

EMNLP 2010

**Conference on  
Empirical Methods in Natural Language Processing**

**Proceedings of the Conference**

9-11 October 2010

USB memory sticks produced by  
*Omnipress Inc.*  
2600 Anderson Street  
Madison, WI 53707  
USA

©2010 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-932432-86-2 / 1-932432-86-8

Dedicated to  
**Fred Jelinek**  
(18 November 1932 – 14 September 2010)



## Preface

Welcome to EMNLP 2010, Conference on Empirical Methods in Natural Language Processing! The conference is organized by SIGDAT, the Association for Computational Linguistics' special interest group on linguistic data and corpus-based approaches to NLP, and is held this year as a stand-alone conference at the MIT Stata Center, Massachusetts, USA on October 9-11. <sup>1</sup>

EMNLP 2010 received 500 submissions, a new record for the conference. The program committee was able to accept 125 papers in total (an acceptance rate of 25%). Among them, 70 of the papers (14%) were accepted for oral presentations, and 55 (11%) for poster presentations. The PC, which consists of 18 area chairs and 460 PC members from Asia, Europe, and North America, worked together to create a strong program with high quality oral and poster presentations and enlightening invited talks.

First and foremost, we would like to thank the authors who submitted their work to EMNLP 2010. The sheer number of submissions reflects how broad and active our field is. We are deeply indebted to the area chairs and the PC members for their hard work. They enabled us to make a wonderful program and to provide valuable feedback to the authors. We are very grateful to our invited speakers Kevin Knight, Andrew Ng and Amit Singhal, who kindly agreed to give talks at EMNLP. Many thanks to local arrangements chair, Regina Barzilay, who has made the conference smoothly held at the wonderful venue of MIT; the publications chair, Eric Fosler-Lussier, who put this volume together with assistance from Preethi Jyothi and Rohit Prabhavalkar; best paper award committee chair, Jason Eisner, who lead the effort of selecting the best papers. Special thanks to David Yarowsky and Ken Church of SIGDAT, as well as Jason Eisner, Philipp Koehn, Rada Mihalcea, who provided much valuable advice and assistance in the past months. David Yarowsky also worked on the important issues of invitation letters and travel grants. We are most grateful to Priscilla Rasmussen who helped us with various logistic and organizational aspects of the conference. Rich Gerber and the START team responded to our questions quickly, and helped us manage the large number of submissions smoothly; we would like to thank them as well.

To enhance the quality of the conference, the program committee made a number of new efforts and continued some existing good practices in PC management, paper selection, and conference participation.

First, a strict process for selecting PC members was set up. PC members were first nominated by the area chairs; PC co-chairs then carefully checked whether they were qualified as reviewers by looking at their publications and academic experience. Second, a best reviewer award was created for the first time in this conference. After the reviewing process, 86 outstanding PC members were selected as the "best reviewers" as noted in the proceedings: they submitted all the reviews on time, made detailed, constructive, and helpful comments, and actively participated in the paper discussion when necessary. These recipients were first nominated by the area chairs and then endorsed by the PC co-chairs. The awards are a recognition of the reviewers who really did hard work in reviews.

In paper selection, we created an author response period, as in the two previous EMNLP conferences. Authors were able to read and respond to the reviews of their papers before the program committee made a final decision. They were asked to correct factual errors in the reviews and answer questions

---

<sup>1</sup>Conference web site: <http://www.lsi.upc.edu/events/emnlp2010/>.

raised in the reviewer comments. In some cases, reviewers changed their scores in view of the authors' response; the area chairs read all responses carefully prior to making recommendations for acceptance. As PC co-chairs, we did our best effort in final paper selection. The area chairs made accept/reject suggestions to us on the papers in their areas. We carefully examined all the cases, discussing with the area chairs the submissions on which we had divergent opinions. In some cases, the original suggestions by the area chairs were reversed after the discussions. The final paper selection was based on the consensus of the program committee. The accepted papers were further classified into oral and poster types, based on recommendations by the area chairs and discussions between the area chairs and us. Those papers that are suitable for presentations for the general audience (novel, inspiring, widely applicable, etc.) were selected as oral. After the paper notification, an independent committee for best paper awards led by Jason Eisner was created. From the candidates nominated by the PC, the committee performed independent reviews of the papers and made the best paper award selection, as listed in the proceedings.

Since EMNLP is held in the US, many participants needed visas to attend the conference. In order to help authors in the application for visas, a new procedure was implemented to anticipate the submission of invitation letters. The information on whether the authors needed a visa was collected at the paper submission time. The authors whose papers had average scores above a certain threshold were sent invitation letters by David Yarowsky, even before paper acceptance notification. In this way, the authors were given more time for their visa applications.

The success of a conference is really a result of the great efforts of everybody involved. We hope that you enjoy the conference at the fantastic building of MIT Stata Center!

Hang Li and Lluís Màrquez  
EMNLP 2010 Program Co-Chairs



**Program Co-chairs**

Hang Li, Microsoft Research Asia  
Lluís Màrquez, Technical University of Catalonia

**Area Chairs**

Eneko Agirre, University of the Basque Country  
Kalina Bontcheva, University of Sheffield  
Antal van den Bosch, Tilburg University  
David Chiang, USC Information Sciences Institute  
Jennifer Chu-Carroll, IBM Research  
Evgeniy Gabrilovich, Yahoo! Research  
Jianfeng Gao, Microsoft Research  
Julio Gonzalo, UNED  
Nancy Ide, Vassar College  
Haizhou Li, Institute for Infocomm Research  
Qun Liu, ICT, Chinese Academy of Sciences  
Yusuke Miyao, National Institute of Informatics  
Bo Pang, Yahoo! Research  
Dragomir Radev, University of Michigan  
Noah Smith, Carnegie Mellon University  
Kristina Toutanova, Microsoft Research  
Yoshimasa Tsuruoka, JAIST  
Yi Zhang, University of California, Santa Cruz

**Local Arrangements Chair**

Regina Barzilay, Massachusetts Institute of Technology

**Publications Chair**

Eric Fosler-Lussier, The Ohio State University

**Best Paper Award Committee**

Jason Eisner, Johns Hopkins University (Chair)  
Sharon Goldwater, University of Edinburgh  
Kevin Knight, University of Southern California  
Dan Roth, University of Illinois, Urbana-Champaign  
Ben Taskar, University of Pennsylvania  
Jun'ichi Tsujii, University of Tokyo  
Hanna Wallach, University of Massachusetts, Amherst

