

NAACL HLT 2007

**Human Language Technologies 2007:
The Conference of the
North American Chapter
of the
Association for Computational Linguistics**

Proceedings of the Main Conference

Candace Sidner, General Chair
Tanja Schultz, Matthew Stone, and ChengXiang Zhai
Program Committee Chairs

22–27 April 2007
Rochester, New York, USA

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

©2007 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: +1-507-476-0860
acl@aclweb.org

We Thank Our Sponsors



TM



Preface from the General Chair

This year the annual conference organized by the North American chapter of the Association for Computational Linguistics (NAACL) has undergone a name change to NAACL HLT. This change reflects the integral part that all of Human Language Technology plays in the NAACL. It is symbolic of the focus of the conference, which is represented by the collection of submitted and accepted papers. They span our community's emphasis on speech processing, information retrieval and language processing techniques and applications.

The yearly NAACL conference is always the result of the volunteer contributions of a great many people from the NAACL community who put in many hours to make the conference possible. Most of the sub-committees of the organizing committee include researchers from the areas of language processing, speech processing and information retrieval, again reflecting the diversity of expertise and interests in the NAACL world.

Each year the general chair calls on a new group of members to serve as the organizing committee. They learn, from scratch, with advice from the previous organizing committee, the tasks needed to make the conference happen. I want to thank each of them for their hard work and good-natured spirit through this process. I thank the program chairs, Tanja Schultz, Matthew Stone, and ChengXiang Zhai; the local arrangement chairs, James Allen, Dan Gildea, and Lenhart Schubert; the tutorial and workshop chairs, James Allan, Marti Hearst, and Gina Levow; the publications chairs, Yang Liu, Ronnie Smith, and Ellen Voorhees; the sponsorship chair David Day, and exhibits chair, Tim Paek; the publicity chairs, Dilek Hakkani-Tür, Miles Osbourne, and Tomek Strzalkowski; the demos chairs, Bob Carpenter, Amanda Stent, and Jason Williams; and the doctoral consortium chairs, Jackson Liscombe, Phil Michalak, and the consortium faculty advisor, Julia Hirschberg.

In addition to the organizing committee, thanks are due to the senior program committee, all the paper reviewers, and the students who volunteered during the conference. A special thank you to our conference sponsors, whose contributions made this conference possible: the Eastman Kodak company, Microsoft Research, Powerset, Thomson, the Association For Machine Translation in the Americas, IBM, and Language Weaver.

Finally, I would also like to thank the NAACL executive committee and the Advisory Board for their advice in preparation for the conference. I especially want to thank Priscilla Rasmussen, who served as Treasurer for this year's meeting, as well as in her normal role as Business Manager for the ACL office. Her knowledge, willingness to make all the ahead-details for the conference go smoothly, and her skills as NAACL corporate memory were invaluable to this conference and its organizers.

Candace L. Sidner
BAE Systems AIT
General Chair, NAACL HLT 2007

Preface from the Program Co-Chairs

We welcome you to NAACL HLT 2007, Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics. NAACL HLT 2007 continues to attract high quality submissions across three broad topic areas – natural language processing, information retrieval, and speech recognition. This year, 298 full papers were submitted and 72 accepted (24% acceptance rate); 150 late-breaking (short) papers were submitted and 55 accepted (37% acceptance rate). The numbers of submissions for both full and short papers continue to grow compared to those of last year.

Reviewing of the submissions was double blind and was handled using a two tiered reviewing system. The PC Chairs selected 30 internationally recognized experts as senior program committee (SPC) members. Each SPC member then selected a group of experts in specific areas to review both the full and short submitted papers. The complete PC numbered around 340. Three (or two in the case of short papers) reviewers and one SPC member were assigned per paper. The SPC oversaw the reviewing process, helped resolve any disputes, and at the end produced, for each paper, an overview of the reviewers' comments along with a preliminary acceptance decision. The final decisions were made by the program chairs based on online discussions among the SPC members.

Three award papers were chosen by the program chairs based on reviews, recommendations, the papers themselves, and our sense that the research efforts epitomize the interactions and opportunities across HLT that the conference aims to foster. The award for the best paper goes to: "Combining Outputs from Multiple Machine Translation Systems" by Antti-Veikko Rosti, Bing Xiang, Spyros Matsoukas, Richard Schwartz, Necip Fazil Ayan and Bonnie Dorr. The award for the best student paper goes to: "Global, Joint Determination of Anaphoricity and Coreference Resolution using Integer Programming", by Pascal Denis and Jason Baldrige. The award for the best late-breaking news paper goes to "Exploring Affect-Context Dependencies for Adaptive System Development" by Kate Forbes-Riley, Mihai Rotaru, Diane Litman and Joel Tetreault. Congratulations to all the authors!

Reflecting its multi-disciplinary nature, the NAACL HLT 2007 Program consists of oral/poster presentations of full and short papers and software demonstrations that cover a broad spectrum of topics in natural language processing, information retrieval, and speech recognition. We are honored to have two prominent keynote speakers, Franz Josef Och (Google Inc.) and Luis von Ahn (Carnegie Mellon University) for what will undoubtedly be thought-provoking and enjoyable keynote talks. In addition, the program also features a special panel on high impact future research directions for HLT (thanks to Donna Harman).

