

NAACL HLT 2010

**Human Language Technologies:
The 2010 Annual Conference of the
North American Chapter of the
Association for
Computational Linguistics**

Proceedings of the Main Conference

June 2–4, 2010
Los Angeles, California

Conference Sponsors:

- AT&T Interactive (Platinum Level)
- Microsoft Research (Gold Level)
- USC Information Sciences Institute (Gold Level)
- Google (Silver Level)
- J.D. Power and Associates (Silver Level)
- Yahoo! Labs (Silver Level)
- Basis Technology (Bronze Level)
- Factual (Bronze Level)
- IBM Research (Best Student Paper Award)
- Language Weaver (Conference Bag Sponsor)

USB memory sticks produced by

Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2010 The Association for Computational Linguistics

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-10: 1-932432-65-5

ISBN-13: 978-1-932432-65-7

Preface: General Chair

It is with great pleasure that I welcome you to the 2010 Human Language Technologies conference of the North American chapter of the Association for Computational Linguistics. An enormous amount of effort has gone into organizing the conference, and the result is the rich set of intellectual and social experiences that you will enjoy this week.

The NAACL HLT 2010 conference is an orchestrated production of many events. The centerpiece is the collection of carefully chosen papers, posters, and demonstrations that will be presented during the three days of the main conference. This includes the papers and posters of the Student Research Workshop, a setting that gives encouragement and opportunity for new members of our community to present their work. The main conference is preceded by a day of tutorials on topics of current interest in the field, and it is followed by a very full program of specialized workshops.

Many people have volunteered their time and energy to ensure the success of the conference. Indeed, the community as a whole has contributed by submitting papers and proposals for workshops, demos, and tutorials, by making thoughtful judgments as members of the many review committees, and for coming to participate in the events this week. These contributions are very much appreciated. The heaviest burdens have been shared by the people who agreed to lead the major subcommittees of the conference, and I want to recognize them here and explicitly thank them for their service.

To begin, I want to express my gratitude to Jill Burstein, Mary Harper, and Gerald Penn, the co-Chairs of the Program Committee. They supervised the selection of papers and scheduling for the main conference, recruiting the Area Chairs and managing the overall process. They chose the Noisy Genre theme of the conference, collaborated with the Area Chairs to select the Invited Speakers, administered the START reviewing system, and worked with the Area Chairs and the Local Arrangements Committee to assemble the final program. They also added a new feature to the program, a "one-minute madness" session to raise awareness and excitement about the poster and demo presentations.

The Program Committee co-Chairs and Area Chairs recommended a small number of papers for further consideration by the Best Papers Committee, chaired by Aravind Joshi. Tremendous thanks go to Aravind, Eugene Charniak, Michael Collins, Diane Litman, Daniel Marcu, and Drago Radev for their work in identifying the best examples of current work in the field.

The main-conference days include some other important activities. I appreciate Kristina Toutanova's effort to organize a second lunch-time Industry Panel, a follow-on to the successful panel at NAACL HLT 2009. I thank Carolyn Penstein Rosé, Chair of the Demonstration Committee, for soliciting candidates for the Demo Session and overseeing the rigorous process of review. For organizing the Student Research Workshop I thank Adriane Boyd, Mahesh Joshi, and Frank Rudzicz, the student co-Chairs, and also Julia Hockenmaier and Diane Litman for the assistance and guidance they provided as Faculty Advisors. Financial support for the student workshop came from a grant from the National Science Foundation.

I thank Richard Sproat and David Traum for serving as co-Chairs of the Workshop Committee. The workshop proposal and selection process was a joint activity of committees representing all three of this year's computational linguistics conferences, the ACL conference in Uppsala, Coling in Beijing, and NAACL HLT. Their challenge was to honor the location requests of the workshop proposers while

arranging for each conference to have a broad and balanced collection of workshops. A full complement of 16 workshops were chosen for NAACL HLT, and Richard and David then became the interface with the organizers of the individual workshops as they put out their calls, reviewed submissions, and assembled final papers for the Publications Committee.

I am grateful to Jason Baldrige, Peter Clark, and Gokhan Tur for volunteering to co-Chair the Tutorial Committee and for attracting and selecting tutorials that introduce a range of novel techniques and address a number of theoretical and practical problems.

The Publications Committee is responsible for constructing the written Proceedings of the conference from the materials that come from the many sources. The Publications Committee, with Claudia Leacock and Richard Wicentowski as co-Chairs, comes in at the end of the preparation stage and must operate on a strict schedule if the Proceedings are to be available by the time the conference begins. This year Claudia and Richard have been the pioneers for two major departures from past procedure. ACL proceedings in the past have been produced by a stand-alone community-developed software package, ACLPUB; this year we are using a version of that package that has been integrated directly into the START conference-management system. This should simplify the publication process for future conferences, but Claudia and Richard endured a fair amount of first-user suffering. The second departure from past practices is that for the first time we are providing the proceedings only on a USB memory stick, unaccompanied by a hardcopy volume. I deeply appreciate the care and attention with which Claudia and Richard have carried out their responsibilities.

We are all very much indebted to David Chiang, Jason Riesa, Jonathan May, and Eduard Hovy, the co-Chairs of the Local Arrangements Committee, and to the other members of the committee from the USC Information Sciences Institute. They have been hard at work for the longest time, and the comfort and elegance of the conference setting is the result of the many decisions they have made. To highlight just a few, they selected the hotel, made sure that the session rooms are properly laid out and equipped, and planned for the reception and banquet. They also designed and maintained the very attractive and informative conference web site. Liang Huang served as Coordinator of the Student Volunteers, and Jillian Gerten and Kenji Sagae served as Exhibit Coordinators.

I am grateful to the sponsoring organizations listed above for their financial support of the conference, and to the members of the Sponsorship Committee for identifying and nurturing those important relationships. The Sponsorship Committee—Srinivas Bangalore, Christy Doran, Eduard Hovy, Stephen Pulman, and Frédérique Segond—raised money jointly for NAACL HLT and ACL this year.

The NAACL executive committee, chaired by Owen Rambow and Rebecca Hwa, provides overall supervision and maintains the continuity and traditions of the conference from year to year. They helped by identifying candidates for many of the committee positions, by resolving several significant issues of conference policy, and by overseeing the budgeting process.

Finally, my heartfelt thanks go to Priscilla Rasmussen, the Business Manager of the ACL office and for the ACL conferences, for her steady guidance through all the stages of conference preparation. Priscilla, as always, played a critical role as the voice of experience and the source of wise advice that kept everything moving in the right direction.

Ronald M. Kaplan
Powerset division of Microsoft Bing
General Chair, NAACL HLT 2010

Preface: Program Chairs

Thanks for attending NAACL HLT 2010! As can be seen in the Conference Program, the NAACL HLT program contains innovative, high-quality work spanning a large array of disciplines within computational linguistics and human language technology. This year, we have included a special Noisy Genre theme to acknowledge the significant work in that area across several disciplines. We would like to thank L.Venkata Subramaniam of IBM Research-India for this suggestion.

NAACL HLT 2010 consists of oral and poster presentations of full and short papers, tutorials, application demonstrations, and workshops. A one-minute madness session has been introduced this year to highlight the poster and demo contributions. We are excited to have two very interesting and diverse keynote speakers: David Temperley, University of Rochester, whose talk is entitled, “Music, Language, and Computational Modeling: Lessons from the Key-Finding Problem”, and Steve Renals, University of Edinburgh, “Recognition and Understanding of Meetings.” In addition, we have a panel session Recent and Future HLT Challenges in Industry, chaired by Kristina Toutanova, which very appropriately reflects the conference theme.

This year, 291 full papers were submitted and reviewed, of which 90 papers were accepted (a 30.9% acceptance rate); 159 short papers were submitted and reviewed, of which 56 were accepted (a 35.2% acceptance rate). One of the accepted short papers was withdrawn to appear elsewhere, leaving a total of 55 short papers in the proceedings. Eighty-six full papers and 25 short papers will be presented orally. The remaining 34 accepted full and short papers will be presented as posters. We would like to thank all of the authors for submitting such remarkable papers to the conference.

Two best papers have been selected this year, and will be presented at the conference in a special awards session. The Best Full Paper, entitled Coreference Resolution in a Modular, Entity-Centered Model, was written by Aria Haghighi and Dan Klein. The Best Short Paper, entitled “cba to check the spelling”: Investigating Parser Performance on Discussion Forum Posts, was written by Jennifer Foster. Congratulations to the authors! The selection process worked as follows. The Senior Program Committee (SPC) nominated an initial set of best paper candidates for the awards; the final decisions were then made by a separate committee: Aravind Joshi (chair), Eugene Charniak, Michael Collins, Diane Litman, Daniel Marcu, and Dragomir Radev. We would like to thank the committee for reading, discussing, and contributing to the final selection process. The committee was selected to handle the breadth of expertise required to review the nominated best papers. Once the camera-ready versions of the papers were submitted by the authors, the Best Paper committee chair used the following process: (a) Committee members were asked to notify the chair about conflicts of interest. Members with conflicts of interest did not rank or discuss those papers with which they were conflicted. (b) Short papers were ranked first by committee members, and additional discussions were held to make a final decision. (c) Next, long papers were ranked by members and a smaller set of top papers were identified. Discussions were then held to make the final decision. (d) The committee chair was available to break any ties.

