NAACL HLT 2009

Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics

Proceedings of the Conference

May 31 – June 5, 2009
Boulder, Colorado

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

Sponsors:

- Rosetta Stone
- CNGL
- Microsoft Research
- Google
- AT&T
- Language Weaver
- J.D. Power
- IBM Research
- The Linguistic Data Consortium
- The Human Language Technology Center of Excellence at the Johns Hopkins University
- The Computational Language and Education Research Center at the University of Colorado at Boulder

©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

ISBN: 978-1-932432-41-1

Preface: General Chair

I am honored that the North American Chapter of the Association of Computational Linguistics (NAACL) has given me the opportunity, as General Conference Chair, to continue the NAACL HLT tradition of covering topics from all areas of Human Language Technology, which makes it possible for researchers to discuss algorithms and applications that cut across the fields of natural language processing (NLP), speech processing, and information retrieval (IR).

I have been very fortunate to work with a terrific group of Technical Program Co-Chairs: Michael Collins (NLP), Shri Narayanan (speech), Douglas W. Oard (IR), and Lucy Vanderwende (NLP). This year the technical program emphasizes the breadth and interdisciplinary nature of human language processing research. The plenary talks will stretch our thinking about how language is used by considering the application of language to vision in one case, and language as it relates to food in another. There are two special sessions with themes that cut across multiple sub-areas of HLT: Large Scale Language Processing and Speech Information Retrieval. We also recognize the increasing importance of industry in our field with a lunchtime panel discussion on the Next Big Applications in Industry, with thanks to Bill Dolan for organizing and moderating the discussion. Finally, we have a breadth of excellent technical papers in lecture and poster sessions, thanks to the efforts of our Senior Program Committee members, the many reviewers on the Program Committee who helped us keep to our schedule, and the Paper Awards Committee. Together they have done a great job in putting together an interesting technical program. It has also been a pleasure to work with Local Organizers Martha Palmer and Jim Martin, who have done a terrific job in hosting a meeting that shows us Colorado's character as well as offering a great technical program. I hope you enjoy your stay in beautiful Boulder, as you are learning about new ideas and networking with valued colleagues.

The tradition of NAACL HLT is that it incorporates many events, including tutorials and workshops that have expanded in scope such that they are almost as big as the main conference. As a result, many other people have played important roles in making the overall conference a success and representative of the breadth of HLT. Specifically, I thank Matthew Stone, Gokhan Tur and Diana Inkpen for their work as Publicity Chairs; Christy Doran and Eric Ringger for their work as Publications Chairs; Fred Popowich and Michael Johnston for serving as Demo Chairs; Tutorial Chairs Ciprian Chelba, Paul Kantor and Brian Roark for bringing us an outstanding slate of tutorials; Workshop Chairs Nizar Habash and Mark Hasegawa-Johnson for their efforts in choosing and supporting the 12 workshops that extend our program by two days; and the Student Co-Chairs of the Doctoral Consortium organizers Svetlana Stenchikova, Ulrich Germann and Chirag Shah working with faculty advisors Carolyn Rosé and Anoop Sarkar. Thanks also to Nicolas Nicolov for his efforts as NAACL HLT Sponsorship Chair, working in coordination with Sponsorship Chairs from other ACL regions. Of course, we greatly appreciate the support of our sponsors: Rosetta Stone, CNGL, Microsoft Research, Google, AT&T, Language Weaver, J.D. Power, IBM Research, the Linguistic Data Consortium, the Human Language Technology Center of Excellence at the Johns Hopkins University, and the Computational Language and Education Research Center at the University of Colorado at Boulder.

In organizing this conference, we have had a lot of support from the NAACL Board and the HLT Advisory Board. I would particularly like to thank Owen Rambow, Jennifer Chu-carroll, Chris Manning and Graeme Hirst for their help and advice. Last, but certainly not least, we are indebted to Priscilla Rasmussen for her expertise and support in running the conference.

Mari Ostendorf, University of Washington

Preface: Program Chairs

We welcome you to NAACL HLT 2009! The NAACL HLT program continues to include high-quality work in the areas of computational linguistics, information retrieval, and speech technology. This year, 260 full papers were submitted, of which 75 papers were accepted (giving a 29% acceptance rate); and 178 short papers were submitted, of which 71 were accepted (giving a 40% acceptance rate).

Two best paper awards were given at the conference, to "Unsupervised Morphological Segmentation with Log-Linear Models", by Hoifung Poon, Colin Cherry and Kristina Toutanova (this paper also received the best student paper award), and "11,001 New Features for Statistical Machine Translation", by David Chiang, Kevin Knight and Wei Wang. The senior program committee members for the conference nominated an initial set of papers that were candidates for the awards; the final decisions were then made by a committee chaired by Candace Sidner, and with Hal Daume III, Roland Kuhn, Ryan McDonald, and Mark Steedman as its other members. We would like to congratulate the authors, and thank the committee for their work in choosing these papers.

NAACL HLT 2009 consists of oral presentations of all full papers, oral or poster presentations of short papers, and tutorials and software demonstrations. We are delighted to have two keynote speakers: Antonio Torralba, with a talk "Understanding Visual Scenes", and Dan Jurafsky, with a talk "The Language of Food". In addition, we have a panel on emerging application areas in computational linguistics, chaired by Bill Dolan.

We would like to thank the authors for submitting a remarkable set of papers to the conference. The review process was organized through a two-tier system, with eighteen senior program committee (SPC) members, and 352 reviewers. The SPC members managed the review process for both the full and short paper submissions: each full paper received at least three reviews, and each short paper received at least two reviews. We are thoroughly indebted to the reviewers for all their work, and to the SPC members for the long hours they spent in evaluating the submissions. In addition, we would like to thank Rich Gerber and the START team for their help with the system that managed paper submissions and reviews; the local arrangement chairs, James Martin and Martha Palmer, for their help with organizing the program; and the publication chairs, Christy Doran and Eric Ringger, for putting together these proceedings. Finally, we are incredibly grateful to the general chair, Mari Ostendorf, for the invaluable advice and support that she provided throughout every step of the process.

We hope that you enjoy the conference!

