ACL-HLT 2011

Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies

Volume 1: Long Papers

We wish to thank our sponsors

PLATINUM SPONSORS

research at Google

Baidu 百度
百度一下 你就知道

GOLD SPONSOR

Microsoft Research

SILVER SPONSORS

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

YAHOO! LABS

BRONZE SPONSORS

Raytheon BBN Technologies

Appen ButlerHill

ETS *Listening. Learning. Leading.®*

UNIVERSITY of WASHINGTON

IBM Research

OREGON HEALTH & SCIENCE UNIVERSITY
CENTER FOR SPOKEN LANGUAGE UNDERSTANDING

OREGON HEALTH & SCIENCE UNIVERSITY
DEPARTMENT OF MEDICAL INFORMATICS AND CLINICAL EPIDEMIOLOGY

SUPPORTERS

Rakuten

WIDMER BROTHERS

viaLanguage. say what you mean®

Preferred Infrastructure.

at&t

SDL

ACL HLT 2011

The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies

Proceedings of the Conference

19-24 June, 2011
Portland, Oregon, USA

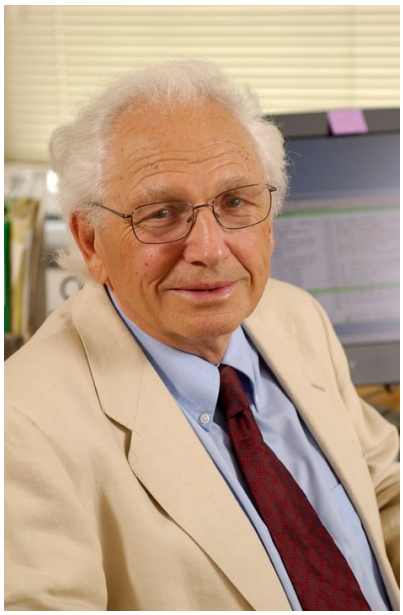
Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

©2011 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

ISBN 978-1-932432-87-9



We dedicate the ACL 2011 proceedings to the memory of Fred Jelinek (1932-2010), who received ACL's Lifetime Achievement Award in 2009. His award acceptance speech can be found in Computational Linguistics 35(4), and an obituary by Mark Liberman appeared in Computational Linguistics 36(4). Several other newspaper and professional society obituaries have described his extraordinary personal life and career.

Fred's influence on computational linguistics is almost impossible to overstate. In the 1970s and 1980s, he and his colleagues at IBM developed the statistical paradigm that dominates our field today, including a great many specific techniques for modeling, parameter estimation, and search that continue to enjoy wide use. Even more fundamentally, as Mark Liberman recounts in his obituary, Fred led the field away from a mode where lone inventors defended their designs by appealing to aesthetics and anecdotes, to a more communal and transparent process of evaluating methods objectively through controlled comparisons on training and test sets.

Under Fred's visionary leadership, the IBM group revolutionized speech recognition by adopting a statistical, data-driven perspective that was deeply at odds with the rationalist ethos of the time. The group began with Fred's information-theoretic reconceptualization of the task as recovering a source signal (text) after it had passed through a noisy channel. They then worked out the many components needed for a full speech recognizer, along with the training algorithms for each component and global decoding algorithms. Steve Young, in an obituary in the IEEE SLTC Newsletter, describes Fred as not a pioneer but the pioneer of speech recognition.

In the 1980s, the IBM speech group's work on language modeling drew them toward deeper analysis of text. Fred and his colleagues introduced NLP methods such as word clustering, HMM part-of-speech tagging, history-based parsing, and prefix probability computation in PCFGs. They famously turned their noisy-channel lens on machine translation, founding the field of statistical MT with a series of ingenious and highly influential models.

After Fred moved to Johns Hopkins University in 1993, he worked tirelessly to improve language modeling by incorporating syntactic and other long-range dependencies as well as semantic classes. He also presided for 16 years over the Johns Hopkins Summer Workshops, whose 51 teams from 1995-2010 attacked a wide range of topics in human language technology, many making groundbreaking advances in the field.

There is a popular conception that Fred was somehow hostile to linguistics. Certainly he liked to entertain others by repeating his 1988 quip that "Any time a linguist leaves the group, the recognition rate goes up." Yet he had tried to leave information theory for linguistics as early as 1962, influenced by Noam Chomsky's lectures and his wife Milena's earlier studies with Roman Jakobson. He always strove for clean formal models just as linguists do. He was deeply welcoming toward any attempt to improve models through better linguistics, as long as they had a large number of parameters. Indeed, it was one of the major frustrations of his career that it was so difficult to beat n-gram language models, when humans were evidently using additional linguistic and world knowledge to obtain much better predictive performance. As he said in an award acceptance speech in 2004, "My colleagues and I always hoped that linguistics will eventually allow us to strike gold."

Fred was skeptical only about the relevance of armchair linguistics to engineering, believing that there was far more variation in the data than could be described compactly by humans. For this reason, while he was quite interested in recovering or exploiting latent linguistic structure, he trusted human-annotated linguistic data to be a better description of that structure than human-conceived linguistic rules. Statistical models could be aided even by imperfect or incomplete annotations, such as unaligned orthographic transcriptions, bilingual corpora, or syntactic analyses furnished by ordinary speakers. Fred pushed successfully for the development of such resources, notably the IBM/Lancaster Treebank and its successor, the Penn Treebank.

Fred influenced many of us personally. He was warm-hearted, witty, cultured, thoughtful about the scientific process, a generous mentor, and always frank, honest, and unpretentious. The changes that he brought to our field are largely responsible for its recent empirical progress and commercial success. They have also helped make it attractive to many bright, technically sophisticated young researchers. This proceedings volume, which is dedicated to his memory, testifies to the overwhelming success of his leadership and vision.

By Jason Eisner, on behalf of ACL 2011 Organizing Committee

Preface: General Chair

Welcome to the 49th Annual Meeting of the Association for Computational Linguistics in Portland, Oregon. ACL is perhaps the longest-running conference series in computer science. Amazingly, it is still growing. We expect this year's ACL to attract an even larger number of participants than usual, since 2011 happens to be an off-year for COLING, EACL and NAACL.

The yearly success of ACL results from the dedication and hard work of many people. This year is no exception. I would like to thank all of them for volunteering their time and energy in service to our community.

I thank the Program Co-Chairs Rada Mihalcea and Yuji Matsumoto for putting together a wonderful main conference program, including 164 long papers, 128 short papers and much anticipated keynote speeches by David Ferrucci and Lera Boroditsky. Tutorial Co-Chairs, Patrick Pantel and Andy Way solicited proposals and selected six fascinating tutorials in a wide range of topics. The Workshop Co-Chairs, Hal Daume III and John Carroll, organized a joint selection process with EMNLP 2011. The program consists of 3 two-day workshops and 13 one-day workshops, a new record number for ACL. Sadao Kurohashi, Chair of System Demonstrations, assembled a committee and oversaw the review of 46 demo submissions.

The Student Session is organized by Co-Chairs, Sasa Petrovic, Emily Pitler, Ethan Selfridge and Faculty Advisors: Miles Osborne, Thamar Solorio. They introduced a new, poster-only format to be held in conjunction with the main ACL poster session. They also obtained NSF funding to provide travel support for all student session authors.

Special thank goes to Publication Chair, Guodong Zhou and his assistant Hong Yu. They produced the entire proceedings of the conference.

We are indebted to Brain Roark and the local arrangement committee for undertaking a phenomenal amount detailed work over the course of two years to host this conference, such as allocating appropriate space to meet all the needs of the scientific program, compiling and printing of the conference handbook, arranging a live tango band for the banquet and dance, to name just a few. The local arrangement committee consists of: Nate Bodenstab (webmeister), Peter Heeman (exhibitions), Christian Monson (student volunteers), Zak Shafran and Meg Mitchell (social), Richard Sproat (local sponsorship), Mahsa Yarmohammadi and Masoud Rouhizadeh (student housing coordinators) and Aaron Dunlop (local publications coordinator).

I want to express my gratitude to Ido Dagan, Chair of the ACL Conference Coordination Committee, Dragomir Radev, ACL Secretary, and Priscilla Rasmussen, ACL Business Manager, for their advice and guidance throughout the process.

ACL 2011 has two Platinum Sponsors (Google and Baidu), one Gold Sponsor (Microsoft), two Silver sponsors (Pacific Northwest National Lab and Yahoo!), and seven Bronze Sponsors and six Supporters. We are grateful for the financial support from these organizations. I would like to thank and applaud the tremendous effort by the ACL sponsorship committee: Srinivas Bangalore (AT&T), Massimiliano Ciaramita (Google), Kevin Duh (NTT), Michael Gamon (Microsoft), Stephen Pulman (Oxford), Priscilla Rasmussen (ACL), and Haifeng Wang (Baidu).

Finally, I would like to thank all the area chairs, workshop organizers, tutorial presenters, authors, reviewers and conference attendees for their participation and contribution. I hope everyone will have a great time sharing ideas and inspiring one another at this conference.

ACL 2011 General Chair
Dekang Lin, Google, Inc.

Preface: Program Committee Co-Chairs

Welcome to the program of the 2011 Conference of the Association for Computational Linguistics! ACL continues to grow, and this year the number of paper submissions broke once again the record set by previous years. We received a total of 1,146 papers, out of which 634 were submitted as long papers and 512 were submitted as short papers. 25.7

To achieve the goal of a broad technical program, we followed the initiative from last year and solicited papers under four main different categories: *theoretical computational linguistics*, *empirical/data-driven approaches*, *resources/evaluation*, and *applications/tools*. We also continued to accept other types of papers (e.g., surveys or challenge papers), although unlike the previous year, no separate category was created for these papers. The papers falling under one of the four categories were reviewed using specialized reviewed forms; we also had a general review form that was used to review the papers that did not fall under one of the four main categories.