## Reviewers

Hua Ai, Jan Alexandersson, Enrique Alfonseca, Enrique Amigó, Ion Androustopoulos, Shlomo Argamon, Javier Artilles, Necip Fazil Ayan, Jing Bai, Collin Baker, Timothy Baldwin, Rafael Banchs, Carmen Banea, Shenghua Bao, Marco Baroni, Roberto Basili, Beata Beigman Klebanov, Núria Bel, Anja Belz, José-Miguel Benedí, Sabine Bergler, Mikhail Bilenko, John Blitzer, Phil Blunsom, Elizabeth Boschee, Jordan Boyd-Graber, SRK Branavan, Martin Braschler, Eric Breck, Christopher Brewster, Chris Brockett, Razvan Bunescu, Bill Byrne, Michael Cafarella, Aoife Cahill, Lynne Cahill, Chris Callison-Burch, Nicoletta Calzolari, Yunbo Cao, Claire Cardie, Andrew Carlson, Marine Carpuat, Xavier Carreras, John Carroll, Francisco Casacuberta, Asli Celikyilmaz, Joyce Chai, Raman Chandrasekar, Ming-Wei Chang, Yi Chang, Ciprian Chelba, Wenliang Chen, Colin Cherry, Jen-Tzung Chien, Hori Chiori, Yejin Choi, Ken Church, Massimiliano Ciaramita, Chris Cieri, Alex Clark, Stephen Clark, James Clarke, Paul Clough, Shay Cohen, Trevor Cohn, Nigel Collier, Michael Collins, Anna Corazza, Mark Core, Marta Costa-jussà, Josep M. Crego, Dan Cristea, Andras Csomai, Silviu-Petru Cucerzan, Hang Cui, Walter Daelemans, Robert Dale, Bob Damer, Cristian Danescu-Niculescu-Mizil, Hoa Dang, Van Dang, Sajib Dasgupta, Hal Daumé III, Adrià de Gispert, Guy De Pauw, Maarten de Rijke, Steve DeNeefe, John DeNero, Thierry Declerck, Pascal Denis, Tejaswini Deoskar, Mona Diab, Fernando Diaz, Anna Divoli, Bill Dolan, Pinar Donmez, Doug Downey, Markus Dreyer, Rebecca Dridan, Kevin Duh, Chris Dyer, Koji Eguchi, Maud Ehrmann, Jacob Eisenstein, Jason Eisner, Noemie Elhadad, Micha Elsner, Stephanie Elzer, Tomaz Erjavec, Hui Fang, Anna Feldman, Christiane Fellbaum, Denis Filimonov, Jenny Finkel, Dan Flickinger, Radu Florian, Mikel Forcada, Victoria Fossum, George Foster, Jennifer Foster, Alex Fraser, Marjorie Freedman, Maria Fuentes, Pascale Fung, Michel Galley, Michael Gamon, Wei Gao, Yuqing Gao, Claire Gardent, Nikesh Garera, Eric Gaussier, Gary Geunbae Lee, Dafydd Gibbon, Kevin Gimpel, Jesús Giménez, Roxana Girju, Natalie Glance, Randy Goebel, Genevieve Gorrell, Nancy Green, Stephan Greene, Gregory Grefenstette, Ralph Grishman, Jiafeng Guo, Iryna Gurevych, Le An Ha, Ben Hachey, Reza Haffari, Aria Haghighi, Udo Hahn, Dilek Hakkani-Tur, Keith Hall, Kazuo Hara, Sanda Harabagiu, Ahmed Hassan, Hany Hassan, Samer Hassan, Xiaodong He, James Henderson, Julia Hirschberg, Graeme Hirst, Kristy Hollingshead, Tracy Holloway King, Mark Hopkins, Veronique Hoste, Eduard Hovy, Paul Hsu, Pei-Yun Sabrina Hsueh, Yunhua Hu, Chu-Ren Huang, Liang Huang, Minlie Huang, Songfang Huang, Xuanjing Huang, Zhongqiang Huang, Diana Inkpen, Hideki Isozaki, Abe Ittycheriah, Heng Ji, Jing Jiang, Nitin Jindal, Hongyan Jing, Richard Johansson, Mark Johnson, Rie Johnson, Aravind Joshi, Laura Kallmeyer, Lauri Karttunen, Daisuke Kawahara, Junichi Kazama, Genichiro Kikui, Alexandre Klementiev, Kevin Knight, Philipp Koehn, Alek Kolcz, Greg Kondrak, Terry Koo, Moshe Koppel, Anna Korhonen, Zornitsa Kozareva, Geert-Jan Kruijff, Roland Kuhn, Ravi Kumar, Shankar Kumar, Oren Kurland, Sadao Kurohashi, Wai Lam, Guy Lapalme, Matt Lease, Guy Lebanon, James Lester, Lori Levin, Gina-Anne Levow, Mu Li, Shoushan Li, Wenjie Li, Xiao Li, Zhifei Li, Percy Liang, Chin-Yew Lin, Frank Lin, Ken Litkowski, Jingjing Liu, Yan Liu, Yang Liu, Yi Liu, Zhanyi Liu, Adam Lopez, Oier Lopez de Lacalle, Yajuan Lu, Yue Lu, Bin Ma, Yanjun Ma, Bill MacCartney, Craig Macdonald, Wolfgang Macherey, Nitin Madhani, Bernardo Magnini, Suresh Manandhar, Christopher Manning, Daniel Marcu, Mitch Marcus, James Martin, David Martinez, Andre Martins, Yuji Matsumoto, Takuya Matsuzaki, Mausam, Mike Maxwell, Jonathan May, Diana Maynard, Diana McCarthy, David McClosky, Kathy McCoy, Ryan McDonald, Kathy McKeown, Paul McNamee, Michael McTear, Qianzhu Mei, Donald Metzler, Haitao Mi, Tim Miller, Dunja Mladenic, Mary-Francine Moens, Saif Mohammad, Dan Moldovan, Ray Mooney, Alessandro Moschitti, Dragos Munteanu, Gabriel Murray, Masaaki Nagata, Tetsuji Nakagawa, Ramesh Nallapati, Jason Naradowsky, Tahira Naseem, Vivi Nastase, Tetsuya Nasukawa, Roberto Navigli, Ani Nenkova, Hermann Ney, Hwee Tou Ng, Vincent Ng, Jian-Yun Nie, Takashi Ninomiya, Kemal Oflazer, Daisuke Okanohara, Naoaki Okazaki, Manabu Okumura, Miles Osborne, Sebastian Padó, Lluís Padró, Martha Palmer, Sinno Jialin Pan, SooMin Pantel, Ivandre Paraboni, Cecile Paris, Jong Park, Kristen P. Parton, Jon Patrick, Siddharth Patwardhan, Adam Pauls, Fuchun Peng, Gerald Penn, Marco Pennacchiotti, Carol Peters, Slav Petrov, Emily Pitler, Ferran Pla, Massimo Poesio, Joe Polifroni, Simone Paolo Ponzetto, Hoifung Poon, Maja Popovic, John Prager, Stephen Pulman, James Pustejovsky, Sampo Pyysalo, Vahed Qazvinian, Chris Quirk, Stephan Raaijmakers, Ari Rappoport, Manny Rayner, Joseph Reisinger, Philip Resnik, Martin Reynaert,



## **Reviewers (continued)**

Giuseppe Riccardi, Sebastian Riedel, Jason Riesa, Stefan Riezler, German Rigau, Ellen Riloff, Laura Rimell, Christoph Ringlstetter, Alan Ritter, Brian Roark, Mats Rooth, Barbara Rosario, Paolo Rosso, Antti-Veikko Rosti, Kenji Sagae, Horacio Saggion, Rajeev Sangal, Murat Saraclar, Ruhi Sarikaya, Anoop Sarkar, Manabu Sassano, Helmut Schmid, Lane Schwartz, Holger Schwenk, Hendra Setiawan, Burr Settles, Dou Shen, Libin Shen, Wade Shen, Shuming Shi, Luo Si, Advaith Siddharthan, Michel Simard, Khalil Sima'an, Matthew Snover, Ben Snyder, Marina Sokolova, Swapna Somasundaran, Youngin Song, Lucia Specia, Caroline Sporleder, Amanda Stent, Mark Stevenson, Veselin Stoyanov, Michael Strube, Jian Su, Keh-Yih Su, Eiichiro Sumita, Jun Sun, Maosong Sun, Mihai Surdeanu, Charles Sutton, Hisami Suzuki, Krysta Svore, Stan Szpakowicz, Idan Szpektor, Maite Taboada, Hiroya Takamura, Jie Tang, Joel Tetreault, Mariet Theune, Christoph Tillmann, Reut Tsarfaty, Benjamin Tsou, Hajime Tsukada, Dan Tufis, Gokhan Tur, Jordi Turmo, Raghavendra Udupa, Takehito Utsuro, Tim Van de Cruys, Josef van Genabith, Gertjan van Noord, Lucy Vanderwende, Vasudeva Varma, Olga Vechtomova, Ashish Venugopal, Juan Vilar, Andreas Vlachos, Stephen Wan, Xiaojun Wan, Dingding Wang, Haifeng Wang, Wei Wang, Xuanhui Wang, Ye-Yi Wang, Taro Watanabe, Bonnie Webber, Ralph Weischedel, Ji-Rong Wen, Jan Wiebe, Theresa Wilson, Shuly Wintner, Tak-Lam Wong, Dekai Wu, Fei Wu, Hua Wu, Fei Xia, Deyi Xiong, Gu Xu, Jinxi Xu, Peng Xu, Nianwen Xue, Xiaobing Xue, Hirofumi Yamamoto, Kazuhide Yamamoto, Hui Yang, Tao Yang, Roman Yangarber, Alexander Yates, Xing Yi, Scott Yih, Bei Yu, Kun Yu, François Yvon, Jun Zhao, Annie Zaenen, Fabio Massimo Zanzotto, Vanni Zavarella, Jakub Zavrel, Richard Zens, Torsten Zesch, Luke Zettlemoyer, ChengXiang Zhai, Duo Zhang, Hao Zhang, Hui Zhang, Min Zhang, Ying Zhang, Yue Zhang, Guodong Zhou, Jun Zhu, Imed Zitouni, Andreas Zollmann, Chengqing Zong