Michael Collins, Massachusetts Institute of Technology
Shri Narayanan, University of Southern California
Douglas W. Oard, University of Maryland
Lucy Vanderwende, Microsoft Research

Organizers

General Chair:

Mari Ostendorf, University of Washington

Local Arrangements:

James Martin, University of Colorado
Martha Palmer, University of Colorado

Program Committee Chairs:

Michael Collins, Massachusetts Institute of Technology
Shri Narayanan, University of Southern California
Douglas W. Oard, University of Maryland
Lucy Vanderwende, Microsoft Research

Publicity Chairs:

Matthew Stone, Rutgers University
Gokhan Tur, SRI International
Diana Inkpen, University of Ottawa

Publications Chairs:

Christy Doran, MITRE
Eric Ringger, Brigham Young University

Tutorials Chairs:

Ciprian Chelba, Google
Paul Kantor, Rutgers University
Brian Roark, Oregon Health and Science University

Workshops Chairs:

Nizar Habash, Columbia University
Mark Hasegawa-Johnson, University of Illinois

Doctoral Consortium Organizers:

Carolyn Rosé, Faculty Chair, CMU
Anoop Sarkar, Faculty Chair, Simon Fraser University

Svetlana Stoyachev, Student Co-Chair, Stony Brook University
Ulrich Germann, Student Co-Chair, University of Toronto
Chirag Shah, Student Co-Chair, University of North Carolina

Demo Chairs:

Fred Popowich, Simon Fraser University
Michael Johnston, AT&T

Sponsorship Committee:

Nicolas Nicolov (Local Chair)
Hitoshi Isahara and Kim-Teng Lua (Asian ACL Rrepresentatives)
Philipp Koehn and Josef van Genabith (European ACL Representatives)
Srinivas Bangalore and Christy Doran (American ACL Representatives)

Program Committee

Senior Program Committee Members:

Michiel Bacchiani, Google
Regina Barzilay, Massachusetts Institute of Technology
Kenneth W. Church, Microsoft Research
Charles L. A. Clarke, University of Waterloo
Eric Fosler-Lussier, Ohio State University
Sharon Goldwater, University of Edinburgh
Julia Hirschberg, Columbia University
Jimmy Huang, York University
Mark Johnson, Brown University
Philipp Koehn, University of Edinburgh
Roland Kuhn, National Research Council of Canada, IIT
Gina-Anne Levow, University of Manchester
Dekang Lin, Google
Ryan McDonald, Google
Premkumar Natarajan, BBN Technologies
Patrick Pantel, Yahoo! Labs
Kristina Toutanova, Microsoft Research
Geoff Zweig, Microsoft Research

Paper Award Committee:

Candace Sidner, Chair, BAE Systems AIT
Hal Daumé III, University of Utah
Roland Kuhn, NRC Institute for Information Technology
Ryan McDonald, Google Inc.
Mark Steedman, University of Edinburgh

Program Committee Members:

Stephen Abney
Meni Adler
Eugene Agichtein
Eneko Agirre
Lars Ahrenberg
Adam Albright
Enrique Alfonseca
Afra Alishahi
Sophia Ananiadou
Shankar Ananthakrishnan
Bill Andreopoulos
Galen Andrew
Walter Andrews
Masayuki Asahara
Necip Fazil Ayan
Mark Baillie
Timothy Baldwin
Roberto Basili
Ron Bekkerman
Sabine Bergler
Shane Bergsma
Rahul Bhagat
Dan Bikel
Mikhail Bilenko

Alexandra Birch
Alan Black
Sasha Blair-Goldensohn
John Blitzer
Paul Boersma
Johan Bos
Alexandre Bouchard-Côté
S.R.K. Branavan
Chris Brew
Ted Briscoe
Chris Brockett
Stefan Buettcher
Razvan Bunescu
Jill Burstein
Cory Butz
William Byrne
Chris Callison-Burch
Claire Cardie
Giuseppe Carenini
Marine Carpuat
Xavier Carreras
Francisco Casacuberta
Joyce Chai
Yllias Chali
Nate Chambers
Jason Chang
Eugene Charniak
Ciprian Chelba
Harr Chen
Colin Cherry
David Chiang
Tat-Seng Chua
Grace Chung
Massimiliano Ciaramita
Stephen Clark
Peter Clark
Mark Craven
Mathias Creutz
Aron Culotta
James Cussens
Robert Dale
Cristian Danescu Niculescu-Mizil
Hal Daumé III
Guy De Pauw
John DeNero
Barbara Di Eugenio
Mona Diab
Bill Dolan
Christy Doran
Doug Downey
Mark Dredze
Markus Dreyer
Rebecca Dridan
Kevin Duh
Chris Dyer
Andreas Eisele
Jacob Eisenstein
Jason Eisner
Michael Elhadad
Noemie Elhadad
Mark Ellison
Micha Elsner
Dominique Estival
Oren Etzioni
Hui Fang
Marcello Federico
Paolo Ferragina
Jenny Finkel
Erin Fitzgerald
Radu Florian
George Foster
Dayne Freitag
Pascale Fung
Robert Gaizauskas
Michael Gamon
Kuzman Ganchev
Jianfeng Gao
Claire Gardent
Stuart Geman
Ulrich Germann
Shlomo Geva
Mazin Gilbert
Daniel Gildea
Jesus Gimenez
Roxana Girju
Randy Goebel
John Goldsmith
Ralph Grishman
Asela Gunawardana
Gholamreza Haffari
Aria Haghighi
Udo Hahn

Dilek Hakkani-Tür
Keith Hall
Hyoil Han
Mary Harper
Saša Hasan
Mark Hasegawa-Johnson
Timothy J. Hazen
Xiaodong He
William Headden
Peter Heeman
James Henderson
Iris Hendrickx
Graeme Hirst
Hieu Hoang
Kristy Hollingshead
Mark Hopkins
Vronique Hoste
Chu-Ren Huang
Liang Huang
Rebecca Hwa
Diana Inkpen
Abe Ittycheriah
Gaja Jarosz
Heng Ji
Richard Johansson
Howard Johnson
Rie Johnson
Doug Jones
Gareth Jones
Aravind Joshi
Min-Yen Kan
Chia-lin Kao
Nikiforos Karamanis
Rohit Kate
Vlado Keselj
Shahram Khadivi
Sanjeev Khudanpur
Adam Kilgarriff
Jin-Dong Kim
Owen Kimball
Dan Klein
Kevin Knight
Mamoru Komachi
Grzegorz Kondrak
Terry Koo
Anna Korhonen
Kimmo Koskenniemi
Emiel Krahmer
Jonas Kuhn
Shankar Kumar
Christian König
Philippe Langlais
Mirella Lapata
Alex Lascarides
Alon Lavie
Claudia Leacock
Lillian Lee
Yoong Keok Lee
James Lester
Gregor Leusch
Roger Levy
David Lewis
Wei Li
Xiao Li
Haizhou Li
Hang Li
Ping Li
Percy Liang
Hank Liao
Jimmy Lin
Chin-Yew Lin
Bing Liu
Yang Liu
Tie-Yan Liu
Andrej Ljolje
Adam Lopez
Alex Lopez-Ortiz
Bill MacCartney
Nitin Madnani
Bernardo Magnini
Jonathan Mamou
Suresh Manandhar
Lidia Mangu
Gideon Mann
Chris Manning
Daniel Marcu
Evgeny Matusov
Arne Mauser
David McAllester
Andrew McCallum
Diana McCarthy
David McClosky

