

EMNLP 2014

**The 2014 Conference on Empirical Methods
In Natural Language Processing**

Proceedings of the Conference

October 25-29, 2014
Doha, Qatar

Sponsors

Diamond



Platinum



Gold



Silver



Bronze



Supporter

IBM Research

Sponsor of Student Volunteers



©2014 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 978-1-937284-96-1

Preface by the General Chair

Welcome to the 2014 Conference on Empirical Methods in Natural Language Processing.

The EMNLP conference series is annually organized by SIGDAT, the Association for Computational Linguistics' special interest group on linguistic data and corpus-based approaches to NLP. This year the conference is being held from October 25, 2014 (Sat.) to October 29, 2014 (Wed.) in Doha, Qatar.

In the past five years, the EMNLP conference attendance has been continuously growing, reaching just over 500 paying attendees in 2013, and it is nowadays considered as one of the leading conferences in Computational Linguistics and Natural Language Processing.

Given the growing trend, we believed it was the right time to lead EMNLP into an organization structure typical of large and important conferences. Therefore, we proposed several novelties: first of all, a large organization committee consisting of twenty (plus twenty-six area chairs) well-known members of the ACL community, who carried out several tasks required by the new achieved scale.

Secondly, as this is the first conference edition spanning five days, in addition to six workshops, we also selected and included for the first time an excellent selection of eight tutorials. We defined a registration policy that allows the participants to attend any of the tutorials and workshops (held on October 25th and 29th) by just paying a low flat rate on top of the registration fee for the main conference. We believe this can greatly increase the spread of advanced technology and promote a unified view of the techniques and foundations of our research field.

Thirdly, as a standalone conference, EMNLP required the definition of new administrative procedures and policies, regarding sponsorship booklets, double submission, scholarship assignment, and the joint EACL-ACL-EMNLP call for workshop proposals.

Next, EMNLP is finding new ways to foster the dissemination of research work by facing the increasing number of papers to be presented at the conference. Our new approach consisted in presenting posters in nine sessions each proposing a small numbers of papers: this way poster presentations can receive the space and consideration that they deserve. Then, we are adding a surprise in terms of paper presentation and dissemination, which will be unveiled only few days before the start of the conference.

Finally, this is the first time that an ACL conference is largely supported by a government research foundation. The Qatar National Research Foundation (QNRF) has included EMNLP 2014 as one of its local funding events. This enabled EMNLP and SIGDAT to perform unprecedented student scholarship support: more than 30 students were sponsored (partially or entirely) for participating in the conference. The obtained funds also allowed for offering a social dinner free of charge to all the attendees and still closing the conference budget in active, thus creating additional resources that SIGDAT can use to support the upcoming conferences.

The novelties above as well as the traditional activities that the EMNLP conference series proposes to its members could not have been organized without the work of our large committee. In this respect, I would like to thank our PC co-chairs Walter Daelemans and Bo Pang, who greatly used their large experience with program committees of our community for selecting an excellent program.

Special thanks go to our publication chair Yuval Marton, who did a terrific job in organizing and preparing the proceedings. As a side effect of his proactive action, workshop organizers and future publication chairs using the SoftConf START/ACL PUB systems can now streamline the inclusion of workshops and conference schedules in the proceedings, without heavy manual customization.

We are very grateful to Enrique Alfonseca and Eric Gaussier for selecting interesting and successful

workshops and to Lucia Specia and Xavier Carreras, who, for the first time, carried out the new task of selecting tutorials for an EMNLP conference. The workshops and tutorials nicely filled the additional two days of EMNLP, making our conference even more valuable.

Many thanks are due to Katrin Erk and Sebastian Padó, who were challenged by the new activity (for EMNLP) of defining policy for the selection and assignment of participation scholarships to the most deserving students. The uncertainty over the final amount of funds and their diverse nature made this task particularly difficult. Nevertheless, they were able to find appropriate and successful solutions.

As any large conference, we could count on the help of publicity co-chairs to advertise the old and new EMNLP features. We give our gratitude to Mona Diab and Irina Matveeva for their professional work.

Fund hunting is a very important activity for conferences, in this respect, I would like to thank our sponsorship co-chairs, Jochen Leidner, Veselin Stoyanov and Min Zhang, for helping us to look for sponsors in three different continents.

Regarding the SIGDAT side, a special thank is devoted to Noah Smith, who promptly answered any question I came out with. I am also grateful to the other SIGDAT officers (past and new): Eugene Charniak, Mark Johnson, Philipp Koehn, Mark Steedman, who were always there to give suggestions and solutions to critical issues that inevitably arise in any large event.

Many thanks also to Tim Baldwin, Anna Korhonen, Graeme Hirst and David Yarowsky who provided much useful information from past conferences. Last but not least, I would like to thank Priscilla Rasmussen for her help and advice, and her undoubtful qualities of soothsayer regarding the estimation of conference numbers.

Coming back to the sponsor topic, we are enormously thankful to QNRF, for accepting our proposal to fund EMNLP: this has made it possible to sponsor an unprecedented number of students and offer a banquet free of charge to all participants (we needed to create a new level of sponsorship for them, namely, Diamond). We are very grateful to The Qatar Computing Research Institute, which in addition to providing the very valuable Platinum sponsorship, also provided the required man power for organizing the event.

In particular, EMNLP could not be organized in Qatar without the work of Kareem Darwish, the local organization chair. We are also very grateful to Kemal Oflazer, local co-chair and Francisco Guzman Herrera, local sponsorship chair, whose work was determinant to obtain the QNRF sponsorship. We are deeply in debt with the other local organizers, Lluís Màrquez, who also edited the conference booklet, Preslav Nakov, Fabrizio Sebastiani and Stephan Vogel for their help with the daily big and little issues.

Special thanks go to The Carnegie Mellon University in Qatar for helping us with the proposal preparation and management of the QNRF funds and also for supporting us with a Gold sponsorship. Additionally, many thanks go to our silver sponsors, Facebook and Yandex and our bronze sponsor iHorizons, who show the increasing interest of industry in the technology of our community for the design of real-world and high-societal impact applications. In this respect, we sincerely thank Google Inc. and IBM Watson, New York, for supporting the student participation with their scholarships.

Finally, and foremost, thanks to all the authors and conference attendees who are the main actors of this event, bringing the real value to it and determining its success. My personal thanks also go to the entire SIGDAT committee, for choosing me as the chair of this fantastic conference, held in a fascinating venue.

Alessandro Moschitti

General Chair of EMNLP 2014

Preface by the Program Committee Co-Chairs

We welcome you to the 2014 Conference on Empirical Methods in Natural Language Processing.