## **Secondary Reviewers**

Sadaf Abdul-Rauf, Masayuki Asahara, Shen Bin, Yassine Benajiba, Shane Bergsma, Aditya Bhargava, Jiang Bian, Eduardo Blanco, Kristy Elizabeth Boyer, Pere R. Comas, Xiangyu Duan, Aaron Dunlop, Ji Fang, Yi Fang, Wei Gao, Edgar González, Charles Greenbacker, Eun Ha, Stefan Hahn, Zhongjun He, Yun Huang, Adrian Iftene, Tomoya Iwakura, Jiarong Jiang, Bilal Khaliq, Mi-Young Kim, Ioannis Korkontzelos, Sai Krishna, Gregor Leban, Sophia Yat Mei Lee, Shoushan Li, Shuguang Li, Changsong Liu, Shui Liu, Bin Lu, Cynthia Matuszek, Blaz Novak, Antoni Olivé, Elias Ponvert, Rob Phillips, Long Qiu, Longhua Qian, Rodrygo Santos, Kiran Sarvabhotla, Daniel Stein, Qi Su, György Szarvas, David Talbot, Diana Trandabăţ, Emily Tucker Prud'hommeaux, Damiano Spina Valenti, Cristina Vertan, Ying Xu, Shuang-Hong Yang, Limin Yao, Hui Zhang, Lanbo Zhang, Jiajun Zhang

## Invited talks

### “Why do we call it decoding?”

**Kevin Knight, Information Sciences Institute, University of Southern California**

The first natural language processing systems had a straightforward goal — decipher coded messages sent by the enemy. Sixty years later, we have many more applications! These include web search, question answering, summarization, speech recognition, and language translation. This talk explores connections between early decipherment research and today’s work. We find that many ideas from the earlier era have become core to the field, while others still remain to be picked up and developed.

### “Unsupervised feature learning and Deep Learning”

**Andrew Ng, Computer Science Department, Stanford University**

Machine learning has seen numerous successes, but applying learning algorithms today often means spending a long time laboriously hand-engineering the input feature representation. This is often true for learning in NLP, vision, audio, and many other problems. To address this, recently in machine learning there has been significant interest in unsupervised feature learning algorithms, including “deep learning” algorithms, that can automatically learn rich feature representations from unlabeled data. These algorithms build on such ideas as sparse coding, ICA, and deep belief networks, and have proved very effective for learning good feature representations in many problems. Since these algorithms mostly learn from unlabeled data, they also have the potential to learn from vastly increased amounts of data (since unlabeled data is cheap), and therefore perhaps also achieving vastly improved performance. In this talk, I’ll survey the key ideas in this nascent area of unsupervised feature learning and deep learning. I’ll outline a few algorithms, and describe a few successful applications of these ideas to problems in NLP, audio/speech, vision, and other problems.

### “Challenges in running a commercial search engine”

**Amit Singhal, Google, Inc.**

These are exciting times for Information Retrieval and NLP. Web search engines have brought IR to the masses. It now affects the lives of hundreds of millions of people, and growing, as Internet search companies launch ever more products based on techniques developed in IR and NLP research. The real world poses unique challenges for search algorithms. They operate at unprecedented scales, and over a wide diversity of information. In addition, we have entered an unprecedented world of “Adversarial Information Retrieval.” The lure of billions of dollars of commerce, guided by search engines, motivates all kinds of people to try all kinds of tricks to get their sites to the top of the search results. What techniques do people use to defeat IR algorithms? What are the evaluation challenges for a web search engine? How much impact has IR had on search engines? How does Google serve over 250 Million queries a day, often with sub-second response times? This talk will show that the world of algorithm and system design for commercial search engines can be described by two of Murphy’s Laws: a) If anything can go wrong, it will, and b) If anything cannot go wrong, it will anyway.

## Fred Jelinek Best Paper Award

The EMNLP-2010 Best Paper Award is named in memory of Fred Jelinek (18 Nov. 1932 - 14 Sept. 2010).

### “Dual Decomposition for Parsing with Non-Projective Head Automata,”

Terry Koo, Alexander M. Rush, Michael Collins, Tommi Jaakkola and David Sontag

This paper introduces algorithms for non-projective parsing based on *dual decomposition*. We focus on parsing algorithms for *non-projective head automata*, a generalization of head-automata models to non-projective structures. The dual decomposition algorithms are simple and efficient, relying on standard dynamic programming and minimum spanning tree algorithms. They provably solve an LP relaxation of the non-projective parsing problem. Empirically the LP relaxation is very often tight: for many languages, exact solutions are achieved on over 98% of test sentences. The accuracy of our models is higher than previous work on a broad range of datasets.

### Best Reviewer Awards

Alex Clark	François Yvon	Maria Fuentes	Rebecca Dridan
Amanda Stent	Greg Grefenstette	Mariet Theune	Rie Johnson
André Martins	Greg Kondrak	Martin Reynaert	Robert Dale
Andreas Vlachos	Helmut Schmid	Matt Lease	Roland Kuhn
Anja Belz	Ion Androutsopoulos	Maud Ehrmann	Sabine Bergler
Anna Korhonen	Jan Wiebe	Mausam	Sampo Pyysalo
Aoife Cahill	James Lester	Michael Cafarella	Sebastian Riedel
Beata Beigman Klebanov	Jason Eisner	Michael Collins	Shuly Wintner
Ben Hachey	John Blitzer	Michael Gamon	Simone Paolo Ponzetto
Bill MacCartney	John Prager	Michael Strube	Songfang Huang
Bob Dampier	Jonathan May	Mike Maxwell	Stephen Wan
Cecile Paris	Kathy McCoy	Mikel Forcada	Steve DeNeefe
ChengXiang Zhai	Kemal Oflazer	Misha Bilenko	Swapna Somasundaran
Chin-Yew Lin	Kenji Sagae	Natalie Glance	Torsten Zesch
Chris Brockett	Kevin Duh	Nigel Collier	Tracy Holloway King
Chris Cieri	Kristen P. Parton	Nitin Madnani	Veselin Stoyanov
Craig Macdonald	Laura Kallmeyer	Oren Kurland	Vincent Ng
Diana Maynard	Laura Rimell	Paul Hsu	Wei Gao
Don Metzler	Liang Huang	Paul McNamee	Xiaojun Wan
Ellen Riloff	Lucy Vanderwende	Radu Florian	Zhongqiang Huang
Enrique Alfonseca	Luke Zettlemoyer	Raghavendra Udupa	
Eric Breck	Marco Baroni	Raman Chandrasekar	



## Table of Contents

<i>On Dual Decomposition and Linear Programming Relaxations for Natural Language Processing</i> Alexander M Rush, David Sontag, Michael Collins and Tommi Jaakkola .....	1
<i>Self-Training with Products of Latent Variable Grammars</i> Zhongqiang Huang, Mary Harper and Slav Petrov .....	12
<i>Utilizing Extra-Sentential Context for Parsing</i> Jackie Chi Kit Cheung and Gerald Penn .....	23
<i>Turbo Parsers: Dependency Parsing by Approximate Variational Inference</i> Andre Martins, Noah Smith, Eric Xing, Pedro Aguiar and Mario Figueiredo .....	34
<i>Holistic Sentiment Analysis Across Languages: Multilingual Supervised Latent Dirichlet Allocation</i> Jordan Boyd-Graber and Philip Resnik .....	45
<i>Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid</i> Xin Zhao, Jing Jiang, Hongfei Yan and Xiaoming Li .....	56
<i>Summarizing Contrastive Viewpoints in Opinionated Text</i> Michael Paul, ChengXiang Zhai and Roxana Girju .....	66
<i>Automatically Producing Plot Unit Representations for Narrative Text</i> Amit Goyal, Ellen Riloff and Hal Daume III .....	77
<i>Handling Noisy Queries in Cross Language FAQ Retrieval</i> Danish Contractor, Govind Kothari, Tanveer Faruque, L V Subramaniam and Sumit Negi ....	87
<i>Learning the Relative Usefulness of Questions in Community QA</i> Razvan Bunescu and Yunfeng Huang .....	97
<i>Positional Language Models for Clinical Information Retrieval</i> Florian Boudin, Jian-Yun Nie and Martin Dawes .....	108
<i>Inducing Word Senses to Improve Web Search Result Clustering</i> Roberto Navigli and Giuseppe Crisafulli .....	116
<i>Improving Translation via Targeted Paraphrasing</i> Philip Resnik, Olivia Buzek, Chang Hu, Yakov Kronrod, Alex Quinn and Benjamin B. Bederson	127
<i>Soft Syntactic Constraints for Hierarchical Phrase-Based Translation Using Latent Syntactic Distributions</i> Zhongqiang Huang, Martin Cmejrek and Bowen Zhou .....	138
<i>A Hybrid Morpheme-Word Representation for Machine Translation of Morphologically Rich Languages</i> Minh-Thang Luong, Preslav Nakov and Min-Yen Kan .....	148