Kathy McCoy
Kathleen McKeown
Susan McRoy
Qiaozhu Mei
Paola Merlo
Rada Mihalcea
Yusuke Miyao
Saif Mohammad
Dan Moldovan
Bob Moore
Richard Moot
Pedro Moreno
Dragos Munteanu
Smaranda Muresan
Muthu Muthukrishnan
Tetsuji Nakagawa
Preslav Nakov
Ani Nenkova
Hermann Ney
Hwee Tou Ng
Vincent Ng
Raymond Ng
Patrick Nguyen
Jian-Yun Nie
Joakim Nivre
Franz Och
Kemal Oflazer
Scott Olsson
Luca Onnis
Miles Osborne
Tim Paek
Bo Pang
Marius Pasca
Rebecca Passonneau
Matthias Paulik
Ted Pedersen
Marco Pennacchiotti
Mati Pentus
Amy Perfors
Slav Petrov
Joseph Picone
Janet Pierrehumbert
Livia Polanyi
Hoifung Poon
Ana-Maria Popescu
Maja Popovic
Fred Popowich
John Prager
Rohit Prasad
Partha Pratim Talukdar
Matthew Purver
Chris Quirk
Drago Radev
Rajat Raina
Daniel Ramage
Owen Rambow
Vivek Kumar Rangarajan Sridhar
Deepak Ravichandran
Stefan Riezler
Ellen Riloff
Eric Ringger
Brian Roark
Barbara Rosario
Dan Roth
Alex Rudnicky
Marta Ruiz
Anton Rytting
Kenji Sagae
Johan Schalkwyk
David Schlangen
Tanja Schultz
Petr Schwarz
Holger Schwenk
Satoshi Sekine
Mike Seltzer
Stephanie Seneff
Wade Shen
Stuart Shieber
Luo Si
Michel Simard
Olivier Siohan
Kevin Small
David Smith
Noah Smith
Mark Smucker
Rion Snow
Ben Snyder
Radu Soricut
Richard Sproat
Amit Srivastava
David Stallard
Mark Steedman

Mark Stevenson
Michael Strube
Amarnag Subramanya
Torsten Suel
Eiichiro Sumita
Charles Sutton
David Talbot
Ben Taskar
Yee Whye Teh
Simone Teufel
Joerg Tiedemann
Christoph Tillmann
Ivan Titov
Isabel Trancoso
David Traum
Andrew Trotman
Peter Turney
Nicola Ueffing
Jay Urbain
Antal van den Bosch
Benjamin van Durme
Olga Vechtomova
Dimitra Vergyri
Evelyne Viegas
David Vilar
Ye-Yi Wang
Qin Wang
Nigel Ward
Taro Watanabe
Bonnie Webber
MIchael White
Richard Wicentowski
Jason Williams
Shuly Wintner
Dekai Wu
Mingfang Wu
Peng Xu
Roman Yangarber
Alex Yates
Zheng Ye
Scott Wen-tau Yih
Chen Yu
Dong Yu
Fabio Massimo Zanzotto
Richard Zens
Luke Zettlemoyer
Hao Zhang
Ming Zhou
Wei Zhou
Bowen Zhou
Jerry Zhu
Jianhan Zhu
Andreas Zollmann

Table of Contents

Subjectivity Recognition on Word Senses via Semi-supervised Mincuts
Fangzhong Su and Katja Markert 1

Integrating Knowledge for Subjectivity Sense Labeling
Yaw Gyamfi, Janyce Wiebe, Rada Mihalcea and Cem Akkaya 10

A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches
Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Pasca and Aitor Soroa . 19

A Fully Unsupervised Word Sense Disambiguation Method Using Dependency Knowledge
Ping Chen, Wei Ding, Chris Bowes and David Brown 28

Learning Phoneme Mappings for Transliteration without Parallel Data
Sujith Ravi and Kevin Knight 37

A Corpus-Based Approach for the Prediction of Language Impairment in Monolingual English and Spanish-English Bilingual Children
Keyur Gabani, Melissa Sherman, Thamar Solorio, Yang Liu, Lisa Bedore and Elizabeth Peña . 46

A Discriminative Latent Variable Chinese Segmenter with Hybrid Word/Character Information
Xu Sun, Yaozhong Zhang, Takuya Matsuzaki, Yoshimasa Tsuruoka and Jun'ichi Tsujii 56

Improved Reconstruction of Protolanguage Word Forms
Alexandre Bouchard-Côté, Thomas L. Griffiths and Dan Klein 65

Shared Logistic Normal Distributions for Soft Parameter Tying in Unsupervised Grammar Induction
Shay Cohen and Noah A. Smith 74

Adding More Languages Improves Unsupervised Multilingual Part-of-Speech Tagging: a Bayesian Non-Parametric Approach
Benjamin Snyder, Tahira Naseem, Jacob Eisenstein and Regina Barzilay 83

Efficiently Parsable Extensions to Tree-Local Multicomponent TAG
Rebecca Nesson and Stuart Shieber 92

Improving Unsupervised Dependency Parsing with Richer Contexts and Smoothing
William P. Headden III, Mark Johnson and David McClosky 101

Context-Dependent Alignment Models for Statistical Machine Translation
Jamie Brunning, Adrià de Gispert and William Byrne 110

Graph-based Learning for Statistical Machine Translation
Andrei Alexandrescu and Katrin Kirchhoff 119

Intersecting Multilingual Data for Faster and Better Statistical Translations
Yu Chen, Martin Kay and Andreas Eisele 128

Without a 'doubt'? Unsupervised Discovery of Downward-Entailing Operators
Cristian Danescu-Niculescu-Mizil, Lillian Lee and Richard Ducott 137

The Role of Implicit Argumentation in Nominal SRL
Matthew Gerber, Joyce Chai and Adam Meyers 146

Jointly Identifying Predicates, Arguments and Senses using Markov Logic
Ivan Meza-Ruiz and Sebastian Riedel 155

Structured Generative Models for Unsupervised Named-Entity Clustering
Micha Elsner, Eugene Charniak and Mark Johnson 164

Hierarchical Dirichlet Trees for Information Retrieval
Gholamreza Haffari and Yee Whye Teh 173

Phrase-Based Query Degradation Modeling for Vocabulary-Independent Ranked Utterance Retrieval
J. Scott Olsson and Douglas W. Oard 182

Japanese Query Alteration Based on Lexical Semantic Similarity
Masato Hagiwara and Hisami Suzuki 191

Context-based Message Expansion for Disentanglement of Interleaved Text Conversations
Lidan Wang and Douglas W. Oard 200

Unsupervised Morphological Segmentation with Log-Linear Models
Hoifung Poon, Colin Cherry and Kristina Toutanova 209

11,001 New Features for Statistical Machine Translation
David Chiang, Kevin Knight and Wei Wang 218

Efficient Parsing for Transducer Grammars
John DeNero, Mohit Bansal, Adam Pauls and Dan Klein 227

Preference Grammars: Softening Syntactic Constraints to Improve Statistical Machine Translation
Ashish Venugopal, Andreas Zollmann, Noah A. Smith and Stephan Vogel 236

