

EMNLP-CoNLL 2012

**2012 Joint Conference on  
Empirical Methods in Natural Language Processing  
and Computational Natural Language Learning**

**Proceedings of the Conference**

July 12–14, 2012  
Jeju Island, Korea

We wish to thank our sponsors



©2012 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 978-1-937284-43-5

## Preface by General and Program Chairs

It is our pleasure to welcome you to the EMNLP-CoNLL 2012 conference, a joint meeting of the Conference on Empirical Methods in Natural Language Learning (EMNLP) and the Conference on Computational Natural Language Learning (CoNLL). After the successful first collaboration in 2007, EMNLP-CoNLL 2012 is being jointly organized by the SIGDAT and SIGNLL special interest groups of the Association of Computational Linguistics.

This time, EMNLP-CoNLL is co-located with, and immediately after ACL's 50th anniversary conference. The choice of the location is an opportunity for the ACL community to return to the beautiful Jeju Island, Korea, following a seven-year hiatus since the Second International Joint Conference on Natural Language Processing (IJCNLP 2005) was held here.

Out of 606 submissions received by EMNLP-CoNLL this year, a total of 36 submissions were eventually withdrawn or rejected without review. From the remaining submissions, 99 were accepted for oral presentation and 40 for poster presentation, for a combined acceptance rate of 24.8%. As in recent editions of EMNLP, authors were given the opportunity to provide supplementary material in conjunction with their submissions, which the program committee could but was not required to take into account during reviewing. Also as in recent editions, authors of accepted papers were offered an additional page in the camera-ready version of their submissions, so that comments received from reviewers could be more easily addressed.

The papers submitted to the conference were subject to a rigorous reviewing process, made possible by efforts of a team of 525 primary and 66 secondary reviewers, acting under the guidance of 22 area chairs. Presence of unsupported claims, or failure to properly compare with previous work, were likely serious obstacles, on the path from initial submission to acceptance and then publication in our proceedings. Luckily, our preface has a guaranteed placement in the proceedings. Therefore, without access to insider data and impressions from previous editions of the conference, we will still go on a limb here, and make the unsupported claim that our team of area chairs has been the greatest. That their expertise, dedication and willingness to go beyond the call of duty had a positive impact on a timely reviewing process and a high-quality conference program, would be an understatement. It has been a pleasure to interact and work with our area chairs.

The schedule of our conference is strengthened by two invited speakers, Eric Xing and Patrick Pantel, who we were very happy to have accept our invitation; and by the CoNLL Shared Task, an annual tradition for the CoNLL conferences. This year's CoNLL Shared Task is Modeling Multilingual Unrestricted Coreference in OntoNotes, and its proceedings and detailed schedule are available separately.

We would like to thank all authors who submitted to our conference, for their willingness to share their knowledge with the rest of us. It may take a few weeks or many years, for the knowledge distilled into the present proceedings to have a measurable impact on our field and beyond. An impact that would not be possible without countless hours spent by authors, from developing ideas to running experiments to building usable systems - steps that often fail, and sometimes succeed.

We would also like to thank all members of the program committee, for their willingness to offer

feedback that sometimes reaches extraordinary levels of detail and value to authors. To recognise some of the most dedicated reviewers, we include the Best Reviewer awards a little later in these proceedings.

Naoaki Okazaki, Publications Chair, deserves our special thanks. He brought in a healthy dose of rigor to the planning and preparation of not only this proceedings but also conference materials, matched only by his dedication to deliver under tight scheduling constraints.

If the combination of oral presentations, posters and invited talks that make up EMNLP-CoNLL 2012 is considered a success, it is because it benefited from the touch of many people. Francesco Figari easily kept tabs on our salvos of large and small requests for updates to our conference website. Rich Gerber, Paolo Gai and the larger team managing the conference submission system were quick to offer answers to all our questions. The publication chairs and local arrangements committee of ACL 2012, including Michael White, Maggie Li, Jong Park and especially Gary Geunbae Lee, covered significant tasks on behalf of EMNLP-CoNLL, all with a smile. Chin-Yew Lin, Miles Osborne, Eric Fosler-Lussier, Dekang Lin, Rada Mihalcea, Regina Barzilay, Ulrich Germann and David Yarowsky offered high-level advice or answered detailed questions, drawing upon their experience as organizers of previous ACL-sponsored conferences.

We are grateful to our sponsors (Baidu, Google and Microsoft), for their support of best paper awards and support of student travel in particular, and the financial well-being of the conference in general. It has been an honor to be of service to the conference, for which we would like to thank the community and those who offered us this opportunity. We hope that you enjoy the conference, and have a productive and pleasant stay in South Korea!

Jun'ichi Tsujii, General Chair

James Henderson and Marius Paşca, Program Chairs

**General Chair:**

Jun'ichi Tsujii, Microsoft Research Asia

**Program Chairs:**

James Henderson, Xerox Research Centre Europe

Marius Paşca, Google

**Area Chairs:**

Eugene Agichtein, Emory University

Yejin Choi, Stony Brook University

Hal Daumé III, University of Maryland

Doug Downey, Northwestern University

Chris Dyer, Carnegie Mellon University

Adrià de Gispert, University of Cambridge

Julia Hirschberg, Columbia University

Anna Korhonen, University of Cambridge

Bing Liu, University of Illinois at Chicago

Hwee Tou Ng, National University of Singapore

Patrick Pantel, Microsoft Research

Marco Pennacchiotti, Yahoo! Labs

Slav Petrov, Google

Simone Paolo Ponzetto, Sapienza University of Rome

John Prager, IBM Research

Chris Quirk, Microsoft Research

Sebastian Riedel, University of Massachusetts at Amherst

Hiroya Takamura, Tokyo Institute of Technology

Partha Talukdar, Carnegie Mellon University

Kristina Toutanova, Microsoft Research

Reut Tsarfaty, Uppsala University

Marilyn Walker, University of California at Santa Cruz

**Publication Chair:**

Naoaki Okazaki, Tohoku University

## **Program Committee: Reviewers:**

Ahmed Abbasi, Robert Abbott, Omri Abend, Meni Adler, Mikhail Ageev, Lars Ahrenberg, Hua Ai, Cem Akkaya, Pranav Anand, Alina Andreevskaia, Ion Androustopoulos, Eiji Aramaki, Yoav Artzi, Abhishek Arun, Nicholas Asher, Giuseppe Attardi, Necip Fazil Ayan,

Anton Bakalov, Alexandra Balahur, Niranjan Balasubramanian, Timothy Baldwin, Carmen Banea, Ritwik Banerjee, Mohit Bansal, Roy Bar-Haim, Kedar Belare, Anja Belz, Paul Bennett, Stefan Benus, Taylor Berg-Kirkpatrick, Nicola Bertoldi, Justin Betteridge, Chandra Sekhar Bhagavathula, Aditya Bhargava, Jiang Bian, Chris Biemann, Dan Bikel, Alexandra Birch, Arianna Bisazza, Graeme Blackwood, Sasha Blair-Goldensohn, Jim Blevins, Michael Bloodgood, Phil Blunsom, Bernd Bohnet, Ester Boldrini, Antal van den Bosch, Elizabeth Boschee, Jan Botha, Alexandre Bouchard-Côté, Jordan Boyd-Graber, Thorsten Brants, Ulf Brefeld, Chris Brew, Ted Briscoe, Razvan Bunescu, David Burkett, Bill Byrne, Donna Byron,

Chris Callison-Burch, Nicola Cancedda, Marie Candito, Yunbo Cao, Giuseppe Carenini, Michael Carl, Andrew Carlson, Marine Carpuat, Xavier Carreras, Francisco Casacuberta, Dan Cer, Ozlem Cetinoglu, Mauro Cettolo, Joyce Chai, Soumen Chakrabarti, Jason Chang, Ming-Wei Chang, Ciprian Chelba, Hsin-Hsi Chen, Stanley Chen, Colin Cherry, Jackie Cheung, David Chiang, Munmun De Choudhury, Grzegorz Chrupala, Philipp Cimiano, Alex Clark, Steve Clark, Shay Cohen, Trevor Cohn, Bonaventura Coppola, Marta Ruiz Costa-Jussa, Danilo Croce, Tim Van de Cruys, Silviu Cucerzan, Aron Culotta,

Walter Daelemans, Daniel Dahlmeier, Bhavana Dalvi, Dana Dannells, Dipanjan Das, Steve DeNeefe, John DeNero, Vera Demberg, Michael Denkowski, Tejaswini Deoskar, Barry Devereux, Ann Devitt, Mona Diab, Laura Dietz, Mike Dillinger, Nikhil Dinesh, Xiaowen Ding, Bill Dolan, Mark Dras, Mark Dredze, Markus Dreyer, Greg Druck, Kevin Duh, Benjamin van Durme, Marc Dymetman,

Koji Eguchi, Andreas Eisele, Jacob Eisenstein, Jason Eisner, Micha Eisner, Michael Elhadad, Charles Elkan, David Elson, Katrin Erk, Oren Etzioni,