<i>”Poetic” Statistical Machine Translation: Rhyme and Meter</i>	
Dmitriy Genzel, Jakob Uszkoreit and Franz Och . . . . .	158
<i>Efficient Graph-Based Semi-Supervised Learning of Structured Tagging Models</i>	
Amarnag Subramanya, Slav Petrov and Fernando Pereira . . . . .	167
<i>Better Punctuation Prediction with Dynamic Conditional Random Fields</i>	
Wei Lu and Hwee Tou Ng . . . . .	177
<i>Joint Training and Decoding Using Virtual Nodes for Cascaded Segmentation and Tagging Tasks</i>	
Xian Qian, Qi Zhang, Yaqian Zhou, Xuanjing Huang and Lide Wu . . . . .	187
<i>Crouching Dirichlet, Hidden Markov Model: Unsupervised POS Tagging with Context Local Tag Generation</i>	
Taesun Moon, Katrin Erk and Jason Baldridge . . . . .	196
<i>Improving Gender Classification of Blog Authors</i>	
Arjun Mukherjee and Bing Liu . . . . .	207
<i>Negative Training Data Can be Harmful to Text Classification</i>	
Xiao-Li Li, Bing Liu and See-Kiong Ng . . . . .	218
<i>Modeling Organization in Student Essays</i>	
Isaac Persing, Alan Davis and Vincent Ng . . . . .	229
<i>Evaluating Models of Latent Document Semantics in the Presence of OCR Errors</i>	
Daniel Walker, William B. Lund and Eric K. Ringger . . . . .	240
<i>Translingual Document Representations from Discriminative Projections</i>	
John Platt, Kristina Toutanova and Wen-tau Yih . . . . .	251
<i>Storing the Web in Memory: Space Efficient Language Models with Constant Time Retrieval</i>	
David Guthrie and Mark Hepple . . . . .	262
<i>Efficient Incremental Decoding for Tree-to-String Translation</i>	
Liang Huang and Haitao Mi . . . . .	273
<i>Modeling Perspective Using Adaptor Grammars</i>	
Eric Hardisty, Jordan Boyd-Graber and Philip Resnik . . . . .	284
<i>Predicting the Semantic Compositionality of Prefix Verbs</i>	
Shane Bergsma, Aditya Bhargava, Hua He and Grzegorz Kondrak . . . . .	293
<i>Joint Inference for Bilingual Semantic Role Labeling</i>	
Tao Zhuang and Chengqing Zong . . . . .	304
<i>Automatic Discovery of Manner Relations and its Applications</i>	
Eduardo Blanco and Dan Moldovan . . . . .	315

<i>Tense Sense Disambiguation: A New Syntactic Polysemy Task</i> Roi Reichart and Ari Rappoport .....	325
<i>Improving Mention Detection Robustness to Noisy Input</i> Radu Florian, John Pitrelli, Salim Roukos and Imed Zitouni .....	335
<i>Clustering-Based Stratified Seed Sampling for Semi-Supervised Relation Classification</i> Longhua Qian and Guodong Zhou .....	346
<i>Unsupervised Discovery of Negative Categories in Lexicon Bootstrapping</i> Tara McIntosh .....	356
<i>Automatic Keyphrase Extraction via Topic Decomposition</i> Zhiyuan Liu, Wenyi Huang, Yabin Zheng and Maosong Sun .....	366
<i>Incorporating Content Structure into Text Analysis Applications</i> Christina Sauper, Aria Haghighi and Regina Barzilay .....	377
<i>Exploiting Conversation Structure in Unsupervised Topic Segmentation for Emails</i> Shafiq Joty, Giuseppe Carenini, Gabriel Murray and Raymond T. Ng .....	388
<i>A Semi-Supervised Approach to Improve Classification of Infrequent Discourse Relations Using Feature Vector Extension</i> Hugo Hernault, Danushka Bollegala and Mitsuru Ishizuka .....	399
<i>A Game-Theoretic Approach to Generating Spatial Descriptions</i> Dave Golland, Percy Liang and Dan Klein .....	410
<i>Facilitating Translation Using Source Language Paraphrase Lattices</i> Jinhua Du, Jie Jiang and Andy Way .....	420
<i>Mining Name Translations from Entity Graph Mapping</i> Gae-won You, Seung-won Hwang, Young-In Song, Long Jiang and Zaiqing Nie .....	430
<i>Non-Isomorphic Forest Pair Translation</i> Hui Zhang, Min Zhang, Haizhou Li and Eng Siong Chng .....	440
<i>Discriminative Instance Weighting for Domain Adaptation in Statistical Machine Translation</i> George Foster, Cyril Goutte and Roland Kuhn .....	451
<i>NLP on Spoken Documents Without ASR</i> Mark Dredze, Aren Jansen, Glen Coppersmith and Ken Church .....	460
<i>Fusing Eye Gaze with Speech Recognition Hypotheses to Resolve Exophoric References in Situated Dialogue</i> Zahar Prasov and Joyce Y. Chai .....	471
<i>Multi-Document Summarization Using A* Search and Discriminative Learning</i> Ahmet Aker, Trevor Cohn and Robert Gaizauskas .....	482

<i>A Multi-Pass Sieve for Coreference Resolution</i>	
Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nate Chambers, Mihai Surdeanu, Dan Jurafsky and Christopher Manning .....	492
<i>A Simple Domain-Independent Probabilistic Approach to Generation</i>	
Gabor Angeli, Percy Liang and Dan Klein .....	502
<i>Title Generation with Quasi-Synchronous Grammar</i>	
Kristian Woodsend, Yansong Feng and Mirella Lapata .....	513
<i>Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation</i>	
Erica Greene, Tugba Bodrumlu and Kevin Knight .....	524
<i>Discriminative Word Alignment with a Function Word Reordering Model</i>	
Hendra Setiawan, Chris Dyer and Philip Resnik .....	534
<i>Hierarchical Phrase-Based Translation Grammars Extracted from Alignment Posterior Probabilities</i>	
Adrià de Gispert, Juan Pino and William Byrne .....	545
<i>Maximum Entropy Based Phrase Reordering for Hierarchical Phrase-Based Translation</i>	
Zhongjun He, Yao Meng and Hao Yu .....	555
<i>Further Meta-Evaluation of Broad-Coverage Surface Realization</i>	
Dominic Espinosa, Rajakrishnan Rajkumar, Michael White and Shoshana Berleant .....	564
<i>Two Decades of Unsupervised POS Induction: How Far Have We Come?</i>	
Christos Christodoulopoulos, Sharon Goldwater and Mark Steedman .....	575
<i>We're Not in Kansas Anymore: Detecting Domain Changes in Streams</i>	
Mark Dredze, Tim Oates and Christine Piatko .....	585
<i>A Fast Fertility Hidden Markov Model for Word Alignment Using MCMC</i>	
Shaojun Zhao and Daniel Gildea .....	596
<i>Minimum Error Rate Training by Sampling the Translation Lattice</i>	
Samidh Chatterjee and Nicola Cancedda .....	606
<i>Statistical Machine Translation with a Factorized Grammar</i>	
Libin Shen, Bing Zhang, Spyros Matsoukas, Jinxi Xu and Ralph Weischedel .....	616
<i>Discriminative Sample Selection for Statistical Machine Translation</i>	
Sankaranarayanan Ananthakrishnan, Rohit Prasad, David Stallard and Prem Natarajan .....	626
<i>Effects of Empty Categories on Machine Translation</i>	
Tagyoung Chung and Daniel Gildea .....	636
<i>SCFG Decoding Without Binarization</i>	
Mark Hopkins and Greg Langmead .....	646