As in the previous EMNLP, we invited both long and short papers with a single submission deadline. Short papers encourage the submission of smaller and more preliminary contributions.

We received 790 submissions (after initial withdrawals of unfinished submissions and removal of duplicates), of which 28 were rejected before review for not adhering to the instructions in the call for papers regarding paper length or anonymity. The remaining 510 long and 252 short papers were allocated to one of the fourteen areas. The most popular areas this year were Machine Translation, Semantics, and Syntax (Tagging, Chunking, and Parsing).

Reviewing for a conference this size involves an army of dedicated professionals volunteering to donate their valuable and scarce time to make sure that the highest possible reviewing standards are reached. We are very grateful to our 26 area chairs and a programme committee of more than 500 for their efforts. We accepted 155 long and 70 short papers, representing a global acceptance rate of just under 30%. Nine papers accepted by the ACL journal TACL were added to the program.

Based on the reviews and on nominations by the area chairs, 5 long papers were shortlisted for the best paper award. The best paper will be presented in a plenary best paper award ceremony. We would like to thank Mark Johnson and Claire Cardie for their willingness to serve in the best paper award committee that was set up and for providing excellent advice and motivation for their choice.

We are grateful to the authors for selecting EMNLP as the venue for their work. Congratulations to the authors of accepted submissions. To the authors of rejected submissions we would like to offer as consolation the fact that because of the competitive nature of the conference and the inevitable time and space limitations, many worthwhile papers could not be included in the program. We hope the feedback of the reviewers will be considered worthwhile by them and lead to successful future submissions.

We are very grateful to our invited speakers Thorsten Joachims and Salim Roukos. Thorsten Joachims is professor at the Computer Science and Information Science departments at Cornell University and shows how integrating microeconomic models of human behavior into the learning process leads to new interaction models and learning algorithms, in turn leading to better performing systems. Salim Roukos is senior manager of multilingual NLP and CTO of Translation Technologies at IBM T.J. Watson research Center and addresses IBM's approach to cognitive computing for building systems and solutions that enable and support richer human-machine interactions, and remaining opportunities in this area for novel statistical models for natural language processing. We thank them for their inspiring talks and presence at the conference.

We would also like to thank our general chair Alessandro Moschitti for his leadership, advice, encouragement, and support, Kareem Darwish and his colleagues for impeccable cooperation from local organization, and Yuval Marton for doing an excellent job assembling these proceedings.

It was an honour to serve as Programme Chairs of EMNLP 2014, and we hope that you will enjoy the conference and be able to think back later and remember a scientifically stimulating conference and a pleasant time in Doha, Qatar.

Bo Pang and Walter Daelemans

EMNLP 2014 Program Chairs

Organizers:

General Conference Chair

Alessandro Moschitti, Qatar Computing Research Institute

Program Committee Co-Chairs

Walter Daelemans, University of Antwerp

Bo Pang, Google

Workshops Co-Chairs

Enrique Alfonseca, Google Research at Zurich

Eric Gaussier, Université Joseph Fourier (Grenoble I)

Tutorial Co-Chairs

Lucia Specia, University of Sheffield

Xavier Carreras, Universitat Politècnica de Catalunya

Publication Chair

Yuval Marton, Microsoft Corporation

Publicity Co-Chairs

Mona Diab, George Washington University

Irina Matveeva, NexLP

Sponsorship Co-Chairs

Jochen Leidner, Thomson Reuters

Veselin Stoyanov, Facebook

Min Zhang, Soochow University

Student Scholarship Co-Chairs

Katrin Erk, University of Texas at Austin

Sebastian Padó, University of Stuttgart

Reviewing Coordinators

Mark Dredze, Johns Hopkins University

Jiang Guo (Student Volunteer), Harbin Institute of Technology

Area Chairs

Phonology, Morphology, and Segmentation

Tomaž Erjavec, Jožef Stefan Institute

Tagging, Chunking, Syntax and Parsing

Gosse Bouma, University of Groningen

Yuji Matsumoto, Nara Institute of Science and Technology

Discourse, Dialogue, and Pragmatics

Jennifer Chu-Carroll, IBM Watson Research Center

Olga Uryupina, University of Trento

Semantics

Rada Mihalcea, University of Michigan

Sameer Pradhan, Harvard Medical School

Summarization and Generation

Anja Belz, University of Brighton

Dilek Hakkani-Tür, Microsoft Research

NLP-related Machine Learning: theory, methods and algorithms

Ivan Titov, University of Amsterdam

Jerry Zhu, University of Wisconsin-Madison

Machine Translation

Chris Callison-Burch, University of Pennsylvania

Dan Gildea, University of Rochester

Information Retrieval, Text Categorization, and Question Answering

Sien Moens, Katholieke Universiteit Leuven

Hinrich Schütze, Ludwig Maximilian University of Munich

Information Extraction

Doug Downey, Northwestern University

Marius Pasca, Google

Text Mining and Natural Language Processing Applications

Massimiliano Ciaramita, Google

Hwee Tou Ng, National University of Singapore

Sentiment Analysis and Opinion Mining

Yejin Choi, Stony Brook University

Minlie Huang, Tsinghua University

NLP for the Web and Social Media

Irwin King, The Chinese University of Hong Kong

Qiaozhu Mei, University of Michigan

Spoken Language Processing

Pascale Fung, Hong Kong University of Science and Technology

Hugo Van hamme, Katholieke Universiteit Leuven

Computational Psycholinguistics

Sharon Goldwater, University of Edinburgh

Local Organization

Local Arrangements Co-Chairs

Kareem Darwish, Qatar Computing Research Institute

Kemal Oflazer, Carnegie Mellon University – Qatar

Local Sponsorship Chair

Francisco Guzmán, Qatar Computing Research Institute

Conference Handbook Editor

Lluís Màrquez, Qatar Computing Research Institute

Local Organizing Committee

Preslav Nakov, Qatar Computing Research Institute

Fabrizio Sebastiani, Qatar Computing Research Institute

Local QCRI Administration

Kimberly Mathern, Qatar Computing Research Institute

Lawrence Tingson, Qatar Computing Research Institute

Jacqueline Caparas, Qatar Computing Research Institute

Program Committee:

Omri Abend, The University of Edinburgh; Amjad Abu-Jbara, University of Michigan; Eneko Agirre, University of the Basque Country; Cem Akkaya, University of Pittsburgh; Iñaki Alegria, University of the Basque Country (UPV/EHU); Nikolaos Aletras, University of Sheffield; Enrique