Anthony Fader, Katja Filippova, Radu Florian, George Foster, Jennifer Foster, Anette Frank, Marjorie Freedman, Dayne Freitag, Atsushi Fujita, Sadaoki Furui,

Michael Gamon, Kuzman Ganchev, Juri Ganitkevitch, Jianfeng Gao, Claire Gardent, Matt Gardner, Milica Gasic, Josef van Genabith, Matthew Gerber, Daniel Gildea, Daniel Gillick, Kevin Gimpel, Roxana Girju, Amir Globerson, Yoav Goldberg, Dan Goldwasser, Cyril Goutte, Amit Goyal, João Graça, Spence Green, Ralph Grishman, Joakim Gustafson,

Ryuichiro Higashinaka, Viet Ha-Thuc, Barry Haddow, Gholamreza Haffari, Patrick Haffner, Jan Hajic, Dilek Hakkani-Tur, David Hall, Keith Hall, Greg Hanneman, Christian Hardmeier, Saša Hasan, Helen Hastie, Jun Hatori, Zhongjun He, Kenneth Heafield, John Henderson, Tsutomu Hirao, Graeme Hirst, Anna Hjalmarsson, Hieu Hoang, Julia Hockenmaier, Liangjie Hong, Mark Hopkins, Veronique Hoste, Edward Hovy, Estevam Hruschka, Yunhua Hu, Fei Huang, Liang Huang, Minlie Huang, Samar Husain,

Francisco Iacobelli, Gonzalo Iglesias, Diana Inkpen, Ann Irvine, Tatsuya Izuha,

Jagadeesh Jagarlamud, Heng Ji, Jing Jiang, Wen Bin Jiang, Richard Johansson, Howard Johnson, Nobuhiro Kaji, Pallika Kanani, Hiroshi Kanayama, Rohit Kate, Junichi Kazama, Shahram Khadivi, Jin Dong Kim, Irwin King, Tracy King, Brian Kingsbury, Katrin Kirchhoff, Kevin Knight, Philipp

Koehn, Mamoru Komachi, Greg Kondrak, Christian Konig, Terry Koo, Moshe Koppel, Alexander Kotov, Zornitsa Kozareva, Jayant Krishnamurthy, Lun-Wei Ku, Marco Kuhlmann, Roland Kuhn, Seth Kulick, Shankar Kumar, Oren Kurland,

Oier Lopez de Lacalle, Philippe Langlais, Mirella Lapata, Claudia Leacock, John Lee, Yoong Keok Lee, Alessandro Lenci, Gregor Leusch, Abby Levenberg, Chi-Ho Li, Fangtao Li, Mu Li, Shoushan Li, Hui Lin, Thomas Lin, Xiao Ling, Jing Liu, Qun Liu, Yang Liu, Zhanyi Liu, Andrej Ljoje, Adam Lopez, Annie Louis, Bin Lu, Wei Lu, Yue Lu, Michael Lucas, Stephanie Lukin, Yajuan Lv,

Yanjun Ma, Klaus Macherey, Wolfgang Macherey, Nitin Madnani, Wolfgang Maier, Suresh Manandhar, Daniel Marcu, Anna Margolis, Katja Markert, Marie-Catherine de Marneffe, Craig Martell, Andre Martins, Yuval Marton, Spyros Matsoukas, Takuya Matsuzaki, Evgeny Matusov, Mausam, Arne Mauser, Jon May, Andrew McCallum, Diana McCarthy, David McClosky, Ryan McDonald, Kathleen McKeown, Beata Megyesi, Gerard De Melo, Paola Merlo, Haitao Mi, Rada Mihalcea, David Mimno, Yasuhiro Minami, Teru Misu, Yusuke Miyao, Saif Mohammad, Behrang Mohit, Karo Moilanen, Dan Moldovan, Martin Molina, Robert Moore, Paul Morarescu, Pedro Moreno, Shinsuke Mori, Hajime Morita, Alessandro Moschitti, Dana Movshovitz-Attias, Arjun Mukherjee, Vanessa Murdock, Markos Mylonakis, Lluís Mfriquez,

Tetsuji Nakagawa, Mikio Nakano, Toshiaki Nakazawa, Preslav Nakov, Jason Naradowsky, Vivi Nastase, Roberto Navigli, Mark-Jan Nederhof, Ani Nenkova, Graham Neubig, Alena Neviarouskaya, Vincent Ng, Jian-Yun Nie, John Niekrasz, Hitoshi Nishikawa, Joakim Nivre, Tadashi Nomoto, Scott Nowson,

Brendan O'Connor, Stephan Oepen, Kemal Oflazer, Naoaki Okazaki, Manabu Okumura, Constantin Orasan, Myle Ott,

Sebastian Pado, Shimei Pan, Sinno Jialin Pan, Soo-Min Pantel, Becky Passonneau, Alexandre Passos, Siddharth Patwardhan, Michael Paul, Matthias Paulik, Adam Pauls, Gerald Penn, Sasa Petrovic, Elias Ponvert, Hoifung Poon, Ana-Maria Popescu, Andrei Popescu-Belis, Maja Popovic, Fred Popowich, Matt Post, Peter Prettenhofer, Emily Mower Provost, Sampo Pyysalo, Yanjun Qi, Silvia Quarteroni,

Dragomir Radev, Altaf Rahman, Bhuvana Ramabhadran, Maya Ramanath, Owen Rambow, Ari Rappoport, Sujith Ravi, Deepak Ravichandran, Emmanuel Rayner, Sravana Reddy, Ines Rehbein, Roi Reichart, Joseph Reisinger, David Reitter, Jason Riesa, Verena Rieser, Stefan Riezler, German Rigau, Ellen Riloff, Laura Rimell, Eric Ringger, Alan Ritter, Brian Roark, Andrew Rosenberg, Paolo Rosso, Dan Roth, Wang Rui, Josef Ruppenhofer, Alexander Rush, Graham Russell,

Stijn De Saeger, Kenji Sagae, Mark Sammons, Sunita Sarawagi, Anoop Sarkar, Christina Sauper, Hassan Sawaf, Frank Schilder, Hinrich Schuetze, Bjoern Schuller, Holger Schwenk, Djame Seddah, Satoshi Sekine, Hendra Setiawan, Burr Settles, Fei Sha, Libin Shen, Shuming Shi, Ekaterina Shutova, Mario J. Silva, Fabrizio Silvestri, Michel Simard, Sameer Singh, Kairit Sirts, Gabriel Skantze, Jason Smith, Noah Smith, Ronnie Smith, Ben Snyder, Stephen Soderland, Radu Soricut, Lucia Specia, Valentin Spitkovsky, Caroline Sporleder, Vivek Srikumar, Mark Steedman, Benno Stein, Amanda Stent, Svetlana Stoyanchev, Veselin Stoyanov, Carlo Strapparava, Michael Strube, Keh-Yih Su, Michael Subotin, Fabian Suchanek, Katsuhito Sudoh, Ang Sun, Mihai Surdeanu, Idan Szpektor, Diarmuid  Seaghda,

David Talbot, Songbo Tan, Joseph Tepperman, Joel Tetreault, Blaise Thomson, Joerg Tiedemann,

Christoph Tillmann, Ivan Titov, Cigdem Toprak, Kentaro Torisawa, Roy Tromble, Oren Tsur, Yoshimasa Tsuruoka, Peter Turney, Oscar Täckström,

Raghavendra Udupa, Jakob Uszkoreit, Masao Utiyama,

Lucy Vanderwende, Tony Veale, Paola Velardi, Rossano Venturini, Yannick Versley, Jette Viethen, David Vilar, Aline Villavicencio, Andreas Vlachos, Martin Volk,

Sabine Schulte im Walde, Stephen Wan, Xiaojun Wan, Haifeng Wang, Taro Watanabe, Furu Wei, Gerhard Weikum, Ralph Weischedel, Ji-Rong Wen, Chris Wendt, Dominic Widdows, Janyce Wiebe, Derry Wijaya, Shuly Wintner, Kamfai Wong, Frank Wood, Fei Wu, Joern Wuebker,

Yunqing Xia, Deyi Xiong, Peng Xu, Nianwen Xue,

Grace Yang, Muyun Yang, Yi Yang, Alexander Yates, Mark Yatskar, Ainur Yessenalina, Elad Yom-Tov, Jianxing Yu, Yisong Yue,

Wlodek Zadrozny, Fabio Massimo Zanzotto, Richard Zens, Torsten Zesch, Luke Zettlemoyer, Bing Zhang, Congle Zhang, Hao Zhang, Hui Zhang, Joy Zhang, Lei Zhang, Min Zhang, Qi Zhang, Yue Zhang, Bing Zhao, Shiqi Zhao, Tiejun Zhao, Wayne Xin Zhao, Jing Zheng, Liu Zhiyuan, Bowen Zhou, Guodong Zhou, Ming Zhou, Jing-Bo Zhu, Xiaodan Zhu, Chengqing Zong, Geoff Zweig