We are indebted to all the authors who submitted papers to the conference and all those who helped us put together the conference program, especially all the reviewers and SPC members who volunteered their time and worked many long hours reviewing and, later, discussing the submissions. We are also grateful to our General Conference Chair Candy Sidner and Chair of the NAACL Board Owen Rambow for their advice and support, and to Rich Gerber for his help with using the START reviewing system. The NAACL HLT conference has a PC chair for each of its three disciplines. Although work tasks were shared between the three chairs equally, as natural language processing received by far the greatest number of submissions, Matthew Stone ended up having to oversee many more papers and recruit many more SPC members than the other two chairs. He has also taken the primary responsibility of

managing the review process and coordinating our effort. Therefore, the two other chairs of NAACL HLT 2007 (Tanja Schultz & ChengXiang Zhai), wish to thank Matthew for all of his additional work in pulling this conference together.

Once again, we welcome you to NAACL HLT 2007 and hope that you enjoy the conference!

Tanja Schultz, Carnegie Mellon University
Matthew Stone, Rutgers University
ChengXiang Zhai, University of Illinois at Urbana-Champaign

Conference Organizers

General Chair:

Candace Sidner, BAE Systems Inc-AIT

Local Arrangements:

James Allen, University of Rochester
Daniel Gildea, University of Rochester
Lenhart Schubert, University of Rochester

Program Committee Chairs:

Tanja Schultz, Carnegie Mellon University
Matthew Stone, Rutgers University
ChengXiang Zhai, University of Illinois at Urbana-Champaign

Publicity Chairs:

Dilek Hakkani-Tür, ICSI
Miles Osbourne, Edinburgh University
Tomek Strzalkowski, SUNY Albany

Workshops and Tutorials Chairs:

James Allan, University of Massachusetts
Marti Hearst, UC Berkeley
Gina Levow, University of Chicago

Demonstrations Chairs:

Bob Carpenter, Alias I
Amanda Stent, SUNY Stony Brook
Jason Williams, AT&T

Exhibits Chair:

Tim Paek, Microsoft

Sponsorship Chair:

David Day, MITRE

Publications:

Yang Liu, UT Dallas
Ronnie Smith, East Carolina University
Ellen Voorhees, NIST

Doctoral Consortium:

Julia Hirschberg, Columbia University (Faculty Advisor)

Jackson Liscombe, Columbia University

Phil Michalak, University of Rochester

Treasurer:

Priscilla Rasmussen, ACL

NAACL Chair:

Owen Rambow, Columbia University

Senior Program Members:

Emily Bender, University of Washington

Alan Black, Carnegie Mellon University

Andrei Broder, Yahoo! Inc.

Donna Byron, Ohio State University

Jamie Callan, Carnegie Mellon University

Li Deng, Microsoft Research

Oren Etzioni, University of Washington

Jianfeng Gao, Microsoft Research

Nizar Habash, Columbia University

Sanda Harabagiu, University of Texas, Dallas

Julia Hockenmaier, University of Pennsylvania

Rebecca Hwa, University of Pittsburgh

Dietrich Klakow, Saarland University, Germany

Dan Klein, University of California, Berkeley

Philipp Koehn, University of Edinburgh

Alexander Koller, Columbia University

Liz Liddy, Syracuse University

Dekang Lin, Google Inc.

Ryan McDonald, Google Inc.

Ani Nenkova, Stanford University

Jian-yun Nie, University of Montreal

Patrick Pantel, University of Southern California

John Prager, IBM TJ Watson Research Center

Philip Resnik, University of Maryland, College Park

Eric Ringger, Brigham Young University

Brian Roark, Oregon Graduate Institute of Science and Technology

Suzanne Stevenson, University of Toronto

David Traum, University of Southern California

Gokhan Tur, SRI International

Michael White, Ohio State University

Program Committee Members:

Jeff Adams	Eugene Agichtein	Gregory Aist
Yaser Al-Onaizan	Rie Ando	Galen Andrew
Doug Arnold	Necip Fazil Ayan	Saliha Azzam
Michiel Bacchiani	Jason Baldridge	Timothy Baldwin
Srinivas Bangalore	Michele Banko	Regina Barzilay
Frederic Bechet	Linda Bell	Robert Belvin
Paul Bennett	Shane Bergsma	Nicola Bertoldi
Daniel Bikel	John Blitzer	Dan Bohus
Gary Borchardt	Johan Bos	Chris Brew
Eric Brown	Ralf Brown	Sabine Buchholz
Miriam Butt	Aoife Cahill	Ellen Campana
Yunbo Cao	Jaime Carbonell	Violetta Cavalli-Sforza
Joyce Chai	Jason Chang	Eugene Charniak
Ciprian Chelba	Hsin-Hsi Chen	Jiangping Chen
Colin Cherry	David Chiang	Timothy Chklovski
Ken Church	Alexander Clark	Robert Clark
Stephen Clark	William Cohen	Trevor Cohn
Michael Collins	Alistair Conkie	Paul Cook
Ann Copestake	Mark Core	Richard Crouch
Berthold Crysmann	Silviu-Petru Cucerzan	Aron Culotta
James Curran	Robert Dale	Hoa Dang
Hal Daume	Renato De Mori	Nikhil Dinesh
Christy Doran	John Dowding	Amit Dubey
Kevin Duh	Jakob Elming	Katrin Erk
David Farwell	Afsaneh Fazly	Junlan Feng
Dan Flickinger	Martin Forst	Mary Ellen Foster
George Foster	Alexander Fraser	Robert Frederking
Andrew Freeman	Dayne Freitag	Mark Fuhs
Pascale Fung	Sadaoki Furui	Michel Galley
Kallirroi Georgila	Roxana Girju	Gregory Grefenstette
Michelle Gregory	Teg Grenager	Stephanie Haas
Dilek Hakkani-Tur	Keith Hall	Jeff Hancock
Donna Harman	Mary Harper	Anthony Hartley
Xiaodong He	James Henderson	Andrew Hickl
Graeme Hirst	Liang Huang	Chu-Ren Huang
Diana Inkpen	Kentaro Inui	Pierre Isabelle
Martin Jansche	Mark Johnson	Kristiina Jokinen
Gareth Jones	Rosie Jones	Joemon Jose
Aravind Joshi	Stan Jou	Jeremy Kahn
Nanda Kambhatla	Min Yen Kan	Lauri Karttunen
Boris Katz	John Kelleher	Frank Keller
Sanjeev Khudanpur	Simon King	Tracy King