Alfonseca, Google; James Allan, University of Massachusetts Amherst; Alexander Allauzen, Université Paris-Sud / LIMSI-CNRS; Waleed Ammar, CMU; David Andrzejewski, Sumo Logic; Gabor Angeli, Stanford University; Stephen Anthony, The University of New South Wales; Jordi Atserias Batalla, Yahoo! Research; Giuseppe Attardi, Università di Pisa; Michael Auli, Microsoft Research; Wilker Aziz, University of Sheffield;

Alexandra Balahur, Joint Research Centre, European Commission; Timothy Baldwin, The University of Melbourne; Borja Balle, UPC; David Bamman, Carnegie Mellon University; Carmen Banea, University of Michigan; Roberto Basili, University of Roma, Tor Vergata; Daniel Beck, University of Sheffield; Kedar Bellare, Facebook; Emily M. Bender, University of Washington; Michael Bendersky, Google; Fabrício Benevenuto, Federal University of Minas Gerais (UFMG); Jonathan Berant, Stanford University; Taylor Berg-Kirkpatrick, UC Berkeley; Nicola Bertoldi, FBK; Steven Bethard, University of Alabama at Birmingham; Chandra Bhagavatula, Northwestern University; Pushpak Bhattacharyya, CSE Department, IIT Bombay; Klinton Bicknell, Northwestern University; Chris Biemann, TU Darmstadt; Alexandra Birch, University of Edinburgh; Arianna Bisazza, University of Amsterdam; Yonatan Bisk, University of Illinois at Urbana-Champaign; Anders Björkelund, IMS, Stuttgart; Graeme Blackwood, IBM Research; Eduardo Blanco, Lymba Corporation; Roi Blanco, Yahoo! Labs; Phil Blunsom, University of Oxford; Bernd Bohnet, University Birmingham; Kalina Bontcheva, University of Sheffield; Houda Bouamor, Carnegie Mellon University; Jordan Boyd-Graber, University of Maryland; Pavel Braslavski, Kontur labs / Ural federal university; Chris Brew, Nuance Communications; Chris Brockett, Microsoft Research; Eric Brown, IBM Research; Paul Buitelaar, INSIGHT, National University of Ireland, Galway; Donna Byron, IBM Watson Solutions;

Aoife Cahill, Educational Testing Service; Erik Cambria, Nanyang Technological University; Marie Candito, Univ Paris Diderot - INRIA - Alpage; Hailong Cao, HIT; Sandra Carberry, University of Delaware; Marine Carpuat, National Research Council; Xavier Carreras, Universitat Politècnica de Catalunya; Francisco Casacuberta, Universitat Politècnica de València; Taylor Cassidy, IBM TJ Watson Research Center; Carlos Castillo, Qatar Computing Research Institute; Asli Celikyilmaz, Microsoft; Daniel Cer, Google; Özlem Çetinoğlu, IMS, University of Stuttgart; Jonathan Chang, Facebook; Berlin Chen, National Taiwan Normal University; Boxing Chen, NRC; Chen Chen, University of Texas at Dallas; Hsin-Hsi Chen, National Taiwan University; John Chen, AT&T Labs-Research; Lin Chen, University of Illinois at Chicago; Colin Cherry, NRC; Jackie Chi Kit Cheung, University of Toronto; David Chiang, USC/ISI; Jinho D. Choi, Emory University; Christos Christodoulopoulos, University of Illinois at Urbana Champaign; Grzegorz Chrupała, Tilburg University; Stephen Clark, University of Cambridge; Shay B. Cohen, University of Edinburgh; Trevor Cohn, University of Melbourne; Gao Cong, Nanyang Technological University; Paul Cook, The University of Melbourne; Bonaventura Coppola, IBM Research;

Bhavana Dalvi, Carnegie Mellon University; bharath dandala, University of North Texas; Dipanjan Das, Google Inc.; Martine De Cock, Ghent University; Adrià de Gispert, SDL Research; Steve DeNeefe, SDL Language Weaver; Pascal Denis, INRIA; Michael Denkowski, Carnegie Mellon University; Leon Derczynski, University of Sheffield; Ann Devitt, Trinity College Dublin; Nicholas Diakopoulos, Columbia University; Markus Dickinson, Indiana University; Michelangelo Diligenti, University of Siena; Georgiana Dinu, University of Trento; Doug Downey, Northwestern University; Eduard Dragut, Temple University; Mark Dredze, Johns Hopkins University; Markus Dreyer, SDL Language Weaver; Gregory Druck, Yummlly; Lan Du, Macquarie University; Kevin Duh, Nara Institute of Science and Technology; Greg Durrett, UC Berkeley; Chris Dyer, Carnegie Mellon University; Marc Dymetman, Xerox Research Centre Europe;

Koji Eguchi, Kobe University; Patrick Ehlen, AT&T; Andreas Eisele, DGT, European Commission; Jacob Eisenstein, Georgia Institute of Technology; Jason Eisner, Johns Hopkins University; Micha Elsner, The Ohio State University; Tomaz Erjavec, Dept. of Knowledge Technologies, Jožef Stefan Institute;

Angela Fahrni, HITS gGmbH; Hui Fang, University of Delaware; Lei Fang, Tsinghua University; Benoit Favre, Aix-Marseille University LIF/CNRS; Anna Feldman, Montclair State University; Naomi Feldman, University of Maryland; Minwei Feng, RWTH Aachen University; Song Feng, Stony Brook University; Yang Feng, USC/ISI; Eraldo Fernandes, UFMS; Katja Filippova, Google; Andrew Finch, NICT; Margaret Fleck, Univ. Illinois, Urbana-Champaign; George Foster, NRC; Jennifer Foster, Dublin City University; Anette Frank, Heidelberg University; Stefan L. Frank, Radboud University Nijmegen;

Michel Galley, Microsoft Research; Michael Gamon, Microsoft Research; Kuzman Ganchev, Google; Kavita Ganesan, Univ. Illinois, Urbana-Champaign; Wei Gao, Qatar Computing Research Institute; Claire Gardent, CNRS/LORIA, Nancy; Matt Gardner, Carnegie Mellon University; Dan Garrette, University of Texas at Austin; Guillermo Garrido, NLP & IR Group at UNED; Albert Gatt, University of Malta; Kallirroi Georgila, University of Southern California Institute for Creative Technologies; Shima Gerani, University of British Columbia; Ulrich Germann, University of Edinburgh; Andrea Gesmundo, Google Inc.; Daniel Gillick, Google; Kevin Gimpel, Toyota Technological Institute at Chicago; Filip Ginter, University of Turku; Roxana Girju, University of Illinois, Urbana-Champaign; Dan Goldwasser, University of Maryland; Sharon Goldwater, University of Edinburgh; Juan Carlos Gomez, KU Leuven; Carlos Gómez-Rodríguez, Universidade da Coruña; Matthew R. Gormley, Johns Hopkins University; Joao Graca, Inesc-Id; Spence Green, Stanford University; Edward Grefenstette, University of Oxford; Gregory Grefenstette, INRIA; Weiwei Guo, Columbia University;