**Program Committee: Secondary Reviewers:**

Gabor Angeli, Gerlof Bouma, Steven Burrows, Paula Carvalho, Diego Ceccarelli, Janara Christensen, Eric Corlett, Bart Desmet, Frank Ferraro, Tiziano Flati, Francisco Guzman, Yifan He, Dirk Hovy, Rubén Izquierdo, Jiarong Jiang, Tetsuo Kiso, Effi Levi, Baichuan Li, Nedim Lipka, Lemao Liu, Jeff Lund, Pierre Magistry, Thang Luong Minh, Makoto Miwa, Abdelrahman Mohamed, Taesun Moon, Tanmoy Mukherjee, Ryo Nagata, Franco Maria Nardini, Viet-An Nguyen, Gozde Ozbal, Wang Pidong, Barbara Plank, Quentin Pleple, Natalia Ponomareva, Vladimir Popescu, Daniel Preotiuc, Ju Qi, Kyle Rawlins, Majid Razmara, Nils Reiter, Joseph Le Roux, Benoît Sagot, Mohammad Salameh, Baskaran Sankaran, Uma Sawant, Roy Schwartz, Aliaksei Severyn, Kathrin Spreyer, Gabriele Tolomei, Olga Uryopina, Rakesh Varna, Zhiyang Wang, Zhuoran Wang, Jonny Weese, Michael Wick, Wei Xu, Haiqin Yang, Xuchen Yao, Mo Yu, Yintao Yu, Jiajun Zhang, Bo Zhao, Shanheng Zhao, Tu Zhaopeng, Chao Zhou



## **Invited Talks:**

### **“On Learning Sparse Structured Input-Output Models”**

**Eric Xing, Carnegie Mellon University**

In many modern problems across areas such as natural language processing, computer vision, and social media inference, one is often interested in learning a Sparse Structured Input-Output Model (SIOM), in which the input variables of the model such as lexicons in a document bear rich structures due to the syntactic and semantic dependences between them in the text; and the output variables such as the elements in a multi-way classification, a parse, or a topic representation are also structured because of their interrelatedness. A SIOM can nicely capture rich structural properties in the data and in the problem, but it also raises severe computational and theoretical challenge on sparse, consistent, and tractable model identification and inference.

In this talk, I will present models, algorithms, and theories that learn Sparse SIOMs of various kinds in very high dimensional input/output space, with fast and highly scalable optimization procedures, and strong statistical guarantees. I will demonstrate application of our approach to problems in large-scale text classification, topic modeling, and dependency parsing.

### **“The Appification of the Web and the Renaissance of Conversational User Interfaces”**

**Patrick Pantel, Microsoft Research**

The appification of the Web is triggering a fundamental shift in how users access information. We are moving from centralized access points, such as search engines, towards highly specialized, and yet fragmented, functionalities in disconnected apps. This talk explores an entity-centric conversational interface as a mechanism to overcome this fragmentation, highlighting the numerous associated NLP challenges and opportunities that lie ahead.

Consider mobile scenarios, where the traditional search engine paradigm is being cannibalized by search and browse functionalities built directly into specialized apps. For example, while users can search for restaurants and products using their mobile browser, they are increasingly turning directly to applications such as Yelp, Urbanspoon and Amazon. However, interoperability between applications and lacking generalized interfaces to their functionalities pose serious scalability challenges. In this talk, we argue for an entity-centric conversational interface in which natural user interactions with entities are paired with actions that can be performed on the entities, thus enabling the brokering of web pages and applications that can satisfy the intended action. In this vision, the broker is aware of all entities and actions of interest to its users, understands the intent of the user, and provides direct actionable results through APIs with external providers satisfying the intent. The user saves clicks and time to accomplish her intended action and can discover related actions. New revenue streams open up from paid action placement and lead generation opportunities. At the forefront of this direction are a number of NLP challenges in the areas of entity recognition, entity linking, knowledge extraction, intent recognition, and dialog modeling, to name a few.

We end by proposing one particular technique for learning and mapping user intents in a search interface. In an annotation study conducted over a traffic sample of web usage logs, we found that a large proportion of user queries involve actions on entities, calling for an automatic approach to identifying relevant actions for entity-bearing queries. We pose the problem of finding actions

that can be performed on entities as the problem of doing probabilistic inference in a graphical model that captures how entity-bearing information requests are generated. Given a large collection of real-world queries and clicks from a commercial search engine, the models are learned efficiently through maximum likelihood estimation using an EM algorithm. Given a new query, inference enables the recommendation of a set of pertinent actions and providers. We propose an evaluation methodology for measuring the relevance of our recommended actions, and show empirical evidence of the quality and the diversity of the discovered actions.

**Best Reviewer Awards:**

Hua Ai	Ion Androutsopoulos	Yoav Artzi	Mohit Bansal
Aditya Bhargava	Graeme Blackwood	Sasha Blair-Goldensohn	David Burkett
Marie Candito	Marine Carpuat	Xavier Carreras	Soumen Chakrabarti
Colin Cherry	Mark Dredze	Kevin Duh	David Elson
Katja Filippova	Matthew Gerber	Kevin Gimpel	Yoav Goldberg
Spence Green	Greg Hanneman	Saša Hasan	Alexander Kotov
Jayant Krishnamurthy	Marco Kuhlmann	Roland Kuhn	Mirella Lapata
Annie Louis	Andre Martins	Ryan McDonald	Robert Moore
Arjun Mukherjee	Vanessa Murdock	Jason Naradowsky	Roberto Navigli
Vincent Ng	John Niekrasz	Joakim Nivre	Becky Passonneau
Hoifung Poon	Andrei Popescu-Belis	Owen Rambow	Ines Rehbein
Alan Ritter	Alexander Rush	Noah Smith	Oren Tsur
Oscar Täckström	Lucy Vanderwende	Jette Viethen	Gerhard Weikum
Janyce Wiebe	Joern Wuebker	Yue Zhang	



## Table of Contents

<i>Syntactic Transfer Using a Bilingual Lexicon</i> Greg Durrett, Adam Pauls and Dan Klein .....	1
<i>Regularized Interlingual Projections: Evaluation on Multilingual Transliteration</i> Jagadeesh Jagarlamudi and Hal Daume III .....	12
<i>Bilingual Lexicon Extraction from Comparable Corpora Using Label Propagation</i> Akihiro Tamura, Taro Watanabe and Eiichiro Sumita .....	24
<i>Lexical Differences in Autobiographical Narratives from Schizophrenic Patients and Healthy Controls</i> Kai Hong, Christian G. Kohler, Mary E. March, Amber A. Parker and Ani Nenkova .....	37
<i>Streaming Analysis of Discourse Participants</i> Benjamin Van Durme .....	48
<i>Detecting Subgroups in Online Discussions by Modeling Positive and Negative Relations among Participants</i> Ahmed Hassan, Amjad Abu-Jbara and Dragomir Radev .....	59
<i>Generative Goal-Driven User Simulation for Dialog Management</i> Aciel Eshky, Ben Allison and Mark Steedman .....	71
<i>Optimising Incremental Dialogue Decisions Using Information Density for Interactive Systems</i> Nina Dethlefs, Helen Hastie, Verena Rieser and Oliver Lemon .....	82
<i>Mixed Membership Markov Models for Unsupervised Conversation Modeling</i> Michael J. Paul .....	94
<i>An Entity-Topic Model for Entity Linking</i> Xianpei Han and Le Sun .....	105
<i>Linking Named Entities to Any Database</i> Avirup Sil, Ernest Cronin, Penghai Nie, Yinfei Yang, Ana-Maria Popescu and Alexander Yates .....	116
<i>Towards Efficient Named-Entity Rule Induction for Customizability</i> Ajay Nagesh, Ganesh Ramakrishnan, Laura Chiticariu, Rajasekar Krishnamurthy, Ankush Dharkar and Pushpak Bhattacharyya .....	128
<i>Active Learning for Imbalanced Sentiment Classification</i> Shoushan Li, Shengfeng Ju, Guodong Zhou and Xiaojun Li .....	139
<i>A Weakly Supervised Model for Sentence-Level Semantic Orientation Analysis with Multiple Experts</i> Lizhen Qu, Rainer Gemulla and Gerhard Weikum .....	149