Program Committee Members (continued):

Katrin Kirchoff	Esther Klabbers	Alexandre Klementiev
Kevin Knight	Greg Kondrak	Valia Kordoni
Anna Korhonen	Sandra Kuebler	Jonas Kuhn
Shankar Kumar	Mikko Kurimo	KL Kwok
Namhee Kwon	Mounia Lalmas	Ian Lane
Philippe Langlais	Guy Lapalme	Mirella Lapata
Ray Larson	Staffan Larsson	Alex Lascarides
Alon Lavie	Matthew Lease	Gary Lee
Jong-Hyeok Lee	Lillian Lee	Oliver Lemon
Gina Levow	Roger Levy	Hang Li
Xiaolong Li	Chin-Yew Lin	Jimmy Lin
Yang Liu	Deryle Lonsdale	Adam Lopez
Jie Lu	Xiaoqiang Luo	Dick Lyon
Bill MacCartney	Elliott Macklovitch	Bente Maegaard
Bernardo Magnini	Rob Malouf	Gideon Mann
Chris Manning	Daniel Marcu	Evgeny Matusov
Nicolas Maudet	Andrew McCallum	Diana McCarthy
Kathleen McKeown	Paul McNamee	Dan Melamed
Arul Menezes	Donald Metzler	Rada Mihalcea
Teruko Mitamura	Yusuke Miyao	Mehryar Mohri
Dan Moldovan	Christof Monz	Alessandro Moschitti
Srini Narayanan	Mark-Jan Nederhof	Hwee Tou Ng
Vincent Ng	Sergei Nirenburg	Joakim Nivre
Stephan Oepen	Iadh Ounis	Martha Palmer
Bo Pang	Rebecca Passoneau	Fuchun Peng
Gerald Penn	Marco Pennacchiotti	Paul Piwek
Sameer Pradhan	John Prange	Partha Pratim Talukdar
Stephen Pulman	Chris Quirk	Dragomir Radev
Owen Rambow	Deepak Ravichandran	Florence Reeder
David Reitter	Giuseppe Riccardi	Stephen Richardson
Korin Richmond	Sebastian Riedel	Ellen Riloff
Jin Rong	Salim Roukos	Fatiha Sadat
Kenji Sagae	Mark Sanderson	Murat Saraclar
Ruhi Sarikaya	Thomas Schaaf	Charles Schafer
Helmut Schmid	Hinrich Schuetze	Sabine Schulte im Walde
Satoshi Sekine	Zak Shafran	James Shaw
Libin Shen	Chilin Shih	Luo Si
Malcolm Slaney	Noah Smith	Ronnie Smith
David Smith	Rion Snow	Stephen Soderland
Thamar Solorio	Harold Somers	Dawei Song
Radu Soricut	Richard Sproat	Mark Steedman
Amanda Stent	Laura Stoia	Nicola Stokes
Michael Strube	Tomek Strzalkowski	Jian Su
Mihai Surdeanu	Charles Sutton	Ann Syrdal

Program Committee Members (continued):

Joel Tetreault	Simone Teufel	Stefan Thater
Ye Tian	Christoph Tillmann	Tomoki Toda
Takenobu Tokunaga	Elaine Toms	Arthur Toth
Kristina Toutanova	Peter Turney	Andrew Turpin
Nicola Ueffing	Kees Van Deemter	Lucy Vanderwende
Sebastian Varges	Renata Vieira	David Vilar
Stephan Vogel	Ellen Voorhees	Ye-Yi Wang
Karen Ward	Taro Watanabe	Andy Way
Bonnie Webber	Eric Wehrli	Ben Wellington
Fuliang Weng	Martin Westphal	Ed Whittaker
Janyce Wiebe	Jason Williams	Shuly Wintner
Dekai Wu	Fei Xia	David Yarowsky
Alex Yates	Chen Yu	Hugo Zaragoza
Daniel Zeman	Richard Zens	Jian Zhang
Joy Zhang	Yi Zhang	Bing Zhao
Liang Zhou	Ming Zhou	Imed Zitouni