A new initiative this year was to also accept papers accompanied by supplemental materials (software and/or datasets). In addition to the regular review of the research quality of the paper, the accompanied resources were also reviewed for their quality, and the acceptance or rejection decisions were made based on the quality of both the paper and the supplemental materials. Among all the submissions, a total of 84 papers were accompanied by a software package and 117 papers were accompanied by a dataset. Among all the accepted papers, 30 papers are accompanied by software and 35 papers are accompanied by a dataset. These materials will be hosted on the ACL web site under http://www.aclweb.org/supplementals.

We are delighted to have two distinguished invited speakers: Dr. David Ferrucci (Principal Investigator, IBM Research), who will talk about his team's work on building *Watson* – a deep question answering system that achieved champion-level performance at Jeopardy!, and Lera Boroditsky (Assistant Professor, Stanford University), who will give a presentation on her research on how the languages we speak shape the way we think. In addition, the recipient of the ACL Lifetime Achievement Award will present a plenary lecture during the final day of the conference.

As in previous years, there will be three awards, one for the best long paper, one for the best long paper by a student, and one for the best short paper. The candidates for the best paper awards were nominated by the area chairs, who took into consideration the feedback they received from the reviewers on whether a paper might merit a best paper prize. From among the nominations we received, we selected the top five candidates for the long and short papers, and the final awards were then selected by the area chairs together with the program co-chairs. The recipients of the best paper awards will present their papers in a plenary session during the second day of the conference.

There are many individuals to thank for their contributions to the conference program. First and foremost, we would like to thank the authors who submitted their work to ACL. The growing number of submissions reflects how broad and active our field is. We are deeply indebted to the area chairs and the reviewers for their hard work. They enabled us to select an exciting program and to provide valuable feedback to the authors. We thank the general conference chair Dekang Lin and the local arrangements committee headed by Brian Roark for their help and advice, as well as last year's program committee co-chairs, Stephen Clark and Sandra Carberry, for sharing their experiences. Additional thanks go to

the publications chair, Guodong Zhang, who put this volume together, and Yu Hong, who helped him with this task.

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.

Enjoy the conference!

ACL 2011 Program Co-Chairs
Yuji Matsumoto, Nara Institute of Science and Technology
Rada Mihalcea, University of North Texas

Organizing Committee

General Chair

Dekang Lin, Google

Local Arrangements Chair

Brian Roark, Oregon Health & Science University

Program Co-Chairs

Yuji Matsumoto, Nara Institute of Science and Technology
Rada Mihalcea, University of North Texas

Local Arrangements Committee

Nate Bodenstab, Oregon Health & Science University
Aaron Dunlop, Oregon Health & Science University
Peter Heeman, Oregon Health & Science University
Meg Mitchell, Oregon Health & Science University
Christian Monson, Nuance
Zak Shafran, Oregon Health & Science University
Richard Sproat, Oregon Health & Science University
Masoud Rouhizadeh, Oregon Health & Science University
Mahsa Yarmohammadi, Oregon Health & Science University

Publications Chair

Guodong Zhou, Suzhou University

Sponsorship Chairs

Haifeng Wang, Baidu
Kevin Duh, National Inst. of Information and Communications Technology
Massimiliano Ciaramita, Google
Michael Gamon, Microsoft
Priscilla Rasmussen, Association for Computational Linguistics
Srinivas Bangalore, AT&T
Stephen Pulman, Oxford University

Tutorial Co-chairs

Patrick Pantel, Microsoft Research
Andy Way, Dublin City University

Workshop Co-chairs

Hal Daume III, University of Maryland
John Carroll, University of Sussex

Demo Chair

Sadao Kurohashi, Kyoto University

Mentoring

Chair
- Tim Baldwin, University of Melbourne

Committee
- Chris Biemann, TU Darmstadt
- Mark Dras, Macquarie University
- Jeremy Nicholson, University of Melbourne

Student Research Workshop

Student Co-chairs
- Sasa Petrovic, University of Edinburgh
- Emily Pitler, University of Pennsylvania
- Ethan Selfridge, Oregon Health & Science University

Faculty Advisors
- Miles Osborne, University of Edinburgh
- Thamar Solorio, University of Alabama at Birmingham

ACL Conference Coordination Committee

Ido Dagan, Bar Ilan University (chair)
Chris Brew, Ohio State University
Graeme Hirst, University of Toronto
Lori Levin, Carnegie Mellon University
Christopher Manning, Stanford University
Dragomir Radev, University of Michigan
Owen Rambow, Columbia University
Priscilla Rasmussen, Association for Computational Linguistics
Suzanne Stevenson, University of Toronto

ACL Business Manager

Priscilla Rasmussen, Association for Computational Linguistics

Program Committee

Program Co-chairs

Yuji Matsumoto, Nara Institute of Science and Technology
Rada Mihalcea, University of North Texas

Area Chairs

Razvan Bunescu, Ohio University
Xavier Carreras, Technical University of Catalonia
Anna Feldman, Montclair University
Pascale Fung, Hong Kong University of Science and Technology
Chu-Ren Huang, Hong Kong Polytechnic University
Kentaro Inui, Tohoku University
Greg Kondrak, University of Alberta
Shankar Kumar, Google
Yang Liu, University of Texas at Dallas
Bernardo Magnini, Fondazione Bruno Kessler
Elliott Macklovitch, Marque d'Or
Katja Markert, University of Leeds
Lluis Marquez, Technical University of Catalonia
Diana McCarthy, Lexical Computing Ltd
Ryan McDonald, Google
Alessandro Moschitti, University of Trento
Vivi Nastase, Heidelberg Institute for Theoretical Studies
Manabu Okumura, Tokyo Institute of Technology
Vasile Rus, University of Memphis
Fabrizio Sebastiani, National Research Council of Italy
Michel Simard, National Research Council of Canada
Thamar Solorio, University of Alabama at Birmingham
Svetlana Stoyanchev, Open University
Carlo Strapparava, Fondazione Bruno Kessler
Dan Tufis, Romanian Academy of Artificial Intelligence
Xiaojun Wan, Peking University
Taro Watanabe, National Inst. of Information and Communications Technology
Alexander Yates, Temple University
Deniz Yuret, Koc University

Program Committee

Ahmed Abbasi, Eugene Agichtein, Eneko Agirre, Lars Ahrenberg, Gregory Aist, Enrique Alfonseca, Laura Alonso i Alemany, Gianni Amati, Alina Andreevskaia, Ion Androutsopoulos, Abhishek Arun, Masayuki Asahara, Nicholas Asher, Giuseppe Attardi, Necip Fazil Ayan

Collin Baker, Jason Baldridge, Tim Baldwin, Krisztian Balog, Carmen Banea, Verginica Barbu

Mititelu, Marco Baroni, Regina Barzilay, Roberto Basili, John Bateman, Tilman Becker, Lee Becker, Beata Beigman-Klebanov, Cosmin Bejan, Ron Bekkerman, Daisuke Bekki, Kedar Bellare, Anja Belz, Sabine Bergler, Shane Bergsma, Raffaella Bernardi, Nicola Bertoldi, Pushpak Bhattacharyya, Archana Bhattarai, Tim Bickmore, Chris Biemann, Dan Bikel, Alexandra Birch, Maria Biryukov, Alan Black, Roi Blanco, John Blitzer, Phil Blunsom, Gemma Boleda, Francis Bond, Kalina Bontcheva, Johan Bos, Gosse Bouma, Kristy Boyer, S.R.K. Branavan, Thorsten Brants, Eric Breck, Ulf Brefeld, Chris Brew, Ted Briscoe, Samuel Brody

Michael Cafarella, Aoife Cahill, Chris Callison-Burch, Rafael Calvo, Nicoletta Calzolari, Nicola Cancedda, Claire Cardie, Giuseppe Carenini, Claudio Carpineto, Marine Carpuat, Xavier Carreras, John Carroll, Ben Carterette, Francisco Casacuberta, Helena Caseli, Julio Castillo, Mauro Cettolo, Hakan Ceylan, Joyce Chai, Pi-Chuan Chang, Vinay Chaudhri, Berlin Chen, Ying Chen, Hsin-Hsi Chen, John Chen, Colin Cherry, David Chiang, Yejin Choi, Jennifer Chu-Carroll, Grace Chung, Kenneth Church, Massimiliano Ciaramita, Philipp Cimiano, Stephen Clark, Shay Cohen, Trevor Cohn, Nigel Collier, Michael Collins, John Conroy, Paul Cook, Ann Copestake, Bonaventura Coppola, Fabrizio Costa, Koby Crammer, Dan Cristea, Montse Cuadros, Silviu-Petru Cucerzan, Aron Culotta, James Curran

Walter Daelemans, Robert Damper, Hoa Dang, Dipanjan Das, Hal Daume, Adria de Gispert, Marie-Catherine de Marneffe, Gerard de Melo, Maarten de Rijke, Vera Demberg, Steve DeNeefe, John DeNero, Pascal Denis, Ann Devitt, Giuseppe Di Fabbrizio, Mona Diab, Markus Dickinson, Mike Dillinger, Bill Dolan, Doug Downey, Markus Dreyer, Greg Druck, Kevin Duh, Chris Dyer, Marc Dymetman

Markus Egg, Koji Eguchi, Andreas Eisele, Jacob Eisenstein, Jason Eisner, Michael Elhadad, Tomaz Erjavec, Katrin Erk, Hugo Escalante, Andrea Esuli

Hui Fang, Alex Chengyu Fang, Benoit Favre, Anna Feldman, Christiane Fellbaum, Donghui Feng, Raquel Fernandez, Nicola Ferro, Katja Filippova, Jenny Finkel, Seeger Fisher, Margaret Fleck, Dan Flickinger, Corina Forascu, Kate Forbes-Riley, Mikel L. Forcada, Eric Fosler-Lussier, Jennifer Foster, George Foster, Anette Frank, Alex Fraser, Dayne Freitag, Guohong Fu, Hagen Fuerstenau, Pascale Fung, Sadaoki Furui