<i>Collocation Polarity Disambiguation Using Web-based Pseudo Contexts</i> Yanyan Zhao, Bing Qin and Ting Liu .....	160
<i>Aligning Predicates across Monolingual Comparable Texts using Graph-based Clustering</i> Michael Roth and Anette Frank .....	171
<i>Local and Global Context for Supervised and Unsupervised Metonymy Resolution</i> Vivi Nastase, Alex Judea, Katja Markert and Michael Strube .....	183
<i>Learning Verb Inference Rules from Linguistically-Motivated Evidence</i> Hila Weisman, Jonathan Berant, Idan Szpektor and Ido Dagan .....	194
<i>Spectral Dependency Parsing with Latent Variables</i> Paramveer Dhillon, Jordan Rodu, Michael Collins, Dean Foster and Lyle Ungar .....	205
<i>A Phrase-Discovering Topic Model Using Hierarchical Pitman-Yor Processes</i> Robert Lindsey, William Headden and Michael Stipicevic .....	214
<i>A Bayesian Model for Learning SCFGs with Discontiguous Rules</i> Abby Levenberg, Chris Dyer and Phil Blunsom .....	223
<i>Multiple Aspect Summarization Using Integer Linear Programming</i> Kristian Woodsend and Mirella Lapata .....	233
<i>Minimal Dependency Length in Realization Ranking</i> Michael White and Rajakrishnan Rajkumar .....	244
<i>Framework of Automatic Text Summarization Using Reinforcement Learning</i> Seonggi Ryang and Takeshi Abekawa .....	256
<i>Large Scale Decipherment for Out-of-Domain Machine Translation</i> Qing Dou and Kevin Knight .....	266
<i>N-gram-based Tense Models for Statistical Machine Translation</i> Zhengxian Gong, Min Zhang, Chew Lim Tan and Guodong Zhou .....	276
<i>Source Language Adaptation for Resource-Poor Machine Translation</i> Pidong Wang, Preslav Nakov and Hwee Tou Ng .....	286
<i>Exploiting Reducibility in Unsupervised Dependency Parsing</i> David Mareček and Zdeněk Žabokrtský .....	297
<i>Improving Transition-Based Dependency Parsing with Buffer Transitions</i> Daniel Fernández-González and Carlos Gómez-Rodríguez .....	308
<i>Generalized Higher-Order Dependency Parsing with Cube Pruning</i> Hao Zhang and Ryan McDonald .....	320
<i>Universal Grapheme-to-Phoneme Prediction Over Latin Alphabets</i> Young-Bum Kim and Benjamin Snyder .....	332

<i>Name Phylogeny: A Generative Model of String Variation</i> Nicholas Andrews, Jason Eisner and Mark Dredze .....	344
<i>Syntactic Surprisal Affects Spoken Word Duration in Conversational Contexts</i> Vera Demberg, Asad Sayeed, Philip Gorinski and Nikolaos Engonopoulos .....	356
<i>Why Question Answering using Sentiment Analysis and Word Classes</i> Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Takuya Kawada, Stijn De Saeger, Jun'ichi Kazama and You Wang .....	368
<i>Natural Language Questions for the Web of Data</i> Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, Maya Ramanath, Volker Tresp and Gerhard Weikum .....	379
<i>Answering Opinion Questions on Products by Exploiting Hierarchical Organization of Consumer Reviews</i> Jianxing Yu, Zheng-Jun Zha and Tat-Seng Chua .....	391
<i>Locally Training the Log-Linear Model for SMT</i> Lemao Liu, Hailong Cao, Taro Watanabe, Tiejun Zhao, Mo Yu and Conghui Zhu .....	402
<i>Iterative Annotation Transformation with Predict-Self Reestimation for Chinese Word Segmentation</i> Wenbin Jiang, Fandong Meng, Qun Liu and Yajuan Lü .....	412
<i>Automatically Constructing a Normalisation Dictionary for Microblogs</i> Bo Han, Paul Cook and Timothy Baldwin .....	421
<i>Unsupervised PCFG Induction for Grounded Language Learning with Highly Ambiguous Supervision</i> Joohyun Kim and Raymond Mooney .....	433
<i>Forced Derivation Tree based Model Training to Statistical Machine Translation</i> Nan Duan, Mu Li and Ming Zhou .....	445
<i>Multi-instance Multi-label Learning for Relation Extraction</i> Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati and Christopher D. Manning .....	455
<i>An “AI readability” Formula for French as a Foreign Language</i> Thomas François and Cédric Fairon .....	466
<i>Dynamic Programming for Higher Order Parsing of Gap-Minding Trees</i> Emily Pitler, Sampath Kannan and Mitchell Marcus .....	478
<i>Joint Entity and Event Coreference Resolution across Documents</i> Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu and Dan Jurafsky .....	489
<i>Joint Chinese Word Segmentation, POS Tagging and Parsing</i> Xian Qian and Yang Liu .....	501

<i>Translation Model Based Cross-Lingual Language Model Adaptation: from Word Models to Phrase Models</i>	
Shixiang Lu, Wei Wei, Xiaoyin Fu and Bo Xu .....	512
<i>Open Language Learning for Information Extraction</i>	
Mausam, Michael Schmitz, Stephen Soderland, Robert Bart and Oren Etzioni .....	523
<i>Modelling Sequential Text with an Adaptive Topic Model</i>	
Lan Du, Wray Buntine and Huidong Jin .....	535
<i>A Comparison of Vector-based Representations for Semantic Composition</i>	
William Blacoe and Mirella Lapata .....	546
<i>Exploiting Chunk-level Features to Improve Phrase Chunking</i>	
Junsheng Zhou, Weiguang Qu and Fen Zhang .....	557
<i>A Beam-Search Decoder for Grammatical Error Correction</i>	
Daniel Dahlmeier and Hwee Tou Ng .....	568
<i>A Statistical Relational Learning Approach to Identifying Evidence Based Medicine Categories</i>	
Mathias Verbeke, Vincent Van Asch, Roser Morante, Paolo Frasconi, Walter Daelemans and Luc De Raedt .....	579
<i>Lyrics, Music, and Emotions</i>	
Rada Mihalcea and Carlo Strapparava .....	590
<i>Assessment of ESL Learners' Syntactic Competence Based on Similarity Measures</i>	
Su-Youn Yoon and Suma Bhat .....	600
<i>A Unified Approach to Transliteration-based Text Input with Online Spelling Correction</i>	
Hisami Suzuki and Jianfeng Gao .....	609
<i>Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web</i>	
Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jong-Hoon Oh and Jun'ichi Kazama .....	619
<i>Enlarging Paraphrase Collections through Generalization and Instantiation</i>	
Atsushi Fujita, Pierre Isabelle and Roland Kuhn .....	631
<i>Concurrent Acquisition of Word Meaning and Lexical Categories</i>	
Afra Alishahi and Grzegorz Chrupala .....	643
<i>Do Neighbours Help? An Exploration of Graph-based Algorithms for Cross-domain Sentiment Classification</i>	
Natalia Ponomareva and Mike Thelwall .....	655
<i>Learning Lexicon Models from Search Logs for Query Expansion</i>	
Jianfeng Gao, Shasha Xie, Xiaodong He and Alnur Ali .....	666



<i>Joint Inference for Event Timeline Construction</i>	
Quang Do, Wei Lu and Dan Roth .....	677
<i>Three Dependency-and-Boundary Models for Grammar Induction</i>	
Valentin I. Spitzkovsky, Hiyan Alshawi and Daniel Jurafsky .....	688
<i>Exploring Adaptor Grammars for Native Language Identification</i>	
Sze-Meng Jojo Wong, Mark Dras and Mark Johnson .....	699
<i>Discovering Diverse and Salient Threads in Document Collections</i>	
Jennifer Gillenwater, Alex Kulesza and Ben Taskar .....	710
<i>Generalizing Sub-sentential Paraphrase Acquisition across Original Signal Type of Text Pairs</i>	
Aurélien Max, Houda Bouamor and Anne Vilnat .....	721
<i>Parse, Price and Cut—Delayed Column and Row Generation for Graph Based Parsers</i>	
Sebastian Riedel, David Smith and Andrew McCallum .....	732
<i>Domain Adaptation for Coreference Resolution: An Adaptive Ensemble Approach</i>	
Jian Bo Yang, Qi Mao, Qiao Liang Xiang, Ivor Wai-Hung Tsang, Kian Ming Adam Chai and Hai Leong Chieu .....	744
<i>Weakly Supervised Training of Semantic Parsers</i>	
Jayant Krishnamurthy and Tom Mitchell .....	754
<i>Cross-Lingual Language Modeling with Syntactic Reordering for Low-Resource Speech Recognition</i>	
Ping Xu and Pascale Fung .....	766
<i>Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge</i>	
Altaf Rahman and Vincent Ng .....	777
<i>A Sequence Labelling Approach to Quote Attribution</i>	
Timothy O’Keefe, Silvia Pareti, James R. Curran, Irena Koprinska and Matthew Honnibal ...	790
<i>SSHLDA: A Semi-Supervised Hierarchical Topic Model</i>	
Xian-Ling Mao, Zhao-Yan Ming, Tat-Seng Chua, Si Li, Hongfei Yan and Xiaoming Li .....	800
<i>Improving NLP through Marginalization of Hidden Syntactic Structure</i>	
Jason Naradowsky, Sebastian Riedel and David Smith .....	810
<i>Type-Supervised Hidden Markov Models for Part-of-Speech Tagging with Incomplete Tag Dictionaries</i>	
Dan Garrette and Jason Baldridge .....	821
<i>Explore Person Specific Evidence in Web Person Name Disambiguation</i>	
Liwei Chen, Yansong Feng, Lei Zou and Dongyan Zhao .....	832
<i>Inducing a Discriminative Parser to Optimize Machine Translation Reordering</i>	
Graham Neubig, Taro Watanabe and Shinsuke Mori .....	843