Using a Dependency Parser to Improve SMT for Subject-Object-Verb Languages
Peng Xu, Jaeho Kang, Michael Ringgaard and Franz Och 245

Learning Bilingual Linguistic Reordering Model for Statistical Machine Translation
Han-Bin Chen, Jian-Cheng Wu and Jason S. Chang 254

May All Your Wishes Come True: A Study of Wishes and How to Recognize Them
Andrew B. Goldberg, Nathanael Fillmore, David Andrzejewski, Zhiting Xu, Bryan Gibson and Xiaojin Zhu 263

Predicting Risk from Financial Reports with Regression
Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi and Noah A. Smith 272

Domain Adaptation with Latent Semantic Association for Named Entity Recognition
Honglei Guo, Huijia Zhu, Zhili Guo, Xiaoxun Zhang, Xian Wu and Zhong Su 281

Semi-Automatic Entity Set Refinement
Vishnu Vyas and Patrick Pantel 290

Unsupervised Constraint Driven Learning For Transliteration Discovery
Ming-Wei Chang, Dan Goldwasser, Dan Roth and Yuancheng Tu 299

On the Syllabification of Phonemes
Susan Bartlett, Grzegorz Kondrak and Colin Cherry 308

Improving nonparameteric Bayesian inference: experiments on unsupervised word segmentation with adaptor grammars
Mark Johnson and Sharon Goldwater 317

Joint Parsing and Named Entity Recognition
Jenny Rose Finkel and Christopher D. Manning 326

Minimal-length linearizations for mildly context-sensitive dependency trees
Y. Albert Park and Roger Levy 335

Positive Results for Parsing with a Bounded Stack using a Model-Based Right-Corner Transform
William Schuler 344

Hierarchical Text Segmentation from Multi-Scale Lexical Cohesion
Jacob Eisenstein 353

Exploring Content Models for Multi-Document Summarization
Aria Haghighi and Lucy Vanderwende 362

Global Models of Document Structure using Latent Permutations
Harr Chen, S.R.K. Branavan, Regina Barzilay and David R. Karger 371

Assessing and Improving the Performance of Speech Recognition for Incremental Systems
Timo Baumann, Michaela Atterer and David Schlangen 380

Geo-Centric Language Models for Local Business Voice Search
Amanda Stent, Ilija Zeljkovic, Diamantino Caseiro and Jay Wilpon 389

Improving the Arabic Pronunciation Dictionary for Phone and Word Recognition with Linguistically-Based Pronunciation Rules
Fadi Biadsy, Nizar Habash and Julia Hirschberg 397

Using a maximum entropy model to build segmentation lattices for MT
Chris Dyer 406

Active Learning for Statistical Phrase-based Machine Translation
Gholamreza Haffari, Maxim Roy and Anoop Sarkar 415

Semi-Supervised Lexicon Mining from Parenthetical Expressions in Monolingual Web Pages
Xianchao Wu, Naoaki Okazaki and Jun'ichi Tsujii 424

Hierarchical Phrase-Based Translation with Weighted Finite State Transducers
Gonzalo Iglesias, Adrià de Gispert, Eduardo R. Banga and William Byrne 433

Improved pronunciation features for construct-driven assessment of non-native spontaneous speech
Lei Chen, Klaus Zechner and Xiaoming Xi 442

Performance Prediction for Exponential Language Models
Stanley Chen 450

Tied-Mixture Language Modeling in Continuous Space
Ruhi Sarikaya, Mohamed Afify and Brian Kingsbury 459

Shrinking Exponential Language Models
Stanley Chen 468

Predicting Response to Political Blog Posts with Topic Models
Tae Yano, William W. Cohen and Noah A. Smith 477

An Iterative Reinforcement Approach for Fine-Grained Opinion Mining
Weifu Du and Songbo Tan 486

For a few dollars less: Identifying review pages sans human labels
Luciano Barbosa, Ravi Kumar, Bo Pang and Andrew Tomkins 494

More than Words: Syntactic Packaging and Implicit Sentiment
Stephan Greene and Philip Resnik 503

Streaming for large scale NLP: Language Modeling
Amit Goyal, Hal Daume III and Suresh Venkatasubramanian 512

The Effect of Corpus Size on Case Frame Acquisition for Discourse Analysis
Ryohei Sasano, Daisuke Kawahara and Sadao Kurohashi 521

Semantic-based Estimation of Term Informativeness
Kirill Kireyev 530

Optimal Reduction of Rule Length in Linear Context-Free Rewriting Systems
Carlos Gómez-Rodríguez, Marco Kuhlmann, Giorgio Satta and David Weir 539

Inducing Compact but Accurate Tree-Substitution Grammars
Trevor Cohn, Sharon Goldwater and Phil Blunsom 548

Hierarchical Search for Parsing
Adam Pauls and Dan Klein 557

An effective Discourse Parser that uses Rich Linguistic Information
Rajen Subba and Barbara Di Eugenio 566

Graph-Cut-Based Anaphoricity Determination for Coreference Resolution
Vincent Ng 575

Using Citations to Generate surveys of Scientific Paradigms
Saif Mohammad, Bonnie Dorr, Melissa Egan, Ahmed Hassan, Pradeep Muthukrishan, Vahed Qazvinian, Dragomir Radev and David Zajic 584

Non-Parametric Bayesian Areal Linguistics
Hal Daume III 593

Hierarchical Bayesian Domain Adaptation
Jenny Rose Finkel and Christopher D. Manning 602

Online EM for Unsupervised Models
Percy Liang and Dan Klein 611

Unsupervised Approaches for Automatic Keyword Extraction Using Meeting Transcripts
Feifan Liu, Deana Pennell, Fei Liu and Yang Liu 620

A Finite-State Turn-Taking Model for Spoken Dialog Systems
Antoine Raux and Maxine Eskenazi 629

Extracting Social Meaning: Identifying Interactional Style in Spoken Conversation
Dan Jurafsky, Rajesh Ranganath and Dan McFarland 638

Linear Complexity Context-Free Parsing Pipelines via Chart Constraints
Brian Roark and Kristy Hollingshead 647

Improved Syntactic Models for Parsing Speech with Repairs
Tim Miller 656

A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs
Klinton Bicknell and Roger Levy 665

Conference Program Overview

Monday, June 1, 2009

9:00–10:10 Plenary Session – Invited Talk by Antonio Torralba: *Understanding Visual Scenes*

10:40–11:20 Session 1A: Semantics
Session 1B: Multilingual Processing / Morphology and Phonology
Session 1C: Syntax and Parsing
Student Research Workshop Session 1

2:00–3:30 Short Paper Presentations:
Session 2A: Machine Translation
Session 2B: Information Retrieval / Information Extraction / Sentiment
Session 2C: Dialog / Speech / Semantics
Student Research Workshop Session 2

4:00–5:40 Session 3A: Machine Translation
Session 3B: Semantics
Session 3C: Information Retrieval
Student Research Workshop Session 3