Evgeniy Gabrilovich, Robert Gaizauskas, Michel Galley, Michael Gamon, Kuzman Ganchev, Jianfeng Gao, Claire Gardent, Thomas Gärtner, Albert Gatt, Dmitriy Genzel, Kallirroi Georgila, Carlo Geraci, Pablo Gervas, Shlomo Geva, Daniel Gildea, Alastair Gill, Dan Gillick, Jesus Gimenez, Kevin Gimpel, Roxana Girju, Claudio Giuliano, Amir Globerson, Yoav Goldberg, Sharon Goldwater, Carlos Gomez Rodriguez, Julio Gonzalo, Brigitte Grau, Stephan Greene, Ralph Grishman, Tunga Gungor, Zhou GuoDong, Iryna Gurevych, David Guthrie

Nizar Habash, Ben Hachey, Barry Haddow, Gholamreza Haffari, Aria Haghighi, Udo Hahn, Jan Hajic, Dilek Hakkani-Tür, Keith Hall, Jirka Hana, John Hansen, Sanda Harabagiu, Mark Hasegawa-Johnson, Koiti Hasida, Ahmed Hassan, Katsuhiko Hayashi, Ben He, Xiaodong He, Ulrich Heid, Michael Heilman, Ilana Heintz, Jeff Heinz, John Henderson, James Henderson, Iris

Hendrickx, Aurelie Herbelot, Erhard Hinrichs, Tsutomu Hirao, Julia Hirschberg, Graeme Hirst, Julia Hockenmaier, Tracy Holloway King, Bo-June (Paul) Hsu, Xuanjing Huang, Liang Huang, Jimmy Huang, Jian Huang, Chu-Ren Huang, Juan Huerta, Rebecca Hwa

Nancy Ide, Gonzalo Iglesias, Gabriel Infante-López, Diana Inkpen, Radu Ion, Elena Irimia, Pierre Isabelle, Mitsuru Ishizuka, Aminul Islam, Abe Ittycheriah, Tomoharu Iwata

Martin Jansche, Sittichai Jiampojamarn, Jing Jiang, Valentin Jijkoun, Richard Johansson, Mark Johnson, Aravind Joshi

Nanda Kambhatla, Min-Yen Kan, Kyoko Kanzaki, Rohit Kate, Junichi Kazama, Bill Keller, Andre Kempe, Philipp Keohn, Fazel Keshtkar, Adam Kilgarriff, Jin-Dong Kim, Su Nam Kim, Brian Kingsbury, Katrin Kirchhoff, Ioannis Klapaftis, Dan Klein, Alexandre Klementiev, Kevin Knight, Rob Koeling, Oskar Kohonen, Alexander Kolcz, Alexander Koller, Kazunori Komatani, Terry Koo, Moshe Koppel, Valia Kordoni, Anna Korhonen, Andras Kornai, Zornitsa Kozareva, Lun-Wei Ku, Sandra Kuebler, Marco Kuhlmann, Roland Kuhn, Mikko Kurimo, Oren Kurland, Olivia Kwong

Krista Lagus, Philippe Langlais, Guy Lapalme, Mirella Lapata, Dominique Laurent, Alberto Lavelli, Matthew Lease, Gary Lee, Kiyong Lee, Els Lefever, Alessandro Lenci, James Lester, Gina-Anne Levow, Tao Li, Shoushan LI, Fangtao Li, Zhifei Li, Haizhou Li, Hang Li, Wenjie Li, Percy Liang, Chin-Yew Lin, Frank Lin, Mihai Lintean, Ken Litkowski, Diane Litman, Marina Litvak, Yang Liu, Bing Liu, Qun Liu, Jingjing Liu, Elena Lloret, Birte Loenneker-Rodman, Adam Lopez, Annie Louis, Xiaofei Lu, Yue Lu

Tengfei Ma, Wolfgang Macherey, Klaus Macherey, Elliott Macklovitch, Nitin Madnani, Bernardo Magnini, Suresh Manandhar, Gideon Mann, Chris Manning, Daniel Marcu, David Martínez, Andre Martins, Yuval Marton, Sameer Maskey, Spyros Matsoukas, Mausam, Arne Mauser, Jon May, David McAllester, Andrew McCallum, David McClosky, Ryan McDonald, Bridget McInnes, Tara McIntosh, Kathleen McKeown, Paul McNamee, Yashar Mehdad, Qiaozhu Mei, Arul Menezes, Paola Merlo, Donald Metzler, Adam Meyers, Haitao Mi, Jeff Mielke, Einat Minkov, Yusuke Miyao, Dunja Mladenic, Marie-Francine Moens, Saif Mohammad, Dan Moldovan, Diego Molla, Christian Monson, Manuel Montes y Gomez, Raymond Mooney, Robert Moore, Tatsunori Mori, Glyn Morrill, Sara Morrissey, Alessandro Moschitti, Jack Mostow, Smaranda Muresan, Gabriel Murray, Gabriele Musillo, Sung-Hyon Myaeng

Tetsuji Nakagawa, Mikio Nakano, Preslav Nakov, Ramesh Nallapati, Vivi Nastase, Borja Navarro-Colorado, Roberto Navigli, Mark-Jan Nederhof, Matteo Negri, Ani Nenkova, Graham Neubig, Guenter Neumann, Vincent Ng, Hwee Tou Ng, Patrick Nguyen, Jian-Yun Nie, Rodney Nielsen, Joakim Nivre, Tadashi Nomoto, Scott Nowson

Diarmuid Ó Séaghdha, Sharon O'Brien, Franz Och, Stephan Oepen, Kemal Oflazer, Jong-Hoon Oh, Constantin Orasan, Miles Osborne, Gozde Ozbal

Sebastian Pado, Tim Paek, Bo Pang, Patrick Pantel, Soo-Min Pantel, Ivandre Paraboni, Cecile Paris, Marius Pasca, Gabriella Pasi, Andrea Passerini, Rebecca J. Passonneau, Siddharth Patwardhan, Adam Pauls, Adam Pease, Ted Pedersen, Anselmo Penas, Anselmo Peñas, Jing Peng, Fuchun Peng, Gerald Penn, Marco Pennacchiotti, Wim Peters, Slav Petrov, Emanuele Pianta, Michael Picheny, Daniele Pighin, Manfred Pinkal, David Pinto, Stelios Piperidis, Paul Piwek, Benjamin Piwowarski, Massimo Poesio, Livia Polanyi, Simone Paolo Ponzetto, Hoifung Poon, Ana-Maria Popescu, Andrei Popescu-Belis, Maja Popovic, Martin Potthast, Richard Power, Sameer Pradhan, John Prager, Rashmi Prasad, Partha Pratim Talukdar, Adam Przepiórkowski, Vasin Punyakanok, Matthew Purver, Sampo Pyysalo

Silvia Quarteroni, Ariadna Quattoni, Chris Quirk

Stephan Raaijmakers, Dragomir Radev, Filip Radlinski, Bhuvana Ramabhadran, Ganesh Ramakrishnan, Owen Rambow, Aarne Ranta, Delip Rao, Ari Rappoport, Lev Ratinov, Antoine Raux, Emmanuel Rayner, Roi Reichart, Ehud Reiter, Steve Renals, Philip Resnik, Giuseppe Riccardi, Sebastian Riedel, Stefan Riezler, German Rigau, Ellen Riloff, Laura Rimell, Eric Ringger, Horacio Rodríguez, Paolo Rosso, Antti-Veikko Rosti, Rachel Edita Roxas, Alex Rudnicky, Marta Ruiz Costa-Jussa, Vasile Rus, Graham Russell, Anton Rytting

Rune Sætre, Kenji Sagae, Horacio Saggion, Tapio Salakoski, Agnes Sandor, Sudeshna Sarkar, Anoop Sarkar, Giorgio Satta, Hassan Sawaf, Frank Schilder, Anne Schiller, David Schlangen, Sabine Schulte im Walde, Tanja Schultz, Holger Schwenk, Donia Scott, Yohei Seki, Satoshi Sekine, Stephanie Seneff, Jean Senellart, Violeta Seretan, Burr Settles, Serge Sharoff, Dou Shen, Wade Shen, Libin Shen, Kiyoaki Shirai, Luo Si, Grigori Sidorov, Mário Silva, Fabrizio Silvestri, Khalil Simaan, Michel Simard, Gabriel Skantze, Noah Smith, Matthew Snover, Rion Snow, Benjamin Snyder, Stephen Soderland, Marina Sokolova, Thamar Solorio, Swapna Somasundaran, Lucia Specia, Valentin Spitkovsky, Richard Sproat, Manfred Stede, Mark Steedman, Amanda Stent, Mark Stevenson, Svetlana Stoyanchev, Veselin Stoyanov, Michael Strube, Sara Stymne, Keh-Yih Su, Fangzhong Su, Jian Su, L Venkata Subramaniam, David Suendermann, Maosong Sun, Mihai Surdeanu, Richard Sutcliffe, Charles Sutton, Jun Suzuki, Stan Szpakowicz, Idan Szpektor

Hiroya Takamura, David Talbot, Irina Temnikova, Michael Tepper, Simone Teufel, Stefan Thater, Allan Third, Jörg Tiedemann, Christoph Tillmann, Ivan Titov, Takenobu Tokunaga, Kentaro Torisawa, Kristina Toutanova, Isabel Trancoso, Richard Tsai, Vivian Tsang, Dan Tufis

Takehito Utsuro

Shivakumar Vaithyanathan, Alessandro Valitutti, Antal van den Bosch, Hans van Halteren, Gertjan van Noord, Lucy Vanderwende, Vasudeva Varma, Tony Veale, Olga Vechtomova, Paola Velardi, Rene Venegas, Ashish Venugopal, Jose Luis Vicedo, Evelyne Viegas, David Vilar, Begona Villada Moiron, Sami Virpioja, Andreas Vlachos, Stephan Vogel, Piek Vossen

Michael Walsh, Xiaojun Wan, Xinglong Wang, Wei Wang, Haifeng Wang, Justin Washtell, Andy

Way, David Weir, Ben Wellner, Ji-Rong Wen, Chris Wendt, Michael White, Ryen White, Richard Wicentowski, Jan Wiebe, Sandra Williams, Jason Williams, Theresa Wilson, Shuly Wintner, Kam-Fai Wong, Fei Wu