Table of Contents

<i>Exploiting Acoustic and Syntactic Features for Prosody Labeling in a Maximum Entropy Framework</i> Vivek Kumar Rangarajan Sridhar, Srinivas Bangalore and Shrikanth Narayanan	1
<i>To Memorize or to Predict: Prominence labeling in Conversational Speech</i> Ani Nenkova, Jason Brenier, Anubha Kothari, Sasha Calhoun, Laura Whitton, David Beaver and Dan Jurafsky	9
<i>Avoiding and Resolving Initiative Conflicts in Dialogue</i> Fan Yang and Peter A. Heeman	17
<i>What Decisions Have You Made?: Automatic Decision Detection in Meeting Conversations</i> Pei-yun Hsueh and Johanna D. Moore	25
<i>Automatic Evaluation of Machine Translation Based on Rate of Accomplishment of Sub-Goals</i> Kiyotaka Uchimoto, Katsunori Kotani, Yujie Zhang and Hitoshi Isahara	33
<i>Source-Language Features and Maximum Correlation Training for Machine Translation Evaluation</i> Ding Liu and Daniel Gildea	41
<i>Generating Case Markers in Machine Translation</i> Kristina Toutanova and Hisami Suzuki	49
<i>Direct Translation Model 2</i> Abraham Ittycheriah and Salim Roukos	57
<i>Structured Local Training and Biased Potential Functions for Conditional Random Fields with Application to Coreference Resolution</i> Yejin Choi and Claire Cardie	65
<i>Coreference or Not: A Twin Model for Coreference Resolution</i> Xiaoqiang Luo	73
<i>First-Order Probabilistic Models for Coreference Resolution</i> Aron Culotta, Michael Wick and Andrew McCallum	81
<i>Information Retrieval On Empty Fields</i> Victor Lavrenko, Xing Yi and James Allan	89
<i>Improving Diversity in Ranking using Absorbing Random Walks</i> Xiaojin Zhu, Andrew Goldberg, Jurgen Van Gael and David Andrzejewski	97
<i>A Random Text Model for the Generation of Statistical Language Invariants</i> Chris Biemann	105
<i>A Systematic Exploration of the Feature Space for Relation Extraction</i> Jing Jiang and ChengXiang Zhai	113

<i>Unsupervised Resolution of Objects and Relations on the Web</i>	
Alexander Yates and Oren Etzioni	121
<i>The Domain Restriction Hypothesis: Relating Term Similarity and Semantic Consistency</i>	
Alfio Massimiliano Gliozzo, Marco Pennacchiotti and Patrick Pantel	131
<i>Bayesian Inference for PCFGs via Markov Chain Monte Carlo</i>	
Mark Johnson, Thomas Griffiths and Sharon Goldwater	139
<i>Worst-Case Synchronous Grammar Rules</i>	
Daniel Gildea and Daniel Stefankovic	147
<i>High-Performance, Language-Independent Morphological Segmentation</i>	
Sajib Dasgupta and Vincent Ng	155
<i>Probabilistic Generation of Weather Forecast Texts</i>	
Anja Belz	164
<i>Generation by Inverting a Semantic Parser that Uses Statistical Machine Translation</i>	
Yuk Wah Wong and Raymond Mooney	172
<i>Lexicalized Markov Grammars for Sentence Compression</i>	
Michel Galley and Kathleen McKeown	180
<i>Hybrid Models for Semantic Classification of Chinese Unknown Words</i>	
Xiaofei Lu	188
<i>Using Wikipedia for Automatic Word Sense Disambiguation</i>	
Rada Mihalcea	196
<i>Data-Driven Graph Construction for Semi-Supervised Graph-Based Learning in NLP</i>	
Andrei Alexandrescu and Katrin Kirchhoff	204
<i>Is Question Answering Better than Information Retrieval? Towards a Task-Based Evaluation Framework for Question Series</i>	
Jimmy Lin	212
<i>A Case For Shorter Queries, and Helping Users Create Them</i>	
Giridhar Kumaran and James Allan	220
<i>Combining Outputs from Multiple Machine Translation Systems</i>	
Antti-Veikko Rosti, Necip Fazil Ayan, Bing Xiang, Spyros Matsoukas, Richard Schwartz and Bonnie Dorr	228
<i>Joint Determination of Anaphoricity and Coreference Resolution using Integer Programming</i>	
Pascal Denis and Jason Baldridge	236
<i>Automating Creation of Hierarchical Faceted Metadata Structures</i>	
Emilia Stoica, Marti Hearst and Megan Richardson	244

<i>Cross-Instance Tuning of Unsupervised Document Clustering Algorithms</i>	
Damianos Karakos, Jason Eisner, Sanjeev Khudanpur and Carey Priebe	252
<i>Using “Annotator Rationales” to Improve Machine Learning for Text Categorization</i>	
Omar Zaidan, Jason Eisner and Christine Piatko	260
<i>Combining Reinforcement Learning with Information-State Update Rules</i>	
Peter Heeman	268
<i>Estimating the Reliability of MDP Policies: a Confidence Interval Approach</i>	
Joel Tetreault, Dan Bohus and Diane Litman	276
<i>An Exploration of Eye Gaze in Spoken Language Processing for Multimodal Conversational Interfaces</i>	
Shaolin Qu and Joyce Chai	284
<i>Extracting Semantic Orientations of Phrases from Dictionary</i>	
Hiroya Takamura, Takashi Inui and Manabu Okumura	292
<i>Multiple Aspect Ranking Using the Good Grief Algorithm</i>	
Benjamin Snyder and Regina Barzilay	300
<i>Extracting Appraisal Expressions</i>	
Kenneth Bloom, Navendu Garg and Shlomo Argamon	308
<i>Whose Idea Was This, and Why Does it Matter? Attributing Scientific Work to Citations</i>	
Advait Siddharthan and Simone Teufel	316
<i>Combining Probability-Based Rankers for Action-Item Detection</i>	
Paul N. Bennett and Jaime G. Carbonell	324
<i>Multi-Document Relationship Fusion via Constraints on Probabilistic Databases</i>	
Gideon Mann	332
<i>An Integrated Approach to Measuring Semantic Similarity between Words Using Information Available on the Web</i>	
Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka	340
<i>An Information Retrieval Approach to Sense Ranking</i>	
Mirella Lapata and Frank Keller	348
<i>Near-Synonym Choice in an Intelligent Thesaurus</i>	
Diana Inkpen	356
<i>A Log-Linear Block Transliteration Model based on Bi-Stream HMMs</i>	
Bing Zhao, Nguyen Bach, Ian Lane and Stephan Vogel	364
<i>Applying Many-to-Many Alignments and Hidden Markov Models to Letter-to-Phoneme Conversion</i>	
Sittichai Jiampojarn, Grzegorz Kondrak and Tarek Sherif	372