The review process was organized as a two-tier system with 18 SPC members and 382 reviewers. The SPC members managed the review process for both the full and short paper submissions: each paper received at least three reviews. Strict conflict-of-interest policies were in place during the review process. Specifically, authors who served on the Program Committee in any capacity were removed

from any responsibility related to their paper. Any such conflict-of-interest paper was handled by another committee member. In addition, the author was prohibited from participating in any discussion or decision making related to the paper. A similar policy also applied if a Program Committee member had an institutional or personal conflict with an author. All decisions on papers with these types of conflicts were made by members without conflicts. We would not have been able to produce such an interesting conference program without the dedicated SPC members who spent many weeks handling and evaluating the submission reviews, nor without the reviewers who provided such thoughtful evaluations. The full list of the SPC members and reviewers is listed elsewhere in these Proceedings. We would also like to thank the SPC members for their best paper recommendations and suggestions for the program, and Chris Dyer for his suggestions on the Machine Translation sessions.

There were, of course, a number of additional people with whom we directly interacted, and who made significant contributions to the success of this conference. So, here are some well-deserved acknowledgements. We would like to thank Rich Gerber and the START team for their help with the system that managed paper submissions and reviews, and the Local Arrangement co-chairs, David Chiang, Eduard Hovy, Jonathan May, and Jason Riesa, for their help with organizing the program, the preparation and publication of the conference handbook, handling the NAACL HLT 2010 website, and various other tasks – a list too lengthy to name here. We would also like to thank the Publication Co-chairs, Claudia Leacock and Richard Wicentowski, for forging through a number of obstacles, and doing an excellent job of handling the preparation and printing of these proceedings. We would like to thank the Workshop Co-chairs, David Traum and Richard Sproat, the Demo Chair, Carolyn Penstein Rosé, the Tutorial Co-chairs, Jason Baldridge, Peter Clark and Gokhan Tur, and the Student Research Workshop Faculty Co-chairs, Julia Hockenmeier and Diane Litman, and Student Co-chairs, Adriane Boyd, Mahesh Joshi, and Frank Rudzicz. The hard work of all of these co-chairs contributed to the quality of the entire conference program. We would also like to thank Chris Callison-Burch for handling Student Author Support, and to Liang Huang for serving as the Student Volunteer Coordinator. We are grateful to Priscilla Rasmussen for supporting the visa application process, as well as answering a variety of our questions and concerns around general conference logistics and event planning. Finally, we would like to express our deepest thanks to the General Chair, Ron Kaplan, for his continual support and patience throughout this process.

We hope that you will have a unique and enriching conference experience!

Jill Burstein, Educational Testing Service
Mary Harper, University of Maryland; Johns Hopkins HLT COE
Gerald Penn, University of Toronto

Organizers

General Chair:

Ronald Kaplan, Powerset Division of Microsoft Bing

Program Co-Chairs:

Jill Burstein (NLP), Educational Testing Service

Mary Harper (Speech), University of Maryland and The Johns Hopkins University HLTCOE

Gerald Penn (NLP), University of Toronto

Workshop Chairs:

Richard Sproat, Oregon Health & Science University

David Traum, University of Southern California, Institute for Creative Technologies

Demo Chair:

Carolyn Penstein Rosé, Carnegie Mellon University

Tutorial Chairs:

Jason Baldridge, The University of Texas at Austin

Peter Clark, The Boeing Company

Gokhan Tur, SRI International

Publications Chairs:

Claudia Leacock, Butler Hill Group

Richard Wicentowski, Swarthmore College

Student Research Workshop Chairs:

Julia Hockenmaier (Faculty Advisor), University of Illinois at Urbana-Champaign

Diane Litman (Faculty Advisor), University of Pittsburgh

Adriane Boyd (NLP Co-Chair), University of Ohio

Mahesh Joshi (NLP Co-Chair), Carnegie Mellon University

Frank Rudzicz (Speech Co-Chair), University of Toronto

Local Arrangements Chairs:

David Chiang, University of Southern California, Information Sciences Institute

Eduard Hovy, University of Southern California, Information Sciences Institute

Jonathan May, University of Southern California, Information Sciences Institute

Jason Riesa, University of Southern California, Information Sciences Institute

Exhibits Chairs:

Jillian Gerten, University of Southern California, Institute for Creative Technologies
Kenji Sagae, University of Southern California, Information Sciences Institute

Student Volunteers Coordinator:

Liang Huang, University of Southern California, Information Sciences Institute

Sponsorship Chairs:

Srinivas Bangalore (North America), AT&T
Christy Doran (North America), MITRE
Eduard Hovy (Local), University of Southern California, Information Sciences Institute
Stephen Pulman (Europe), Oxford University
Frédérique Segond (Europe), Xerox Research Centre Europe

ACL Business Manager:

Priscilla Rasmussen

Program Committee

Senior Program Committee Members (Area Chairs):

Eugene Agichtein, Emory University
Yaser Al-Onaizan, IBM
Ciprian Chelba, Google
Mona Diab, Columbia University
Barbara Di Eugenio, University of Illinois at Chicago
Eric Fosler-Lussier, Ohio State University
Makoto Kanazawa, National Institute of Informatics, Tokyo
Damianos Karakos, Johns Hopkins University
Philip Koehn, University of Edinburgh
Mike Maxwell, University of Maryland
Diana McCarthy, Lexical Computing Ltd
Ani Nenkova, University of Pennsylvania
Stefan Oepen, University of Oslo
Dan Roth, University of Illinois at Urbana-Champaign
Noah Smith, Carnegie Mellon University
Amanda Stent, AT&T
Joel Tetreault, Educational Testing Service
Jan Wiebe, University of Pittsburgh

Paper Award Committee:

Aravind Joshi (Chair), University of Pennsylvania
Daniel Marcu, University of Southern California, Information Sciences Institute
Diane Litman, University of Pittsburgh
Drago Radev, University of Michigan
Eugene Charniak, Brown University
Michael Collins, Massachusetts Institute of Technology

Program Committee Members:

Steve Abney	Abhishek Arun	Yassine Benajiba
Mikhail Ageev	Necip Fazil Ayan	Emily M. Bender
Eneko Agirre	Jason Baldridge	Shane Bergsma
David Ahn	Tim Baldwin	Steven Bethard
Gregory Aist	Carmen Banea	Pushpak Bhattacharyya
Salah Ait-Mokhtar	Srinivas Bangalore	Jiang Bian
Cem Akkaya	Regina Barzilay	Dan Bikel
Ingunn Amdal	Roberto Basili	Alexandra Birch
Ron Artstein	Anja Belz	Steven Bird

Philippe Blache	Kordula De Kuthy	Nancy Green
Bernd Bohnet	Eric de la Clergerie	Ralph Grishman
Kristy Boyer	Dina Demner Fushman	Qi Guo
SRK Branavan	John DeNero	Weiwei Guo
Antnio Branco	David DeVault	Iryna Gurevych
Thorsten Brants	Doug Downey	Aria Haghighi
Chris Brew	Mark Dredze	Keith Hall
Sam Brody	Jasha Droppo	Mike Hammond
Razvan Bunescu	Amit Dubey	Sanda Harabagiu
Harry Bunt	Kevin Duh	Mary Harper
Lukas Burget	Markus Egg	Mark Hasegawa-Johnson
Bill Byrne	Andreas Eisele	Laura Hasler
Donna Byron	Jacob Eisenstein	Hany Hassan
Lynne Cahill	Jason Eisner	Xiaodong He
Aoife Cahill	Noemie Elhadad	Jeff Heinz
Chris Callison-Burch	Jonathan Elsas	James Henderson
Claire Cardie	Micha Elsner	Andy Hickl
Giuseppe Carenini	Ahmad Emami	Silja Hildebrand
Michael Carl	Andrea Esuli	Graeme Hirst
Marine Carpuat	Tanveer Faruque	Hieu Hoang
John Carroll	Benoit Favre	Julia Hockenmaier
Francisco Casacuberta	Elena Filatova	Kristy Hollingshead
Joyce Chai	Katja Filippova	Rebecca Hwa
Eugene Charniak	Jenny Finkel	Nancy Ide
Ciprian Chelba	Dan Flickinger	Diana Inkpen
Jinying Chen	Kate Forbes Riley	Abraham Ittycheriah
John Chen	Eric Fosler-Lussier	Aren Jansen
Harr Chen	Davide Fossati	Heng Ji
Yejin Choi	Jennifer Foster	Kristiina Jokinen
Alexander Clark	George Foster	Doug Jones
Stephen Clark	Alexander Fraser	Laura Kallmeyer
James Clarke	Maria Fuentes	Hiroshi Kanayama
Martin Cmejrek	Pascale Fung	Damianos Karakos
Shay Cohen	Ryan Gabbard	Lauri Karttunen
Trevor Cohn	Michel Galley	Athanassios Katsamanis
Michael Collins	Michael Gamon	Frank Keller
Michael Connor	Albert Gatt	Cynthia Kersey
John Conroy	Panayiotis Georgiou	Tracy King
Mark Core	Dan Gildea	Brian Kingsbury
Mathias Creutz	Jesus Gimenez	Dan Klein
Aron Culotta	Roxana Girju	Kevin Knight
Robert Dale	Claudio Giuliano	Greg Kobele
Cristian Danescu-Niculescu-	Vaibhava Goel	Rob Koeling
Mizil	Dan Goldwasser	Greg Kondrak
Hal Daume	Carlos Gomez Rodriguez	Terry Koo
Anne David	Jeff Good	Stefan Kopp