Deyi Xiong, Peng Xu, Jinxi Xu, Nianwen Xue

Scott Wen-tau Yih, Emine Yilmaz

David Zajic, Fabio Zanzotto, Richard Zens, Torsten Zesch, Hao Zhang, Bing Zhang, Min Zhang, Huarui Zhang, Jun Zhao, Bing Zhao, Jing Zheng, Li Hai Zhou, Michael Zock, Andreas Zollmann, Geoffrey Zweig, Pierre Zweigenbaum

Secondary Reviewers

Omri Abend, Rodrigo Agerri, Paolo Annesi, Wilker Aziz, Tyler Baldwin, Verginica Barbu Mititelu, David Batista, Delphine Bernhard, Stephen Boxwell, Janez Brank, Chris Brockett, Tim Buckwalter, Wang Bukang, Alicia Burga, Steven Burrows, Silvia Calegari, Marie Candito, Marina Cardenas, Bob Carpenter, Paula Carvalho, Diego Ceccarelli, Asli Celikyilmaz, Soumaya Chaffar, Bin Chen, Danilo Croce, Daniel Dahlmeier, Hong-Jie Dai, Mariam Daoud, Steven DeNeefe, Leon Derczynski, Elina Desypri, Sobha Lalitha Devi, Gideon Dror, Loic Dugast, Eraldo Fernandes, Jody Foo, Kotaro Funakoshi, Jing Gao, Wei Gao, Diman Ghazi, Julius Goth, Joseph Grafsgaard, Eun Young Ha, Robbie Haertel, Matthias Hagen, Enrique Henestroza, Hieu Hoang, Maria Holmqvist, Dennis Hoppe, Yunhua Hu, Yun Huang, Radu Ion, Elena Irimia, Jagadeesh Jagarlamudi, Antonio Juárez-González, Sun Jun, Evangelos Kanoulas, Aaron Kaplan, Caroline Lavecchia, Lianhau Lee, Michael Levit, Ping Li, Thomas Lin, Wang Ling, Ying Liu, José David Lopes, Bin Lu, Jia Lu, Saab Mansour, Raquel Martinez-Unanue, Haitao Mi, Simon Mille, Teruhisa Misu, Behrang Mohit, Sílvio Moreira, Rutu Mulkar-Mehta, Jason Naradowsky, Sudip Naskar, Heung-Seon Oh, You Ouyang, Lluís Padró, Sujith Ravi, Marta Recasens, Luz Rello, Stefan Rigo, Alan Ritter, Alvaro Rodrigo, Hasim Sak, Kevin Seppi, Aliaksei Severyn, Chao Shen, Shuming Shi, Laurianne Sitbon, Jun Sun, György Szarvas, Eric Tang, Alberto Téllez-Valero, Luong Minh Thang, Gabriele Tolomei, David Tomás, Diana Trandabat, Zhaopeng Tu, Gokhan Tur, Kateryna Tymoshenko, Fabienne Venant, Esaú Villatoro-Tello, Joachim Wagner, Dan Walker, Wei Wei, Xinyan Xiao, Jun Xie, Hao Xiong, Gu Xu, Jun Xu, Huichao Xue, Taras Zagibalov, Beñat Zapirain, Kalliopi Zervanou, Renxian Zhang, Daqi Zheng, Arkaitz Zubiaga

Table of Contents

A Word-Class Approach to Labeling PSCFG Rules for Machine Translation
Andreas Zollmann and Stephan Vogel 1

Deciphering Foreign Language
Sujith Ravi and Kevin Knight 12

Effective Use of Function Words for Rule Generalization in Forest-Based Translation
Xianchao Wu, Takuya Matsuzaki and Jun'ichi Tsujii 22

Combining Morpheme-based Machine Translation with Post-processing Morpheme Prediction
Ann Clifton and Anoop Sarkar 32

Evaluating the Impact of Coder Errors on Active Learning
Ines Rehbein and Josef Ruppenhofer 43

A Fast and Accurate Method for Approximate String Search
Ziqi Wang, Gu Xu, Hang Li and Ming Zhang 52

Domain Adaptation by Constraining Inter-Domain Variability of Latent Feature Representation
Ivan Titov 62

Exact Decoding of Syntactic Translation Models through Lagrangian Relaxation
Alexander M. Rush and Michael Collins 72

Jigs and Lures: Associating Web Queries with Structured Entities
Patrick Pantel and Ariel Fuxman 83

Semi-Supervised SimHash for Efficient Document Similarity Search
Qixia Jiang and Maosong Sun 93

Joint Annotation of Search Queries
Michael Bendersky, W. Bruce Croft and David A. Smith 102

Query Weighting for Ranking Model Adaptation
Peng Cai, Wei Gao, Aoying Zhou and Kam-Fai Wong 112

Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification
Yulan He, Chenghua Lin and Harith Alani 123

Using Multiple Sources to Construct a Sentiment Sensitive Thesaurus for Cross-Domain Sentiment Classification
Danushka Bollegala, David Weir and John Carroll 132

Learning Word Vectors for Sentiment Analysis
Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts 142

Target-dependent Twitter Sentiment Classification
Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu and Tiejun Zhao 151

A Comprehensive Dictionary of Multiword Expressions
Kosho Shudo, Akira Kurahone and Toshifumi Tanabe 161

Multi-Modal Annotation of Quest Games in Second Life
Sharon Gower Small, Jennifer Strommer-Galley and Tomek Strzalkowski 171

A New Dataset and Method for Automatically Grading ESOL Texts
Helen Yannakoudakis, Ted Briscoe and Ben Medlock 180

Collecting Highly Parallel Data for Paraphrase Evaluation
David Chen and William Dolan 190

A Large Scale Distributed Syntactic, Semantic and Lexical Language Model for Machine Translation
Ming Tan, Wenli Zhou, Lei Zheng and Shaojun Wang 201

Goodness: A Method for Measuring Machine Translation Confidence
Nguyen Bach, Fei Huang and Yaser Al-Onaizan 211

MEANT: An inexpensive, high-accuracy, semi-automatic metric for evaluating translation utility based on semantic roles
Chi-kiu Lo and Dekai Wu 220

An exponential translation model for target language morphology
Michael Subotin 230

Bayesian Inference for Zodiac and Other Homophonic Ciphers
Sujith Ravi and Kevin Knight 239

Interactive Topic Modeling
Yuening Hu, Jordan Boyd-Graber and Brianna Satinoff 248

Faster and Smaller N-Gram Language Models
Adam Pauls and Dan Klein 258

Learning to Win by Reading Manuals in a Monte-Carlo Framework
S.R.K Branavan, David Silver and Regina Barzilay 268

Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity
Tony Veale 278

Local Histograms of Character N-grams for Authorship Attribution
Hugo Jair Escalante, Thamar Solorio and Manuel Montes-y-Gomez 288

Word Maturity: Computational Modeling of Word Knowledge
Kirill Kireyev and Thomas K Landauer 299

Finding Deceptive Opinion Spam by Any Stretch of the Imagination
Myle Ott, Yejin Choi, Claire Cardie and Jeffrey T. Hancock 309

Joint Bilingual Sentiment Classification with Unlabeled Parallel Corpora
Bin Lu, Chenhao Tan, Claire Cardie and Benjamin K. Tsou 320

A Pilot Study of Opinion Summarization in Conversations
Dong Wang and Yang Liu 331

Contrasting Opposing Views of News Articles on Contentious Issues
Souneil Park, Kyung Soon Lee and Junehwa Song 340

Content Models with Attitude
Christina Sauper, Aria Haghighi and Regina Barzilay 350

Recognizing Named Entities in Tweets
Xiaohua LIU, Shaodian ZHANG, Furu WEI and Ming ZHOU 359

Lexical Normalisation of Short Text Messages: Makn Sens a #twitter
Bo Han and Timothy Baldwin 368

Topical Keyphrase Extraction from Twitter
Xin Zhao, Jing Jiang, Jing He, Yang Song, Palakorn Achanauparp, Ee-Peng Lim and Xiaoming Li
379

Event Discovery in Social Media Feeds
Edward Benson, Aria Haghighi and Regina Barzilay 389

How do you pronounce your name? Improving G2P with transliterations
Aditya Bhargava and Grzegorz Kondrak 399

Unsupervised Word Alignment with Arbitrary Features
Chris Dyer, Jonathan H. Clark, Alon Lavie and Noah A. Smith 409

Model-Based Aligner Combination Using Dual Decomposition
John DeNero and Klaus Macherey 420

An Algorithm for Unsupervised Transliteration Mining with an Application to Word Alignment
Hassan Sajjad, Alexander Fraser and Helmut Schmid 430

Beam-Width Prediction for Efficient Context-Free Parsing
Nathan Bodenstab, Aaron Dunlop, Keith Hall and Brian Roark 440

Optimal Head-Driven Parsing Complexity for Linear Context-Free Rewriting Systems
Pierluigi Crescenzi, Daniel Gildea, Andrea Marino, Gianluca Rossi and Giorgio Satta 450

Prefix Probability for Probabilistic Synchronous Context-Free Grammars
Mark-Jan Nederhof and Giorgio Satta 460

A Comparison of Loopy Belief Propagation and Dual Decomposition for Integrated CCG Supertagging and Parsing
Michael Auli and Adam Lopez 470

Jointly Learning to Extract and Compress
Taylor Berg-Kirkpatrick, Dan Gillick and Dan Klein 481

Discovery of Topically Coherent Sentences for Extractive Summarization
Asli Celikyilmaz and Dilek Hakkani-Tur 491

Coherent Citation-Based Summarization of Scientific Papers
Amjad Abu-Jbara and Dragomir Radev 500

A Class of Submodular Functions for Document Summarization
Hui Lin and Jeff Bilmes 510

Semi-supervised Relation Extraction with Large-scale Word Clustering
Ang Sun, Ralph Grishman and Satoshi Sekine 521

In-domain Relation Discovery with Meta-constraints via Posterior Regularization
Harr Chen, Edward Benson, Tahira Naseem and Regina Barzilay 530

Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations
Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer and Daniel S. Weld 541

Exploiting Syntactico-Semantic Structures for Relation Extraction
Yee Seng Chan and Dan Roth 551

Together We Can: Bilingual Bootstrapping for WSD
Mitesh M. Khapra, Salil Joshi, Arindam Chatterjee and Pushpak Bhattacharyya 561