<i>Analysis of Morph-Based Speech Recognition and the Modeling of Out-of-Vocabulary Words Across Languages</i>	
Mathias Creutz, Teemu Hirsimäki, Mikko Kurimo, Antti Puurula, Janne Pylkkänen, Vesa Siivola, Matti Varjokallio, Ebru Arisoy, Murat Saraclar and Andreas Stolcke	380
<i>Tree Revision Learning for Dependency Parsing</i>	
Giuseppe Attardi and Massimiliano Ciaramita	388
<i>Incremental Non-Projective Dependency Parsing</i>	
Joakim Nivre	396
<i>Improved Inference for Unlexicalized Parsing</i>	
Slav Petrov and Dan Klein	404
<i>Approximate Factoring for A* Search</i>	
Aria Haghighi, John DeNero and Dan Klein	412
<i>A Cascaded Machine Learning Approach to Interpreting Temporal Expressions</i>	
David Ahn, Joris van Rantwijk and Maarten de Rijke	420
<i>Building and Refining Rhetorical-Semantic Relation Models</i>	
Sasha Blair-Goldensohn, Kathleen McKeown and Owen Rambow	428
<i>A Unified Local and Global Model for Discourse Coherence</i>	
Micha Elsner, Joseph Austerweil and Eugene Charniak	436
<i>Randomized Decoding for Selection-and-Ordering Problems</i>	
Pawan Deshpande, Regina Barzilay and David Karger	444
<i>Multilingual Structural Projection across Interlinear Text</i>	
Fei Xia and William Lewis	452
<i>Combining Lexical and Grammatical Features to Improve Readability Measures for First and Second Language Texts</i>	
Michael Heilman, Kevyn Collins-Thompson, Jamie Callan and Maxine Eskenazi	460
<i>Automatic Assessment of Student Translations for Foreign Language Tutoring</i>	
Chao Wang and Stephanie Seneff	468
<i>Automatic and Human Scoring of Word Definition Responses</i>	
Kevyn Collins-Thompson and Jamie Callan	476
<i>A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation</i>	
Masao Utiyama and Hitoshi Isahara	484
<i>Efficient Phrase-Table Representation for Machine Translation with Applications to Online MT and Speech Translation</i>	
Richard Zens and Hermann Ney	492

<i>An Efficient Two-Pass Approach to Synchronous-CFG Driven Statistical MT</i>	
Ashish Venugopal, Andreas Zollmann and Vogel Stephan	500
<i>Statistical Phrase-Based Post-Editing</i>	
Michel Simard, Cyril Goutte and Pierre Isabelle	508
<i>Automatic Answer Typing for How-Questions</i>	
Christopher Pinchak and Shane Bergsma	516
<i>A Probabilistic Framework for Answer Selection in Question Answering</i>	
Jeongwoo Ko, Luo Si and Eric Nyberg	524
<i>Question Answering Using Integrated Information Retrieval and Information Extraction</i>	
Barry Schiffman, Kathleen McKeown, Ralph Grishman and James Allan	532
<i>Toward Multimedia: A String Pattern-Based Passage Ranking Model for Video Question Answering</i>	
Yu-Chieh Wu and Jie-Chi Yang	540
<i>Can Semantic Roles Generalize Across Genres?</i>	
Szu-ting Yi, Edward Loper and Martha Palmer	548
<i>Towards Robust Semantic Role Labeling</i>	
Sameer Pradhan, Wayne Ward and James Martin	556
<i>ISP: Learning Inferential Selectional Preferences</i>	
Patrick Pantel, Rahul Bhagat, Bonaventura Coppola, Timothy Chklovski and Eduard Hovy ..	564
<i>Computing Semantic Similarity between Skill Statements for Approximate Matching</i>	
Feng Pan and Robert Farrell	572

Conference Program

Monday, April 23, 2007

9:00-9:10 Opening Session

9:10-10:10 Invited talk by Franz Josef Och, Google

10:10-10:40 Break

10:40-12:20 Paper Sessions

Conversational Speech

10:40-11:05 *Exploiting Acoustic and Syntactic Features for Prosody Labeling in a Maximum Entropy Framework*
Vivek Kumar Rangarajan Sridhar, Srinivas Bangalore and Shrikanth Narayanan

11:05-11:30 *To Memorize or to Predict: Prominence labeling in Conversational Speech*
Ani Nenkova, Jason Brenier, Anubha Kothari, Sasha Calhoun, Laura Whitton, David Beaver and Dan Jurafsky

11:30-11:55 *Avoiding and Resolving Initiative Conflicts in Dialogue*
Fan Yang and Peter A. Heeman

11:55-12:20 *What Decisions Have You Made?: Automatic Decision Detection in Meeting Conversations*
Pei-yun Hsueh and Johanna D. Moore

Machine Translation 1

10:40-11:05 *Automatic Evaluation of Machine Translation Based on Rate of Accomplishment of Sub-Goals*
Kiyotaka Uchimoto, Katsunori Kotani, Yujie Zhang and Hitoshi Isahara