Valia Kordoni	Mike Maxwell	Daniele Pighin
Anna Korhonen	James Mayfield	Emily Pitler
Sandra Kuebler	David McClosky	Massimo Poesio
Marco Kuhlmann	Mark McConville	Joe Polifroni
Roland Kuhn	Nancy McCracken	Simone Ponzetto
Jonas Kuhn	David McDonald	Maja Popovic
Seth Kulick	Ryan McDonald	Dan Povey
Mikko Kurimo	Kathy McKeown	Sameer Pradhan
Oren Kurland	Susan McRoy	Rashmi Prasad
Sadao Kurohashi	Florian Metze	Chris Quirk
Arnd Christian Knig	Eleni Miltsakaki	Stephan Raaijmakers
Philippe Langlais	Gilad Mishne	Bhuvana Ramabhadran
Mirella Lapata	Yusuke Miyao	Allan Ramsay
Alberto Lavelli	Saif Mohammad	Ari Rappoport
Matt Lease	Bob Moore	Ariya Rastrow
Lin-Shan Lee	Alessandro Moschitti	Lev Ratinov
Els Lefever	Dragos Munteanu	Roi Reichart
James Lester	Smaranda Muresan	Ehud Reiter
Gregor Leusch	Gabriel Murray	German Rigau
Haizhou Li	Tor Andre Myrvoll	Michael Riley
Zhifei Li	Roberto Navigli	Laura Rimell
Henry Li	Mark-Jan Nederhof	Nicholas Rizzolo
Mu Li	Sumit Negi	Brian Roark
Percy Liang	Hwee Tou Ng	Horacio Rodriguez
Jimmy Lin	Vincent Ng	Victoria Rosen
Diane Litman	Patrick Nguyen	Antti-Veikko Rosti
Qun Liu	Takashi Ninomiya	Dan Roth
Yang Liu	Joakim Nivre	Shourya Roy
Zhanyi Liu	Doug Oard	Marta Ruiz Costa-juss
Adam Lopez	Franz Och	Mark Sammons
Annie Louis	Kemal Oflazer	George Saon
Bernardo Magnini	Alice Oh	Murat Saraclar
Wolfgang Maier	Constantin Orasan	Kevin Scannell
Rob Malouf	Miles Osborne	Odette Scharenborg
Arindam Mandal	Lilja Ovreid	David Schlangen
Lidia Mangu	Deepak Padmanabhan	Helmut Schmid
Christopher Manning	Ulrike Pado	Hinrich Schuetze
Daniel Marcu	Alexis Palmer	Mike Seltzer
Lluis Marquez Villodre	Martha Palmer	Hendra Setiawan
David Martinez	Bo Pang	Fei Sha
Andre Martins	Marius Pasca	Izhak Shafran
Yuval Marton	Becky Passonneau	Wade Shen
Sameer Maskey	Sid Patwardhan	Libin Shen
Yuji Matsumoto	Ted Pedersen	Nobuyuki Shimizu
Evgeny Matusov	Slav Petrov	Advaith Siddharthan
Arne Mauser	Christine Piatko	Khalil Simaan

Michel Simard
Kevin Small
David Smith
Mark Smucker
Swapna Somasundaran
Radu Soricut
Richard Sproat
Vivek Srikumar
Yaji Sripada
Mark Steedman
Josef Steinberger
Amanda Stent
Mark Stevenson
Carlo Strapparava
Kristina Striegnitz
Michael Strube
LV Subramaniam
Mihai Surdeanu
Hiroya Takamura
David Talbot
Simone Teufel
Joerg Tiedemann
Christoph Tillmann
Ivan Titov
Kristina Toutanova
Roy Tromble
Yoshimasa Tsuruoka

Ielka van der Sluis
Josef van Genabith
Keith Vander Linden
Balakrishnan Varadarajan
Ashish Venugopal
Yannick Versley
David Vilar
Stephan Vogel
Piek Vossen
Marilyn Walker
Stephen Wan
Wen Wang
Haifeng Wang
XingLong Wang
Qin Wang
Ye-Yi Wang
Andy Way
Bonnie Webber
David Weir
Fuiliang Weng
Michael White
Chris White
Richard Wicentowski
Theresa Wilson
Shuly Wintner
Yunfang Wu
Dekai Wu

Hua Wu
Holger Wunsch
Bing Xiang
Deyi Xiong
Puyang Xu
Jinxi Xu
Peng Xu
Wen-tau Yih
Dong Yu
Deniz Yuret
David Zajic
Richard Zens
Hao Zhang
Yue Zhang
Dongdong Zhang
Yi Zhang
Min Zhang
Ying Zhang
Tiejun Zhao
Bing Zhao
Bowen Zhou
Xiaodan Zhu
Imed Zitouni
Andreas Zollmann
Chengqing Zong
Geoff Zweig

Table of Contents

<i>Invited Talk: Recognition and Understanding of Meetings</i>	
Steve Renals	1
<i>Chart Mining-based Lexical Acquisition with Precision Grammars</i>	
Yi Zhang, Timothy Baldwin, Valia Kordoni, David Martinez and Jeremy Nicholson	10
<i>Products of Random Latent Variable Grammars</i>	
Slav Petrov	19
<i>Automatic Domain Adaptation for Parsing</i>	
David McClosky, Eugene Charniak and Mark Johnson	28
<i>Appropriately Handled Prosodic Breaks Help PCFG Parsing</i>	
Zhongqiang Huang and Mary Harper	37
<i>Using Confusion Networks for Speech Summarization</i>	
Shasha Xie and Yang Liu	46
<i>Qme! : A Speech-based Question-Answering system on Mobile Devices</i>	
Taniya Mishra and Srinivas Bangalore	55
<i>Dialogue-Oriented Review Summary Generation for Spoken Dialogue Recommendation Systems</i>	
Jingjing Liu, Stephanie Seneff and Victor Zue	64
<i>Minimally-Supervised Extraction of Entities from Text Advertisements</i>	
Sameer Singh, Dustin Hillard and Chris Leggetter	73
<i>Taxonomy Learning Using Word Sense Induction</i>	
Ioannis P. Klapaftis and Suresh Manandhar	82
<i>Visual Information in Semantic Representation</i>	
Yansong Feng and Mirella Lapata	91
<i>Automatic Evaluation of Topic Coherence</i>	
David Newman, Jey Han Lau, Karl Grieser and Timothy Baldwin	100
<i>Multi-Prototype Vector-Space Models of Word Meaning</i>	
Joseph Reisinger and Raymond J. Mooney	109
<i>Unsupervised Syntactic Alignment with Inversion Transduction Grammars</i>	
Adam Pauls, Dan Klein, David Chiang and Kevin Knight	118
<i>Joint Parsing and Alignment with Weakly Synchronized Grammars</i>	
David Burkett, John Blitzer and Dan Klein	127
<i>Learning Translation Boundaries for Phrase-Based Decoding</i>	
Deyi Xiong, Min Zhang and Haizhou Li	136