Which Noun Phrases Denote Which Concepts?
Jayant Krishnamurthy and Tom Mitchell 570

Semantic Representation of Negation Using Focus Detection
Eduardo Blanco and Dan Moldovan 581

Learning Dependency-Based Compositional Semantics
Percy Liang, Michael Jordan and Dan Klein 590

Unsupervised Part-of-Speech Tagging with Bilingual Graph-Based Projections
Dipanjan Das and Slav Petrov 600

Global Learning of Typed Entailment Rules
Jonathan Berant, Ido Dagan and Jacob Goldberger 610

Incremental Syntactic Language Models for Phrase-based Translation
Lane Schwartz, Chris Callison-Burch, William Schuler and Stephen Wu 620

An Unsupervised Model for Joint Phrase Alignment and Extraction
Graham Neubig, Taro Watanabe, Eiichiro Sumita, Shinsuke Mori and Tatsuya Kawahara 632

Learning Hierarchical Translation Structure with Linguistic Annotations
Markos Mylonakis and Khalil Sima'an 642

Phrase-Based Translation Model for Question Retrieval in Community Question Answer Archives
Guangyou Zhou, Li Cai, Jun Zhao and Kang Liu 653

Neutralizing Linguistically Problematic Annotations in Unsupervised Dependency Parsing Evaluation
Roy Schwartz, Omri Abend, Roi Reichart and Ari Rappoport 663

Dynamic Programming Algorithms for Transition-Based Dependency Parsers
Marco Kuhlmann, Carlos Gómez-Rodríguez and Giorgio Satta 673

Shift-Reduce CCG Parsing
Yue Zhang and Stephen Clark 683

Web-Scale Features for Full-Scale Parsing
Mohit Bansal and Dan Klein 693

The impact of language models and loss functions on repair disfluency detection
Simon Zwarts and Mark Johnson 703

Learning Sub-Word Units for Open Vocabulary Speech Recognition
Carolina Parada, Mark Dredze, Abhinav Sethy and Ariya Rastrow 712

Computing and Evaluating Syntactic Complexity Features for Automated Scoring of Spontaneous Non-Native Speech
Miao Chen and Klaus Zechner 722

N-Best Rescoring Based on Pitch-accent Patterns
Je Hun Jeon, Wen Wang and Yang Liu 732

Lexically-Triggered Hidden Markov Models for Clinical Document Coding
Svetlana Kiritchenko and Colin Cherry 742

Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments
Michael Mohler, Razvan Bunescu and Rada Mihalcea 752

Age Prediction in Blogs: A Study of Style, Content, and Online Behavior in Pre- and Post-Social Media Generations
Sara Rosenthal and Kathleen McKeown 763

Extracting Social Power Relationships from Natural Language
Philip Bramsen, Martha Escobar-Molano, Ami Patel and Rafael Alonso 773

Bootstrapping coreference resolution using word associations
Hamidreza Kobdani, Hinrich Schuetze, Michael Schiehlen and Hans Kamp 783

Large-Scale Cross-Document Coreference Using Distributed Inference and Hierarchical Models
Sameer Singh, Amarnag Subramanya, Fernando Pereira and Andrew McCallum 793

A Cross-Lingual ILP Solution to Zero Anaphora Resolution
Ryu Iida and Massimo Poesio 804

Coreference Resolution with World Knowledge
Altaf Rahman and Vincent Ng 814

How to train your multi bottom-up tree transducer
Andreas Maletti 825

Binarized Forest to String Translation
Hao Zhang, Licheng Fang, Peng Xu and Xiaoyun Wu 835

Learning to Transform and Select Elementary Trees for Improved Syntax-based Machine Translations
Bing Zhao, Young-Suk Lee, Xiaoqiang Luo and Liu Li 846

Rule Markov Models for Fast Tree-to-String Translation
Ashish Vaswani, Haitao Mi, Liang Huang and David Chiang 856

A Hierarchical Pitman-Yor Process HMM for Unsupervised Part of Speech Induction
Phil Blunsom and Trevor Cohn 865

Using Deep Morphology to Improve Automatic Error Detection in Arabic Handwriting Recognition
Nizar Habash and Ryan Roth 875

A Discriminative Model for Joint Morphological Disambiguation and Dependency Parsing
John Lee, Jason Naradowsky and David A. Smith 885

Unsupervised Bilingual Morpheme Segmentation and Alignment with Context-rich Hidden Semi-Markov Models
Jason Naradowsky and Kristina Toutanova 895

A Graph Approach to Spelling Correction in Domain-Centric Search
Zhuowei Bao, Benny Kimelfeld and Yunyao Li 905

Grammatical Error Correction with Alternating Structure Optimization
Daniel Dahlmeier and Hwee Tou Ng 915

Algorithm Selection and Model Adaptation for ESL Correction Tasks
Alla Rozovskaya and Dan Roth 924

Automated Whole Sentence Grammar Correction Using a Noisy Channel Model
Y. Albert Park and Roger Levy 934

A Generative Entity-Mention Model for Linking Entities with Knowledge Base
Xianpei Han and Le Sun 945

Simple supervised document geolocation with geodesic grids
Benjamin Wing and Jason Baldridge 955

Piggyback: Using Search Engines for Robust Cross-Domain Named Entity Recognition
Stefan Rüd, Massimiliano Ciaramita, Jens Müller and Hinrich Schütze 965

Template-Based Information Extraction without the Templates
Nathanael Chambers and Dan Jurafsky 976

Classifying arguments by scheme
Vanessa Wei Feng and Graeme Hirst 987

Automatically Evaluating Text Coherence Using Discourse Relations
Ziheng Lin, Hwee Tou Ng and Min-Yen Kan 997

Underspecifying and Predicting Voice for Surface Realisation Ranking
Sina Zarrieß, Aoife Cahill and Jonas Kuhn 1007

Recognizing Authority in Dialogue with an Integer Linear Programming Constrained Model
Elijah Mayfield and Carolyn Penstein Rosé 1018

Reordering Metrics for MT
Alexandra Birch and Miles Osborne 1027

Reordering with Source Language Collocations
Zhanyi Liu, Haifeng Wang, Hua Wu, Ting Liu and Sheng Li 1036

A Joint Sequence Translation Model with Integrated Reordering
Nadir Durrani, Helmut Schmid and Alexander Fraser 1045

Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results
Roger Levy 1055

Metagrammar engineering: Towards systematic exploration of implemented grammars
Antske Fokkens 1066

Simple Unsupervised Grammar Induction from Raw Text with Cascaded Finite State Models
Elias Ponvert, Jason Baldridge and Katrin Erk 1077

Extracting Paraphrases from Definition Sentences on the Web
Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jun'ichi Kazama and Sadao Kurohashi 1087

Learning From Collective Human Behavior to Introduce Diversity in Lexical Choice
Vahed Qazvinian and Dragomir R. Radev 1098

Ordering Prenominal Modifiers with a Reranking Approach
Jenny Liu and Aria Haghighi 1109

Unsupervised Semantic Role Induction via Split-Merge Clustering
Joel Lang and Mirella Lapata 1117

Using Cross-Entity Inference to Improve Event Extraction
Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou and Qiaoming Zhu 1127

Peeling Back the Layers: Detecting Event Role Fillers in Secondary Contexts
Ruihong Huang and Ellen Riloff 1137

Knowledge Base Population: Successful Approaches and Challenges
Heng Ji and Ralph Grishman 1148

Nonlinear Evidence Fusion and Propagation for Hyponymy Relation Mining
Fan Zhang, Shuming Shi, Jing Liu, Shuqi Sun and Chin-Yew Lin 1159

A Pronoun Anaphora Resolution System based on Factorial Hidden Markov Models
Dingcheng Li, Tim Miller and William Schuler 1169

Disentangling Chat with Local Coherence Models
Micha Elsner and Eugene Charniak 1179

An Affect-Enriched Dialogue Act Classification Model for Task-Oriented Dialogue
Kristy Boyer, Joseph Grafsgaard, Eun Young Ha, Robert Phillips and James Lester 1190

Fine-Grained Class Label Markup of Search Queries
Joseph Reisinger and Marius Pasca 1200

Creating a manually error-tagged and shallow-parsed learner corpus
Ryo Nagata, Edward Whittaker and Vera Sheinman 1210

Crowdsourcing Translation: Professional Quality from Non-Professionals
Omar F. Zaidan and Chris Callison-Burch 1220

A Statistical Tree Annotator and Its Applications
Xiaoqiang Luo and Bing Zhao 1230

Consistent Translation using Discriminative Learning - A Translation Memory-inspired Approach
Yanjun Ma, Yifan He, Andy Way and Josef van Genabith 1239

Machine Translation System Combination by Confusion Forest
Taro Watanabe and Eiichiro Sumita 1249

Hypothesis Mixture Decoding for Statistical Machine Translation
Nan Duan, Mu Li and Ming Zhou 1258

Minimum Bayes-risk System Combination
Jesús González-Rubio, Alfons Juan and Francisco Casacuberta 1268

Adjoining Tree-to-String Translation
Yang Liu, Qun Liu and Yajuan Lü 1278

Enhancing Language Models in Statistical Machine Translation with Backward N-grams and Mutual Information Triggers
Deyi Xiong, Min Zhang and Haizhou Li 1288

Translating from Morphologically Complex Languages: A Paraphrase-Based Approach
Preslav Nakov and Hwee Tou Ng 1298

Gappy Phrasal Alignment By Agreement
Mohit Bansal, Chris Quirk and Robert Moore 1308

Translationese and Its Dialects
Moshe Koppel and Noam Ordan 1318

Rare Word Translation Extraction from Aligned Comparable Documents
Emmanuel Prochasson and Pascale Fung 1327

Using Bilingual Parallel Corpora for Cross-Lingual Textual Entailment
Yashar Mehdad, Matteo Negri and Marcello Federico 1336

Using Large Monolingual and Bilingual Corpora to Improve Coordination Disambiguation
Shane Bergsma, David Yarowsky and Kenneth Church 1346