<i>Example-Based Paraphrasing for Improved Phrase-Based Statistical Machine Translation</i>	
Aurélien Max .....	656
<i>Combining Unsupervised and Supervised Alignments for MT: An Empirical Study</i>	
Jinxi Xu and Antti-Veikko Rosti .....	667
<i>Top-Down Nearly-Context-Sensitive Parsing</i>	
Eugene Charniak .....	674
<i>Improved Fully Unsupervised Parsing with Zoomed Learning</i>	
Roi Reichart and Ari Rappoport .....	684
<i>Unsupervised Parse Selection for HPSG</i>	
Rebecca Dridan and Timothy Baldwin .....	694
<i>Uptraining for Accurate Deterministic Question Parsing</i>	
Slav Petrov, Pi-Chuan Chang, Michael Ringgaard and Hiyan Alshawi .....	705
<i>A Unified Framework for Scope Learning via Simplified Shallow Semantic Parsing</i>	
Qiaoming Zhu, Junhui Li, Hongling Wang and Guodong Zhou .....	714
<i>A New Approach to Lexical Disambiguation of Arabic Text</i>	
Rushin Shah, Paramveer S. Dhillon, Mark Liberman, Dean Foster, Mohamed Maamouri and Lyle Ungar .....	725
<i>What a Parser Can Learn from a Semantic Role Labeler and Vice Versa</i>	
Stephen Boxwell, Dennis Mehay and Chris Brew .....	736
<i>Word Sense Induction &amp; Disambiguation Using Hierarchical Random Graphs</i>	
Ioannis Klapaftis and Suresh Manandhar .....	745
<i>Towards Conversation Entailment: An Empirical Investigation</i>	
Chen Zhang and Joyce Chai .....	756
<i>The Necessity of Combining Adaptation Methods</i>	
Ming-Wei Chang, Michael Connor and Dan Roth .....	767
<i>Training Continuous Space Language Models: Some Practical Issues</i>	
Hai Son Le, Alexandre Allauzen, Guillaume Wisniewski and François Yvon .....	778
<i>Enhancing Domain Portability of Chinese Segmentation Model Using Chi-Square Statistics and Bootstrapping</i>	
Baobao Chang and Dongxu Han .....	789
<i>Latent-Descriptor Clustering for Unsupervised POS Induction</i>	
Michael Lamar, Yariv Maron and Elie Bienenstock .....	799
<i>A Probabilistic Morphological Analyzer for Syriac</i>	
Peter McClanahan, George Busby, Robbie Haertel, Kristian Heal, Deryle Lonsdale, Kevin Seppi and Eric Ringger .....	810

<i>Lessons Learned in Part-of-Speech Tagging of Conversational Speech</i> Vladimir Eidelman, Zhongqiang Huang and Mary Harper .....	821
<i>An Efficient Algorithm for Unsupervised Word Segmentation with Branching Entropy and MDL</i> Valentin Zhikov, Hiroya Takamura and Manabu Okumura .....	832
<i>A Fast Decoder for Joint Word Segmentation and POS-Tagging Using a Single Discriminative Model</i> Yue Zhang and Stephen Clark .....	843
<i>Simple Type-Level Unsupervised POS Tagging</i> Yoong Keok Lee, Aria Haghighi and Regina Barzilay .....	853
<i>Classifying Dialogue Acts in One-on-One Live Chats</i> Su Nam Kim, Lawrence Cavedon and Timothy Baldwin .....	862
<i>Resolving Event Noun Phrases to Their Verbal Mentions</i> Bin Chen, Jian Su and Chew Lim Tan .....	872
<i>A Tree Kernel-Based Unified Framework for Chinese Zero Anaphora Resolution</i> Fang Kong and Guodong Zhou .....	882
<i>Automatic Comma Insertion for Japanese Text Generation</i> Masaki Murata, Tomohiro Ohno and Shigeki Matsubara .....	892
<i>Using Unknown Word Techniques to Learn Known Words</i> Kostadin Cholakov and Gertjan van Noord .....	902
<i>WikiWars: A New Corpus for Research on Temporal Expressions</i> Pawel Mazur and Robert Dale .....	913
<i>PEM: A Paraphrase Evaluation Metric Exploiting Parallel Texts</i> Chang Liu, Daniel Dahlmeier and Hwee Tou Ng .....	923
<i>Assessing Phrase-Based Translation Models with Oracle Decoding</i> Guillaume Wisniewski, Alexandre Allauzen and François Yvon .....	933
<i>Automatic Evaluation of Translation Quality for Distant Language Pairs</i> Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh and Hajime Tsukada .....	944
<i>An Approach of Generating Personalized Views from Normalized Electronic Dictionaries : A Practical Experiment on Arabic Language</i> Aida Khemakhem, Bilel Gargouri and Abdelmajid Ben Hamadou .....	953
<i>Generating Confusion Sets for Context-Sensitive Error Correction</i> Alla Rozovskaya and Dan Roth .....	961
<i>Confidence in Structured-Prediction Using Confidence-Weighted Models</i> Avihai Mejer and Koby Crammer .....	971

<i>Evaluating the Impact of Alternative Dependency Graph Encodings on Solving Event Extraction Tasks</i> Ekaterina Buyko and Udo Hahn.....	982
<i>Enhancing Mention Detection Using Projection via Aligned Corpora</i> Yassine Benajiba and Imed Zitouni.....	993
<i>Domain Adaptation of Rule-Based Annotators for Named-Entity Recognition Tasks</i> Laura Chiticariu, Rajasekar Krishnamurthy, Yunyao Li, Frederick Reiss and Shivakumar Vaithyanathan	1002
<i>Collective Cross-Document Relation Extraction Without Labelled Data</i> Limin Yao, Sebastian Riedel and Andrew McCallum.....	1013
<i>Automatic Detection and Classification of Social Events</i> Apoorv Agarwal and Owen Rambow.....	1024
<i>Extracting Opinion Targets in a Single and Cross-Domain Setting with Conditional Random Fields</i> Niklas Jakob and Iryna Gurevych.....	1035
<i>Multi-Level Structured Models for Document-Level Sentiment Classification</i> Ainur Yessenalina, Yisong Yue and Claire Cardie.....	1046
<i>Cross Language Text Classification by Model Translation and Semi-Supervised Learning</i> Lei Shi, Rada Mihalcea and Mingjun Tian.....	1057
<i>SRL-Based Verb Selection for ESL</i> Xiaohua Liu, Bo Han, Kuan Li, Stephan Hyeonjun Stiller and Ming Zhou.....	1068
<i>Context Comparison of Bursty Events in Web Search and Online Media</i> Yunliang Jiang, Cindy Xide Lin and Qiaozhu Mei.....	1077
<i>Learning First-Order Horn Clauses from Web Text</i> Stefan Schoenmackers, Jesse Davis, Oren Etzioni and Daniel Weld.....	1088
<i>Constraints Based Taxonomic Relation Classification</i> Quang Do and Dan Roth.....	1099
<i>A Semi-Supervised Method to Learn and Construct Taxonomies Using the Web</i> Zornitsa Kozareva and Eduard Hovy.....	1110
<i>Function-Based Question Classification for General QA</i> Fan Bu, Xingwei Zhu, Yu Hao and Xiaoyan Zhu.....	1119
<i>Learning Recurrent Event Queries for Web Search</i> Ruiqiang Zhang, Yuki Konda, Anlei Dong, Pranam Kolari, Yi Chang and Zhaohui Zheng...	1129
<i>Staying Informed: Supervised and Semi-Supervised Multi-View Topical Analysis of Ideological Perspective</i> Amr Ahmed and Eric Xing.....	1140