Keith Hall, Google Research; Jirka Hana, Charles University; Greg Hanneman, Carnegie Mellon University; Sanda Harabagiu, University of Texas at Dallas; Christian Hardmeier, Uppsala universitet; Kazi Saidul Hasan, University of Texas at Dallas; Katsuhiko Hayashi, NTT CS Lab; Yifan He, New York University; Jeffrey Heinz, University of Delaware; James Henderson, Xerox Research Centre Europe; John Henderson, MITRE; Andreas Henrich, University of Bamberg; Aurélie Herbelot, Universität Potsdam; Tsutomu Hirao, NTT Communication Science Labs.; Graeme Hirst, University of Toronto; Hieu Hoang, University of Edinburgh; Julia Hockenmaier, University of Illinois Urbana-Champaign; Johannes Hoffart, Max-Planck-Institute for Informatics; Ales Horak, Masaryk University; Estevam Hruschka, Federal University of São Carlos; Fei Huang, Temple University; Liang Huang, City University of New York (CUNY); Xuanjing Huang, Fudan University; Mans Hulden, University of Helsinki;

Gonzalo Iglesias, SDL; Diana Inkpen, University of Ottawa; Kai Ishikawa, NEC Corporation; Abe Ittycheriah, IBM;

Heng Ji, Rensselaer Polytechnic Institute; Sittichai Jiampojarn, Google Inc.; Jing Jiang, Singapore Management University; Yunliang Jiang, Twitter Inc.; Charles Jochim, IBM Research – Dublin; Richard Johansson, University of Gothenburg; Mark Johnson, Macquarie University;

Mijail Kabadjov, DaXtra Technologies Ltd.; Nobuhiro Kaji, University of Tokyo; Min-Yen Kan, National University of Singapore; Pallika Kanani, Oracle Labs; Rohit Kate, University of Wisconsin-Milwaukee; Andre Kempe, Nuance Communications; Jin-Dong Kim, Database Center for Life Science; Kevin Knight, USC/ISI; Philipp Koehn, University of Edinburgh; Oleksandr Kolomiyets, KU Leuven; Mamoru Komachi, Tokyo Metropolitan University; Fang Kong, Soochow University, National University of Singapore; Moshe Koppel, Bar-Ilan University; Anna Korhonen, University of Cambridge; Zornitsa Kozareva, Yahoo!; Mikhail Kozhevnikov, MMCI Cluster of Excellence, University of Saarland; Jayant Krishnamurthy, Carnegie Mellon University; Lun-Wei Ku, Academia Sinica; Sandra Kübler, Indiana University; Marco Kuhlmann, Linköping University; Roland Kuhn, National Research Council of Canada; Shankar Kumar, Google;

Shibamouli Lahiri, University of North Texas; Mathias Lambert, Amazon.com; Man Lan, East China Normal University; Hugo Larochelle, Université de Sherbrooke; Jey Han Lau, King's College London; Florian Laws, University of Stuttgart; Joseph Le Roux, Université Paris Nord; Gi-

anluca Lebani, University of Pisa; Jung-Tae Lee, Naver Corp.; Yoong Keok Lee, MIT; Jochen Leidner, Thomson Reuters; Alessandro Lenci, University of Pisa; Chee Wee Leong, Educational Testing Service; Omer Levy, Bar-Ilan University; Fangtao Li, Google Research; Hang Li, Huawei Technologies; Lishuang Li, Dalian University of Technology; Peng Li, State Key Laboratory of Intelligent Technology and Systems; Tsinghua National Laboratory for Information Science and Technology; Department of Computer Sci. and Tech., Tsinghua University, Beijing, China; Qi Li, Computer Science, Rensselaer Polytechnic Institute; Shoushan Li, Soochow University; Wenjie Li, The Hong Kong Polytechnic University; Xiao Ling, University of Washington; Christina Lioma, University of Copenhagen; Marina Litvak, Shamoon College of Engineering; Kang Liu, Chinese Academy of Sciences; Qun Liu, Dublin City University; Yang Liu, University of Texas at Dallas; Nikola Ljubešić, University of Zagreb; Oier Lopez de Lacalle, IKERBASQUE, Basque Foundation for Science - University of Edinburgh; Bin Lu, City University of Hong Kong; Wei Lu, Singapore University of Technology and Design; Xiaofei Lu, Pennsylvania State University; Marco Lui, University of Melbourne / NICTA VRL; Susann Luperfoy, MIT;

Wolfgang Macherey, Google; Suresh Manandhar, University of York; Gideon Mann, Google Inc; Daniel Marcu, SDL; André F. T. Martins, Priberam, Instituto de Telecomunicacoes; Cettolo Mauro, FBK; Arne Mauser, Google, Inc; Jonathan May, USC Information Sciences Institute; Elijah Mayfield, Carnegie Mellon University; Diana Maynard, University of Sheffield; Julian McAuley, UCSD; David McClosky, IBM Research; Ryan McDonald, Google; Louise McNally, Universitat Pompeu Fabra; Edgar Meij, Yahoo! Research; Adam Meyers, New York University; Haitao Mi, IBM Watson Research Center; Lukas Michelbacher, Center for Information and Language Processing, LMU, Munich; David Mimno, Cornell University; Bonan Min, Raytheon BBN Technologies; Shachar Mirkin, Xerox Research Centre Europe; Margaret Mitchell, Microsoft Research; Samaneh Moghaddam, Simon Fraser University; Saif Mohammad, National Research Council Canada; Behrang Mohit, Carnegie Mellon University; Mitra Mohtarami, National University of Singapore; Karo Moilanen, Dpt of Computer Science, University of Oxford; Christian Monson, Nuance Communications Inc.; Roser Morante, University of Antwerp; Andrea Moro, Sapienza University of Rome; Dana Movshovitz-Attias, Carnegie Mellon University; Thomas Mueller, CIS, University of Munich; Arjun Mukherjee, University of Illinois at Chicago; Dragos Munteanu, SDL Language Technologies; Smaranda Muresan, Columbia University; Gabriel Murray, University of the Fraser Valley; sung-hyon myaeng, Korea Advanced Institute of Science and Technology; Markos Mylonakis, Lexis Research;