<i>Hitting the Right Paraphrases in Good Time</i>	
Stanley Kok and Chris Brockett	145
<i>Training Paradigms for Correcting Errors in Grammar and Usage</i>	
Alla Rozovskaya and Dan Roth	154
<i>Using Mostly Native Data to Correct Errors in Learners' Writing</i>	
Michael Gamon	163
<i>Unsupervised Modeling of Twitter Conversations</i>	
Alan Ritter, Colin Cherry and Bill Dolan	172
<i>Streaming First Story Detection with application to Twitter</i>	
Saša Petrović, Miles Osborne and Victor Lavrenko	181
<i>Unsupervised Model Adaptation using Information-Theoretic Criterion</i>	
Ariya Rastrow, Frederick Jelinek, Abhinav Sethy and Bhuvana Ramabhadran	190
<i>Formatting Time-Aligned ASR Transcripts for Readability</i>	
Maria Shugrina	198
<i>Cheap, Fast and Good Enough: Automatic Speech Recognition with Non-Expert Transcription</i>	
Scott Novotney and Chris Callison-Burch	207
<i>Contextual Information Improves OOV Detection in Speech</i>	
Carolina Parada, Mark Dredze, Denis Filimonov and Frederick Jelinek	216
<i>Improved Extraction Assessment through Better Language Models</i>	
Arun Ahuja and Doug Downey	225
<i>Language Identification: The Long and the Short of the Matter</i>	
Timothy Baldwin and Marco Lui	229
<i>Inducing Synchronous Grammars with Slice Sampling</i>	
Phil Blunsom and Trevor Cohn	238
<i>Task-based Evaluation of Multiword Expressions: a Pilot Study in Statistical Machine Translation</i>	
Marine Carpuat and Mona Diab	242
<i>Improving Semantic Role Labeling with Word Sense</i>	
Wanxiang Che, Ting Liu and Yongqiang Li	246
<i>Extending the METEOR Machine Translation Evaluation Metric to the Phrase Level</i>	
Michael Denkowski and Alon Lavie	250
<i>Testing a Grammar Customization System with Sahaptin</i>	
Scott Drellishak	254
<i>Two monolingual parses are better than one (synchronous parse)</i>	
Chris Dyer	263

<i>Fast Query for Large Treebanks</i> Sumukh Ghodke and Steven Bird	267
<i>Efficient Parsing of Well-Nested Linear Context-Free Rewriting Systems</i> Carlos Gómez-Rodríguez, Marco Kuhlmann and Giorgio Satta	276
<i>Utility Evaluation of Cross-document Information Extraction</i> Heng Ji, Zheng Chen, Jonathan Feldman, Antonio Gonzalez, Ralph Grishman and Vivek Upadhyay	285
<i>Evaluation Metrics for the Lexical Substitution Task</i> Sanaz Jabbari, Mark Hepple and Louise Guthrie	289
<i>Movie Reviews and Revenues: An Experiment in Text Regression</i> Mahesh Joshi, Dipanjan Das, Kevin Gimpel and Noah A. Smith	293
<i>Using Gaussian Mixture Models to Detect Figurative Language in Context</i> Linlin Li and Caroline Sporleder	297
<i>Improving Phrase-Based Translation with Prototypes of Short Phrases</i> Frank Liberato, Behrang Mohit and Rebecca Hwa	301
<i>Putting the User in the Loop: Interactive Maximal Marginal Relevance for Query-Focused Summarization</i> Jimmy Lin, Nitin Madnani and Bonnie Dorr	305
<i>Improving Blog Polarity Classification via Topic Analysis and Adaptive Methods</i> Feifan Liu, Dong Wang, Bin Li and Yang Liu	309
<i>Creating Local Coherence: An Empirical Assessment</i> Annie Louis and Ani Nenkova	313
<i>Time-Efficient Creation of an Accurate Sentence Fusion Corpus</i> Kathleen McKeown, Sara Rosenthal, Kapil Thadani and Coleman Moore	317
<i>Towards Cross-Lingual Textual Entailment</i> Yashar Mehdad, Matteo Negri and Marcello Federico	321
<i>A Comparative Study of Word Co-occurrence for Term Clustering in Language Model-based Sentence Retrieval</i> Saeedeh Momtazi, Sanjeev Khudanpur and Dietrich Klakow	325
<i>Information Content Measures of Semantic Similarity Perform Better Without Sense-Tagged Text</i> Ted Pedersen	329
<i>Generating Expository Dialogue from Monologue: Motivation, Corpus and Preliminary Rules</i> Paul Piwek and Svetlana Stoyanchev	333
<i>The Simple Truth about Dependency and Phrase Structure Representations: An Opinion Piece</i> Owen Rambow	337

<i>Word Alignment with Stochastic Bracketing Linear Inversion Transduction Grammar</i> Markus Saers, Joakim Nivre and Dekai Wu	341
<i>Crowdsourcing the evaluation of a domain-adapted named entity recognition system</i> Asad B. Sayeed, Timothy J. Meyer, Hieu C. Nguyen, Olivia Buzek and Amy Weinberg.....	345
<i>Generalizing Hierarchical Phrase-based Translation using Rules with Adjacent Nonterminals</i> Hendra Setiawan and Philip Resnik	349
<i>The Effect of Ambiguity on the Automated Acquisition of WSD Examples</i> Mark Stevenson and Yikun Guo	353
<i>Word Sense Subjectivity for Cross-lingual Lexical Substitution</i> Fangzhong Su and Katja Markert	357
<i>Query Ambiguity Revisited: Clickthrough Measures for Distinguishing Informational and Ambiguous Queries</i> Yu Wang and Eugene Agichtein.....	361
<i>For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia</i> Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil and Lillian Lee.....	365
<i>Predicting Human-Targeted Translation Edit Rate via Untrained Human Annotators</i> Omar F. Zaidan and Chris Callison-Burch.....	369
<i>Improving Semantic Role Classification with Selectional Preferences</i> Beñat Zapirain, Eneko Agirre, Lluís Màrquez and Mihai Surdeanu	373
<i>Generalizing Syntactic Structures for Product Attribute Candidate Extraction</i> Yanyan Zhao, Bing Qin, Shen Hu and Ting Liu.....	377
<i>“cba to check the spelling”: Investigating Parser Performance on Discussion Forum Posts</i> Jennifer Foster	381
<i>Coreference Resolution in a Modular, Entity-Centered Model</i> Aria Haghighi and Dan Klein	385
<i>Stream-based Translation Models for Statistical Machine Translation</i> Abby Levenberg, Chris Callison-Burch and Miles Osborne	394
<i>Extracting Parallel Sentences from Comparable Corpora using Document Level Alignment</i> Jason R. Smith, Chris Quirk and Kristina Toutanova	403
<i>Statistical Machine Translation of Texts with Misspelled Words</i> Nicola Bertoldi, Mauro Cettolo and Marcello Federico	412
<i>Everybody loves a rich cousin: An empirical study of transliteration through bridge languages</i> Mitesh M. Khapra, A Kumaran and Pushpak Bhattacharyya	420
<i>Discriminative Learning over Constrained Latent Representations</i> Ming-Wei Chang, Dan Goldwasser, Dan Roth and Vivek Srikumar	429

<i>Some Empirical Evidence for Annotation Noise in a Benchmarked Dataset</i>	
Beata Beigman Klebanov and Eyal Beigman	438
<i>Bayesian Inference for Finite-State Transducers</i>	
David Chiang, Jonathan Graehl, Kevin Knight, Adam Pauls and Sujith Ravi	447
<i>Distributed Training Strategies for the Structured Perceptron</i>	
Ryan McDonald, Keith Hall and Gideon Mann	456
<i>Term Weighting Schemes for Latent Dirichlet Allocation</i>	
Andrew T. Wilson and Peter A. Chew	465
<i>Learning Dense Models of Query Similarity from User Click Logs</i>	
Fabio De Bona, Stefan Riezler, Keith Hall, Massimiliano Ciaramita, Amaç Herdağdelen and Maria Holmqvist	474
<i>Learning to Link Entities with Knowledge Base</i>	
Zhicheng Zheng, Fangtao Li, Minlie Huang and Xiaoyan Zhu	483
<i>Improving the Multilingual User Experience of Wikipedia Using Cross-Language Name Search</i>	
Raghavendra Udupa and Mitesh M. Khapra	492
<i>Learning Words and Their Meanings from Unsegmented Child-directed Speech</i>	
Bevan K. Jones, Mark Johnson and Michael C. Frank	501
<i>Subword Variation in Text Message Classification</i>	
Robert Munro and Christopher D. Manning	510
<i>Automatic Diacritization for Low-Resource Languages Using a Hybrid Word and Consonant CMM</i>	
Robbie Haertel, Peter McClanahan and Eric K. Ringger	519
<i>Urdu Word Segmentation</i>	
Nadir Durrani and Sarmad Hussain	528
<i>Enabling Monolingual Translators: Post-Editing vs. Options</i>	
Philipp Koehn	537
<i>Online Learning for Interactive Statistical Machine Translation</i>	
Daniel Ortiz-Martínez, Ismael García-Varea and Francisco Casacuberta	546
<i>The Best Lexical Metric for Phrase-Based Statistical MT System Optimization</i>	
Daniel Cer, Christopher D. Manning and Daniel Jurafsky	555
<i>Variational Inference for Adaptor Grammars</i>	
Shay B. Cohen, David M. Blei and Noah A. Smith	564
<i>Type-Based MCMC</i>	
Percy Liang, Michael I. Jordan and Dan Klein	573
<i>Painless Unsupervised Learning with Features</i>	
Taylor Berg-Kirkpatrick, Alexandre Bouchard-Côté, John DeNero and Dan Klein	582