<i>Word-Based Dialect Identification with Georeferenced Rules</i>	
Yves Scherrer and Owen Rambow .....	1151
<i>Measuring Distributional Similarity in Context</i>	
Georgiana Dinu and Mirella Lapata .....	1162
<i>A Mixture Model with Sharing for Lexical Semantics</i>	
Joseph Reisinger and Raymond Mooney .....	1173
<i>Nouns are Vectors, Adjectives are Matrices: Representing Adjective-Noun Constructions in Semantic Space</i>	
Marco Baroni and Roberto Zamparelli .....	1183
<i>Practical Linguistic Steganography Using Contextual Synonym Substitution and Vertex Colour Coding</i>	
Ching-Yun Chang and Stephen Clark .....	1194
<i>Unsupervised Induction of Tree Substitution Grammars for Dependency Parsing</i>	
Phil Blunsom and Trevor Cohn .....	1204
<i>It Depends on the Translation: Unsupervised Dependency Parsing via Word Alignment</i>	
Samuel Brody .....	1214
<i>Inducing Probabilistic CCG Grammars from Logical Form with Higher-Order Unification</i>	
Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater and Mark Steedman .....	1223
<i>Using Universal Linguistic Knowledge to Guide Grammar Induction</i>	
Tahira Naseem, Harr Chen, Regina Barzilay and Mark Johnson .....	1234
<i>What's with the Attitude? Identifying Sentences with Attitude in Online Discussions</i>	
Ahmed Hassan, Vahed Qazvinian and Dragomir Radev .....	1245
<i>Hashing-Based Approaches to Spelling Correction of Personal Names</i>	
Raghavendra Udupa and Shaishav Kumar .....	1256
<i>Identifying Functional Relations in Web Text</i>	
Thomas Lin, Mausam and Oren Etzioni .....	1266
<i>A Latent Variable Model for Geographic Lexical Variation</i>	
Jacob Eisenstein, Brendan O'Connor, Noah A. Smith and Eric P. Xing .....	1277
<i>Dual Decomposition for Parsing with Non-Projective Head Automata</i>	
Terry Koo, Alexander M. Rush, Michael Collins, Tommi Jaakkola and David Sontag .....	1288

# Conference Program

**Saturday, October 9, 2010**

8:50–9:00      Opening Remarks

9:00–10:00    Invited talk: Kevin Knight, “Why do we call it decoding?”  
CHAIR: HANG LI

10:00–10:30   Break

## **Session 1A: Syntactic Parsing and Machine Learning**

CHAIR: XAVIER CARRERAS

10:30–10:55   *On Dual Decomposition and Linear Programming Relaxations for Natural Language Processing*

Alexander M Rush, David Sontag, Michael Collins and Tommi Jaakkola

10:55–11:20   *Self-Training with Products of Latent Variable Grammars*

Zhongqiang Huang, Mary Harper and Slav Petrov

11:20–11:45   *Utilizing Extra-Sentential Context for Parsing*

Jackie Chi Kit Cheung and Gerald Penn

11:45–12:10   *Turbo Parsers: Dependency Parsing by Approximate Variational Inference*

Andre Martins, Noah Smith, Eric Xing, Pedro Aguiar and Mario Figueiredo

## **Session 1B: Sentiment Analysis and Opinion Mining**

CHAIR: HANNA WALLACH

10:30–10:55   *Holistic Sentiment Analysis Across Languages: Multilingual Supervised Latent Dirichlet Allocation*

Jordan Boyd-Graber and Philip Resnik

10:55–11:20   *Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid*

Xin Zhao, Jing Jiang, Hongfei Yan and Xiaoming Li

11:20–11:45   *Summarizing Contrastive Viewpoints in Opinionated Text*

Michael Paul, ChengXiang Zhai and Roxana Girju

**Saturday, October 9, 2010 (continued)**

11:45–12:10 *Automatically Producing Plot Unit Representations for Narrative Text*  
Amit Goyal, Ellen Riloff and Hal Daume III

**Session 1C: Information Retrieval and Question Answering**

CHAIR: JING JIANG

10:30–10:55 *Handling Noisy Queries in Cross Language FAQ Retrieval*  
Danish Contractor, Govind Kothari, Tanveer Faruque, L V Subramaniam and Sumit Negi

10:55–11:20 *Learning the Relative Usefulness of Questions in Community QA*  
Razvan Bunescu and Yunfeng Huang

11:20–11:45 *Positional Language Models for Clinical Information Retrieval*  
Florian Boudin, Jian-Yun Nie and Martin Dawes

11:45–12:10 *Inducing Word Senses to Improve Web Search Result Clustering*  
Roberto Navigli and Giuseppe Crisafulli

12:10–14:10 Lunch

**Session 2A: Machine Translation I**

CHAIR: DANIEL MARCU

14:10–14:35 *Improving Translation via Targeted Paraphrasing*  
Philip Resnik, Olivia Buzek, Chang Hu, Yakov Kronrod, Alex Quinn and Benjamin B. Bederson

14:35–15:00 *Soft Syntactic Constraints for Hierarchical Phrase-Based Translation Using Latent Syntactic Distributions*  
Zhongqiang Huang, Martin Cmejrek and Bowen Zhou

15:00–15:25 *A Hybrid Morpheme-Word Representation for Machine Translation of Morphologically Rich Languages*  
Minh-Thang Luong, Preslav Nakov and Min-Yen Kan

15:25–15:50 *"Poetic" Statistical Machine Translation: Rhyme and Meter*  
Dmitriy Genzel, Jakob Uszkoreit and Franz Och

**Saturday, October 9, 2010 (continued)**

**Session 2B: Tagging, Chunking and Segmentation**

CHAIR: SHARON GOLDWATER

- 14:10–14:35 *Efficient Graph-Based Semi-Supervised Learning of Structured Tagging Models*  
Amarnag Subramanya, Slav Petrov and Fernando Pereira
- 14:35–15:00 *Better Punctuation Prediction with Dynamic Conditional Random Fields*  
Wei Lu and Hwee Tou Ng
- 15:00–15:25 *Joint Training and Decoding Using Virtual Nodes for Cascaded Segmentation and Tagging Tasks*  
Xian Qian, Qi Zhang, Yaqian Zhou, Xuanjing Huang and Lide Wu
- 15:25–15:50 *Crouching Dirichlet, Hidden Markov Model: Unsupervised POS Tagging with Context Local Tag Generation*  
Taesun Moon, Katrin Erk and Jason Baldridge

**Session 2C: Text Mining**

CHAIR: HAL DAUME III

- 14:10–14:35 *Improving Gender Classification of Blog Authors*  
Arjun Mukherjee and Bing Liu
- 14:35–15:00 *Negative Training Data Can be Harmful to Text Classification*  
Xiao-Li Li, Bing Liu and See-Kiong Ng
- 15:00–15:25 *Modeling Organization in Student Essays*  
Isaac Persing, Alan Davis and Vincent Ng
- 15:25–15:50 *Evaluating Models of Latent Document Semantics in the Presence of OCR Errors*  
Daniel Walker, William B. Lund and Eric K. Ringger

**Saturday, October 9, 2010 (continued)**

15:50–16:20 Break

**Session 3A: Machine Learning for NLP**

CHAIR: NOAH SMITH

16:20–16:45 *Translingual Document Representations from Discriminative Projections*  
John Platt, Kristina Toutanova and Wen-tau Yih

16:45–17:10 *Storing the Web in Memory: Space Efficient Language Models with Constant Time Retrieval*  
David Guthrie and Mark Hepple

17:10–17:35 *Efficient Incremental Decoding for Tree-to-String Translation*  
Liang Huang and Haitao Mi

17:35–18:00 *Modeling Perspective Using Adaptor Grammars*  
Eric Hardisty, Jordan Boyd-Graber and Philip Resnik

**Session 3B: Semantics**

CHAIR: MARK STEEDMAN

16:20–16:45 *Predicting the Semantic Compositionality of Prefix Verbs*  
Shane Bergsma, Aditya Bhargava, Hua He and Grzegorz Kondrak

16:45–17:10 *Joint Inference for Bilingual Semantic Role Labeling*  
Tao Zhuang and Chengqing Zong

17:10–17:35 *Automatic Discovery of Manner Relations and its Applications*  
Eduardo Blanco and Dan Moldovan