<i>Re-training Monolingual Parser Bilingually for Syntactic SMT</i>	
Shujie Liu, Chi-Ho Li, Mu Li and Ming Zhou .....	854
<i>Transforming Trees to Improve Syntactic Convergence</i>	
David Burkett and Dan Klein .....	863
<i>Learning Constraints for Consistent Timeline Extraction</i>	
David McClosky and Christopher D. Manning .....	873
<i>Identifying Constant and Unique Relations by using Time-Series Text</i>	
Yohei Takaku, Nobuhiro Kaji, Naoki Yoshinaga and Masashi Toyoda .....	883
<i>No Noun Phrase Left Behind: Detecting and Typing Unlinkable Entities</i>	
Thomas Lin, Mausam and Oren Etzioni .....	893
<i>A Novel Discriminative Framework for Sentence-Level Discourse Analysis</i>	
Shafiq Joty, Giuseppe Carenini and Raymond Ng .....	904
<i>Using Discourse Information for Paraphrase Extraction</i>	
Michaela Regneri and Rui Wang .....	916
<i>Generating Non-Projective Word Order in Statistical Linearization</i>	
Bernd Bohnet, Anders Björkelund, Jonas Kuhn, Wolfgang Seeker and Sina Zarriess .....	928
<i>Learning Syntactic Categories Using Paradigmatic Representations of Word Context</i>	
Mehmet Ali Yatbaz, Enis Sert and Deniz Yuret .....	940
<i>Exploring Topic Coherence over Many Models and Many Topics</i>	
Keith Stevens, Philip Kegelmeyer, David Andrzejewski and David Buttler .....	952
<i>Entropy-based Pruning for Phrase-based Machine Translation</i>	
Wang Ling, João Graça, Isabel Trancoso and Alan Black .....	962
<i>A Systematic Comparison of Phrase Table Pruning Techniques</i>	
Richard Zens, Daisy Stanton and Peng Xu .....	972
<i>Probabilistic Finite State Machines for Regression-based MT Evaluation</i>	
Mengqiu Wang and Christopher D. Manning .....	984
<i>An Empirical Investigation of Statistical Significance in NLP</i>	
Taylor Berg-Kirkpatrick, David Burkett and Dan Klein .....	995
<i>Employing Compositional Semantics and Discourse Consistency in Chinese Event Extraction</i>	
Peifeng Li, Guodong Zhou, Qiaoming Zhu and Libin Hou .....	1006
<i>Reading The Web with Learned Syntactic-Semantic Inference Rules</i>	
Ni Lao, Amarnag Subramanya, Fernando Pereira and William W. Cohen .....	1017
<i>Ensemble Semantics for Large-scale Unsupervised Relation Extraction</i>	
Bonan Min, Shuming Shi, Ralph Grishman and Chin-Yew Lin .....	1027

<i>Forest Reranking through Subtree Ranking</i>	
Richard Farkas and Helmut Schmid .....	1038
<i>Parser Showdown at the Wall Street Corral: An Empirical Investigation of Error Types in Parser Output</i>	
Jonathan K. Kummerfeld, David Hall, James R. Curran and Dan Klein .....	1048
<i>Extending Machine Translation Evaluation Metrics with Lexical Cohesion to Document Level</i>	
Billy T. M. Wong and Chunyu Kit .....	1060
<i>Fast Large-Scale Approximate Graph Construction for NLP</i>	
Amit Goyal, Hal Daume III and Raul Guerra .....	1069
<i>Building a Lightweight Semantic Model for Unsupervised Information Extraction on Short Listings</i>	
Doo Soon Kim, Kunal Verma and Peter Yeh .....	1081
<i>Sketch Algorithms for Estimating Point Queries in NLP</i>	
Amit Goyal, Hal Daume III and Graham Cormode .....	1093
<i>Monte Carlo MCMC: Efficient Inference by Approximate Sampling</i>	
Sameer Singh, Michael Wick and Andrew McCallum .....	1104
<i>On Amortizing Inference Cost for Structured Prediction</i>	
Vivek Srikumar, Gourab Kundu and Dan Roth .....	1114
<i>Exact Sampling and Decoding in High-Order Hidden Markov Models</i>	
Simon Carter, Marc Dymetman and Guillaume Bouchard .....	1125
<i>PATTY: A Taxonomy of Relational Patterns with Semantic Types</i>	
Ndapandula Nakashole, Gerhard Weikum and Fabian Suchanek .....	1135
<i>Training Factored PCFGs with Expectation Propagation</i>	
David Hall and Dan Klein .....	1146
<i>A Coherence Model Based on Syntactic Patterns</i>	
Annie Louis and Ani Nenkova .....	1157
<i>Language Model Rest Costs and Space-Efficient Storage</i>	
Kenneth Heafield, Philipp Koehn and Alon Lavie .....	1169
<i>Document-Wide Decoding for Phrase-Based Statistical Machine Translation</i>	
Christian Hardmeier, Joakim Nivre and Jörg Tiedemann .....	1179
<i>Left-to-Right Tree-to-String Decoding with Prediction</i>	
Yang Feng, Yang Liu, Qun Liu and Trevor Cohn .....	1191
<i>Semantic Compositionality through Recursive Matrix-Vector Spaces</i>	
Richard Socher, Brody Huval, Christopher D. Manning and Andrew Y. Ng .....	1201
<i>Polarity Inducing Latent Semantic Analysis</i>	
Wen-tau Yih, Geoffrey Zweig and John Platt .....	1212

<i>First Order vs. Higher Order Modification in Distributional Semantics</i>	
Gemma Boleda, Eva Maria Vecchi, Miquel Cornudella and Louise McNally .....	1223
<i>Learning-based Multi-Sieve Co-reference Resolution with Knowledge</i>	
Lev Ratinov and Dan Roth .....	1234
<i>Joint Learning for Coreference Resolution with Markov Logic</i>	
Yang Song, Jing Jiang, Wayne Xin Zhao, Sujian Li and Houfeng Wang .....	1245
<i>Resolving “This-issue” Anaphora</i>	
Varada Kolhatkar and Graeme Hirst .....	1255
<i>Entity based Q&amp;A Retrieval</i>	
Amit Singh .....	1266
<i>Constructing Task-Specific Taxonomies for Document Collection Browsing</i>	
Hui Yang .....	1278
<i>Besting the Quiz Master: Crowdsourcing Incremental Classification Games</i>	
Jordan Boyd-Graber, Brianna Satinoff, He He and Hal Daume III .....	1290
<i>Multi-Domain Learning: When Do Domains Matter?</i>	
Mahesh Joshi, Mark Dredze, William W. Cohen and Carolyn Rose .....	1302
<i>Biased Representation Learning for Domain Adaptation</i>	
Fei Huang and Alexander Yates .....	1313
<i>Unambiguity Regularization for Unsupervised Learning of Probabilistic Grammars</i>	
Kewei Tu and Vasant Honavar .....	1324
<i>Extracting Opinion Expressions with semi-Markov Conditional Random Fields</i>	
Bishan Yang and Claire Cardie .....	1335
<i>Opinion Target Extraction Using Word-Based Translation Model</i>	
Kang Liu, Liheng Xu and Jun Zhao .....	1346
<i>Word Salad: Relating Food Prices and Descriptions</i>	
Victor Chahuneau, Kevin Gimpel, Bryan R. Routledge, Lily Scherlis and Noah A. Smith ...	1357
<i>Learning to Map into a Universal POS Tagset</i>	
Yuan Zhang, Roi Reichart, Regina Barzilay and Amir Globerson .....	1368
<i>Part-of-Speech Tagging for Chinese-English Mixed Texts with Dynamic Features</i>	
Jiayi Zhao, Xipeng Qiu, Shu Zhang, Feng Ji and Xuanjing Huang .....	1379
<i>Wiki-ly Supervised Part-of-Speech Tagging</i>	
Shen Li, João Graça and Ben Taskar .....	1389
<i>Joining Forces Pays Off: Multilingual Joint Word Sense Disambiguation</i>	
Roberto Navigli and Simone Paolo Ponzetto .....	1399

<i>A New Minimally-Supervised Framework for Domain Word Sense Disambiguation</i> Stefano Faralli and Roberto Navigli .....	1411
<i>Grounded Models of Semantic Representation</i> Carina Silberer and Mirella Lapata .....	1423
<i>Improved Parsing and POS Tagging Using Inter-Sentence Consistency Constraints</i> Alexander Rush, Roi Reichart, Michael Collins and Amir Globerson .....	1434
<i>Unified Dependency Parsing of Chinese Morphological and Syntactic Structures</i> Zhongguo Li and Guodong Zhou .....	1445
<i>A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing</i> Bernd Bohnet and Joakim Nivre .....	1455
<i>Identifying Event-related Bursts via Social Media Activities</i> Xin Zhao, Baihan Shu, Jing Jiang, Yang Song, Hongfei Yan and Xiaoming Li .....	1466
<i>User Demographics and Language in an Implicit Social Network</i> Katja Filippova .....	1478
<i>Revisiting the Predictability of Language: Response Completion in Social Media</i> Bo Pang and Sujith Ravi .....	1489
<i>Supervised Text-based Geolocation Using Language Models on an Adaptive Grid</i> Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing and Jason Baldrige ...	1500
<i>A Discriminative Model for Query Spelling Correction with Latent Structural SVM</i> Huizhong Duan, Yanen Li, ChengXiang Zhai and Dan Roth .....	1511
<i>Characterizing Stylistic Elements in Syntactic Structure</i> Song Feng, Ritwik Banerjee and Yejin Choi .....	1522