Tetsuji Nakagawa, Google Japan Inc.; Preslav Nakov, Qatar Computing Research Institute; Jason Naradowsky, UMass Amherst; Tahira Naseem, MIT-CSAIL; Vivi Nastase, FBK; Roberto Navigli, Sapienza University of Rome; Mark-Jan Nederhof, University of St Andrews; Ani Nenkova, University of Pennsylvania; Graham Neubig, Nara Institute of Science and Technology; Guenter Neumann, DFKI; Vincent Ng, University of Texas at Dallas; Jian-Yun NIE, University of Montreal; Rodney Nielsen, University of North Texas ; University of Colorado; Malvina Nissim, University of Bologna;

Diarmuid Ó Séaghdha, University of Cambridge; Stephan Oepen, Universitetet i Oslo; Kemal Oflazer, Carnegie Mellon University - Qatar; Alice Oh, KAIST; Jong-Hoon Oh, NICT; Kouzou Ohara, Aoyama Gakuin University; Naoaki Okazaki, Tohoku University; Constantin Orasan, University of Wolverhampton;

Martha Palmer, University of Colorado; Sinno J. Pan, Institute for Infocomm Research; Kristen Parton, Facebook; Rebecca J. Passonneau, Columbia University; John K Pate, Macquarie University; Siddharth Patwardhan, IBM T. J. Watson Research Center; Mykola Pechenizkiy, TU Eindhoven; Slav Petrov, Google; Steven Piantadosi, University of Rochester, BCS; Barbara Plank, University of Copenhagen; Simone Paolo Ponzetto, University of Mannheim; Hoifung Poon, Microsoft Research; Maja Popović, DFKI; Matt Post, Johns Hopkins University; Sameer Pradhan, Harvard University; John Prager, IBM Research; Rashmi Prasad, University of Wisconsin-

Milwaukee; Adam Przepiórkowski, Institute of Computer Science at the Polish Academy of Sciences; Stephen Pulman, Oxford University; Sampo Pyysalo, NaCTeM;

Vahed Qazvinian, Google;

Altaf Rahman, Yahoo Labs; Maya Ramanath, IIT-Delhi; Owen Rambow, Columbia University; Sujith Ravi, Google Inc.; Marta Recasens, Google Inc.; Ines Rehbein, Potsdam University; Roi Reichart, Technion - Israel Institute of Technology; Sebastian Riedel, UCL; Korbinian Riedhammer, Int'l Computer Science Institute; Verena Rieser, Heriot-Watt University; Stefan Riezler, Heidelberg University; German Rigau, UPV/EHU; Laura Rimell, University of Cambridge; Alan Ritter, Carnegie Mellon University; Andrew Rosenberg, CUNY Queens College; Michael Roth, The University of Edinburgh; Alla Rozovskaya, Columbia University; Josef Ruppenhofer, Hildesheim University; Vasile Rus, The University of Memphis; Alexander M. Rush, MIT; Delia Rusu, Jozef Stefan Institute;

Markus Saers, Hong Kong University of Science and Technology; Mehdi Samadi, Carnegie Mellon University; Federico Sangati, FBK, Trento; Sunita Sarawagi, IIT Bombay; Anoop Sarkar, Simon Fraser University; David Schlangen, Bielefeld University; Helmut Schmid, CIS, Ludwig-Maximilians-Universität; Tobias Schnabel, Cornell University; Sabine Schulte im Walde, University of Stuttgart; H. Andrew Schwartz, University of Pennsylvania; Holger Schwenk, University of Le Mans; Pavel Serdyukov, Yandex; Hendra Setiawan, IBM T.J. Watson Research Center; Burr Settles, Duolingo, Inc.; Kashif Shah, University of Sheffield; Azadeh Shakeri, University of Tehran; Shuming Shi, Microsoft Research Asia; Avirup Sil, IBM Research; Carina Silberer, School of Informatics, University of Edinburgh; Mario J. Silva, IST/INESC-ID; Khalil Sima'an, ILLC, University of Amsterdam; Sameer Singh, University of Washington, Seattle; Kevin Small, Amazon; Otakar Smrž, Seznam.cz; Jan Šnajder, University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb; Richard Socher, Stanford University; Stephen Soderland, University of Washington; Anders Søgaard, University of Copenhagen; Thamar Solorio, UAB; Swapna Somasundaran, Educational Testing Services; Linfeng Song, ICT/CAS; Yang Song, Microsoft Research Redmond; Lucia Specia, University of Sheffield; Valentin Spilovskiy, Google Inc.; Vivek Srikumar, Stanford University; Mark Steedman, University of Edinburgh; Benno Stein, Bauhaus Universität Weimar; Michael Strube, Heidelberg Institute for Theoretical Studies; David Suendermann, DHBW Stuttgart; Hisami Suzuki, Microsoft Corporation; Jun Suzuki, NTT CS Lab.; Ben Swanson, Brown University;

Hiroya Takamura, Tokyo Institute of Technology; David Talbot, Google Russia; Partha P. Talukdar, CMU; Chenhao Tan, Cornell University; Joel Tetreault, Nuance Communications; Cindi Thompson, University of San Francisco; Jörg Tiedemann, Uppsala University; Christoph Tillmann, TJ Watson IBM Research; Takenobu Tokunaga, Tokyo Institute of Technology; Kristina Toutanova, Microsoft Research; Ming-Feng Tsai, National Chengchi University; Yoshimasa Tsuruoka, University of Tokyo;

Jakob Uszkoreit, Google, Inc.; Masao Utiyama, NICT;

Andreas van Cranenburgh, University of Amsterdam; Antal van den Bosch, Radboud University Nijmegen; Benjamin Van Durme, JHU; Gertjan van Noord, University of Groningen; Menno van Zaanen, Tilburg University; Yannick Versley, University of Heidelberg; David Vilar, Pixformance GmbH; Aline Villavicencio, Institute of Informatics, Federal University of Rio Grande do Sul; Sami Virpioja, Aalto University; Andreas Vlachos, University College London; Adam Vogel, Stanford University; Vinod Vydiswaran, University of Illinois;

Stephen Wan, CSIRO; Xiaojun Wan, Peking University; Haifeng Wang, Baidu; Lu Wang, Cornell University; Leo Wanner, ICREA and Pompeu Fabra University; Nigel Ward, University of Texas at El Paso; Taro Watanabe, NICT; Bonnie Webber, University of Edinburgh; Furu Wei, Microsoft Research Asia; Albert Weichselbraun, University of Applied Sciences Chur; Gerhard Weikum, Max-Planck Institute for Informatics (MPII); Michael Wiegand, Saarland University; Jason D

Williams, Microsoft Research; Shuly Wintner, University of Haifa; Kam-Fai Wong, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong; Fei Wu, google; Joern Wuebker, RWTH Aachen University;

Yunqing Xia, Tsinghua University; Peng Xu, Google Inc.; Wei Xu, University of Pennsylvania; Nianwen Xue, Brandeis University;