Unsupervised Decomposition of a Document into Authorial Components
Moshe Koppel, Navot Akiva, Idan Dershowitz and Nachum Dershowitz 1356

Discovering Sociolinguistic Associations with Structured Sparsity
Jacob Eisenstein, Noah A. Smith and Eric P. Xing 1365

Local and Global Algorithms for Disambiguation to Wikipedia
Lev Ratinov, Dan Roth, Doug Downey and Mike Anderson 1375

A Stacked Sub-Word Model for Joint Chinese Word Segmentation and Part-of-Speech Tagging
Weiwei Sun 1385

Language-independent compound splitting with morphological operations
Klaus Macherey, Andrew Dai, David Talbot, Ashok Popat and Franz Och 1395

Parsing the Internal Structure of Words: A New Paradigm for Chinese Word Segmentation
Zhongguo Li 1405

A Simple Measure to Assess Non-response
Anselmo Peñas and Alvaro Rodrigo 1415

Improving Question Recommendation by Exploiting Information Need
Shuguang Li and Suresh Manandhar 1425

Semi-Supervised Frame-Semantic Parsing for Unknown Predicates
Dipanjan Das and Noah A. Smith 1435

A Bayesian Model for Unsupervised Semantic Parsing
Ivan Titov and Alexandre Klementiev 1445

Unsupervised Learning of Semantic Relation Composition
Eduardo Blanco and Dan Moldovan 1456

Unsupervised Discovery of Domain-Specific Knowledge from Text
Dirk Hovy, Chunliang Zhang, Eduard Hovy and Anselmo Peñas 1466

Latent Semantic Word Sense Induction and Disambiguation
Tim Van de Cruys and Marianna Apidianaki 1476

Confidence Driven Unsupervised Semantic Parsing
Dan Goldwasser, Roi Reichart, James Clarke and Dan Roth 1486

Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews
Jianxing Yu, Zheng-Jun Zha, Meng Wang and Tat-Seng Chua 1496

Collective Classification of Congressional Floor-Debate Transcripts
Clinton Burfoot, Steven Bird and Timothy Baldwin 1506

Integrating history-length interpolation and classes in language modeling
Hinrich Schütze 1516

Structural Topic Model for Latent Topical Structure Analysis
Hongning Wang, Duo Zhang and ChengXiang Zhai 1526

Automatic Labelling of Topic Models
Jey Han Lau, Karl Grieser, David Newman and Timothy Baldwin 1536

Using Bilingual Information for Cross-Language Document Summarization
Xiaojun Wan 1546

Exploiting Web-Derived Selectional Preference to Improve Statistical Dependency Parsing
Guangyou Zhou, Jun Zhao, Kang Liu and Li Cai 1556

Effective Measures of Domain Similarity for Parsing
Barbara Plank and Gertjan van Noord 1566

Efficient CCG Parsing: A versus Adaptive Supertagging*
Michael Auli and Adam Lopez 1577

Improving Arabic Dependency Parsing with Form-based and Functional Morphological Features
Yuval Marton, Nizar Habash and Owen Rambow 1586

Partial Parsing from Bitext Projections
Prashanth Mannem and Aswarth Dara 1597

Ranking Class Labels Using Query Sessions
Marius Pasca 1607

Insights from Network Structure for Text Mining
Zornitsa Kozareva and Eduard Hovy 1616

Event Extraction as Dependency Parsing
David McClosky, Mihai Surdeanu and Christopher Manning 1626

Extracting Comparative Entities and Predicates from Texts Using Comparative Type Classification
Seon Yang and Youngjoong Ko 1636

Conference Program

Monday, June 20, 2011

(8:45-9:00) Opening Session

(9:00-10:00) Invited Talk 1:IBM Watson Deep QA System (tentative title) by David Ferrucci

(10:00-10:30) Coffee Break

Session 1-A: (10:30-12:10) MT: Methods

A Word-Class Approach to Labeling PSCFG Rules for Machine Translation
Andreas Zollmann and Stephan Vogel

Deciphering Foreign Language
Sujith Ravi and Kevin Knight

Effective Use of Function Words for Rule Generalization in Forest-Based Translation
Xianchao Wu, Takuya Matsuzaki and Jun'ichi Tsujii

Combining Morpheme-based Machine Translation with Post-processing Morpheme Prediction
Ann Clifton and Anoop Sarkar

Session 1-B: (10:30-12:10) Machine Learning Methods 1

Evaluating the Impact of Coder Errors on Active Learning
Ines Rehbein and Josef Ruppenhofer

A Fast and Accurate Method for Approximate String Search
Ziqi Wang, Gu Xu, Hang Li and Ming Zhang

Domain Adaptation by Constraining Inter-Domain Variability of Latent Feature Representation
Ivan Titov

Exact Decoding of Syntactic Translation Models through Lagrangian Relaxation
Alexander M. Rush and Michael Collins

Monday, June 20, 2011 (continued)

Session 1-C: (10:30-12:10) Information Retrieval

Jigs and Lures: Associating Web Queries with Structured Entities
Patrick Pantel and Ariel Fuxman

Semi-Supervised SimHash for Efficient Document Similarity Search
Qixia Jiang and Maosong Sun

Joint Annotation of Search Queries
Michael Bendersky, W. Bruce Croft and David A. Smith

Query Weighting for Ranking Model Adaptation
Peng Cai, Wei Gao, Aoying Zhou and Kam-Fai Wong

Session 1-D: (10:30-12:10) Sentiment Analysis/Opinion Mining 1

Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification
Yulan He, Chenghua Lin and Harith Alani

Using Multiple Sources to Construct a Sentiment Sensitive Thesaurus for Cross-Domain Sentiment Classification
Danushka Bollegala, David Weir and John Carroll

Learning Word Vectors for Sentiment Analysis
Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts

Target-dependent Twitter Sentiment Classification
Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu and Tiejun Zhao

Monday, June 20, 2011 (continued)

Session 1-E: (10:30-12:10) Language Resource

A Comprehensive Dictionary of Multiword Expressions
Kosho Shudo, Akira Kurahone and Toshifumi Tanabe

Multi-Modal Annotation of Quest Games in Second Life
Sharon Gower Small, Jennifer Strommer-Galley and Tomek Strzalkowski

A New Dataset and Method for Automatically Grading ESOL Texts
Helen Yannakoudakis, Ted Briscoe and Ben Medlock

Collecting Highly Parallel Data for Paraphrase Evaluation
David Chen and William Dolan

(12:10 - 2:00) Lunch

Session 2-A: (2:00-3:40) MT: Models & Evaluation

A Large Scale Distributed Syntactic, Semantic and Lexical Language Model for Machine Translation
Ming Tan, Wenli Zhou, Lei Zheng and Shaojun Wang

Goodness: A Method for Measuring Machine Translation Confidence
Nguyen Bach, Fei Huang and Yaser Al-Onaizan

MEANT: An inexpensive, high-accuracy, semi-automatic metric for evaluating translation utility based on semantic roles
Chi-kiu Lo and Dekai Wu

An exponential translation model for target language morphology
Michael Subotin

Monday, June 20, 2011 (continued)

Session 2-B: (2:00-3:40) Machine Learning Methods 2

Bayesian Inference for Zodiac and Other Homophonic Ciphers
Sujith Ravi and Kevin Knight

Interactive Topic Modeling
Yuening Hu, Jordan Boyd-Graber and Brianna Satinoff

Faster and Smaller N-Gram Language Models
Adam Pauls and Dan Klein

Learning to Win by Reading Manuals in a Monte-Carlo Framework
S.R.K Branavan, David Silver and Regina Barzilay

Session 2-C: (2:00-3:40) Linguistic Creativity

Creative Language Retrieval: A Robust Hybrid of Information Retrieval and Linguistic Creativity
Tony Veale

Local Histograms of Character N-grams for Authorship Attribution
Hugo Jair Escalante, Thamar Solorio and Manuel Montes-y-Gomez

Word Maturity: Computational Modeling of Word Knowledge
Kirill Kireyev and Thomas K Landauer

Finding Deceptive Opinion Spam by Any Stretch of the Imagination
Myle Ott, Yejin Choi, Claire Cardie and Jeffrey T. Hancock

Monday, June 20, 2011 (continued)

Session 2-D: (2:00-3:40) Sentiment Analysis/Opinion Mining 2

Joint Bilingual Sentiment Classification with Unlabeled Parallel Corpora
Bin Lu, Chenhao Tan, Claire Cardie and Benjamin K. Tsou

A Pilot Study of Opinion Summarization in Conversations
Dong Wang and Yang Liu

Contrasting Opposing Views of News Articles on Contentious Issues
Souneil Park, Kyung Soon Lee and Junehwa Song

Content Models with Attitude
Christina Sauper, Aria Haghighi and Regina Barzilay

Session 2-E: (2:00-3:40) NLP for Web 2.0

Recognizing Named Entities in Tweets
Xiaohua LIU, Shaodian ZHANG, Furu WEI and Ming ZHOU

Lexical Normalisation of Short Text Messages: Makn Sens a #twitter
Bo Han and Timothy Baldwin

Topical Keyphrase Extraction from Twitter
Xin Zhao, Jing Jiang, Jing He, Yang Song, Palakorn Achanauparp, Ee-Peng Lim and Xiaoming Li

Event Discovery in Social Media Feeds
Edward Benson, Aria Haghighi and Regina Barzilay

Monday, June 20, 2011 (continued)

(3:40-4:10) Coffee Break

Session 3-A: (4:10-5:50) Transliteration/Alignment

How do you pronounce your name? Improving G2P with transliterations
Aditya Bhargava and Grzegorz Kondrak

Unsupervised Word Alignment with Arbitrary Features
Chris Dyer, Jonathan H. Clark, Alon Lavie and Noah A. Smith

Model-Based Aligner Combination Using Dual Decomposition
John DeNero and Klaus Macherey

An Algorithm for Unsupervised Transliteration Mining with an Application to Word Alignment
Hassan Sajjad, Alexander Fraser and Helmut Schmid

Session 3-B: (4:10-5:50) Parsing 1

Beam-Width Prediction for Efficient Context-Free Parsing
Nathan Bodenstab, Aaron Dunlop, Keith Hall and Brian Roark