6:30–9:30 Poster and Demo Session
Student Research Workshop Poster Session

Tuesday, June 2, 2009

9:00-10:10 Plenary Session: Paper Award Presentations

10:10–11:40 Session 4A: Machine Translation
Session 4B: Sentiment Analysis / Information Extraction
Session 4C: Machine Learning / Morphology and Phonology

2:00–3:30 Short Paper Presentations:
Session 5A: Machine Translation / Generation / Semantics
Session 5B: Machine Learning / Syntax
Session 5C: SPECIAL SESSION – Speech Indexing and Retrieval

4:00–5:15 Session 6A: Syntax and Parsing
Session 6B: Discourse and Summarization
Session 6C: Spoken Language Systems

Wednesday, June 3, 2009

9:00–10:10	Plenary Session – Invited Talk by Dan Jurafsky: *Ketchup, Espresso, and Chocolate Chip Cookies: Travels in the Language of Food*
10:40–12:20	Session 7A: Machine Translation Session 7B: Speech Recognition and Language Modeling Session 7C: Sentiment Analysis
12:40-1:40	Panel Discussion: *Emerging Application Areas in Computational Linguistics*
1:40–2:30	NAACL Business Meeting
2:30–3:45	Session 8A: Large-scale NLP Session 8B: Syntax and Parsing Session 8C: Discourse and Summarization
4:15–5:30	Session 9A: Machine Learning Session 9B: Dialog Systems Session 9C: Syntax and Parsing

Conference Program

Monday, June 1, 2009

Plenary Session

9:00–10:10 Welcome and Invited Talk: *Understanding Visual Scenes*
Antonio Torralba

10:10–10:40 **Break**

Session 1A: Semantics

10:40–11:05 *Subjectivity Recognition on Word Senses via Semi-supervised Mincuts*
Fangzhong Su and Katja Markert

11:05–11:30 *Integrating Knowledge for Subjectivity Sense Labeling*
Yaw Gyamfi, Janyce Wiebe, Rada Mihalcea and Cem Akkaya

11:30–11:55 *A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches*
Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Pasca and Aitor Soroa

11:55–12:20 *A Fully Unsupervised Word Sense Disambiguation Method Using Dependency Knowledge*
Ping Chen, Wei Ding, Chris Bowes and David Brown

Session 1B: Multilingual Processing / Morphology and Phonology

10:40–11:05 *Learning Phoneme Mappings for Transliteration without Parallel Data*
Sujith Ravi and Kevin Knight

11:05–11:30 *A Corpus-Based Approach for the Prediction of Language Impairment in Monolingual English and Spanish-English Bilingual Children*
Keyur Gabani, Melissa Sherman, Thamar Solorio, Yang Liu, Lisa Bedore and Elizabeth Peña

11:30–11:55 *A Discriminative Latent Variable Chinese Segmenter with Hybrid Word/Character Information*
Xu Sun, Yaozhong Zhang, Takuya Matsuzaki, Yoshimasa Tsuruoka and Jun'ichi Tsujii

11:55–12:20 *Improved Reconstruction of Protolanguage Word Forms*
Alexandre Bouchard-Côté, Thomas L. Griffiths and Dan Klein

Monday, June 1, 2009 (continued)

Session 1C: Syntax and Parsing

10:40–11:05 *Shared Logistic Normal Distributions for Soft Parameter Tying in Unsupervised Grammar Induction*
Shay Cohen and Noah A. Smith

11:05–11:30 *Adding More Languages Improves Unsupervised Multilingual Part-of-Speech Tagging: a Bayesian Non-Parametric Approach*
Benjamin Snyder, Tahira Naseem, Jacob Eisenstein and Regina Barzilay

11:30–11:55 *Efficiently Parsable Extensions to Tree-Local Multicomponent TAG*
Rebecca Nesson and Stuart Shieber

11:55–12:20 *Improving Unsupervised Dependency Parsing with Richer Contexts and Smoothing*
William P. Headden III, Mark Johnson and David McClosky

Student Research Workshop Session 1:

Note: all student research workshop papers are located in the Companion volume of the proceedings

10:40–11:10 *Classifier Combination Techniques Applied to Coreference Resolution*
Smita Vemulapalli, Xiaoqiang 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

12:20–2:00 **Lunch Break**

Monday, June 1, 2009 (continued)

Session 2A: Short Paper Presentations: Machine Translation

Note: all short papers are located in the Companion volume of the proceedings

2:00–2:15 *Cohesive Constraints in A Beam Search Phrase-based Decoder*
Nguyen Bach, Stephan Vogel and Colin Cherry

2:15–2:30 *Revisiting Optimal Decoding for IBM Machine Translation Model 4*
James Clarke and Sebastian Riedel

2:30–2:45 *Efficient Extraction of Oracle-best Translations from Hypergraphs*
Zhifei Li and Sanjeev Khudanpur

2:45–3:00 *Semantic Roles for SMT: A Hybrid Two-Pass Model*
Dekai Wu and Pascale Fung

3:00–3:15 *Comparison of Extended Lexicon Models in Search and Rescoring for SMT*
Saša Hasan and Hermann Ney

3:15–3:30 *Simplex Armijo Downhill Algorithm for Optimizing Statistical Machine Translation System Parameters*
Bing Zhao and Shengyuan Chen

Session 2B: Short Paper Presentations: Information Retrieval / Information Extraction / Sentiment

Note: all short papers are located in the Companion volume of the proceedings

2:00–2:15 *Translation Corpus Source and Size in Bilingual Retrieval*
Paul McNamee, James Mayfield and Charles Nicholas

2:15–2:30 *Large-scale Computation of Distributional Similarities for Queries*
Enrique Alfonseca, Keith Hall and Silvana Hartmann

2:30–2:45 *Text Categorization from Category Name via Lexical Reference*
Libby Barak, Ido Dagan and Eyal Shnarch

2:45–3:00 *Identifying Types of Claims in Online Customer Reviews*
Shilpa Arora, Mahesh Joshi and Carolyn Rose

Monday, June 1, 2009 (continued)

3:00–3:15 *Towards Automatic Image Region Annotation - Image Region Textual Coreference Resolution*
Emilia Apostolova and Dina Demner-Fushman

3:15–3:30 *TESLA: A Tool for Annotating Geospatial Language Corpora*
Nate Blaylock, Bradley Swain and James Allen

Session 2C: Short Paper Presentations: Dialog / Speech / Semantics

Note: all short papers are located in the Companion volume of the proceedings

2:00–2:15 *Modeling Dialogue Structure with Adjacency Pair Analysis and Hidden Markov Models*
Kristy Elizabeth Boyer, Robert Phillips, Eun Young Ha, Michael Wallis, Mladen Vouk and James Lester

2:15–2:30 *Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue Systems*
Kenji Sagae, Gwen Christian, David DeVault and David Traum

2:30–2:45 *Spherical Discriminant Analysis in Semi-supervised Speaker Clustering*
Hao Tang, Stephen Chu and Thomas Huang