Bishan Yang, Cornell University; Yi Yang, Northwestern University; Limin Yao, University of Massachusetts, Amherst; Wen-tau Yih, Microsoft Research; Hong Yu, Soochow University; Jianxing Yu, Tencent Inc. Shenzhen, China; François Yvon, LIMSI/CNRS;

Fabio Massimo Zanzotto, University of Rome "Tor Vergata"; Richard Zens, Google; Luke Zettlemoyer, University of Washington; Congle Zhang, University of Washington; Hao Zhang, Google; Joy Ying Zhang, Carnegie Mellon University; Lei Zhang, University of Illinois at Chicago; Min Zhang, Soochow University; Qi Zhang, Fudan University; Yi Zhang, German Research Center for Artificial Intelligence; Yue Zhang, Singapore University of Technology and Design; Jun Zhao, Chinese Academy of Sciences; Kai Zhao, Graduate Center, CUNY; Guodong Zhou, Soochow University; Tom Chao Zhou, Baidu Inc.; Chengqing Zong, Institute of Automation, Chinese Academy of Sciences; and Willem Zuidema, University of Amsterdam.

Additional Reviewers:

Aditya Joshi, Alexander Beloborodov, Ali Elkahky, Alvin Grissom II, Antoine Rozenknop, Burcu Can, Caitlin Richter, Chen Li, Christian Gaida, Dave Carter, Dongwoo Kim, Fandong Meng, Feifei Zhai, Frank Ferraro, Georgeta Bordea, Hadi Amiri, Hector Martinez Alonso, Heng Yu, Henning Wachsmuth, Hugo Escalante, Jacob Andreas, Jan Rygl, Janis Dalins, Jeff Mitchell, Jeffrey Flanigan, Jenna Kanerva, Ji Liu, Ji Ma, JinYeong Bak, Jiri Materna, Jiwei Li, John Wieting, Juhani Luotolahti, Jun Xie, Kai Hakala, Kai Liu, Kepa Sarasola, Kristian Woodsend, Li Dong, Li Wang, Likun Qiu, Loic Barrault, Mark Cusick, Martin Riedl, Michael Voelske, Miikka Silfverberg, Mikael Kågebäck, Milos Jakubicek, Mitchell Koch, Mohamed Yahya, Mohit Iyyer, Muhammad Ibrahim, Muhua Zhu, Nadi Tomeh, Ni Lao, Patrick Lange, Pushpendre Rastogi, Qing Dou, Roy Schwartz, Scott Martin, Shumin Wu, Colorado, Steffen Remus, Suwisa Kaewphan, Tim Rocktaeschel, Tongfei Chen, Travis Wolfe, Veronica Perez-Rosas, Vojtech Kovar, Wei He, Wenbin Jiang, Xutao Li, Zhanyi Liu, Zhaohui Wu, and Zhen Hai.

Invited Speakers:

Salim Roukos, IBM T. J. Watson Research Center

Thorsten Joachims, Cornell University

Table of Contents

<i>Invited Talk: IBM Cognitive Computing - An NLP Renaissance!</i> Salim Roukos	1
<i>Modeling Interestingness with Deep Neural Networks</i> Jianfeng Gao, Patrick Pantel, Michael Gamon, Xiaodong He and Li Deng	2
<i>Translation Modeling with Bidirectional Recurrent Neural Networks</i> Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker and Hermann Ney	14
<i>A Neural Network Approach to Selectional Preference Acquisition</i> Tim Van de Cruys	26
<i>Learning Image Embeddings using Convolutional Neural Networks for Improved Multi-Modal Semantics</i> Douwe Kiela and Léon Bottou	36
<i>Identifying Argumentative Discourse Structures in Persuasive Essays</i> Christian Stab and Iryna Gurevych	46
<i>Policy Learning for Domain Selection in an Extensible Multi-domain Spoken Dialogue System</i> Zhuoran Wang, Hongliang Chen, Guanchun Wang, Hao Tian, Hua Wu and Haifeng Wang	57
<i>A Constituent-Based Approach to Argument Labeling with Joint Inference in Discourse Parsing</i> Fang Kong, Hwee Tou Ng and Guodong Zhou	68
<i>Strongly Incremental Repair Detection</i> Julian Hough and Matthew Purver	78
<i>Semi-Supervised Chinese Word Segmentation Using Partial-Label Learning With Conditional Random Fields</i> Fan Yang and Paul Vozila	90
<i>Accurate Word Segmentation and POS Tagging for Japanese Microblogs: Corpus Annotation and Joint Modeling with Lexical Normalization</i> Nobuhiro Kaji and Masaru Kitsuregawa	99
<i>Revisiting Embedding Features for Simple Semi-supervised Learning</i> Jiang Guo, Wanxiang Che, Haifeng Wang and Ting Liu	110
<i>Combining Punctuation and Disfluency Prediction: An Empirical Study</i> Xuancong Wang, Khe Chai Sim and Hwee Tou Ng	121
<i>Submodularity for Data Selection in Machine Translation</i> Katrin Kirchhoff and Jeff Bilmes	131
<i>Improve Statistical Machine Translation with Context-Sensitive Bilingual Semantic Embedding Model</i> Haiyang Wu, Daxiang Dong, Xiaoguang Hu, Dianhai Yu, Wei He, Hua Wu, Haifeng Wang and Ting Liu	142
<i>Transformation from Discontinuous to Continuous Word Alignment Improves Translation Quality</i> Zhongjun He, Hua Wu, Haifeng Wang and Ting Liu	147
<i>Unsupervised Word Alignment Using Frequency Constraint in Posterior Regularized EM</i> Hidetaka Kamigaito, Taro Watanabe, Hiroya Takamura and Manabu Okumura	153