Optimal Head-Driven Parsing Complexity for Linear Context-Free Rewriting Systems
Pierluigi Crescenzi, Daniel Gildea, Andrea Marino, Gianluca Rossi and Giorgio Satta

Prefix Probability for Probabilistic Synchronous Context-Free Grammars
Mark-Jan Nederhof and Giorgio Satta

A Comparison of Loopy Belief Propagation and Dual Decomposition for Integrated CCG Supertagging and Parsing
Michael Auli and Adam Lopez

Monday, June 20, 2011 (continued)

Session 3-C: (4:10-5:50) Summarization

Jointly Learning to Extract and Compress
Taylor Berg-Kirkpatrick, Dan Gillick and Dan Klein

Discovery of Topically Coherent Sentences for Extractive Summarization
Asli Celikyilmaz and Dilek Hakkani-Tur

Coherent Citation-Based Summarization of Scientific Papers
Amjad Abu-Jbara and Dragomir Radev

A Class of Submodular Functions for Document Summarization
Hui Lin and Jeff Bilmes

Session 3-D: (4:10-5:50) Relation Extraction

Semi-supervised Relation Extraction with Large-scale Word Clustering
Ang Sun, Ralph Grishman and Satoshi Sekine

In-domain Relation Discovery with Meta-constraints via Posterior Regularization
Harr Chen, Edward Benson, Tahira Naseem and Regina Barzilay

Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations
Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer and Daniel S. Weld

Exploiting Syntactico-Semantic Structures for Relation Extraction
Yee Seng Chan and Dan Roth

Monday, June 20, 2011 (continued)

Session 3-E: (4:10-5:50) Semantics

Together We Can: Bilingual Bootstrapping for WSD
Mitesh M. Khapra, Salil Joshi, Arindam Chatterjee and Pushpak Bhattacharyya

Which Noun Phrases Denote Which Concepts?
Jayant Krishnamurthy and Tom Mitchell

Semantic Representation of Negation Using Focus Detection
Eduardo Blanco and Dan Moldovan

Learning Dependency-Based Compositional Semantics
Percy Liang, Michael Jordan and Dan Klein

(6:00-8:30) Poster Session (Long papers)

(6:00-8:30) Poster Session (Short papers)

Tuesday, June 21, 2011

Session 4-A: (9:00-10:30) Best Paper Session

Unsupervised Part-of-Speech Tagging with Bilingual Graph-Based Projections
Dipanjan Das and Slav Petrov

Global Learning of Typed Entailment Rules
Jonathan Berant, Ido Dagan and Jacob Goldberger

Tuesday, June 21, 2011 (continued)

(10:30-11:00) Coffee Break

(3:30-4:00) Coffee Break

Session 7-A: (4:00-5:40) SMT: Phrase-based Models

Incremental Syntactic Language Models for Phrase-based Translation
Lane Schwartz, Chris Callison-Burch, William Schuler and Stephen Wu

An Unsupervised Model for Joint Phrase Alignment and Extraction
Graham Neubig, Taro Watanabe, Eiichiro Sumita, Shinsuke Mori and Tatsuya Kawahara

Learning Hierarchical Translation Structure with Linguistic Annotations
Markos Mylonakis and Khalil Sima'an

Phrase-Based Translation Model for Question Retrieval in Community Question Answer Archives
Guangyou Zhou, Li Cai, Jun Zhao and Kang Liu

Session 7-B: (4:00-5:40) Parsing 2

Neutralizing Linguistically Problematic Annotations in Unsupervised Dependency Parsing Evaluation
Roy Schwartz, Omri Abend, Roi Reichart and Ari Rappoport

Dynamic Programming Algorithms for Transition-Based Dependency Parsers
Marco Kuhlmann, Carlos Gómez-Rodríguez and Giorgio Satta

Shift-Reduce CCG Parsing
Yue Zhang and Stephen Clark

Web-Scale Features for Full-Scale Parsing
Mohit Bansal and Dan Klein

Tuesday, June 21, 2011 (continued)

Session 7-C: (4:00-5:40) Spoken Language Processing

The impact of language models and loss functions on repair disfluency detection
Simon Zwarts and Mark Johnson

Learning Sub-Word Units for Open Vocabulary Speech Recognition
Carolina Parada, Mark Dredze, Abhinav Sethy and Ariya Rastrow

Computing and Evaluating Syntactic Complexity Features for Automated Scoring of Spontaneous Non-Native Speech
Miao Chen and Klaus Zechner

N-Best Rescoring Based on Pitch-accent Patterns
Je Hun Jeon, Wen Wang and Yang Liu

Session 7-D: (4:00-5:40) Natural Language Processing Applications

Lexically-Triggered Hidden Markov Models for Clinical Document Coding
Svetlana Kiritchenko and Colin Cherry

Learning to Grade Short Answer Questions using Semantic Similarity Measures and Dependency Graph Alignments
Michael Mohler, Razvan Bunescu and Rada Mihalcea

Age Prediction in Blogs: A Study of Style, Content, and Online Behavior in Pre- and Post-Social Media Generations
Sara Rosenthal and Kathleen McKeown

Extracting Social Power Relationships from Natural Language
Philip Bramsen, Martha Escobar-Molano, Ami Patel and Rafael Alonso

Tuesday, June 21, 2011 (continued)

Session 7-E: (4:00-5:40) Coreference Resolution

Bootstrapping coreference resolution using word associations
Hamidreza Kobdani, Hinrich Schuetze, Michael Schiehlen and Hans Kamp

Large-Scale Cross-Document Coreference Using Distributed Inference and Hierarchical Models
Sameer Singh, Amarnag Subramanya, Fernando Pereira and Andrew McCallum

A Cross-Lingual ILP Solution to Zero Anaphora Resolution
Ryu Iida and Massimo Poesio

Coreference Resolution with World Knowledge
Altaf Rahman and Vincent Ng

(7:00-11:00) Banquet

Wednesday, June 22, 2011

(9:00-10:00) Invited Talk 2: How do the languages we speak shape the ways we think? by Lera Boroditsky

(10:00-10:30) Coffee Break

Session 5-A: (10:30-12:10) SMT: Tree-based Models

How to train your multi bottom-up tree transducer
Andreas Maletti

Binarized Forest to String Translation
Hao Zhang, Licheng Fang, Peng Xu and Xiaoyun Wu

Learning to Transform and Select Elementary Trees for Improved Syntax-based Machine Translations
Bing Zhao, Young-Suk Lee, Xiaoqiang Luo and Liu Li

Rule Markov Models for Fast Tree-to-String Translation
Ashish Vaswani, Haitao Mi, Liang Huang and David Chiang

Wednesday, June 22, 2011 (continued)

Session 5-B: (10:30-12:10) Morphology/POS Induction

A Hierarchical Pitman-Yor Process HMM for Unsupervised Part of Speech Induction
Phil Blunsom and Trevor Cohn

Using Deep Morphology to Improve Automatic Error Detection in Arabic Handwriting Recognition
Nizar Habash and Ryan Roth

A Discriminative Model for Joint Morphological Disambiguation and Dependency Parsing
John Lee, Jason Naradowsky and David A. Smith

Unsupervised Bilingual Morpheme Segmentation and Alignment with Context-rich Hidden Semi-Markov Models
Jason Naradowsky and Kristina Toutanova

Session 5-C: (10:30-12:10) Error Correction

A Graph Approach to Spelling Correction in Domain-Centric Search
Zhuowei Bao, Benny Kimelfeld and Yunyao Li

Grammatical Error Correction with Alternating Structure Optimization
Daniel Dahlmeier and Hwee Tou Ng

Algorithm Selection and Model Adaptation for ESL Correction Tasks
Alla Rozovskaya and Dan Roth

Automated Whole Sentence Grammar Correction Using a Noisy Channel Model
Y. Albert Park and Roger Levy

Wednesday, June 22, 2011 (continued)

Session 5-D: (10:30-12:10) Information Extraction

A Generative Entity-Mention Model for Linking Entities with Knowledge Base
Xianpei Han and Le Sun

Simple supervised document geolocation with geodesic grids
Benjamin Wing and Jason Baldridge

Piggyback: Using Search Engines for Robust Cross-Domain Named Entity Recognition
Stefan Rüd, Massimiliano Ciaramita, Jens Müller and Hinrich Schütze

Template-Based Information Extraction without the Templates
Nathanael Chambers and Dan Jurafsky

Session 5-E: (10:30-12:10) Discourse

Classifying arguments by scheme
Vanessa Wei Feng and Graeme Hirst

Automatically Evaluating Text Coherence Using Discourse Relations
Ziheng Lin, Hwee Tou Ng and Min-Yen Kan

Underspecifying and Predicting Voice for Surface Realisation Ranking
Sina Zarrieß, Aoife Cahill and Jonas Kuhn

Recognizing Authority in Dialogue with an Integer Linear Programming Constrained Model
Elijah Mayfield and Carolyn Penstein Rosé

Wednesday, June 22, 2011 (continued)

(12:10 - 2:00) Lunch

(1:30-3:00) ACL Business Meeting

(3:00-3:30) Coffee Break

Session 6-A: (3:30-4:45) MT: Reordering Models

Reordering Metrics for MT
Alexandra Birch and Miles Osborne

Reordering with Source Language Collocations
Zhanyi Liu, Haifeng Wang, Hua Wu, Ting Liu and Sheng Li

A Joint Sequence Translation Model with Integrated Reordering
Nadir Durrani, Helmut Schmid and Alexander Fraser

Session 6-B: (3:30-4:45) Grammar

Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results
Roger Levy

Metagrammar engineering: Towards systematic exploration of implemented grammars
Antske Fokkens

Simple Unsupervised Grammar Induction from Raw Text with Cascaded Finite State Models
Elias Ponvert, Jason Baldridge and Katrin Erk

Wednesday, June 22, 2011 (continued)

Session 6-C: (3:30-4:45) Generation/Paraphrasing

Extracting Paraphrases from Definition Sentences on the Web
Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jun'ichi Kazama and Sadao Kurohashi