11:05-11:30 *Source-Language Features and Maximum Correlation Training for Machine Translation Evaluation*
Ding Liu and Daniel Gildea

11:30-11:55 *Generating Case Markers in Machine Translation*
Kristina Toutanova and Hisami Suzuki

11:55-12:20 *Direct Translation Model 2*
Abraham Ittycheriah and Salim Roukos

Monday, April 23, 2007 (continued)

Coreference

- 10:40-11:05 *Structured Local Training and Biased Potential Functions for Conditional Random Fields with Application to Coreference Resolution*
Yejin Choi and Claire Cardie
- 11:05-11:30 *Coreference or Not: A Twin Model for Coreference Resolution*
Xiaoqiang Luo
- 11:30-11:55 *First-Order Probabilistic Models for Coreference Resolution*
Aron Culotta, Michael Wick and Andrew McCallum

12:20-2:00 Lunch

2:00-3:15 Paper Sessions

Information Retrieval Models

- 2:00-2:25 *Information Retrieval On Empty Fields*
Victor Lavrenko, Xing Yi and James Allan
- 2:25-2:50 *Improving Diversity in Ranking using Absorbing Random Walks*
Xiaojin Zhu, Andrew Goldberg, Jurgen Van Gael and David Andrzejewski
- 2:50-3:15 *A Random Text Model for the Generation of Statistical Language Invariants*
Chris Biemann

Information Extraction 1

- 2:00-2:25 *A Systematic Exploration of the Feature Space for Relation Extraction*
Jing Jiang and ChengXiang Zhai
- 2:25-2:50 *Unsupervised Resolution of Objects and Relations on the Web*
Alexander Yates and Oren Etzioni
- 2:50-3:15 *The Domain Restriction Hypothesis: Relating Term Similarity and Semantic Consistency*
Alfio Massimiliano GIoZZo, Marco Pennacchiotti and Patrick Pantel

Monday, April 23, 2007 (continued)

Grammatical Inference

- 2:00-2:25 *Bayesian Inference for PCFGs via Markov Chain Monte Carlo*
Mark Johnson, Thomas Griffiths and Sharon Goldwater
- 2:25-2:50 *Worst-Case Synchronous Grammar Rules*
Daniel Gildea and Daniel Stefankovic
- 2:50-3:15 *High-Performance, Language-Independent Morphological Segmentation*
Sajib Dasgupta and Vincent Ng
- 3:15-3:45 Break
- 3:45-5:05 Paper Sessions

Generation

- 3:45-4:10 *Probabilistic Generation of Weather Forecast Texts*
Anja Belz
- 4:10-4:35 *Generation by Inverting a Semantic Parser that Uses Statistical Machine Translation*
Yuk Wah Wong and Raymond Mooney
- 4:35-5:00 *Lexicalized Markov Grammars for Sentence Compression*
Michel Galley and Kathleen McKeown

Word Senses

- 3:45-4:10 *Hybrid Models for Semantic Classification of Chinese Unknown Words*
Xiaofei Lu
- 4:10-4:35 *Using Wikipedia for Automatic Word Sense Disambiguation*
Rada Mihalcea
- 4:35-5:00 *Data-Driven Graph Construction for Semi-Supervised Graph-Based Learning in NLP*
Andrei Alexandrescu and Katrin Kirchhoff

Monday, April 23, 2007 (continued)

Frontiers of Information Retrieval

- 3:45-4:10 *Is Question Answering Better than Information Retrieval? Towards a Task-Based Evaluation Framework for Question Series*
Jimmy Lin
- 4:10-4:35 *A Case For Shorter Queries, and Helping Users Create Them*
Giridhar Kumaran and James Allan
- 4:35-4:50 *Situated Models of Meaning for Sports Video Retrieval*
Michael Fleischman and Deb Roy
- 4:50-5:05 *Speech Summarization Without Lexical Features for Mandarin Broadcast News*
Jian Zhang and Pascale Fung
- 5:30-8:00 Posters and Demos

Tuesday, April 24, 2007

- 9:00-10:40 Plenary Session
- 9:00-9:05 Award Presentations
- 9:05-9:30 *Combining Outputs from Multiple Machine Translation Systems*
Antti-Veikko Rosti, Necip Fazil Ayan, Bing Xiang, Spyros Matsoukas, Richard Schwartz and Bonnie Dorr
- 9:30-9:55 *Joint Determination of Anaphoricity and Coreference Resolution using Integer Programming*
Pascal Denis and Jason Baldridge
- 9:55-10:10 *Exploring Affect-Context Dependencies for Adaptive System Development*
Kate Forbes-Riley, Mihai Rotaru, Diane Litman and Joel Tetreault
- 10:10-10:25 *Demonstration of PLOW: A Dialogue System for One-Shot Task Learning*
James Allen, Nathanael Chambers, George Ferguson, Lucian Galescu, Hyuckchul Jung, Mary Swift and William Taysom
- 10:25-10:40 *Spoken Dialogue Systems for Language Learning*
Stephanie Seneff, Chao Wang and Chih-yu Chao
- 10:40-11:10 Break

11:10-12:25 Paper sessions

Text Classification

11:10-11:35 *Automating Creation of Hierarchical Faceted Metadata Structures*
Emilia Stoica, Marti Hearst and Megan Richardson

11:35-12:00 *Cross-Instance Tuning of Unsupervised Document Clustering Algorithms*
Damianos Karakos, Jason Eisner, Sanjeev Khudanpur and Carey Priebe