<i>Asymmetric Features Of Human Generated Translation</i> Sauleh Eetemadi and Kristina Toutanova	159
<i>Syntax-Augmented Machine Translation using Syntax-Label Clustering</i> Hideya MINO, Taro WATANABE and Eiichiro SUMITA	165
<i>Testing for Significance of Increased Correlation with Human Judgment</i> Yvette Graham and Timothy Baldwin	172
<i>Syntactic SMT Using a Discriminative Text Generation Model</i> Yue Zhang, Kai Song, Linfeng Song, Jingbo Zhu and Qun Liu	177
<i>Learning Hierarchical Translation Spans</i> jingyi zhang, Masao Utiyama, Eiichiro Sumita and Hai Zhao	183
<i>Neural Network Based Bilingual Language Model Growing for Statistical Machine Translation</i> Rui Wang, Hai Zhao, Bao-Liang Lu, Masao Utiyama and Eiichiro Sumita	189
<i>Better Statistical Machine Translation through Linguistic Treatment of Phrasal Verbs</i> Kostadin Cholakov and Valia Kordoni	196
<i>Fitting Sentence Level Translation Evaluation with Many Dense Features</i> Miloš Stanojević and Khalil Sima'an	202
<i>A Human Judgement Corpus and a Metric for Arabic MT Evaluation</i> Houda Bouamor, Hanan Alshikhabobakr, Behrang Mohit and Kemal Oflazer	207
<i>Learning to Differentiate Better from Worse Translations</i> Francisco Guzmán, Shafiq Joty, Lluís Màrquez, Alessandro Moschitti, Preslav Nakov and Massimo Nicosia	214
<i>Two Improvements to Left-to-Right Decoding for Hierarchical Phrase-based Machine Translation</i> Maryam Siahbani and Anoop Sarkar	221
<i>Reordering Model for Forest-to-String Machine Translation</i> Martin Cmejrek	227
<i>Aligning context-based statistical models of language with brain activity during reading</i> Leila Wehbe, Ashish Vaswani, Kevin Knight and Tom Mitchell	233
<i>A Cognitive Model of Semantic Network Learning</i> Aida Nematzadeh, Afsaneh Fazly and Suzanne Stevenson	244
<i>Learning Abstract Concept Embeddings from Multi-Modal Data: Since You Probably Can't See What I Mean</i> Felix Hill and Anna Korhonen	255
<i>Go Climb a Dependency Tree and Correct the Grammatical Errors</i> Longkai Zhang and Houfeng WANG	266
<i>An Unsupervised Model for Instance Level Subcategorization Acquisition</i> Simon Baker, Roi Reichart and Anna Korhonen	278
<i>Parsing low-resource languages using Gibbs sampling for PCFGs with latent annotations</i> Liang Sun, Jason Mielens and Jason Baldridge	290

<i>Incremental Semantic Role Labeling with Tree Adjoining Grammar</i> Ioannis Konstas, Frank Keller, Vera Demberg and Mirella Lapata	301
<i>A Graph-based Approach for Contextual Text Normalization</i> Cagil Sonmez and Arzucan Ozgur	313
<i>ReNoun: Fact Extraction for Nominal Attributes</i> Mohamed Yahya, Steven Whang, Rahul Gupta and Alon Halevy	325
<i>Hierarchical Discriminative Classification for Text-Based Geolocation</i> Benjamin Wing and Jason Baldridge	336
<i>Probabilistic Models of Cross-Lingual Semantic Similarity in Context Based on Latent Cross-Lingual Concepts Induced from Comparable Data</i> Ivan Vulić and Marie-Francine Moens	349
<i>Multi-Predicate Semantic Role Labeling</i> Haitong Yang and Chengqing Zong	363
<i>Werdy: Recognition and Disambiguation of Verbs and Verb Phrases with Syntactic and Semantic Pruning</i> Luciano Del Corro, Rainer Gemulla and Gerhard Weikum	374
<i>Multi-Resolution Language Grounding with Weak Supervision</i> R. Koncel-Kedziorski, Hannaneh Hajishirzi and Ali Farhadi	386
<i>Incorporating Vector Space Similarity in Random Walk Inference over Knowledge Bases</i> Matt Gardner, Partha Talukdar, Jayant Krishnamurthy and Tom Mitchell	397
<i>Composition of Word Representations Improves Semantic Role Labelling</i> Michael Roth and Kristian Woodsend	407
<i>Automatic Domain Assignment for Word Sense Alignment</i> Tommaso Caselli and Carlo Strapparava	414
<i>Nothing like Good Old Frequency: Studying Context Filters for Distributional Thesauri</i> Muntsa Padró, Marco Idiart, Aline Villavicencio and Carlos Ramisch	419
<i>Aligning English Strings with Abstract Meaning Representation Graphs</i> Nima Pourdamghani, Yang Gao, Ulf Hermjakob and Kevin Knight	425
<i>A Shortest-path Method for Arc-factored Semantic Role Labeling</i> Xavier Lluís, Xavier Carreras and Lluís Màrquez	430
<i>Semantic Kernels for Semantic Parsing</i> Iman Saleh, Alessandro Moschitti, Preslav Nakov, Lluís Màrquez and Shafiq Joty	436
<i>An I-vector Based Approach to Compact Multi-Granularity Topic Spaces Representation of Textual Documents</i> Mohamed Morchid, Mohamed Bouallegue, Richard Dufour, Georges Linares, Driss Matrouf and Renato de Mori	443
<i>Explaining the Stars: Weighted Multiple-Instance Learning for Aspect-Based Sentiment Analysis</i> Nikolaos Pappas and Andrei Popescu-Belis	455
<i>Sentiment Analysis on the People's Daily</i> Jiwei Li and Eduard Hovy	467

<i>A Joint Segmentation and Classification Framework for Sentiment Analysis</i> Duyu Tang, Furu Wei, Bing Qin, Li Dong, Ting Liu and Ming Zhou	477
<i>Positive Unlabeled Learning for Deceptive Reviews Detection</i> yafeng ren, donghong ji and hongbin zhang	488
<i>Resolving Shell Nouns</i> Varada Kolhatkar and Graeme Hirst	499
<i>A Comparison of Selectional Preference Models for Automatic Verb Classification</i> Will Roberts and Markus Egg	511
<i>Learning to Solve Arithmetic Word Problems with Verb Categorization</i> Mohammad Javad Hosseini, Hannaneh Hajishirzi, Oren Etzioni and Nate Kushman	523
<i>NaturalLI: Natural Logic Inference for Common Sense Reasoning</i> Gabor Angeli and Christopher D. Manning	534
<i>Modeling Term Translation for Document-informed Machine Translation</i> Fandong Meng, Deyi Xiong, Wenbin Jiang and Qun Liu	546
<i>Beyond Parallel Data: Joint Word Alignment and Decipherment Improves Machine Translation</i> Qing Dou, Ashish Vaswani and Kevin Knight	557
<i>Latent Domain Phrase-based Models for Adaptation</i> Cuong Hoang and Khalil Sima'an	566
<i>Translation Rules with Right-Hand Side Lattices</i> Fabien Cromieres and Sadao Kurohashi	577
<i>Learning to Translate: A Query-Specific Combination Approach for Cross-Lingual Information Retrieval</i> Ferhan Ture and Elizabeth Boschee	589
<i>Semantic-Based Multilingual Document Clustering via Tensor Modeling</i> Salvatore Romeo, Andrea Tagarelli and Dino Ienco	600
<i>Lexical Substitution for the Medical Domain</i> Martin Riedl, Michael Glass and Alfio Gliozzo	610
<i>Question Answering with Subgraph Embeddings</i> Antoine Bordes, Sumit Chopra and Jason Weston	615
<i>Correcting Keyboard Layout Errors and Homoglyphs in Queries</i> Derek Barnes, Mahesh Joshi and Hassan Sawaf	621
<i>Non-linear Mapping for Improved Identification of 1300+ Languages</i> Ralf Brown	627
<i>A Neural Network for Factoid Question Answering over Paragraphs</i> Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher and Hal Daumé III	633
<i>Joint Relational Embeddings for Knowledge-based Question Answering</i> Min-Chul Yang, Nan Duan, Ming Zhou and Hae-Chang Rim	645
<i>Adding High-Precision Links to Wikipedia</i> Thanapon Noraset, Chandra Bhagavatula and Doug Downey	651