# Conference Program

Thursday, July 12, 2012

## Session 1-AM-0P: Opening

09:00-09:15 Opening Remarks

## Session 1-AM-1P: Plenary Session: Invited Talk

Session Chair: James Henderson

09:15-10:30 Invited Talk: On Learning Sparse Structured Input-Output Models (Eric Xing, Carnegie Mellon University)

10:30-11:00 Coffee Break

## Session 1-AM-2A: Machine Translation: Bilingual Lexicons and Alignment

Session Chair: David Chiang

11:00-11:30 *Syntactic Transfer Using a Bilingual Lexicon*  
Greg Durrett, Adam Pauls and Dan Klein

11:30-12:00 *Regularized Interlingual Projections: Evaluation on Multilingual Transliteration*  
Jagadeesh Jagarlamudi and Hal Daume III

12:00-12:30 *Bilingual Lexicon Extraction from Comparable Corpora Using Label Propagation*  
Akihiro Tamura, Taro Watanabe and Eiichiro Sumita

## Session 1-AM-2B: Social Media: Author Style and Attribution

Session Chair: Janyce Wiebe

11:00-11:30 *Lexical Differences in Autobiographical Narratives from Schizophrenic Patients and Healthy Controls*  
Kai Hong, Christian G. Kohler, Mary E. March, Amber A. Parker and Ani Nenkova

11:30-12:00 *Streaming Analysis of Discourse Participants*  
Benjamin Van Durme

12:00-12:30 *Detecting Subgroups in Online Discussions by Modeling Positive and Negative Relations among Participants*  
Ahmed Hassan, Amjad Abu-Jbara and Dragomir Radev

**Thursday, July 12, 2012 (continued)**

**Session 1-AM-2C: Dialogue and Interactive Systems**

Session Chair: Michael White

- 11:00-11:30 *Generative Goal-Driven User Simulation for Dialog Management*  
Aciel Eshky, Ben Allison and Mark Steedman
- 11:30-12:00 *Optimising Incremental Dialogue Decisions Using Information Density for Interactive Systems*  
Nina Dethlefs, Helen Hastie, Verena Rieser and Oliver Lemon
- 12:00-12:30 *Mixed Membership Markov Models for Unsupervised Conversation Modeling*  
Michael J. Paul

**Session 1-AM-2D: Information Extraction: Entity Disambiguation**

Session Chair: Raymond Mooney

- 11:00-11:30 *An Entity-Topic Model for Entity Linking*  
Xianpei Han and Le Sun
- 11:30-12:00 *Linking Named Entities to Any Database*  
Avirup Sil, Ernest Cronin, Penghai Nie, Yinfei Yang, Ana-Maria Popescu and Alexander Yates
- 12:00-12:30 *Towards Efficient Named-Entity Rule Induction for Customizability*  
Ajay Nagesh, Ganesh Ramakrishnan, Laura Chiticariu, Rajasekar Krishnamurthy, Ankush Dharkar and Pushpak Bhattacharyya
- 12:30-14:00 Lunch

**Session 1-PM-1A: Sentiment Analysis**

Session Chair: Bing Liu

- 14:00-14:30 *Active Learning for Imbalanced Sentiment Classification*  
Shoushan Li, Shengfeng Ju, Guodong Zhou and Xiaojun Li
- 14:30-15:00 *A Weakly Supervised Model for Sentence-Level Semantic Orientation Analysis with Multiple Experts*  
Lizhen Qu, Rainer Gemulla and Gerhard Weikum
- 15:00-15:30 *Collocation Polarity Disambiguation Using Web-based Pseudo Contexts*  
Yanyan Zhao, Bing Qin and Ting Liu



**Thursday, July 12, 2012 (continued)**

**Session 1-PM-1B: Semantics: Nouns, Verbs and Predicates**

Session Chair: Rada Mihalcea

- 14:00-14:30 *Aligning Predicates across Monolingual Comparable Texts using Graph-based Clustering*  
Michael Roth and Anette Frank
- 14:30-15:00 *Local and Global Context for Supervised and Unsupervised Metonymy Resolution*  
Vivi Nastase, Alex Judea, Katja Markert and Michael Strube
- 15:00-15:30 *Learning Verb Inference Rules from Linguistically-Motivated Evidence*  
Hila Weisman, Jonathan Berant, Idan Szpektor and Ido Dagan

**Session 1-PM-1C: Machine Learning: Latent Models**

Session Chair: Alexandre Klementiev

- 14:00-14:30 *Spectral Dependency Parsing with Latent Variables*  
Paramveer Dhillon, Jordan Rodu, Michael Collins, Dean Foster and Lyle Ungar
- 14:30-15:00 *A Phrase-Discovering Topic Model Using Hierarchical Pitman-Yor Processes*  
Robert Lindsey, William Headden and Michael Stipicevic
- 15:00-15:30 *A Bayesian Model for Learning SCFGs with Discontiguous Rules*  
Abby Levenberg, Chris Dyer and Phil Blunsom

**Session 1-PM-1D: Summarization**

Session Chair: Dragomir Radev

- 14:00-14:30 *Multiple Aspect Summarization Using Integer Linear Programming*  
Kristian Woodsend and Mirella Lapata
- 14:30-15:00 *Minimal Dependency Length in Realization Ranking*  
Michael White and Rajakrishnan Rajkumar
- 15:00-15:30 *Framework of Automatic Text Summarization Using Reinforcement Learning*  
Seonggi Ryang and Takeshi Abekawa
- 15:30-16:00 Coffee Break

**Thursday, July 12, 2012 (continued)**

**Session 1-PM-2A: Machine Translation**

Session Chair: Xiong Deyi

16:00-16:30 *Large Scale Decipherment for Out-of-Domain Machine Translation*

Qing Dou and Kevin Knight

16:30-17:00 *N-gram-based Tense Models for Statistical Machine Translation*

Zhengxian Gong, Min Zhang, Chew Lim Tan and Guodong Zhou

17:00-17:30 *Source Language Adaptation for Resource-Poor Machine Translation*

Pidong Wang, Preslav Nakov and Hwee Tou Ng

**Session 1-PM-2B: Dependency Parsing**

Session Chair: Emily Pitler

16:00-16:30 *Exploiting Reducibility in Unsupervised Dependency Parsing*

David Mareček and Zdeněk Žabokrtský

16:30-17:00 *Improving Transition-Based Dependency Parsing with Buffer Transitions*

Daniel Fernández-González and Carlos Gómez-Rodríguez

17:00-17:30 *Generalized Higher-Order Dependency Parsing with Cube Pruning*

Hao Zhang and Ryan McDonald

**Session 1-PM-2C: Phonemes, Words and Speech**

Session Chair: Pascale Fung

16:00-16:30 *Universal Grapheme-to-Phoneme Prediction Over Latin Alphabets*

Young-Bum Kim and Benjamin Snyder

16:30-17:00 *Name Phylogeny: A Generative Model of String Variation*

Nicholas Andrews, Jason Eisner and Mark Dredze

17:00-17:30 *Syntactic Surprisal Affects Spoken Word Duration in Conversational Contexts*

Vera Demberg, Asad Sayeed, Philip Gorinski and Nikolaos Engonopoulos

**Thursday, July 12, 2012 (continued)**

**Session 1-PM-2D: Question Answering**

Session Chair: Alessandro Moschitti

- 16:00-16:30 *Why Question Answering using Sentiment Analysis and Word Classes*  
Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Takuya Kawada, Stijn De Saeger, Jun'ichi Kazama and Yiou Wang
- 16:30-17:00 *Natural Language Questions for the Web of Data*  
Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, Maya Ramanath, Volker Tresp and Gerhard Weikum
- 17:00-17:30 *Answering Opinion Questions on Products by Exploiting Hierarchical Organization of Consumer Reviews*  
Jianxing Yu, Zheng-Jun Zha and Tat-Seng Chua
- 17:30-18:00 Break

**Session 1-PM-3P: Poster Session and Reception (18:00-22:00)**

*Locally Training the Log-Linear Model for SMT*

Lemao Liu, Hailong Cao, Taro Watanabe, Tiejun Zhao, Mo Yu and Conghui Zhu

*Iterative Annotation Transformation with Predict-Self Reestimation for Chinese Word Segmentation*

Wenbin Jiang, Fandong Meng, Qun Liu and Yajuan Lü

*Automatically Constructing a Normalisation Dictionary for Microblogs*

Bo Han, Paul Cook and Timothy Baldwin

*Unsupervised PCFG Induction for Grounded Language Learning with Highly Ambiguous Supervision*

Joohyun Kim and Raymond Mooney

*Forced Derivation Tree based Model Training to Statistical Machine Translation*

Nan Duan, Mu Li and Ming Zhou

*Multi-instance Multi-label Learning for Relation Extraction*

Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati and Christopher D. Manning