17:35–18:00 *Tense Sense Disambiguation: A New Syntactic Polysemy Task*  
Roi Reichart and Ari Rappoport



**Saturday, October 9, 2010 (continued)**

**Session 3C: Information Extraction**

CHAIR: ELLEN RILOFF

- 16:20–16:45 *Improving Mention Detection Robustness to Noisy Input*  
Radu Florian, John Pitrelli, Salim Roukos and Imed Zitouni
- 16:45–17:10 *Clustering-Based Stratified Seed Sampling for Semi-Supervised Relation Classification*  
Longhua Qian and Guodong Zhou
- 17:10–17:35 *Unsupervised Discovery of Negative Categories in Lexicon Bootstrapping*  
Tara McIntosh
- 17:35–18:00 *Automatic Keyphrase Extraction via Topic Decomposition*  
Zhiyuan Liu, Wenyi Huang, Yabin Zheng and Maosong Sun

**Sunday, October 10, 2010**

- 9:00–10:00 Invited talk: Andrew Ng, “Unsupervised feature learning and Deep Learning”  
CHAIR: DAVID YAROWSKY
- 10:00–10:30 Break

**Session 4A: Discourse and Dialog**

CHAIR: RADA MIHALCEA

- 10:30–10:55 *Incorporating Content Structure into Text Analysis Applications*  
Christina Sauper, Aria Haghighi and Regina Barzilay
- 10:55–11:20 *Exploiting Conversation Structure in Unsupervised Topic Segmentation for Emails*  
Shafiq Joty, Giuseppe Carenini, Gabriel Murray and Raymond T. Ng
- 11:20–11:45 *A Semi-Supervised Approach to Improve Classification of Infrequent Discourse Relations Using Feature Vector Extension*  
Hugo Hernault, Danushka Bollegala and Mitsuru Ishizuka

**Sunday, October 10, 2010 (continued)**

11:45–12:10 *A Game-Theoretic Approach to Generating Spatial Descriptions*  
Dave Golland, Percy Liang and Dan Klein

**Session 4B: Machine Translation II**

CHAIR: CHRIS QUIRK

10:30–10:55 *Facilitating Translation Using Source Language Paraphrase Lattices*  
Jinhua Du, Jie Jiang and Andy Way

10:55–11:20 *Mining Name Translations from Entity Graph Mapping*  
Gae-won You, Seung-won Hwang, Young-In Song, Long Jiang and Zaiqing Nie

11:20–11:45 *Non-Isomorphic Forest Pair Translation*  
Hui Zhang, Min Zhang, Haizhou Li and Eng Siong Chng

11:45–12:10 *Discriminative Instance Weighting for Domain Adaptation in Statistical Machine Translation*  
George Foster, Cyril Goutte and Roland Kuhn

**Session 4C: NLP Applications**

CHAIR: MANABU OKUMURA

10:30–10:55 *NLP on Spoken Documents Without ASR*  
Mark Dredze, Aren Jansen, Glen Coppersmith and Ken Church

10:55–11:20 *Fusing Eye Gaze with Speech Recognition Hypotheses to Resolve Exophoric References in Situated Dialogue*  
Zahar Prasov and Joyce Y. Chai

11:20–11:45 *Multi-Document Summarization Using A\* Search and Discriminative Learning*  
Ahmet Aker, Trevor Cohn and Robert Gaizauskas

11:45–12:10 *A Multi-Pass Sieve for Coreference Resolution*  
Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nate Chambers, Mihai Surdeanu, Dan Jurafsky and Christopher Manning

12:10–14:10 Lunch

**Sunday, October 10, 2010 (continued)**

**Session 5A: Natural Language Generation**

CHAIR: DRAGOMIR RADEV

- 14:10–14:35 *A Simple Domain-Independent Probabilistic Approach to Generation*  
Gabor Angeli, Percy Liang and Dan Klein
- 14:35–15:00 *Title Generation with Quasi-Synchronous Grammar*  
Kristian Woodsend, Yansong Feng and Mirella Lapata
- 15:00–15:25 *Automatic Analysis of Rhythmic Poetry with Applications to Generation and Translation*  
Erica Greene, Tugba Bodrumlu and Kevin Knight

**Session 5B: Machine Translation III**

CHAIR: DAVID CHIANG

- 14:10–14:35 *Discriminative Word Alignment with a Function Word Reordering Model*  
Hendra Setiawan, Chris Dyer and Philip Resnik
- 14:35–15:00 *Hierarchical Phrase-Based Translation Grammars Extracted from Alignment Posterior Probabilities*  
Adrià de Gispert, Juan Pino and William Byrne
- 15:00–15:25 *Maximum Entropy Based Phrase Reordering for Hierarchical Phrase-Based Translation*  
Zhongjun He, Yao Meng and Hao Yu

**Session 5C: Language Resources**

CHAIR: BENJAMIN TSOU

- 14:10–14:35 *Further Meta-Evaluation of Broad-Coverage Surface Realization*  
Dominic Espinosa, Rajkrishnan Rajkumar, Michael White and Shoshana Berleant
- 14:35–15:00 *Two Decades of Unsupervised POS Induction: How Far Have We Come?*  
Christos Christodoulopoulos, Sharon Goldwater and Mark Steedman
- 15:00–15:25 *We're Not in Kansas Anymore: Detecting Domain Changes in Streams*  
Mark Dredze, Tim Oates and Christine Piatko
- 15:25–15:55 Break

**Sunday, October 10, 2010 (continued)**

15:55–17:30 **Session 6A: Poster Spotlights**  
CHAIR: JASON EISNER

*A Fast Fertility Hidden Markov Model for Word Alignment Using MCMC*

Shaojun Zhao and Daniel Gildea

*Minimum Error Rate Training by Sampling the Translation Lattice*

Samidh Chatterjee and Nicola Cancedda

*Statistical Machine Translation with a Factorized Grammar*

Libin Shen, Bing Zhang, Spyros Matsoukas, Jinxi Xu and Ralph Weischedel

*Discriminative Sample Selection for Statistical Machine Translation*

Sankaranarayanan Ananthakrishnan, Rohit Prasad, David Stallard and Prem Natarajan

*Effects of Empty Categories on Machine Translation*

Tagyoung Chung and Daniel Gildea

*SCFG Decoding Without Binarization*

Mark Hopkins and Greg Langmead

*Example-Based Paraphrasing for Improved Phrase-Based Statistical Machine Translation*

Aurélien Max

*Combining Unsupervised and Supervised Alignments for MT: An Empirical Study*

Jinxi Xu and Antti-Veikko Rosti

*Top-Down Nearly-Context-Sensitive Parsing*

Eugene Charniak

*Improved Fully Unsupervised Parsing with Zoomed Learning*

Roi Reichart and Ari Rappoport

*Unsupervised Parse Selection for HPSG*

Rebecca Dridan and Timothy Baldwin

*Uptraining for Accurate Deterministic Question Parsing*

Slav Petrov, Pi-Chuan Chang, Michael Ringgaard and Hiyan Alshawi

**Sunday, October 10, 2010 (continued)**

*A Unified Framework for Scope Learning via Simplified Shallow Semantic Parsing*  
Qiaoming Zhu, Junhui Li, Hongling Wang and Guodong Zhou

*A New Approach to Lexical Disambiguation of Arabic Text*  
Rushin Shah, Paramveer S. Dhillon, Mark Liberman, Dean Foster, Mohamed Maamouri and Lyle Ungar

*What a Parser Can Learn from a Semantic Role Labeler and Vice Versa*  
Stephen Boxwell, Dennis Mehay and Chris Brew

*Word Sense Induction & Disambiguation Using Hierarchical Random Graphs*  
Ioannis Klapaftis and Suresh Manandhar

*Towards Conversation Entailment: An Empirical Investigation*  
Chen Zhang and Joyce Chai

*The Necessity of Combining Adaptation Methods*  
Ming-Wei Chang, Michael Connor and Dan Roth

*Training Continuous Space Language Models: Some Practical Issues*  
Hai Son Le, Alexandre Allauzen, Guillaume Wisniewski and François Yvon

15:55–17:25 **Session 6B: Poster Spotlights**  
CHAIR: ANDREW MCCALLUM

*Enhancing Domain Portability of Chinese Segmentation Model Using Chi-Square Statistics and Bootstrapping*  
Baobao Chang and Dongxu Han