2:45–3:00 *Learning Bayesian Networks for Semantic Frame Composition in a Spoken Dialog System*
Marie-Jean Meurs, Fabrice Lefvre and Renato De Mori

3:00–3:15 *Evaluation of a System for Noun Concepts Acquisition from Utterances about Images (SINCA) Using Daily Conversation Data*
Yuzu Uchida and Kenji Araki

3:15–3:30 *Web and Corpus Methods for Malay Count Classifier Prediction*
Jeremy Nicholson and Timothy Baldwin

Monday, June 1, 2009 (continued)

Student Research Workshop Session 2

Note: all student research workshop papers are located in the Companion volume of the proceedings

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

3:30–4:00 **Break**

Session 3A: Machine Translation

4:00–4:25 *Context-Dependent Alignment Models for Statistical Machine Translation*
Jamie Brunning, Adrià de Gispert and William Byrne

4:25–4:50 *Graph-based Learning for Statistical Machine Translation*
Andrei Alexandrescu and Katrin Kirchhoff

4:50–5:15 *Intersecting Multilingual Data for Faster and Better Statistical Translations*
Yu Chen, Martin Kay and Andreas Eisele

5:15–5:40 No Presentation

Session 3B: Semantics

4:00–4:25 *Without a 'doubt'? Unsupervised Discovery of Downward-Entailing Operators*
Cristian Danescu-Niculescu-Mizil, Lillian Lee and Richard Ducott

4:25–4:50 *The Role of Implicit Argumentation in Nominal SRL*
Matthew Gerber, Joyce Chai and Adam Meyers

4:50–5:15 *Jointly Identifying Predicates, Arguments and Senses using Markov Logic*
Ivan Meza-Ruiz and Sebastian Riedel

Monday, June 1, 2009 (continued)

5:15–5:40 *Structured Generative Models for Unsupervised Named-Entity Clustering*
Micha Elsner, Eugene Charniak and Mark Johnson

Session 3C: Information Retrieval

4:00–4:25 *Hierarchical Dirichlet Trees for Information Retrieval*
Gholamreza Haffari and Yee Whye Teh

4:25–4:50 *Phrase-Based Query Degradation Modeling for Vocabulary-Independent Ranked Utterance Retrieval*
J. Scott Olsson and Douglas W. Oard

4:50–5:15 *Japanese Query Alteration Based on Lexical Semantic Similarity*
Masato Hagiwara and Hisami Suzuki

5:15–5:40 *Context-based Message Expansion for Disentanglement of Interleaved Text Conversations*
Lidan Wang and Douglas W. Oard

Student Research Workshop Session 3

Note: all student research workshop papers are located in the Companion volume of the proceedings

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

6:30–9:30 **Poster and Demo Session**

Note: all short papers and demo abstracts are located in the Companion volume of the proceedings

Minimum Bayes Risk Combination of Translation Hypotheses from Alternative Morphological Decompositions
Adri de Gispert, Sami Virpioja, Mikko Kurimo and William Byrne

Generating Synthetic Children's Acoustic Models from Adult Models
Andreas Hagen, Bryan Pellom and Kadri Hacioglu

Detecting Pitch Accents at the Word, Syllable and Vowel Level
Andrew Rosenberg and Julia Hirschberg

Shallow Semantic Parsing for Spoken Language Understanding
Bonaventura Coppola, Alessandro Moschitti and Giuseppe Riccardi

Automatic Agenda Graph Construction from Human-Human Dialogs using Clustering Method
Cheongjae Lee, Sangkeun Jung, Kyungduk Kim and Gary Geunbae Lee

A Simple Sentence-Level Extraction Algorithm for Comparable Data
Christoph Tillmann and Jian-ming Xu

Learning Combination Features with L1 Regularization
Daisuke Okanohara and Jun'ichi Tsujii

Multi-scale Personalization for Voice Search
Daniel Bolanos, Geoffrey Zweig and Patrick Nguyen

The Importance of Sub-Utterance Prosody in Predicting Level of Certainty
Heather Pon-Barry and Stuart Shieber

Using Integer Linear Programming for Detecting Speech Disfluencies
Kallirroi Georgila

Contrastive Summarization: An Experiment with Consumer Reviews
Kevin Lerman and Ryan McDonald

Topic Identification Using Wikipedia Graph Centrality
Kino Coursey and Rada Mihalcea

Extracting Bilingual Dictionary from Comparable Corpora with Dependency Heterogeneity
Kun Yu and Junichi Tsujii

Domain Adaptation with Artificial Data for Semantic Parsing of Speech
Lonneke van der Plas, James Henderson and Paola Merlo

Extending Pronunciation Lexicons via Non-phonemic Respellings
Lucian Galescu

A Speech Understanding Framework that Uses Multiple Language Models and Multiple Understanding Models
Masaki Katsumaru, Mikio Nakano, Kazunori Komatani, Kotaro Funakoshi, Tetsuya Ogata and Hiroshi G. Okuno

Taking into Account the Differences between Actively and Passively Acquired Data: The Case of Active Learning with Support Vector Machines for Imbalanced Datasets
Michael Bloodgood and Vijay Shanker

Faster MT Decoding Through Pervasive Laziness
Michael Pust and Kevin Knight

Evaluating the Syntactic Transformations in Gold Standard Corpora for Statistical Sentence Compression
Naman K Gupta, Sourish Chaudhuri and Carolyn P Rose

Incremental Adaptation of Speech-to-Speech Translation
Nguyen Bach, Roger Hsiao, Matthias Eck, Paisarn Charoenpornsawat, Stephan Vogel, Tanja Schultz, Ian Lane, Alex Waibel and Alan Black

Name Perplexity
Octavian Popescu

Answer Credibility: A Language Modeling Approach to Answer Validation
Protima Banerjee and Hyoil Han

Exploiting Named Entity Classes in CCG Surface Realization
Rajakrishnan Rajkumar, Michael White and Dominic Espinosa

Search Engine Adaptation by Feedback Control Adjustment for Time-sensitive Query
Ruiqiang zhang, yi Chang, Zhaohui Zheng, Donald Metzler and Jian-yun Nie

A Local Tree Alignment-based Soft Pattern Matching Approach for Information Extraction
Seokhwan Kim, Minwoo Jeong and Gary Geunbae Lee

Classifying Factored Genres with Part-of-Speech Histograms
Sergey Feldman, Marius Marin, Julie Medero and Mari Ostendorf

Towards Effective Sentence Simplification for Automatic Processing of Biomedical Text
Siddhartha Jonnalagadda, Luis Tari, Jrg Hakenberg, Chitta Baral and Graciela Gonzalez

Improving SCL Model for Sentiment-Transfer Learning
Songbo Tan and Xueqi Cheng

MICA: A Probabilistic Dependency Parser Based on Tree Insertion Grammars (Application Note)
Srinivas Bangalore, Pierre Boullier, Alexis Nasr, Owen Rambow and Benot Sagot