<i>Linguistic Steganography Using Automatically Generated Paraphrases</i> Ching-Yun Chang and Stephen Clark	591
<i>Prenominal Modifier Ordering via Multiple Sequence Alignment</i> Aaron Dunlop, Margaret Mitchell and Brian Roark	600
<i>Good Question! Statistical Ranking for Question Generation</i> Michael Heilman and Noah A. Smith	609
<i>Not All Seeds Are Equal: Measuring the Quality of Text Mining Seeds</i> Zornitsa Kozareva and Eduard Hovy	618
<i>Extracting Glosses to Disambiguate Word Senses</i> Weisi Duan and Alexander Yates	627
<i>Can Recognising Multiword Expressions Improve Shallow Parsing?</i> Ioannis Korkontzelos and Suresh Manandhar	636
<i>A Simple Approach for HPSG Supertagging Using Dependency Information</i> Yao-zhong Zhang, Takuya Matsuzaki and Jun'ichi Tsujii	645
<i>Ensemble Models for Dependency Parsing: Cheap and Good?</i> Mihai Surdeanu and Christopher D. Manning	649
<i>Enlarged Search Space for SITG Parsing</i> Guillem Gascó, Joan-Andreu Sánchez and José-Miguel Benedí	653
<i>Improving Data Driven Dependency Parsing using Clausal Information</i> Phani Gadde, Karan Jindal, Samar Husain, Dipti Misra Sharma and Rajeev Sangal	657
<i>A Treebank Query System Based on an Extracted Tree Grammar</i> Seth Kulick and Ann Bies	661
<i>Reranking the Berkeley and Brown Parsers</i> Mark Johnson and Ahmet Engin Ural	665
<i>An Exploration of Off Topic Conversation</i> Whitney L. Cade, Blair A. Lehman and Andrew Olney	669
<i>Making Conversational Structure Explicit: Identification of Initiation-response Pairs within Online Discussions</i> Yi-Chia Wang and Carolyn P. Rosé	673
<i>Engaging learning groups using Social Interaction Strategies</i> Rohit Kumar and Carolyn P. Rosé	677
<i>Using Entity-Based Features to Model Coherence in Student Essays</i> Jill Burstein, Joel Tetreault and Slava Andreyev	681
<i>Summarizing Microblogs Automatically</i> Beaux Sharifi, Mark-Anthony Hutton and Jugal Kalita	685

<i>Automatic Generation of Personalized Annotation Tags for Twitter Users</i>	
Wei Wu, Bin Zhang and Mari Ostendorf	689
<i>Language identification of names with SVMs</i>	
Aditya Bhargava and Grzegorz Kondrak	693
<i>Integrating Joint n-gram Features into a Discriminative Training Framework</i>	
Sittichai Jiampojamarn, Colin Cherry and Grzegorz Kondrak	697
<i>A Hybrid Morphologically Decomposed Factored Language Models for Arabic LVCSR</i>	
Amr El-Desoky, Ralf Schlüter and Hermann Ney	701
<i>Is Arabic Part of Speech Tagging Feasible Without Word Segmentation?</i>	
Emad Mohamed and Sandra Kübler	705
<i>Arabic Mention Detection: Toward Better Unit of Analysis</i>	
Yassine Benajiba and Imed Zitouni	709
<i>An MDL-based approach to extracting subword units for grapheme-to-phoneme conversion</i>	
Sravana Reddy and John Goldsmith	713
<i>Extracting Phrase Patterns with Minimum Redundancy for Unsupervised Speaker Role Classification</i>	
Bin Zhang, Brian Hutchinson, Wei Wu and Mari Ostendorf	717
<i>Classification of Prosodic Events using Quantized Contour Modeling</i>	
Andrew Rosenberg	721
<i>Investigations into the Crandem Approach to Word Recognition</i>	
Rohit Prabhavalkar, Preethi Jyothi, William Hartmann, Jeremy Morris and Eric Fosler-Lussier	725
<i>Constraint-Driven Rank-Based Learning for Information Extraction</i>	
Sameer Singh, Limin Yao, Sebastian Riedel and Andrew McCallum	729
<i>Softmax-Margin CRFs: Training Log-Linear Models with Cost Functions</i>	
Kevin Gimpel and Noah A. Smith	733
<i>Bitext-Based Resolution of German Subject-Object Ambiguities</i>	
Florian Schwarck, Alexander Fraser and Hinrich Schütze	737
<i>Invited Talk: Music, Language, and Computational Modeling: Lessons from the Key-Finding Problem</i>	
David Temperley	741
<i>An Efficient Algorithm for Easy-First Non-Directional Dependency Parsing</i>	
Yoav Goldberg and Michael Elhadad	742
<i>From Baby Steps to Leapfrog: How “Less is More” in Unsupervised Dependency Parsing</i>	
Valentin I. Spitzkovsky, Hiyan Alshawi and Daniel Jurafsky	751

<i>Relaxed Marginal Inference and its Application to Dependency Parsing</i> Sebastian Riedel and David A. Smith	760
<i>Optimal Parsing Strategies for Linear Context-Free Rewriting Systems</i> Daniel Gildea	769
<i>The viability of web-derived polarity lexicons</i> Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan and Ryan McDonald	777
<i>Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables</i> Tetsuji Nakagawa, Kentaro Inui and Sadao Kurohashi	786
<i>Convolution Kernels for Opinion Holder Extraction</i> Michael Wiegand and Dietrich Klakow	795
<i>An Unsupervised Aspect-Sentiment Model for Online Reviews</i> Samuel Brody and Noemie Elhadad	804
<i>Joint Inference for Knowledge Extraction from Biomedical Literature</i> Hoifung Poon and Lucy Vanderwende	813
<i>Clinical Information Retrieval using Document and PICO Structure</i> Florian Boudin, Jian-Yun Nie and Martin Dawes	822
<i>Topic Models for Image Annotation and Text Illustration</i> Yansong Feng and Mirella Lapata	831
<i>Learning about Voice Search for Spoken Dialogue Systems</i> Rebecca Passonneau, Susan L. Epstein, Tiziana Ligorio, Joshua B. Gordon and Pravin Bhutada	840
<i>A Direct Syntax-Driven Reordering Model for Phrase-Based Machine Translation</i> Niyu Ge	849
<i>Context-free reordering, finite-state translation</i> Chris Dyer and Philip Resnik	858
<i>Improved Models of Distortion Cost for Statistical Machine Translation</i> Spence Green, Michel Galley and Christopher D. Manning	867
<i>Why Synchronous Tree Substitution Grammars?</i> Andreas Maletti	876
<i>An extractive supervised two-stage method for sentence compression</i> Dimitrios Galanis and Ion Androutsopoulos	885
<i>Interpretation and Transformation for Abstracting Conversations</i> Gabriel Murray, Giuseppe Carenini and Raymond Ng	894
<i>Quantifying the Limits and Success of Extractive Summarization Systems Across Domains</i> Hakan Ceylan, Rada Mihalcea, Umut Özertem, Elena Lloret and Manuel Palomar	903

<i>Multi-document Summarization via Budgeted Maximization of Submodular Functions</i>	
Hui Lin and Jeff Bilmes	912
<i>Cross-lingual Induction of Selectional Preferences with Bilingual Vector Spaces</i>	
Yves Peirsman and Sebastian Padó	921
<i>Latent SVMs for Semantic Role Labeling using LTAG Derivation Trees</i>	
Yudong Liu, Gholamreza Haffari and Anoop Sarkar	930
<i>Unsupervised Induction of Semantic Roles</i>	
Joel Lang and Mirella Lapata	939
<i>Probabilistic Frame-Semantic Parsing</i>	
Dipanjan Das, Nathan Schneider, Desai Chen and Noah A. Smith	948
<i>Expected Sequence Similarity Maximization</i>	
Cyril Allauzen, Shankar Kumar, Wolfgang Macherey, Mehryar Mohri and Michael Riley	957
<i>Accurate Non-Hierarchical Phrase-Based Translation</i>	
Michel Galley and Christopher D. Manning	966
<i>Model Combination for Machine Translation</i>	
John DeNero, Shankar Kumar, Ciprian Chelba and Franz Och	975
<i>Detecting Emails Containing Requests for Action</i>	
Andrew Lampert, Robert Dale and Cecile Paris	984
<i>Evaluating Hierarchical Discourse Segmentation</i>	
Lucien Carroll	993
<i>Reformulating Discourse Connectives for Non-Expert Readers</i>	
Advaith Siddharthan and Napoleon Katsos	1002
<i>Tree Edit Models for Recognizing Textual Entailments, Paraphrases, and Answers to Questions</i>	
Michael Heilman and Noah A. Smith	1011
<i>Syntactic/Semantic Structures for Textual Entailment Recognition</i>	
Yashar Mehdad, Alessandro Moschitti and Fabio Massimo Zanzotto	1020
<i>Automatic Metaphor Interpretation as a Paraphrasing Task</i>	
Ekaterina Shutova	1029

Conference Program

Wednesday, June 2, 2010

Plenary Session I

8:45–9:00 Opening Ceremony

9:00–10:10 *Invited Talk: Recognition and Understanding of Meetings*
Steve Renals

10:10–10:40 **Break**

Parsing I

10:40–11:05 *Chart Mining-based Lexical Acquisition with Precision Grammars*
Yi Zhang, Timothy Baldwin, Valia Kordoni, David Martinez and Jeremy Nicholson