*Latent-Descriptor Clustering for Unsupervised POS Induction*  
Michael Lamar, Yariv Maron and Elie Bienenstock

*A Probabilistic Morphological Analyzer for Syriac*  
Peter McClanahan, George Busby, Robbie Haertel, Kristian Heal, Deryle Lonsdale, Kevin Seppi and Eric Ringger

*Lessons Learned in Part-of-Speech Tagging of Conversational Speech*  
Vladimir Eidelman, Zhongqiang Huang and Mary Harper

*An Efficient Algorithm for Unsupervised Word Segmentation with Branching Entropy and MDL*  
Valentin Zhikov, Hiroya Takamura and Manabu Okumura

**Sunday, October 10, 2010 (continued)**

*A Fast Decoder for Joint Word Segmentation and POS-Tagging Using a Single Discriminative Model*

Yue Zhang and Stephen Clark

*Simple Type-Level Unsupervised POS Tagging*

Yoong Keok Lee, Aria Haghighi and Regina Barzilay

*Classifying Dialogue Acts in One-on-One Live Chats*

Su Nam Kim, Lawrence Cavedon and Timothy Baldwin

*Resolving Event Noun Phrases to Their Verbal Mentions*

Bin Chen, Jian Su and Chew Lim Tan

*A Tree Kernel-Based Unified Framework for Chinese Zero Anaphora Resolution*

Fang Kong and Guodong Zhou

*Automatic Comma Insertion for Japanese Text Generation*

Masaki Murata, Tomohiro Ohno and Shigeki Matsubara

*Using Unknown Word Techniques to Learn Known Words*

Kostadin Cholakov and Gertjan van Noord

*WikiWars: A New Corpus for Research on Temporal Expressions*

Pawel Mazur and Robert Dale

*PEM: A Paraphrase Evaluation Metric Exploiting Parallel Texts*

Chang Liu, Daniel Dahlmeier and Hwee Tou Ng

*Assessing Phrase-Based Translation Models with Oracle Decoding*

Guillaume Wisniewski, Alexandre Allauzen and François Yvon

*Automatic Evaluation of Translation Quality for Distant Language Pairs*

Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh and Hajime Tsukada

*An Approach of Generating Personalized Views from Normalized Electronic Dictionaries : A Practical Experiment on Arabic Language*

Aida Khemakhem, Bilel Gargouri and Abdelmajid Ben Hamadou

**Sunday, October 10, 2010 (continued)**

*Generating Confusion Sets for Context-Sensitive Error Correction*

Alla Rozovskaya and Dan Roth

15:55–17:25 **Session 6C: Poster Spotlights**

CHAIR: PHILIP RESNIK

*Confidence in Structured-Prediction Using Confidence-Weighted Models*

Avihai Mejer and Koby Crammer

*Evaluating the Impact of Alternative Dependency Graph Encodings on Solving Event Extraction Tasks*

Ekaterina Buyko and Udo Hahn

*Enhancing Mention Detection Using Projection via Aligned Corpora*

Yassine Benajiba and Imed Zitouni

*Domain Adaptation of Rule-Based Annotators for Named-Entity Recognition Tasks*

Laura Chiticariu, Rajasekar Krishnamurthy, Yunyao Li, Frederick Reiss and Shivakumar Vaithyanathan

*Collective Cross-Document Relation Extraction Without Labelled Data*

Limin Yao, Sebastian Riedel and Andrew McCallum

*Automatic Detection and Classification of Social Events*

Apoorv Agarwal and Owen Rambow

*Extracting Opinion Targets in a Single and Cross-Domain Setting with Conditional Random Fields*

Niklas Jakob and Iryna Gurevych

*Multi-Level Structured Models for Document-Level Sentiment Classification*

Ainur Yessenalina, Yisong Yue and Claire Cardie

*Cross Language Text Classification by Model Translation and Semi-Supervised Learning*

Lei Shi, Rada Mihalcea and Mingjun Tian

*SRL-Based Verb Selection for ESL*

Xiaohua Liu, Bo Han, Kuan Li, Stephan Hyeonjun Stiller and Ming Zhou

*Context Comparison of Bursty Events in Web Search and Online Media*

Yunliang Jiang, Cindy Xide Lin and Qiaozhu Mei

**Sunday, October 10, 2010 (continued)**

*Learning First-Order Horn Clauses from Web Text*

Stefan Schoenmackers, Jesse Davis, Oren Etzioni and Daniel Weld

*Constraints Based Taxonomic Relation Classification*

Quang Do and Dan Roth

*A Semi-Supervised Method to Learn and Construct Taxonomies Using the Web*

Zornitsa Kozareva and Eduard Hovy

*Function-Based Question Classification for General QA*

Fan Bu, Xingwei Zhu, Yu Hao and Xiaoyan Zhu

*Learning Recurrent Event Queries for Web Search*

Ruiqiang Zhang, Yuki Konda, Anlei Dong, Pranam Kolari, Yi Chang and Zhaohui Zheng

*Staying Informed: Supervised and Semi-Supervised Multi-View Topical Analysis of Ideological Perspective*

Amr Ahmed and Eric Xing

*Word-Based Dialect Identification with Georeferenced Rules*

Yves Scherrer and Owen Rambow

17:30–18:00 Break

18:00–21:00 Poster session and Reception

**Monday, October 11, 2010**

9:00–10:00 Invited talk: Amit Singhal, “Challenges in running a commercial search engine”  
CHAIR: KEN CHURCH

10:00–10:30 Break



**Monday, October 11, 2010 (continued)**

**Session 7A: Lexical Semantics**

CHAIR: KRISTINA TOUTANOVA

- 10:30–10:55 *Measuring Distributional Similarity in Context*  
Georgiana Dinu and Mirella Lapata
- 10:55–11:20 *A Mixture Model with Sharing for Lexical Semantics*  
Joseph Reisinger and Raymond Mooney
- 11:20–11:45 *Nouns are Vectors, Adjectives are Matrices: Representing Adjective-Noun Constructions in Semantic Space*  
Marco Baroni and Roberto Zamparelli
- 11:45–12:10 *Practical Linguistic Steganography Using Contextual Synonym Substitution and Vertex Colour Coding*  
Ching-Yun Chang and Stephen Clark

**Session 7B: Syntactic Parsing and Grammar Induction**

CHAIR: CHRIS MANNING

- 10:30–10:55 *Unsupervised Induction of Tree Substitution Grammars for Dependency Parsing*  
Phil Blunsom and Trevor Cohn
- 10:55–11:20 *It Depends on the Translation: Unsupervised Dependency Parsing via Word Alignment*  
Samuel Brody
- 11:20–11:45 *Inducing Probabilistic CCG Grammars from Logical Form with Higher-Order Unification*  
Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater and Mark Steedman
- 11:45–12:10 *Using Universal Linguistic Knowledge to Guide Grammar Induction*  
Tahira Naseem, Harr Chen, Regina Barzilay and Mark Johnson

**Monday, October 11, 2010 (continued)**

**Session 7C: NLP for the Web**

CHAIR: JIANYUN NIE

- 10:30–10:55 *What's with the Attitude? Identifying Sentences with Attitude in Online Discussions*  
Ahmed Hassan, Vahed Qazvinian and Dragomir Radev
- 10:55–11:20 *Hashing-Based Approaches to Spelling Correction of Personal Names*  
Raghavendra Udupa and Shaishav Kumar
- 11:20–11:45 *Identifying Functional Relations in Web Text*  
Thomas Lin, Mausam, and Oren Etzioni
- 11:45–12:10 *A Latent Variable Model for Geographic Lexical Variation*  
Jacob Eisenstein, Brendan O'Connor, Noah A. Smith and Eric P. Xing
- 12:15–13:00 SIGDAT Business Meeting
- 13:00–14:10 Lunch

**Plenary Session: Fred Jelinek Best Paper Award**

CHAIRS: BOB MOORE, JASON EISNER

- 14:10–15:05 *Dual Decomposition for Parsing with Non-Projective Head Automata*  
Terry Koo, Alexander M. Rush, Michael Collins, Tommi Jaakkola and David Sontag

**Closing**

CHAIR: KEN CHURCH