12:00-12:25 *Using “Annotator Rationales” to Improve Machine Learning for Text Categorization*
Omar Zaidan, Jason Eisner and Christine Piatko

Conversational Systems

11:10-11:35 *Combining Reinforcement Learning with Information-State Update Rules*
Peter Heeman

11:35-12:00 *Estimating the Reliability of MDP Policies: a Confidence Interval Approach*
Joel Tetreault, Dan Bohus and Diane Litman

12:00-12:25 *An Exploration of Eye Gaze in Spoken Language Processing for Multimodal Conversational Interfaces*
Shaolin Qu and Joyce Chai

Extracting Sentiment

11:10-11:35 *Extracting Semantic Orientations of Phrases from Dictionary*
Hiroya Takamura, Takashi Inui and Manabu Okumura

11:35-12:00 *Multiple Aspect Ranking Using the Good Grief Algorithm*
Benjamin Snyder and Regina Barzilay

12:00-12:25 *Extracting Appraisal Expressions*
Kenneth Bloom, Navendu Garg and Shlomo Argamon

12:25-2:00 Lunch

Tuesday, April 24, 2007 (continued)

2:00-4:00 Short Paper Sessions

Translation and Tagging

- 2:00-2:15 *Arabic Diacritization through Full Morphological Tagging*
Nizar Habash and Owen Rambow
- 2:15-2:30 *Tagging Icelandic Text Using a Linguistic and a Statistical Tagger*
Hrafn Loftsson
- 2:30-2:45 *Language Modeling for Determiner Selection*
Jenine Turner and Eugene Charniak
- 2:45-3:00 *Joint Morphological-Lexical Language Modeling for Machine Translation*
Ruhi Sarikaya and Yonggang Deng
- 3:00-3:15 *Discriminative Alignment Training without Annotated Data for Machine Translation*
Patrik Lambert, Rafael E. Banchs and Josep M. Crego
- 3:15-3:30 *Generalized Graphical Abstractions for Statistical Machine Translation*
Karim Filali and Jeff Bilmes
- 3:30-3:45 *Kernel Regression Based Machine Translation*
Zhuoran Wang, John Shawe-Taylor and Sandor Szedmak
- 3:45-4:00 *Are Very Large N-Best Lists Useful for SMT?*
Saša Hasan, Richard Zens and Hermann Ney

Language Understanding

- 2:00-2:15 *A Three-Step Deterministic Parser for Chinese Dependency Parsing*
Kun Yu, Sadao Kurohashi and Hao Liu
- 2:15-2:30 *RH: A Retro-Hybrid Parser*
Paula Newman
- 2:30-2:45 *Semi-Supervised Learning for Semantic Parsing Using Support Vector Machines*
Rohit Kate and Raymond Mooney
- 2:45-3:00 *Subtree Mining for Relation Extraction from Wikipedia*
Dat P.T. Nguyen, Yutaka Matsuo and Mitsuru Ishizuka

Tuesday, April 24, 2007 (continued)

3:00-3:15 *Simultaneous Identification of Biomedical Named-Entity and Functional Relation Using Statistical Parsing Techniques*
Zhongmin Shi, Anoop Sarkar and Fred Popowich

3:15-3:30 *Chinese Named Entity Recognition with Cascaded Hybrid Model*
Xiaofeng Yu

3:30-3:45 *Entity Extraction is a Boring Solved Problem—Or is it?*
Marc Vilain, Jennifer Su and Suzi Lubar

3:45-4:00 *A Semi-Automatic Evaluation Scheme: Automated Nuggetization for Manual Annotation*
Liang Zhou, Namhee Kwon and Eduard Hovy

Speech and IR

2:00-2:15 *Joint versus Independent Phonological Feature Models Within CRF Phone Recognition*
Ilana Bromberg, Jeremy Morris and Eric Fosler-Lussier

2:15-2:30 *On Using Articulatory Features for Discriminative Speaker Adaptation*
Florian Metze

2:30-2:45 *Reversible Sound-to-Letter/Letter-to-Sound Modeling Based on Syllable Structure*
Stephanie Seneff

2:45-3:00 *iROVER: Improving System Combination with Classification*
Dustin Hillard, Bjoern Hoffmeister, Mari Ostendorf, Ralf Schlueter and Hermann Ney

3:00-3:15 *Advances in the CMU/Interact Arabic GALE Transcription System*
Mohamed Noamany, Thomas Schaaf and Tanja Schultz

3:15-3:30 *Are Some Speech Recognition Errors Easier to Detect than Others?*
Yongmei Shi and Lina Zhou

3:30-3:45 *Agenda-Based User Simulation for Bootstrapping a POMDP Dialogue System*
Jost Schatzmann, Blaise Thomson, Karl Weilhammer, Hui Ye and Steve Young

3:45-4:00 *Document Similarity Measures to Distinguish Native vs. Non-Native Essay Writers*
Olga Gurevich and Paul Deane

4:00-4:30 Break

Tuesday, April 24, 2007 (continued)

4:30-5:45 Paper Sessions

Information Extraction 2

4:30-4:55 *Whose Idea Was This, and Why Does it Matter? Attributing Scientific Work to Citations*
Advaith Siddharthan and Simone Teufel

4:55-5:20 *Combining Probability-Based Rankers for Action-Item Detection*
Paul N. Bennett and Jaime G. Carbonell

5:20-5:45 *Multi-Document Relationship Fusion via Constraints on Probabilistic Databases*
Gideon Mann