11:05–11:30 *Products of Random Latent Variable Grammars*
Slav Petrov

11:30–11:55 *Automatic Domain Adaptation for Parsing*
David McClosky, Eugene Charniak and Mark Johnson

11:55–12:20 *Appropriately Handled Prosodic Breaks Help PCFG Parsing*
Zhongqiang Huang and Mary Harper

Noisy Genre I

10:40–11:05 *Using Confusion Networks for Speech Summarization*
Shasha Xie and Yang Liu

11:05–11:30 *Qme! : A Speech-based Question-Answering system on Mobile Devices*
Taniya Mishra and Srinivas Bangalore

11:30–11:55 *Dialogue-Oriented Review Summary Generation for Spoken Dialogue Recommendation Systems*
Jingjing Liu, Stephanie Seneff and Victor Zue

Wednesday, June 2, 2010 (continued)

11:55–12:20 *Minimally-Supervised Extraction of Entities from Text Advertisements*
Sameer Singh, Dustin Hillard and Chris Leggetter

Semantics I

10:40–11:05 *Taxonomy Learning Using Word Sense Induction*
Ioannis P. Klapaftis and Suresh Manandhar

11:05–11:30 *Visual Information in Semantic Representation*
Yansong Feng and Mirella Lapata

11:30–11:55 *Automatic Evaluation of Topic Coherence*
David Newman, Jey Han Lau, Karl Grieser and Timothy Baldwin

11:55–12:20 *Multi-Prototype Vector-Space Models of Word Meaning*
Joseph Reisinger and Raymond J. Mooney

Student Research Workshop I

Note: All student research workshop papers are located in a companion volume of the proceedings.

10:40–11:10 *Improving Syntactic Coordination Resolution using Language Modeling*
Philip Ogren

11:10–11:40 *On Automated Evaluation of Readability of Summaries: Capturing Grammaticality, Focus, Structure and Coherence*
Ravikiran Vadlapudi and Rahul Katragadda

11:40–12:10 *Detecting Novelty in the Context of Progressive Summarization*
Praveen Bysani

12:20–2:00 **Lunch**

Wednesday, June 2, 2010 (continued)

Machine Translation I

2:00–2:25 *Unsupervised Syntactic Alignment with Inversion Transduction Grammars*
Adam Pauls, Dan Klein, David Chiang and Kevin Knight

2:25–2:50 *Joint Parsing and Alignment with Weakly Synchronized Grammars*
David Burkett, John Blitzer and Dan Klein

2:50–3:15 *Learning Translation Boundaries for Phrase-Based Decoding*
Deyi Xiong, Min Zhang and Haizhou Li

3:15–3:40 *Hitting the Right Paraphrases in Good Time*
Stanley Kok and Chris Brockett

Noisy Genre II

2:00–2:25 *Training Paradigms for Correcting Errors in Grammar and Usage*
Alla Rozovskaya and Dan Roth

2:25–2:50 *Using Mostly Native Data to Correct Errors in Learners' Writing*
Michael Gamon

2:50–3:15 *Unsupervised Modeling of Twitter Conversations*
Alan Ritter, Colin Cherry and Bill Dolan

3:15–3:40 *Streaming First Story Detection with application to Twitter*
Saša Petrović, Miles Osborne and Victor Lavrenko

Wednesday, June 2, 2010 (continued)

Speech Processing

- 2:00–2:25 *Unsupervised Model Adaptation using Information-Theoretic Criterion*
Ariya Rastrow, Frederick Jelinek, Abhinav Sethy and Bhuvana Ramabhadran
- 2:25–2:50 *Formatting Time-Aligned ASR Transcripts for Readability*
Maria Shugrina
- 2:50–3:15 *Cheap, Fast and Good Enough: Automatic Speech Recognition with Non-Expert Transcription*
Scott Novotney and Chris Callison-Burch
- 3:15–3:40 *Contextual Information Improves OOV Detection in Speech*
Carolina Parada, Mark Dredze, Denis Filimonov and Frederick Jelinek

Student Research Workshop II

Note: All student research workshop papers are located in a companion volume of the proceedings.

- 2:00–2:30 *Extrinsic Parse Selection*
David Goss-Grubbs
- 2:30–3:00 *Towards a Matrix-based Distributional Model of Meaning*
Eugenie Giesbrecht
- 3:00–3:30 *Distinguishing Use and Mention in Natural Language*
Shomir Wilson
- 3:40–4:10 **Break**

Wednesday, June 2, 2010 (continued)

Poster Plenary Session

4:10–5:30 One-Minute Madness: Poster and Demo Previews

5:30–6:30 **Break**

6:30–8:30 **Poster and Demo Session**

Posters

Improved Extraction Assessment through Better Language Models

Arun Ahuja and Doug Downey

Language Identification: The Long and the Short of the Matter

Timothy Baldwin and Marco Lui

Inducing Synchronous Grammars with Slice Sampling

Phil Blunsom and Trevor Cohn

Task-based Evaluation of Multiword Expressions: a Pilot Study in Statistical Machine Translation

Marine Carpuat and Mona Diab

Improving Semantic Role Labeling with Word Sense

Wanxiang Che, Ting Liu and Yongqiang Li

Extending the METEOR Machine Translation Evaluation Metric to the Phrase Level

Michael Denkowski and Alon Lavie

Testing a Grammar Customization System with Sahaptin

Scott Drellishak

Two monolingual parses are better than one (synchronous parse)

Chris Dyer

Fast Query for Large Treebanks

Sumukh Ghodke and Steven Bird

Wednesday, June 2, 2010 (continued)

Efficient Parsing of Well-Nested Linear Context-Free Rewriting Systems

Carlos Gómez-Rodríguez, Marco Kuhlmann and Giorgio Satta

Utility Evaluation of Cross-document Information Extraction

Heng Ji, Zheng Chen, Jonathan Feldman, Antonio Gonzalez, Ralph Grishman and Vivek Upadhyay

Evaluation Metrics for the Lexical Substitution Task

Sanaz Jabbari, Mark Hepple and Louise Guthrie

Movie Reviews and Revenues: An Experiment in Text Regression

Mahesh Joshi, Dipanjan Das, Kevin Gimpel and Noah A. Smith

Using Gaussian Mixture Models to Detect Figurative Language in Context

Linlin Li and Caroline Sporleder

Improving Phrase-Based Translation with Prototypes of Short Phrases

Frank Liberato, Behrang Mohit and Rebecca Hwa

Putting the User in the Loop: Interactive Maximal Marginal Relevance for Query-Focused Summarization

Jimmy Lin, Nitin Madnani and Bonnie Dorr

Improving Blog Polarity Classification via Topic Analysis and Adaptive Methods

Feifan Liu, Dong Wang, Bin Li and Yang Liu

Creating Local Coherence: An Empirical Assessment

Annie Louis and Ani Nenkova

Time-Efficient Creation of an Accurate Sentence Fusion Corpus

Kathleen McKeown, Sara Rosenthal, Kapil Thadani and Coleman Moore

Towards Cross-Lingual Textual Entailment

Yashar Mehdad, Matteo Negri and Marcello Federico

A Comparative Study of Word Co-occurrence for Term Clustering in Language Model-based Sentence Retrieval

Saeedeh Momtazi, Sanjeev Khudanpur and Dietrich Klakow

Wednesday, June 2, 2010 (continued)

Information Content Measures of Semantic Similarity Perform Better Without Sense-Tagged Text

Ted Pedersen

Generating Expository Dialogue from Monologue: Motivation, Corpus and Preliminary Rules

Paul Piwek and Svetlana Stoyanchev

The Simple Truth about Dependency and Phrase Structure Representations: An Opinion Piece

Owen Rambow

Word Alignment with Stochastic Bracketing Linear Inversion Transduction Grammar

Markus Saers, Joakim Nivre and Dekai Wu

Crowdsourcing the evaluation of a domain-adapted named entity recognition system

Asad B. Sayeed, Timothy J. Meyer, Hieu C. Nguyen, Olivia Buzek and Amy Weinberg

Generalizing Hierarchical Phrase-based Translation using Rules with Adjacent Nonterminals

Hendra Setiawan and Philip Resnik

The Effect of Ambiguity on the Automated Acquisition of WSD Examples

Mark Stevenson and Yikun Guo

Word Sense Subjectivity for Cross-lingual Lexical Substitution

Fangzhong Su and Katja Markert

Query Ambiguity Revisited: Clickthrough Measures for Distinguishing Informational and Ambiguous Queries

Yu Wang and Eugene Agichtein

For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia

Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil and Lillian Lee

Predicting Human-Targeted Translation Edit Rate via Untrained Human Annotators

Omar F. Zaidan and Chris Callison-Burch

Improving Semantic Role Classification with Selectional Preferences

Beñat Zepirain, Eneko Agirre, Lluís Màrquez and Mihai Surdeanu

Wednesday, June 2, 2010 (continued)

Generalizing Syntactic Structures for Product Attribute Candidate Extraction

Yanyan Zhao, Bing Qin, Shen Hu and Ting Liu

Demonstrations

Note: Demo abstracts are located in a companion volume of the proceedings.