*An "AI readability" Formula for French as a Foreign Language*

Thomas François and Cédric Fairon

**Thursday, July 12, 2012 (continued)**

*Dynamic Programming for Higher Order Parsing of Gap-Minding Trees*

Emily Pitler, Sampath Kannan and Mitchell Marcus

*Joint Entity and Event Coreference Resolution across Documents*

Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu and Dan Jurafsky

*Joint Chinese Word Segmentation, POS Tagging and Parsing*

Xian Qian and Yang Liu

*Translation Model Based Cross-Lingual Language Model Adaptation: from Word Models to Phrase Models*

Shixiang Lu, Wei Wei, Xiaoyin Fu and Bo Xu

*Open Language Learning for Information Extraction*

Mausam, Michael Schmitz, Stephen Soderland, Robert Bart and Oren Etzioni

*Modelling Sequential Text with an Adaptive Topic Model*

Lan Du, Wray Buntine and Huidong Jin

*A Comparison of Vector-based Representations for Semantic Composition*

William Blacoe and Mirella Lapata

*Exploiting Chunk-level Features to Improve Phrase Chunking*

Junsheng Zhou, Weiguang Qu and Fen Zhang

*A Beam-Search Decoder for Grammatical Error Correction*

Daniel Dahlmeier and Hwee Tou Ng

*A Statistical Relational Learning Approach to Identifying Evidence Based Medicine Categories*

Mathias Verbeke, Vincent Van Asch, Roser Morante, Paolo Frasconi, Walter Daelemans and Luc De Raedt

*Lyrics, Music, and Emotions*

Rada Mihalcea and Carlo Strapparava

*Assessment of ESL Learners' Syntactic Competence Based on Similarity Measures*

Su-Youn Yoon and Suma Bhat

**Thursday, July 12, 2012 (continued)**

*A Unified Approach to Transliteration-based Text Input with Online Spelling Correction*  
Hisami Suzuki and Jianfeng Gao

*Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web*  
Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jong-Hoon Oh and Jun'ichi Kazama

*Enlarging Paraphrase Collections through Generalization and Instantiation*  
Atsushi Fujita, Pierre Isabelle and Roland Kuhn

*Concurrent Acquisition of Word Meaning and Lexical Categories*  
Afra Alishahi and Grzegorz Chrupala

*Do Neighbours Help? An Exploration of Graph-based Algorithms for Cross-domain Sentiment Classification*  
Natalia Ponomareva and Mike Thelwall

*Learning Lexicon Models from Search Logs for Query Expansion*  
Jianfeng Gao, Shasha Xie, Xiaodong He and Alnur Ali

*Joint Inference for Event Timeline Construction*  
Quang Do, Wei Lu and Dan Roth

*Three Dependency-and-Boundary Models for Grammar Induction*  
Valentin I. Spitzkovsky, Hiyan Alshawi and Daniel Jurafsky

*Exploring Adaptor Grammars for Native Language Identification*  
Sze-Meng Jojo Wong, Mark Dras and Mark Johnson

*Discovering Diverse and Salient Threads in Document Collections*  
Jennifer Gillenwater, Alex Kulesza and Ben Taskar

*Generalizing Sub-sentential Paraphrase Acquisition across Original Signal Type of Text Pairs*  
Aurélien Max, Houda Bouamor and Anne Vilnat

*Parse, Price and Cut—Delayed Column and Row Generation for Graph Based Parsers*  
Sebastian Riedel, David Smith and Andrew McCallum

**Thursday, July 12, 2012 (continued)**

*Domain Adaptation for Coreference Resolution: An Adaptive Ensemble Approach*

Jian Bo Yang, Qi Mao, Qiao Liang Xiang, Ivor Wai-Hung Tsang, Kian Ming Adam Chai and Hai Leong Chieu

*Weakly Supervised Training of Semantic Parsers*

Jayant Krishnamurthy and Tom Mitchell

*Cross-Lingual Language Modeling with Syntactic Reordering for Low-Resource Speech Recognition*

Ping Xu and Pascale Fung

*Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge*

Altaf Rahman and Vincent Ng

*A Sequence Labelling Approach to Quote Attribution*

Timothy O'Keefe, Silvia Paretì, James R. Curran, Irena Koprinska and Matthew Honnibal

*SSHLDA: A Semi-Supervised Hierarchical Topic Model*

Xian-Ling Mao, Zhao-Yan Ming, Tat-Seng Chua, Si Li, Hongfei Yan and Xiaoming Li

*Improving NLP through Marginalization of Hidden Syntactic Structure*

Jason Naradowsky, Sebastian Riedel and David Smith

*Type-Supervised Hidden Markov Models for Part-of-Speech Tagging with Incomplete Tag Dictionaries*

Dan Garrette and Jason Baldridge

*Explore Person Specific Evidence in Web Person Name Disambiguation*

Liwei Chen, Yansong Feng, Lei Zou and Dongyan Zhao

**Friday, July 13, 2012**

**Session 2-AM-1P: Plenary Session: Invited Talk**

Session Chair: Marius Pasca

09:15-10:30 Invited Talk: The Application of the Web and the Renaissance of Conversational User Interfaces (Patrick Pantel, Microsoft Research)

10:30-11:00 Coffee Break

**Session 2-AM-2A: Machine Translation: Role of Syntax**

Session Chair: Qun Liu

11:00-11:30 *Inducing a Discriminative Parser to Optimize Machine Translation Reordering*  
Graham Neubig, Taro Watanabe and Shinsuke Mori

11:30-12:00 *Re-training Monolingual Parser Bilingually for Syntactic SMT*  
Shujie Liu, Chi-Ho Li, Mu Li and Ming Zhou

12:00-12:30 *Transforming Trees to Improve Syntactic Convergence*  
David Burkett and Dan Klein

**Session 2-AM-2B: Information Extraction: Temporally-Aware Extraction**

Session Chair: Jong-Hoon Oh

11:00-11:30 *Learning Constraints for Consistent Timeline Extraction*  
David McClosky and Christopher D. Manning

11:30-12:00 *Identifying Constant and Unique Relations by using Time-Series Text*  
Yohei Takaku, Nobuhiro Kaji, Naoki Yoshinaga and Masashi Toyoda

12:00-12:30 *No Noun Phrase Left Behind: Detecting and Typing Unlinkable Entities*  
Thomas Lin, Mausam and Oren Etzioni

**Friday, July 13, 2012 (continued)**

**Session 2-AM-2C: Discourse and Generation**

Session Chair: Anette Frank

- 11:00-11:30 *A Novel Discriminative Framework for Sentence-Level Discourse Analysis*  
Shafiq Joty, Giuseppe Carenini and Raymond Ng
- 11:30-12:00 *Using Discourse Information for Paraphrase Extraction*  
Michaela Regneri and Rui Wang
- 12:00-12:30 *Generating Non-Projective Word Order in Statistical Linearization*  
Bernd Bohnet, Anders Björkelund, Jonas Kuhn, Wolfgang Seeker and Sina Zarriess

**Session 2-AM-2D: CoNLL Shared Task**

Session Chair: Sameer Pradhan

- 11:00-12:30 CoNLL Shared Task Session
- 12:30-13:45 Lunch
- 13:45-14:30 SIGDAT and SIGNLL Business Meetings

**Session 2-PM-1A: Semantics: Words and Topics**

Session Chair: Roi Reichart

- 14:30-15:00 *Learning Syntactic Categories Using Paradigmatic Representations of Word Context*  
Mehmet Ali Yatbaz, Enis Sert and Deniz Yuret
- 15:00-15:30 *Exploring Topic Coherence over Many Models and Many Topics*  
Keith Stevens, Philip Kegelmeyer, David Andrzejewski and David Buttler



**Friday, July 13, 2012 (continued)**

**Session 2-PM-1B: Machine Translation: Pruning**

Session Chair: Kevin Knight

14:30-15:00 *Entropy-based Pruning for Phrase-based Machine Translation*

Wang Ling, João Graça, Isabel Trancoso and Alan Black

15:00-15:30 *A Systematic Comparison of Phrase Table Pruning Techniques*

Richard Zens, Daisy Stanton and Peng Xu

**Session 2-PM-1C: Evaluation**

Session Chair: Billy Tak-Ming Wong

14:30-15:00 *Probabilistic Finite State Machines for Regression-based MT Evaluation*

Mengqiu Wang and Christopher D. Manning

15:00-15:30 *An Empirical Investigation of Statistical Significance in NLP*

Taylor Berg-Kirkpatrick, David Burkett and Dan Klein

**Session 2-PM-1D: CoNLL Shared Task**

Session Chair: Alessandro Moschitti

14:30-15:30 CoNLL Shared Task Session

15:30-16:00 Coffee Break

**Session 2-PM-2A: Information Extraction: Relation and Event Extraction**

Session Chair: Mausam

16:00-16:30 *Employing Compositional Semantics and Discourse Consistency in Chinese Event Extraction*

Peifeng Li, Guodong Zhou, Qiaoming Zhu and Libin Hou

16:30-17:00 *Reading The Web with Learned Syntactic-Semantic Inference Rules*

Ni Lao, Amarnag Subramanya, Fernando Pereira and William W. Cohen

17:00-17:30 *Ensemble Semantics for Large-scale Unsupervised Relation Extraction*