Words and Similarity

4:30-4:55 *An Integrated Approach to Measuring Semantic Similarity between Words Using Information Available on the Web*
Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka

4:55-5:20 *An Information Retrieval Approach to Sense Ranking*
Mirella Lapata and Frank Keller

5:20-5:45 *Near-Synonym Choice in an Intelligent Thesaurus*
Diana Inkpen

Letters and Sounds

4:30-4:55 *A Log-Linear Block Transliteration Model based on Bi-Stream HMMs*
Bing Zhao, Nguyen Bach, Ian Lane and Stephan Vogel

4:55-5:20 *Applying Many-to-Many Alignments and Hidden Markov Models to Letter-to-Phoneme Conversion*
Sittichai Jiampojarn, Grzegorz Kondrak and Tarek Sherif

5:20-5:45 *Analysis of Morph-Based Speech Recognition and the Modeling of Out-of-Vocabulary Words Across Languages*
Mathias Creutz, Teemu Hirsimäki, Mikko Kurimo, Antti Puurula, Janne Pylkkänen, Vesa Siivola, Matti Varjokallio, Ebru Arisoy, Murat Saraclar and Andreas Stolcke

7:00 Banquet

Wednesday, April 25, 2007

- 9:00-10:00 Invited talk by Luis von Ahn, Carnegie Mellon University
- 10:00-10:30 Break
- 10:30-11:15 Panel: MINDS: High Impact Future Research Directions for HLT
- 11:15-12:00 NAACL Business Meeting
- 12:00-1:45 Lunch
- 1:45-3:25 Paper sessions

Parsing

- 1:45-2:10 *Tree Revision Learning for Dependency Parsing*
Giuseppe Attardi and Massimiliano Ciaramita
- 2:10-2:35 *Incremental Non-Projective Dependency Parsing*
Joakim Nivre
- 2:35-3:00 *Improved Inference for Unlexicalized Parsing*
Slav Petrov and Dan Klein
- 3:00-3:25 *Approximate Factoring for A* Search*
Aria Haghighi, John DeNero and Dan Klein

Semantics and Discourse

- 1:45-2:10 *A Cascaded Machine Learning Approach to Interpreting Temporal Expressions*
David Ahn, Joris van Rantwijk and Maarten de Rijke
- 2:10-2:35 *Building and Refining Rhetorical-Semantic Relation Models*
Sasha Blair-Goldensohn, Kathleen McKeown and Owen Rambow
- 2:35-3:00 *A Unified Local and Global Model for Discourse Coherence*
Micha Elsner, Joseph Austerweil and Eugene Charniak
- 3:00-3:25 *Randomized Decoding for Selection-and-Ordering Problems*
Pawan Deshpande, Regina Barzilay and David Karger

Wednesday, April 25, 2007 (continued)

Applications

- 1:45-2:10 *Multilingual Structural Projection across Interlinear Text*
Fei Xia and William Lewis
- 2:10-2:35 *Combining Lexical and Grammatical Features to Improve Readability Measures for First and Second Language Texts*
Michael Heilman, Kevyn Collins-Thompson, Jamie Callan and Maxine Eskenazi
- 2:35-3:00 *Automatic Assessment of Student Translations for Foreign Language Tutoring*
Chao Wang and Stephanie Seneff
- 3:00-3:25 *Automatic and Human Scoring of Word Definition Responses*
Kevyn Collins-Thompson and Jamie Callan
- 3:25-3:55 Break
- 3:55-5:35 Paper sessions

Machine Translation 2

- 3:55-4:20 *A Comparison of Pivot Methods for Phrase-Based Statistical Machine Translation*
Masao Utiyama and Hitoshi Isahara
- 4:20-4:45 *Efficient Phrase-Table Representation for Machine Translation with Applications to On-line MT and Speech Translation*
Richard Zens and Hermann Ney
- 4:45-5:10 *An Efficient Two-Pass Approach to Synchronous-CFG Driven Statistical MT*
Ashish Venugopal, Andreas Zollmann and Vogel Stephan
- 5:10-5:35 *Statistical Phrase-Based Post-Editing*
Michel Simard, Cyril Goutte and Pierre Isabelle

Wednesday, April 25, 2007 (continued)

Question Answering

- 3:55-4:20 *Automatic Answer Typing for How-Questions*
Christopher Pinchak and Shane Bergsma
- 4:20-4:45 *A Probabilistic Framework for Answer Selection in Question Answering*
Jeongwoo Ko, Luo Si and Eric Nyberg
- 4:45-5:10 *Question Answering Using Integrated Information Retrieval and Information Extraction*
Barry Schiffman, Kathleen McKeown, Ralph Grishman and James Allan
- 5:10-5:35 *Toward Multimedia: A String Pattern-Based Passage Ranking Model for Video Question Answering*
Yu-Chieh Wu and Jie-Chi Yang

Semantics

- 3:55-4:20 *Can Semantic Roles Generalize Across Genres?*
Szu-ting Yi, Edward Loper and Martha Palmer
- 4:20-4:45 *Towards Robust Semantic Role Labeling*
Sameer Pradhan, Wayne Ward and James Martin
- 4:45-5:10 *ISP: Learning Inferential Selectional Preferences*
Patrick Pantel, Rahul Bhagat, Bonaventura Coppola, Timothy Chklovski and Eduard Hovy
- 5:10-5:35 *Computing Semantic Similarity between Skill Statements for Approximate Matching*
Feng Pan and Robert Farrell