Camtology: Intelligent Information Access for Science

Ted Briscoe, Karl Harrison, Andrew Naish-Guzman, Andy Parker, Advait Siddharthan, David Sinclair, Mark Slater and Rebecca Watson

Summarizing Textual Information about Locations In a Geo-Spatial Information Display System

Congxing Cai and Eduard Hovy

Phrasal: A Statistical Machine Translation Toolkit for Exploring New Model Features

Daniel Cer, Michel Galley, Daniel Jurafsky and Christopher D. Manning

Multilingual Propbank Annotation Tools: Cornerstone and Jubilee

Jinho Choi, Claire Bonial and Martha Palmer

KSC-PaL: A Peer Learning Agent that Encourages Students to take the Initiative

Cynthia Kersey, Barbara Di Eugenio, Pamela Jordan and Sandra Katz

A Detailed, Accurate, Extensive, Available English Lexical Database

Adam Kilgarriff

An Interactive Tool for Supporting Error Analysis for Text Mining

Elijah Mayfield and Carolyn Penstein-Rosé

Serious Game Environments for Language and Culture Education

Alicia Sagae, W. Lewis Johnson and Rebecca Row

Interpretation of Partial Utterances in Virtual Human Dialogue Systems

Kenji Sagae, David DeVault and David Traum

Interactive Predictive Parsing using a Web-based Architecture

Ricardo Sánchez-Sáez, Luis A. Leiva, Joan-Andreu Sánchez and José-Miguel Benedí

SIMPLIFICA: a tool for authoring simplified texts in Brazilian Portuguese guided by readability assessments

Carolina Scarton, Matheus Oliveira, Arnaldo Candido Jr., Caroline Gasperin and Sandra Aluísio

Wednesday, June 2, 2010 (continued)

An Overview of Microsoft Web N-gram Corpus and Applications

Kuansan Wang, Chris Thrasher, Evelyne Viegas, Xiaolong Li and Bo-june (Paul) Hsu

6:30–8:30 Student Research Workshop Poster Session

Note: All student research workshop papers are located in a companion volume of the proceedings.

A Learning-based Sampling Approach to Extractive Summarization

Vishal Juneja, Sebastian Germesin and Thomas Kleinbauer

Temporal Relation Identification with Endpoints

Chong Min Lee

Identifying Opinion Holders and Targets with Dependency Parser in Chinese News Texts

Bin Lu

A Data Mining Approach to Learn Reorder Rules for SMT

Avinesh PVS

Fine-Tuning in Brazilian Portuguese-English Statistical Transfer Machine Translation: Verbal Tenses

Lucia Silva

Improving Syntactic Coordination Resolution using Language Modeling

Philip Ogren

On Automated Evaluation of Readability of Summaries: Capturing Grammaticality, Focus, Structure and Coherence

Ravikiran Vadlapudi and Rahul Katragadda

Extrinsic Parse Selection

David Goss-Grubbs

Towards a Matrix-based Distributional Model of Meaning

Eugenie Giesbrecht

Distinguishing Use and Mention in Natural Language

Shomir Wilson

Thursday, June 3, 2010

Best Paper Award Plenary Session

9:00–9:10 Best Paper Award: Introduction

9:10–9:40 *“cba to check the spelling”: Investigating Parser Performance on Discussion Forum Posts*
Jennifer Foster

9:40–10:15 *Coreference Resolution in a Modular, Entity-Centered Model*
Aria Haghighi and Dan Klein

10:15–10:45 **Break**

Machine Translation II

10:45–11:10 *Stream-based Translation Models for Statistical Machine Translation*
Abby Levenberg, Chris Callison-Burch and Miles Osborne

11:10–11:35 *Extracting Parallel Sentences from Comparable Corpora using Document Level Alignment*
Jason R. Smith, Chris Quirk and Kristina Toutanova

11:35–12:00 *Statistical Machine Translation of Texts with Misspelled Words*
Nicola Bertoldi, Mauro Cettolo and Marcello Federico

12:00–12:25 *Everybody loves a rich cousin: An empirical study of transliteration through bridge languages*
Mitesh M. Khapra, A Kumaran and Pushpak Bhattacharyya

Thursday, June 3, 2010 (continued)

Machine Learning I

- 10:45–11:10 *Discriminative Learning over Constrained Latent Representations*
Ming-Wei Chang, Dan Goldwasser, Dan Roth and Vivek Srikumar
- 11:10–11:35 *Some Empirical Evidence for Annotation Noise in a Benchmarked Dataset*
Beata Beigman Klebanov and Eyal Beigman
- 11:35–12:00 *Bayesian Inference for Finite-State Transducers*
David Chiang, Jonathan Graehl, Kevin Knight, Adam Pauls and Sujith Ravi
- 12:00–12:25 *Distributed Training Strategies for the Structured Perceptron*
Ryan McDonald, Keith Hall and Gideon Mann

Information Retrieval and Extraction I

- 10:45–11:10 *Term Weighting Schemes for Latent Dirichlet Allocation*
Andrew T. Wilson and Peter A. Chew
- 11:10–11:35 *Learning Dense Models of Query Similarity from User Click Logs*
Fabio De Bona, Stefan Riezler, Keith Hall, Massimiliano Ciaramita, Amaç Herdağdelen and Maria Holmqvist
- 11:35–12:00 *Learning to Link Entities with Knowledge Base*
Zhicheng Zheng, Fangtao Li, Minlie Huang and Xiaoyan Zhu
- 12:00–12:25 *Improving the Multilingual User Experience of Wikipedia Using Cross-Language Name Search*
Raghavendra Udupa and Mitesh M. Khapra

Thursday, June 3, 2010 (continued)

Morphology/Phonology

- 10:45–11:10 *Learning Words and Their Meanings from Unsegmented Child-directed Speech*
Bevan K. Jones, Mark Johnson and Michael C. Frank
- 11:10–11:35 *Subword Variation in Text Message Classification*
Robert Munro and Christopher D. Manning
- 11:35–12:00 *Automatic Diacritization for Low-Resource Languages Using a Hybrid Word and Consonant CMM*
Robbie Haertel, Peter McClanahan and Eric K. Ringger
- 12:00–12:25 *Urdu Word Segmentation*
Nadir Durrani and Sarmad Hussain

Lunch

- 12:40–2:00 Panel: *Recent and Future HLT Challenges in Industry*
- 12:25–2:15 Lunch

Machine Translation III

- 2:15–2:40 *Enabling Monolingual Translators: Post-Editing vs. Options*
Philipp Koehn
- 2:40–3:05 *Online Learning for Interactive Statistical Machine Translation*
Daniel Ortiz-Martínez, Ismael García-Varea and Francisco Casacuberta
- 3:05–3:30 *The Best Lexical Metric for Phrase-Based Statistical MT System Optimization*
Daniel Cer, Christopher D. Manning and Daniel Jurafsky

Thursday, June 3, 2010 (continued)

Machine Learning II

- 2:15–2:40 *Variational Inference for Adaptor Grammars*
Shay B. Cohen, David M. Blei and Noah A. Smith
- 2:40–3:05 *Type-Based MCMC*
Percy Liang, Michael I. Jordan and Dan Klein
- 3:05–3:30 *Painless Unsupervised Learning with Features*
Taylor Berg-Kirkpatrick, Alexandre Bouchard-Côté, John DeNero and Dan Klein

Generation

- 2:15–2:40 *Linguistic Steganography Using Automatically Generated Paraphrases*
Ching-Yun Chang and Stephen Clark
- 2:40–3:05 *Prenominal Modifier Ordering via Multiple Sequence Alignment*
Aaron Dunlop, Margaret Mitchell and Brian Roark
- 3:05–3:30 *Good Question! Statistical Ranking for Question Generation*
Michael Heilman and Noah A. Smith

Lexical Semantics

- 2:15–2:40 *Not All Seeds Are Equal: Measuring the Quality of Text Mining Seeds*
Zornitsa Kozareva and Eduard Hovy
- 2:40–3:05 *Extracting Glosses to Disambiguate Word Senses*
Weisi Duan and Alexander Yates
- 3:05–3:30 *Can Recognising Multiword Expressions Improve Shallow Parsing?*
Ioannis Korkontzelos and Suresh Manandhar

3:30–4:00 **Break**

Thursday, June 3, 2010 (continued)

