Shilpa Arora and Eric Nyberg ....................................................................................... 55

**Loss-Sensitive Discriminative Training of Machine Transliteration Models**  
Kedar Bellare, Koby Crammer and Dayne Freitag ......................................................... 61

**Syntactic Tree-based Relation Extraction Using a Generalization of Collins and Duffy Convolution Tree Kernel**  
Mahdy Khayyamian, Seyed Abolghasem Mirroshandel and Hassan Abolhassani .......... 66

**Towards Building a Competitive Opinion Summarization System: Challenges and Keys**  
Elena Lloret, Alexandra Balahur, Manuel Palomar and Andrés Montoyo ......................... 72

**Domain-Independent Shallow Sentence Ordering**  
Thade Nahnsen .............................................................................................................. 78

**Towards Unsupervised Recognition of Dialogue Acts**  
Nicole Novielli and Carlo Strapparava ......................................................................... 84
Modeling Letter-to-Phoneme Conversion as a Phrase Based Statistical Machine Translation Problem with Minimum Error Rate Training
   Taraka Rama, Anil Kumar Singh and Sudheer Kolachina .......................... 90

Disambiguation of Preposition Sense Using Linguistically Motivated Features
   Stephen Tratz and Dirk Hovy ................................................................. 96
Conference Program

Monday, June 1, 2009

Morning Session

10:40–11:10 Classifier Combination Techniques Applied to Coreference Resolution
Smita Vemulapalli, Xiaqiang Luo, John F. Pitrelli and Imed Zitouni

11:15–11:45 Solving the “Who’s Mark Johnson Puzzle”: Information Extraction Based Cross Document Coreference
Jian Huang, Sarah M. Taylor, Jonathan L. Smith, Konstantinos A. Fotiadis and C. Lee Giles

11:50–12:20 Exploring Topic Continuation Follow-up Questions using Machine Learning
Manuel Kirschner and Raffaella Bernardi

First Afternoon Session

2:00–2:30 Sentence Realisation from Bag of Words with Dependency Constraints
Karthik Gali and Sriram Venkatapathy

2:35–3:05 Using Language Modeling to Select Useful Annotation Data
Dmitriy Dligach and Martha Palmer

Second Afternoon Session

4:00–4:30 Pronunciation Modeling in Spelling Correction for Writers of English as a Foreign Language
Adriane Boyd

4:35–5:05 Building a Semantic Lexicon of English Nouns via Bootstrapping
Ting Qian, Benjamin Van Durme and Lenhart Schubert

5:10–5:40 Multiple Word Alignment with Profile Hidden Markov Models
Aditya Bhargava and Grzegorz Kondrak
Monday, June 1, 2009 (continued)

Poster Session (6:30–9:30)

*Using Emotion to Gain Rapport in a Spoken Dialog System*
Jaime Acosta

*Interactive Annotation Learning with Indirect Feature Voting*
Shilpa Arora and Eric Nyberg

*Loss-Sensitive Discriminative Training of Machine Transliteration Models*
Kedar Bellare, Koby Crammer and Dayne Freitag

*Syntactic Tree-based Relation Extraction Using a Generalization of Collins and Duffy Convolution Tree Kernel*
Mahdy Khayyamian, Seyed Abolghasem Mirroshandel and Hassan Abolhassani

*Towards Building a Competitive Opinion Summarization System: Challenges and Keys*
Elena Lloret, Alexandra Balahur, Manuel Palomar and Andrés Montoyo

*Domain-Independent Shallow Sentence Ordering*
Thade Nahnsen

*Towards Unsupervised Recognition of Dialogue Acts*
Nicole Novielli and Carlo Strapparava

*Modeling Letter-to-Phoneme Conversion as a Phrase Based Statistical Machine Translation Problem with Minimum Error Rate Training*
Taraka Rama, Anil Kumar Singh and Sudheer Kolachina

*Disambiguation of Preposition Sense Using Linguistically Motivated Features*
Stephen Tratz and Dirk Hovy

All papers presented in the morning and afternoon sessions will also be shown as posters.