Lexical and Syntactic Adaptation and Their Impact in Deployed Spoken Dialog Systems
Svetlana Stoyanchev and Amanda Stent

Analysing Recognition Errors in Unlimited-Vocabulary Speech Recognition
Teemu Hirsimäki and Mikko Kurimo

The independence of dimensions in multidimensional dialogue act annotation
Volha Petukhova and Harry Bunt

Improving Coreference Resolution by Using Conversational Metadata
Xiaoqiang Luo, Radu Florian and Todd Ward

Using N-gram based Features for Machine Translation System Combination
Yong Zhao and Xiaodong He

Language Specific Issue and Feature Exploration in Chinese Event Extraction
Zheng Chen and Heng Ji

Improving A Simple Bigram HMM Part-of-Speech Tagger by Latent Annotation and Self-Training
Zhongqiang Huang, Vladimir Eidelman and Mary Harper

6:30–9:30 **Student Research Workshop Poster Session**

Note: all student research workshop papers are located in the Companion volume of the proceedings

Also: All papers presented in the morning and afternoon sessions of the student research workshop will also be shown as posters.

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

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

Monday, June 1, 2009 (continued)

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 Andres 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

Tuesday, June 2, 2009

Plenary Session

9:00–9:10 Paper Awards

9:10–9:40 *Unsupervised Morphological Segmentation with Log-Linear Models*
Hoifung Poon, Colin Cherry and Kristina Toutanova

9:40–10:10 *11,001 New Features for Statistical Machine Translation*
David Chiang, Kevin Knight and Wei Wang

10:10-10:40 **Break**

Session 4A: Machine Translation

10:10–10:35 *Efficient Parsing for Transducer Grammars*
John DeNero, Mohit Bansal, Adam Pauls and Dan Klein

10:35–10:50 *Preference Grammars: Softening Syntactic Constraints to Improve Statistical Machine Translation*
Ashish Venugopal, Andreas Zollmann, Noah A. Smith and Stephan Vogel

10:50–11:15 *Using a Dependency Parser to Improve SMT for Subject-Object-Verb Languages*
Peng Xu, Jaeho Kang, Michael Ringgaard and Franz Och

11:15–11:40 *Learning Bilingual Linguistic Reordering Model for Statistical Machine Translation*
Han-Bin Chen, Jian-Cheng Wu and Jason S. Chang

Session 4B: Sentiment Analysis / Information Extraction

10:10–10:35 *May All Your Wishes Come True: A Study of Wishes and How to Recognize Them*
Andrew B. Goldberg, Nathanael Fillmore, David Andrzejewski, Zhiting Xu, Bryan Gibson and Xiaojin Zhu

10:35–10:50 *Predicting Risk from Financial Reports with Regression*
Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi and Noah A. Smith

10:50–11:15 *Domain Adaptation with Latent Semantic Association for Named Entity Recognition*
Honglei Guo, Huijia Zhu, Zhili Guo, Xiaoxun Zhang, Xian Wu and Zhong Su

11:15–11:40 *Semi-Automatic Entity Set Refinement*
Vishnu Vyas and Patrick Pantel

Tuesday, June 2, 2009 (continued)

Session 4C: Machine Learning / Morphology and Phonology

10:10–10:35 *Unsupervised Constraint Driven Learning For Transliteration Discovery*
Ming-Wei Chang, Dan Goldwasser, Dan Roth and Yuancheng Tu

10:35–10:50 *On the Syllabification of Phonemes*
Susan Bartlett, Grzegorz Kondrak and Colin Cherry

10:50–11:15 *Improving nonparameteric Bayesian inference: experiments on unsupervised word segmentation with adaptor grammars*
Mark Johnson and Sharon Goldwater

11:15–11:40 No Presentation

12:20–2:00 **Lunch Break**

Session 5A: Short Paper Presentations: Machine Translation / Generation / Semantics

Note: all short papers are located in the Companion volume of the proceedings

2:00–2:15 *Statistical Post-Editing of a Rule-Based Machine Translation System*
Antonio-L. Lagarda, Vicent Alabau, Francisco Casacuberta, Roberto Silva and Enrique Daz-de-Liao

2:15–2:30 *On the Importance of Pivot Language Selection for Statistical Machine Translation*
Michael Paul, Hirofumi Yamamoto, Eiichiro Sumita and Satoshi Nakamura

2:30–2:45 *Tree Linearization in English: Improving Language Model Based Approaches*
Katja Filippova and Michael Strube

2:45–3:00 *Determining the position of adverbial phrases in English*
Huayan Zhong and Amanda Stent

3:00–3:15 *Estimating and Exploiting the Entropy of Sense Distributions*
Peng Jin, Diana McCarthy, Rob Koeling and John Carroll

3:15–3:30 *Semantic classification with WordNet Kernels*
Diarmuid Saghdha

Tuesday, June 2, 2009 (continued)

Session 5B: Short Paper Presentations: Machine Learning / Syntax

Note: all short papers are located in the Companion volume of the proceedings

2:00–2:15 *Sentence Boundary Detection and the Problem with the U.S.*
Dan Gillick

2:15–2:30 *Quadratic Features and Deep Architectures for Chunking*
Joseph Turian, James Bergstra and Yoshua Bengio

2:30–2:45 *Active Zipfian Sampling for Statistical Parser Training*
Onur Cobanoglu

2:45–3:00 *Combining Constituent Parsers*
Victoria Fossum and Kevin Knight

3:00–3:15 *Recognising the Predicate-argument Structure of Tagalog*
Meladel Mistica and Timothy Baldwin

3:15–3:30 *Reverse Revision and Linear Tree Combination for Dependency Parsing*
Giuseppe Attardi and Felice Dell'Orletta

Session 5C: Short Paper Presentations: SPECIAL SESSION – Speech Indexing and Retrieval

Note: all short papers are located in the Companion volume of the proceedings

2:00–2:15 *Introduction to the Special Session on Speech Indexing and Retrieval*

2:15–2:30 *Anchored Speech Recognition for Question Answering*
Sibel Yaman, Gokan Tur, Dimitra Vergyri, Dilek Hakkani-Tur, Mary Harper and Wen Wang

2:30–2:45 *Score Distribution Based Term Specific Thresholding for Spoken Term Detection*
Dogan Can and Murat Saraclar

2:45–3:00 *Automatic Chinese Abbreviation Generation Using Conditional Random Field*
Dong Yang, Yi-Cheng Pan and Sadaoki Furui

Tuesday, June 2, 2009 (continued)

3:00–3:15 *Fast decoding for open vocabulary spoken term detection*
Bhuvana Ramabhadran, Abhinav Sethy, Jonathan Mamou, Brian Kingsbury and Upendra Chaudhari

3:15–3:30 *Tightly coupling Speech Recognition and Search*
Taniya Mishra and Srinivas Bangalore