Learning From Collective Human Behavior to Introduce Diversity in Lexical Choice
Vahed Qazvinian and Dragomir R. Radev

Ordering Prenominal Modifiers with a Reranking Approach
Jenny Liu and Aria Haghighi

Session 6-D: (3:30-4:45) Event-Role Extraction

Unsupervised Semantic Role Induction via Split-Merge Clustering
Joel Lang and Mirella Lapata

Using Cross-Entity Inference to Improve Event Extraction
Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou and Qiaoming Zhu

Peeling Back the Layers: Detecting Event Role Fillers in Secondary Contexts
Ruihong Huang and Ellen Riloff

Session 6-E: (3:30-4:20) Knowledge Base Extension

Knowledge Base Population: Successful Approaches and Challenges
Heng Ji and Ralph Grishman

Nonlinear Evidence Fusion and Propagation for Hyponymy Relation Mining
Fan Zhang, Shuming Shi, Jing Liu, Shuqi Sun and Chin-Yew Lin

Wednesday, June 22, 2011 (continued)

(5:00-6:10) Life time achievement award and closing

Monday, June 20, 2011

(6:00-8:30) Poster Session (Long papers)

A Pronoun Anaphora Resolution System based on Factorial Hidden Markov Models
Dingcheng Li, Tim Miller and William Schuler

Disentangling Chat with Local Coherence Models
Micha Elsner and Eugene Charniak

An Affect-Enriched Dialogue Act Classification Model for Task-Oriented Dialogue
Kristy Boyer, Joseph Grafsgaard, Eun Young Ha, Robert Phillips and James Lester

Fine-Grained Class Label Markup of Search Queries
Joseph Reisinger and Marius Pasca

Creating a manually error-tagged and shallow-parsed learner corpus
Ryo Nagata, Edward Whittaker and Vera Sheinman

Crowdsourcing Translation: Professional Quality from Non-Professionals
Omar F. Zaidan and Chris Callison-Burch

A Statistical Tree Annotator and Its Applications
Xiaoqiang Luo and Bing Zhao

Consistent Translation using Discriminative Learning - A Translation Memory-inspired Approach
Yanjun Ma, Yifan He, Andy Way and Josef van Genabith

Machine Translation System Combination by Confusion Forest
Taro Watanabe and Eiichiro Sumita

Hypothesis Mixture Decoding for Statistical Machine Translation
Nan Duan, Mu Li and Ming Zhou

Minimum Bayes-risk System Combination
Jesús González-Rubio, Alfons Juan and Francisco Casacuberta

Adjoining Tree-to-String Translation
Yang Liu, Qun Liu and Yajuan Lü

Enhancing Language Models in Statistical Machine Translation with Backward N-grams and Mutual Information Triggers
Deyi Xiong, Min Zhang and Haizhou Li

Translating from Morphologically Complex Languages: A Paraphrase-Based Approach
Preslav Nakov and Hwee Tou Ng

Gappy Phrasal Alignment By Agreement
Mohit Bansal, Chris Quirk and Robert Moore

Translationese and Its Dialects
Moshe Koppel and Noam Ordan

Rare Word Translation Extraction from Aligned Comparable Documents
Emmanuel Prochasson and Pascale Fung

Using Bilingual Parallel Corpora for Cross-Lingual Textual Entailment
Yashar Mehdad, Matteo Negri and Marcello Federico

Using Large Monolingual and Bilingual Corpora to Improve Coordination Disambiguation
Shane Bergsma, David Yarowsky and Kenneth Church

Unsupervised Decomposition of a Document into Authorial Components
Moshe Koppel, Navot Akiva, Idan Dershowitz and Nachum Dershowitz

Discovering Sociolinguistic Associations with Structured Sparsity
Jacob Eisenstein, Noah A. Smith and Eric P. Xing

Local and Global Algorithms for Disambiguation to Wikipedia
Lev Ratinov, Dan Roth, Doug Downey and Mike Anderson

A Stacked Sub-Word Model for Joint Chinese Word Segmentation and Part-of-Speech Tagging
Weiwei Sun

Language-independent compound splitting with morphological operations
Klaus Macherey, Andrew Dai, David Talbot, Ashok Popat and Franz Och

Parsing the Internal Structure of Words: A New Paradigm for Chinese Word Segmentation
Zhongguo Li

A Simple Measure to Assess Non-response
Anselmo Peñas and Alvaro Rodrigo

Improving Question Recommendation by Exploiting Information Need
Shuguang Li and Suresh Manandhar

Semi-Supervised Frame-Semantic Parsing for Unknown Predicates
Dipanjan Das and Noah A. Smith

A Bayesian Model for Unsupervised Semantic Parsing
Ivan Titov and Alexandre Klementiev

Unsupervised Learning of Semantic Relation Composition
Eduardo Blanco and Dan Moldovan

Unsupervised Discovery of Domain-Specific Knowledge from Text
Dirk Hovy, Chunliang Zhang, Eduard Hovy and Anselmo Peñas

Latent Semantic Word Sense Induction and Disambiguation
Tim Van de Cruys and Marianna Apidianaki

Confidence Driven Unsupervised Semantic Parsing
Dan Goldwasser, Roi Reichart, James Clarke and Dan Roth

Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews
Jianxing Yu, Zheng-Jun Zha, Meng Wang and Tat-Seng Chua

Monday, June 20, 2011 (continued)

Collective Classification of Congressional Floor-Debate Transcripts
Clinton Burfoot, Steven Bird and Timothy Baldwin

Integrating history-length interpolation and classes in language modeling
Hinrich Schütze

Structural Topic Model for Latent Topical Structure Analysis
Hongning Wang, Duo Zhang and ChengXiang Zhai

Automatic Labelling of Topic Models
Jey Han Lau, Karl Grieser, David Newman and Timothy Baldwin

Using Bilingual Information for Cross-Language Document Summarization
Xiaojun Wan

Exploiting Web-Derived Selectional Preference to Improve Statistical Dependency Parsing
Guangyou Zhou, Jun Zhao, Kang Liu and Li Cai

Effective Measures of Domain Similarity for Parsing
Barbara Plank and Gertjan van Noord

Efficient CCG Parsing: A versus Adaptive Supertagging*
Michael Auli and Adam Lopez

Improving Arabic Dependency Parsing with Form-based and Functional Morphological Features
Yuval Marton, Nizar Habash and Owen Rambow

Partial Parsing from Bitext Projections
Prashanth Mannem and Aswarth Dara

Ranking Class Labels Using Query Sessions
Marius Pasca

Insights from Network Structure for Text Mining
Zornitsa Kozareva and Eduard Hovy

Event Extraction as Dependency Parsing
David McClosky, Mihai Surdeanu and Christopher Manning

Extracting Comparative Entities and Predicates from Texts Using Comparative Type Classification
Seon Yang and Youngjoong Ko

Invited Talk 1: Building Watson: An Overview of the DeepQA Project

David Ferrucci, Principal Investigator, IBM Research
Monday, June 20 9:00-10:00

Computer systems that can directly and accurately answer peoples' questions over a broad domain of human knowledge have been envisioned by scientists and writers since the advent of computers themselves. Open domain question answering holds tremendous promise for facilitating informed decision making over vast volumes of natural language content. Applications in business intelligence, healthcare, customer support, enterprise knowledge management, social computing, science and government could all benefit from computer systems capable of deeper language understanding. The DeepQA project is aimed at exploring how advancing and integrating Natural Language Processing (NLP), Information Retrieval (IR), Machine Learning (ML), Knowledge Representation and Reasoning (KR&R) and massively parallel computation can greatly advance the science and application of automatic Question Answering. An exciting proof-point in this challenge was developing a computer system that could successfully compete against top human players at the Jeopardy! quiz show (www.jeopardy.com).

Attaining champion-level performance at Jeopardy! requires a computer to rapidly and accurately answer rich open-domain questions, and to predict its own performance on any given question. The system must deliver high degrees of precision and confidence over a very broad range of knowledge and natural language content with a 3-second response time. To do this, the DeepQA team advanced a broad array of NLP techniques to find, generate, evidence and analyze many competing hypotheses over large volumes of natural language content to build Watson (www.ibmwatson.com). An important contributor to Watson's success is its ability to automatically learn and combine accurate confidences across a wide array of algorithms and over different dimensions of evidence. Watson produced accurate confidences to know when to "buzz in" against its competitors and how much to bet. High precision and accurate confidence computations are critical for real business settings where helping users focus on the right content sooner and with greater confidence can make all the difference. The need for speed and high precision demands a massively parallel computing platform capable of generating, evaluating and combing 1000's of hypotheses and their associated evidence. In this talk, I will introduce the audience to the Jeopardy! Challenge, explain how Watson was built on DeepQA to ultimately defeat the two most celebrated human Jeopardy Champions of all time and I will discuss applications of the Watson technology beyond in areas such as healthcare.

Dr. David Ferrucci is the lead researcher and Principal Investigator (PI) for the Watson/Jeopardy! project. He has been a Research Staff Member at IBM's T.J. Watson's Research Center since 1995 where he heads up the Semantic Analysis and Integration department. Dr. Ferrucci focuses on technologies for automatically discovering valuable knowledge in natural language content and using it to enable better decision making.

Invited Talk 2: How do the languages we speak shape the ways we think?

Lera Boroditsky, Assistant Professor, Stanford University
Wednesday, June 22 9:00-10:00

Do people who speak different languages think differently? Does learning new languages change the way you think? Do polyglots think differently when speaking different languages? Are some thoughts unthinkable without language? I will describe data from experiments conducted around the world that reveal the powerful and often surprising ways that the languages we speak shape the ways we think.

Lera Boroditsky is an assistant professor of psychology at Stanford University and Editor in Chief of Frontiers in Cultural Psychology. Boroditsky's research centers on how knowledge emerges out of the interactions of mind, world, and language, and the ways that languages and cultures shape human thinking. To this end, Boroditsky's laboratory has collected data around the world, from Indonesia to Chile to Turkey to Aboriginal Australia. Her research has been widely featured in the media and has won multiple awards, including the CAREER award from the National Science Foundation, the Searle Scholars award, and the McDonnell Scholars award.