Bonan Min, Shuming Shi, Ralph Grishman and Chin-Yew Lin

**Friday, July 13, 2012 (continued)**

**Session 2-PM-2B: Parsing Models and Evaluation**

Session Chair: Ivan Titov

- 16:00-16:30 *Forest Reranking through Subtree Ranking*  
Richard Farkas and Helmut Schmid
- 16:30-17:00 *Parser Showdown at the Wall Street Corral: An Empirical Investigation of Error Types in Parser Output*  
Jonathan K. Kummerfeld, David Hall, James R. Curran and Dan Klein
- 17:00-17:30 *Extending Machine Translation Evaluation Metrics with Lexical Cohesion to Document Level*  
Billy T. M. Wong and Chunyu Kit

**Session 2-PM-2C: Large-Scale NLP Algorithms**

Session Chair: Benjamin van Durme

- 16:00-16:30 *Fast Large-Scale Approximate Graph Construction for NLP*  
Amit Goyal, Hal Daume III and Raul Guerra
- 16:30-17:00 *Building a Lightweight Semantic Model for Unsupervised Information Extraction on Short Listings*  
Doo Soon Kim, Kunal Verma and Peter Yeh
- 17:00-17:30 *Sketch Algorithms for Estimating Point Queries in NLP*  
Amit Goyal, Hal Daume III and Graham Cormode

**Session 2-PM-2D: Machine Learning: Inference**

Session Chair: Jennifer Gillenwater

- 16:00-16:30 *Monte Carlo MCMC: Efficient Inference by Approximate Sampling*  
Sameer Singh, Michael Wick and Andrew McCallum
- 16:30-17:00 *On Amortizing Inference Cost for Structured Prediction*  
Vivek Srikumar, Gourab Kundu and Dan Roth
- 17:00-17:30 *Exact Sampling and Decoding in High-Order Hidden Markov Models*  
Simon Carter, Marc Dymetman and Guillaume Bouchard

**Saturday, July 14, 2012**

**Session 3-AM-1P: Plenary Session**

Session Chair: Jun'ichi Tsujii

09:00-09:30 *PATY: A Taxonomy of Relational Patterns with Semantic Types*

Ndapandula Nakashole, Gerhard Weikum and Fabian Suchanek

09:30-10:00 *Training Factored PCFGs with Expectation Propagation*

David Hall and Dan Klein

10:00-10:30 *A Coherence Model Based on Syntactic Patterns*

Annie Louis and Ani Nenkova

10:30-11:00 Coffee Break

**Session 3-AM-2A: Machine Translation: Decoding**

Session Chair: Preslav Nakov

11:00-11:30 *Language Model Rest Costs and Space-Efficient Storage*

Kenneth Heafield, Philipp Koehn and Alon Lavie

11:30-12:00 *Document-Wide Decoding for Phrase-Based Statistical Machine Translation*

Christian Hardmeier, Joakim Nivre and Jörg Tiedemann

12:00-12:30 *Left-to-Right Tree-to-String Decoding with Prediction*

Yang Feng, Yang Liu, Qun Liu and Trevor Cohn

**Saturday, July 14, 2012 (continued)**

**Session 3-AM-2B: Distributional and Compositional Semantics**

Session Chair: Bo Pang

- 11:00-11:30 *Semantic Compositionality through Recursive Matrix-Vector Spaces*  
Richard Socher, Brody Huval, Christopher D. Manning and Andrew Y. Ng
- 11:30-12:00 *Polarity Inducing Latent Semantic Analysis*  
Wen-tau Yih, Geoffrey Zweig and John Platt
- 12:00-12:30 *First Order vs. Higher Order Modification in Distributional Semantics*  
Gemma Boleda, Eva Maria Vecchi, Miquel Cornudella and Louise McNally

**Session 3-AM-2C: Discourse: Coreference Resolution**

Session Chair: Mihai Surdeanu

- 11:00-11:30 *Learning-based Multi-Sieve Co-reference Resolution with Knowledge*  
Lev Ratinov and Dan Roth
- 11:30-12:00 *Joint Learning for Coreference Resolution with Markov Logic*  
Yang Song, Jing Jiang, Wayne Xin Zhao, Sujian Li and Houfeng Wang
- 12:00-12:30 *Resolving “This-issue” Anaphora*  
Varada Kolhatkar and Graeme Hirst

**Session 3-AM-2D: Information Retrieval**

Session Chair: Jianfeng Gao

- 11:00-11:30 *Entity based Q&A Retrieval*  
Amit Singh
- 11:30-12:00 *Constructing Task-Specific Taxonomies for Document Collection Browsing*  
Hui Yang
- 12:00-12:30 *Besting the Quiz Master: Crowdsourcing Incremental Classification Games*  
Jordan Boyd-Graber, Brianna Satinoff, He He and Hal Daume III
- 12:30-14:00 Lunch

**Saturday, July 14, 2012 (continued)**

**Session 3-PM-1A: Machine Learning: Transfer and Biases**

Session Chair: Benjamin Snyder

- 14:00-14:30 *Multi-Domain Learning: When Do Domains Matter?*  
Mahesh Joshi, Mark Dredze, William W. Cohen and Carolyn Rose
- 14:30-15:00 *Biased Representation Learning for Domain Adaptation*  
Fei Huang and Alexander Yates
- 15:00-15:30 *Unambiguity Regularization for Unsupervised Learning of Probabilistic Grammars*  
Kewei Tu and Vasant Honavar

**Session 3-PM-1B: Opinion Mining: Discovering Opinion Expressions**

Session Chair: Yejin Choi

- 14:00-14:30 *Extracting Opinion Expressions with semi-Markov Conditional Random Fields*  
Bishan Yang and Claire Cardie
- 14:30-15:00 *Opinion Target Extraction Using Word-Based Translation Model*  
Kang Liu, Liheng Xu and Jun Zhao
- 15:00-15:30 *Word Salad: Relating Food Prices and Descriptions*  
Victor Chahuneau, Kevin Gimpel, Bryan R. Routledge, Lily Scherlis and Noah A. Smith

**Session 3-PM-1C: Part of Speech Tagging**

Session Chair: Slav Petrov

- 14:00-14:30 *Learning to Map into a Universal POS Tagset*  
Yuan Zhang, Roi Reichart, Regina Barzilay and Amir Globerson
- 14:30-15:00 *Part-of-Speech Tagging for Chinese-English Mixed Texts with Dynamic Features*  
Jiayi Zhao, Xipeng Qiu, Shu Zhang, Feng Ji and Xuanjing Huang
- 15:00-15:30 *Wiki-ly Supervised Part-of-Speech Tagging*  
Shen Li, João Graça and Ben Taskar

**Saturday, July 14, 2012 (continued)**

**Session 3-PM-1D: Word Sense Disambiguation**

Session Chair: Marc Dymetman

- 14:00-14:30 *Joining Forces Pays Off: Multilingual Joint Word Sense Disambiguation*  
Roberto Navigli and Simone Paolo Ponzetto
- 14:30-15:00 *A New Minimally-Supervised Framework for Domain Word Sense Disambiguation*  
Stefano Faralli and Roberto Navigli
- 15:00-15:30 *Grounded Models of Semantic Representation*  
Carina Silberer and Mirella Lapata

15:30-16:00 Coffee Break

**Session 3-PM-2A: Syntax and Parsing: Joint Parsing Models**

Session Chair: Jason Eisner

- 16:00-16:30 *Improved Parsing and POS Tagging Using Inter-Sentence Consistency Constraints*  
Alexander Rush, Roi Reichart, Michael Collins and Amir Globerson
- 16:30-17:00 *Unified Dependency Parsing of Chinese Morphological and Syntactic Structures*  
Zhongguo Li and Guodong Zhou
- 17:00-17:30 *A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing*  
Bernd Bohnet and Joakim Nivre

**Session 3-PM-2B: Social Media**

Session Chair: Alan Ritter

- 16:00-16:30 *Identifying Event-related Bursts via Social Media Activities*  
Xin Zhao, Baihan Shu, Jing Jiang, Yang Song, Hongfei Yan and Xiaoming Li
- 16:30-17:00 *User Demographics and Language in an Implicit Social Network*  
Katja Filippova
- 17:00-17:30 *Revisiting the Predictability of Language: Response Completion in Social Media*  
Bo Pang and Sujith Ravi

**Saturday, July 14, 2012 (continued)**

**Session 3-PM-2C: NLP Applications**

Session Chair: Chikara Hashimoto

16:00-16:30 *Supervised Text-based Geolocation Using Language Models on an Adaptive Grid*  
Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing and Jason Baldrige

16:30-17:00 *A Discriminative Model for Query Spelling Correction with Latent Structural SVM*  
Huizhong Duan, Yanen Li, ChengXiang Zhai and Dan Roth

17:00-17:30 *Characterizing Stylistic Elements in Syntactic Structure*  
Song Feng, Ritwik Banerjee and Yejin Choi

**Session 3-PM-3P: Closing**

17:30-17:45 Closing: Closing Remarks