3:30–4:00 **Break**

Session 6A: Syntax and Parsing

4:00–4:25 *Joint Parsing and Named Entity Recognition*
Jenny Rose Finkel and Christopher D. Manning

4:25–4:50 *Minimal-length linearizations for mildly context-sensitive dependency trees*
Y. Albert Park and Roger Levy

4:50–5:15 *Positive Results for Parsing with a Bounded Stack using a Model-Based Right-Corner Transform*
William Schuler

Session 6B: Discourse and Summarization

4:00–4:25 *Hierarchical Text Segmentation from Multi-Scale Lexical Cohesion*
Jacob Eisenstein

4:25–4:50 *Exploring Content Models for Multi-Document Summarization*
Aria Haghighi and Lucy Vanderwende

4:50–5:15 *Global Models of Document Structure using Latent Permutations*
Harr Chen, S.R.K. Branavan, Regina Barzilay and David R. Karger

Tuesday, June 2, 2009 (continued)

Session 6C: Spoken Language Systems

4:00–4:25 *Assessing and Improving the Performance of Speech Recognition for Incremental Systems*
Timo Baumann, Michaela Atterer and David Schlangen

4:25–4:50 *Geo-Centric Language Models for Local Business Voice Search*
Amanda Stent, Ilija Zeljkovic, Diamantino Caseiro and Jay Wilpon

4:50–5:15 *Improving the Arabic Pronunciation Dictionary for Phone and Word Recognition with Linguistically-Based Pronunciation Rules*
Fadi Biadsy, Nizar Habash and Julia Hirschberg

Wednesday, June 3, 2009

Plenary Session

9:00–10:10 Invited Talk: *Ketchup, Espresso, and Chocolate Chip Cookies: Travels in the Language of Food*
Dan Jurafsky

10:10–10:40 **Break**

Session 7A: Machine Translation

10:40–11:05 *Using a maximum entropy model to build segmentation lattices for MT*
Chris Dyer

11:05–11:30 *Active Learning for Statistical Phrase-based Machine Translation*
Gholamreza Haffari, Maxim Roy and Anoop Sarkar

11:30–11:55 *Semi-Supervised Lexicon Mining from Parenthetical Expressions in Monolingual Web Pages*
Xianchao Wu, Naoaki Okazaki and Jun'ichi Tsujii

11:55–12:20 *Hierarchical Phrase-Based Translation with Weighted Finite State Transducers*
Gonzalo Iglesias, Adrià de Gispert, Eduardo R. Banga and William Byrne

Wednesday, June 3, 2009 (continued)

Session 7B: Speech Recognition and Language Modeling

10:40–11:05 *Improved pronunciation features for construct-driven assessment of non-native spontaneous speech*
Lei Chen, Klaus Zechner and Xiaoming Xi

11:05–11:30 *Performance Prediction for Exponential Language Models*
Stanley Chen

11:30–11:55 *Tied-Mixture Language Modeling in Continuous Space*
Ruhi Sarikaya, Mohamed Afify and Brian Kingsbury

11:55–12:20 *Shrinking Exponential Language Models*
Stanley Chen

Session 7C: Sentiment Analysis

10:40–11:05 *Predicting Response to Political Blog Posts with Topic Models*
Tae Yano, William W. Cohen and Noah A. Smith

11:05–11:30 *An Iterative Reinforcement Approach for Fine-Grained Opinion Mining*
Weifu Du and Songbo Tan

11:30–11:55 *For a few dollars less: Identifying review pages sans human labels*
Luciano Barbosa, Ravi Kumar, Bo Pang and Andrew Tomkins

11:55–12:20 *More than Words: Syntactic Packaging and Implicit Sentiment*
Stephan Greene and Philip Resnik

12:20–1:40 **Lunch Break**

12:40-1:40 Panel Discussion: *Emerging Application Areas in Computational Linguistics*
Chaired by Bill Dolan, Microsoft
Panelists: Jill Burstein, Educational Testing Service; Joel Tetreault, Educational Testing Service; Patrick Pantel, Yahoo; Andy Hickl, Language Computer Corporation + Swingly

1:40–2:30 NAACL Business Meeting

Wednesday, June 3, 2009 (continued)

Session 8A: Large-scale NLP

2:30–2:55 *Streaming for large scale NLP: Language Modeling*
Amit Goyal, Hal Daume III and Suresh Venkatasubramanian

2:55–3:20 *The Effect of Corpus Size on Case Frame Acquisition for Discourse Analysis*
Ryohei Sasano, Daisuke Kawahara and Sadao Kurohashi

3:20–3:45 *Semantic-based Estimation of Term Informativeness*
Kirill Kireyev

Session 8B: Syntax and Parsing

2:30–2:55 *Optimal Reduction of Rule Length in Linear Context-Free Rewriting Systems*
Carlos Gómez-Rodríguez, Marco Kuhlmann, Giorgio Satta and David Weir

2:55–3:20 *Inducing Compact but Accurate Tree-Substitution Grammars*
Trevor Cohn, Sharon Goldwater and Phil Blunsom

3:20–3:45 *Hierarchical Search for Parsing*
Adam Pauls and Dan Klein

Session 8C: Discourse and Summarization

2:30–2:55 *An effective Discourse Parser that uses Rich Linguistic Information*
Rajen Subba and Barbara Di Eugenio

2:55–3:20 *Graph-Cut-Based Anaphoricity Determination for Coreference Resolution*
Vincent Ng

3:20–3:45 *Using Citations to Generate surveys of Scientific Paradigms*
Saif Mohammad, Bonnie Dorr, Melissa Egan, Ahmed Hassan, Pradeep Muthukrishan, Vahed Qazvinian, Dragomir Radev and David Zajic

3:45–4:15 **Break**

Wednesday, June 3, 2009 (continued)

Session 9A: Machine Learning

4:15–4:40 *Non-Parametric Bayesian Areal Linguistics*
Hal Daume III

4:40–5:05 *Hierarchical Bayesian Domain Adaptation*
Jenny Rose Finkel and Christopher D. Manning

5:05–5:30 *Online EM for Unsupervised Models*
Percy Liang and Dan Klein

Session 9B: Dialog Systems

4:15–4:40 *Unsupervised Approaches for Automatic Keyword Extraction Using Meeting Transcripts*
Feifan Liu, Deana Pennell, Fei Liu and Yang Liu

4:40–5:05 *A Finite-State Turn-Taking Model for Spoken Dialog Systems*
Antoine Raux and Maxine Eskenazi

5:05–5:30 *Extracting Social Meaning: Identifying Interactional Style in Spoken Conversation*
Dan Jurafsky, Rajesh Ranganath and Dan McFarland

Session 9C: Syntax and Parsing

4:15–4:40 *Linear Complexity Context-Free Parsing Pipelines via Chart Constraints*
Brian Roark and Kristy Hollingshead

4:40–5:05 *Improved Syntactic Models for Parsing Speech with Repairs*
Tim Miller

5:05–5:30 *A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs*
Klinton Bicknell and Roger Levy