Parsing: Short Papers

- 4:00–4:15 *A Simple Approach for HPSG Supertagging Using Dependency Information*
Yao-zhong Zhang, Takuya Matsuzaki and Jun'ichi Tsujii
- 4:15–4:30 *Ensemble Models for Dependency Parsing: Cheap and Good?*
Mihai Surdeanu and Christopher D. Manning
- 4:30–4:45 *Enlarged Search Space for SITG Parsing*
Guillem Gascó, Joan-Andreu Sánchez and José-Miguel Benedí
- 4:45–5:00 *Improving Data Driven Dependency Parsing using Clausal Information*
Phani Gadde, Karan Jindal, Samar Husain, Dipti Misra Sharma and Rajeev Sangal
- 5:00–5:15 *A Treebank Query System Based on an Extracted Tree Grammar*
Seth Kulick and Ann Bies
- 5:15–5:30 *Reranking the Berkeley and Brown Parsers*
Mark Johnson and Ahmet Engin Ural

Noisy Genre: Short Papers

- 4:00–4:15 *An Exploration of Off Topic Conversation*
Whitney L. Cade, Blair A. Lehman and Andrew Olney
- 4:15–4:30 *Making Conversational Structure Explicit: Identification of Initiation-response Pairs within Online Discussions*
Yi-Chia Wang and Carolyn P. Rosé
- 4:30–4:45 *Engaging learning groups using Social Interaction Strategies*
Rohit Kumar and Carolyn P. Rosé
- 4:45–5:00 *Using Entity-Based Features to Model Coherence in Student Essays*
Jill Burstein, Joel Tetreault and Slava Andreyev
- 5:00–5:15 *Summarizing Microblogs Automatically*
Beaux Sharifi, Mark-Anthony Hutton and Jugal Kalita

Thursday, June 3, 2010 (continued)

5:15–5:30 *Automatic Generation of Personalized Annotation Tags for Twitter Users*
Wei Wu, Bin Zhang and Mari Ostendorf

Morphology/Phonology: Short Papers

4:00–4:15 *Language identification of names with SVMs*
Aditya Bhargava and Grzegorz Kondrak

4:15–4:30 *Integrating Joint n-gram Features into a Discriminative Training Framework*
Sittichai Jiampojarn, Colin Cherry and Grzegorz Kondrak

4:30–4:45 *A Hybrid Morphologically Decomposed Factored Language Models for Arabic LVCSR*
Amr El-Desoky, Ralf Schlüter and Hermann Ney

4:45–5:00 *Is Arabic Part of Speech Tagging Feasible Without Word Segmentation?*
Emad Mohamed and Sandra Kübler

5:00–5:15 *Arabic Mention Detection: Toward Better Unit of Analysis*
Yassine Benajiba and Imed Zitouni

5:15–5:30 *An MDL-based approach to extracting subword units for grapheme-to-phoneme conversion*
Sravana Reddy and John Goldsmith

Machine Learning: Short Papers

4:00–4:15 *Extracting Phrase Patterns with Minimum Redundancy for Unsupervised Speaker Role Classification*
Bin Zhang, Brian Hutchinson, Wei Wu and Mari Ostendorf

4:15–4:30 *Classification of Prosodic Events using Quantized Contour Modeling*
Andrew Rosenberg

4:30–4:45 *Investigations into the Crandem Approach to Word Recognition*
Rohit Prabhavalkar, Preethi Jyothi, William Hartmann, Jeremy Morris and Eric Fosler-Lussier

4:45–5:00 *Constraint-Driven Rank-Based Learning for Information Extraction*
Sameer Singh, Limin Yao, Sebastian Riedel and Andrew McCallum

Thursday, June 3, 2010 (continued)

5:00–5:15 *Softmax-Margin CRFs: Training Log-Linear Models with Cost Functions*
Kevin Gimpel and Noah A. Smith

5:15–5:30 *Bitext-Based Resolution of German Subject-Object Ambiguities*
Florian Schwarck, Alexander Fraser and Hinrich Schütze

Friday, June 4, 2010

Plenary Session II

9:00–10:10 *Invited Talk: Music, Language, and Computational Modeling: Lessons from the Key-Finding Problem*
David Temperley

10:10–10:40 **Break**

Parsing II

10:40–11:05 *An Efficient Algorithm for Easy-First Non-Directional Dependency Parsing*
Yoav Goldberg and Michael Elhadad

11:05–11:30 *From Baby Steps to Leapfrog: How “Less is More” in Unsupervised Dependency Parsing*
Valentin I. Spitzkovsky, Hiyan Alshawi and Daniel Jurafsky

11:30–11:55 *Relaxed Marginal Inference and its Application to Dependency Parsing*
Sebastian Riedel and David A. Smith

11:55–12:20 *Optimal Parsing Strategies for Linear Context-Free Rewriting Systems*
Daniel Gildea

Friday, June 4, 2010 (continued)

Sentiment Analysis

- 10:40–11:05 *The viability of web-derived polarity lexicons*
Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan and Ryan McDonald
- 11:05–11:30 *Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables*
Tetsuji Nakagawa, Kentaro Inui and Sadao Kurohashi
- 11:30–11:55 *Convolution Kernels for Opinion Holder Extraction*
Michael Wiegand and Dietrich Klakow
- 11:55–12:20 *An Unsupervised Aspect-Sentiment Model for Online Reviews*
Samuel Brody and Noemie Elhadad

Information Retrieval and Extraction II

- 10:40–11:05 *Joint Inference for Knowledge Extraction from Biomedical Literature*
Hoifung Poon and Lucy Vanderwende
- 11:05–11:30 *Clinical Information Retrieval using Document and PICO Structure*
Florian Boudin, Jian-Yun Nie and Martin Dawes
- 11:30–11:55 *Topic Models for Image Annotation and Text Illustration*
Yansong Feng and Mirella Lapata
- 11:55–12:20 *Learning about Voice Search for Spoken Dialogue Systems*
Rebecca Passonneau, Susan L. Epstein, Tiziana Ligorio, Joshua B. Gordon and Pravin Bhutada

Friday, June 4, 2010 (continued)

Lunch

12:20–2:00 Lunch

1:00–2:00 NAACL Business Meeting

Machine Translation IV

2:00–2:25 *A Direct Syntax-Driven Reordering Model for Phrase-Based Machine Translation*
Niyu Ge

2:25–2:50 *Context-free reordering, finite-state translation*
Chris Dyer and Philip Resnik

2:50–3:15 *Improved Models of Distortion Cost for Statistical Machine Translation*
Spence Green, Michel Galley and Christopher D. Manning

3:15–3:40 *Why Synchronous Tree Substitution Grammars?*
Andreas Maletti

Summarization

2:00–2:25 *An extractive supervised two-stage method for sentence compression*
Dimitrios Galanis and Ion Androutsopoulos

2:25–2:50 *Interpretation and Transformation for Abstracting Conversations*
Gabriel Murray, Giuseppe Carenini and Raymond Ng

2:50–3:15 *Quantifying the Limits and Success of Extractive Summarization Systems Across Domains*
Hakan Ceylan, Rada Mihalcea, Umut Özertem, Elena Lloret and Manuel Palomar

3:15–3:40 *Multi-document Summarization via Budgeted Maximization of Submodular Functions*
Hui Lin and Jeff Bilmes

Friday, June 4, 2010 (continued)

Semantics II

2:00–2:25 *Cross-lingual Induction of Selectional Preferences with Bilingual Vector Spaces*
Yves Peirsman and Sebastian Padó

2:25–2:50 *Latent SVMs for Semantic Role Labeling using LTAG Derivation Trees*
Yudong Liu, Gholamreza Haffari and Anoop Sarkar

2:50–3:15 *Unsupervised Induction of Semantic Roles*
Joel Lang and Mirella Lapata

3:15–3:40 *Probabilistic Frame-Semantic Parsing*
Dipanjan Das, Nathan Schneider, Desai Chen and Noah A. Smith

3:40–4:00 **Break**

Machine Translation V

4:00–4:25 *Expected Sequence Similarity Maximization*
Cyril Allauzen, Shankar Kumar, Wolfgang Macherey, Mehryar Mohri and Michael Riley

4:25–4:50 *Accurate Non-Hierarchical Phrase-Based Translation*
Michel Galley and Christopher D. Manning

4:50–5:15 *Model Combination for Machine Translation*
John DeNero, Shankar Kumar, Ciprian Chelba and Franz Och

Friday, June 4, 2010 (continued)

Discourse

- 4:00–4:25 *Detecting Emails Containing Requests for Action*
Andrew Lampert, Robert Dale and Cecile Paris
- 4:25–4:50 *Evaluating Hierarchical Discourse Segmentation*
Lucien Carroll
- 4:50–5:15 *Reformulating Discourse Connectives for Non-Expert Readers*
Advaith Siddharthan and Napoleon Katsos

Semantics III

- 4:00–4:25 *Tree Edit Models for Recognizing Textual Entailments, Paraphrases, and Answers to Questions*
Michael Heilman and Noah A. Smith
- 4:25–4:50 *Syntactic/Semantic Structures for Textual Entailment Recognition*
Yashar Mehdad, Alessandro Moschitti and Fabio Massimo Zanzotto
- 4:50–5:15 *Automatic Metaphor Interpretation as a Paraphrasing Task*
Ekaterina Shutova
- 5:15–5:30 **End of Main Conference**