<i>Finding Good Enough: A Task-Based Evaluation of Query Biased Summarization for Cross-Language Information Retrieval</i>	
Jennifer Williams, Sharon Tam and Wade Shen	657
<i>Chinese Poetry Generation with Recurrent Neural Networks</i>	
Xingxing Zhang and Mirella Lapata	670
<i>Fear the REAPER: A System for Automatic Multi-Document Summarization with Reinforcement Learning</i>	
Cody Rioux, Sadid A. Hasan and Yllias Chali	681
<i>Improving Multi-documents Summarization by Sentence Compression based on Expanded Constituent Parse Trees</i>	
Chen Li, Yang Liu, Fei Liu, Lin Zhao and Fuliang Weng	691
<i>Analyzing Stemming Approaches for Turkish Multi-Document Summarization</i>	
Muhammed Yavuz Nuzumlalı and Arzucan Özgür	702
<i>Invited Talk: Learning from Rational Behavior</i>	
Thorsten Joachims	707
<i>Evaluating Neural Word Representations in Tensor-Based Compositional Settings</i>	
Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh and Matthew Purver	708
<i>Opinion Mining with Deep Recurrent Neural Networks</i>	
Ozan Irsoy and Claire Cardie	720
<i>The Inside-Outside Recursive Neural Network model for Dependency Parsing</i>	
Phong Le and Willem Zuidema	729
<i>A Fast and Accurate Dependency Parser using Neural Networks</i>	
Danqi Chen and Christopher Manning	740
<i>Why are You Taking this Stance? Identifying and Classifying Reasons in Ideological Debates</i>	
Kazi Saidul Hasan and Vincent Ng	751
<i>Chinese Zero Pronoun Resolution: An Unsupervised Probabilistic Model Rivaling Supervised Resolvers</i>	
Chen Chen and Vincent Ng	763
<i>Unsupervised Sentence Enhancement for Automatic Summarization</i>	
Jackie Chi Kit Cheung and Gerald Penn	775
<i>ReferItGame: Referring to Objects in Photographs of Natural Scenes</i>	
Sahar Kazemzadeh, Vicente Ordonez, Mark Matten and Tamara Berg	787
<i>Unsupervised Template Mining for Semantic Category Understanding</i>	
Lei Shi, Shuming Shi, Chin-Yew Lin, Yi-Dong Shen and Yong Rui	799
<i>Taxonomy Construction Using Syntactic Contextual Evidence</i>	
Tuan Luu Anh, Jung-jae Kim and See Kiong Ng	810
<i>Analysing recall loss in named entity slot filling</i>	
Glen Pink, Joel Nothman and James R. Curran	820

<i>Relieving the Computational Bottleneck: Joint Inference for Event Extraction with High-Dimensional Features</i>	
Deepak Venugopal, Chen Chen, Vibhav Gogate and Vincent Ng	831
<i>Syllable weight encodes mostly the same information for English word segmentation as dictionary stress</i>	
John K Pate and Mark Johnson	844
<i>A Joint Model for Unsupervised Chinese Word Segmentation</i>	
Miaohong Chen, Baobao Chang and Wenzhe Pei	854
<i>Domain Adaptation for CRF-based Chinese Word Segmentation using Free Annotations</i>	
Yijia Liu, Yue Zhang, Wanxiang Che, Ting Liu and Fan Wu	864
<i>Balanced Korean Word Spacing with Structural SVM</i>	
Changki Lee, Edward Choi and Hyunki Kim	875
<i>Morphological Segmentation for Keyword Spotting</i>	
Karthik Narasimhan, Damianos Karakos, Richard Schwartz, Stavros Tsakalidis and Regina Barzilay	880
<i>What Can We Get From 1000 Tokens? A Case Study of Multilingual POS Tagging For Resource-Poor Languages</i>	
Long Duong, Trevor Cohn, Karin Verspoor, Steven Bird and Paul Cook	886
<i>An Experimental Comparison of Active Learning Strategies for Partially Labeled Sequences</i>	
Diego Marcheggiani and Thierry Artières	898
<i>Language Modeling with Functional Head Constraint for Code Switching Speech Recognition</i>	
Ying LI and Pascale Fung	907
<i>A Polynomial-Time Dynamic Oracle for Non-Projective Dependency Parsing</i>	
Carlos Gómez-Rodríguez, Francesco Sartorio and Giorgio Satta	917
<i>Ambiguity Resolution for Vt-N Structures in Chinese</i>	
Yu-Ming Hsieh, Jason S. Chang and Keh-Jiann Chen	928
<i>Neural Networks Leverage Corpus-wide Information for Part-of-speech Tagging</i>	
Yuta Tsuboi	938
<i>System Combination for Grammatical Error Correction</i>	
Raymond Hendy Susanto, Peter Phandi and Hwee Tou Ng	951
<i>Dependency parsing with latent refinements of part-of-speech tags</i>	
Thomas Mueller, Richárd Farkas, Alex Judea, Helmut Schmid and hinrich schuetze	963
<i>Importance weighting and unsupervised domain adaptation of POS taggers: a negative result</i>	
Barbara Plank, Anders Johannsen and Anders Søgaard	968
<i>POS Tagging of English-Hindi Code-Mixed Social Media Content</i>	
Yogarshi Vyas, Spandana Gella, Jatin Sharma, Kalika Bali and Monojit Choudhury	974
<i>Data Driven Grammatical Error Detection in Transcripts of Children’s Speech</i>	
Eric Morley, Anna Eva Hallin and Brian Roark	980
<i>A* CCG Parsing with a Supertag-factored Model</i>	
Mike Lewis and Mark Steedman	990

<i>A Dependency Parser for Tweets</i>	
Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archana Bhatia, Chris Dyer and Noah A. Smith	1001
<i>Greed is Good if Randomized: New Inference for Dependency Parsing</i>	
Yuan Zhang, Tao Lei, Regina Barzilay and Tommi Jaakkola	1013
<i>A Unified Model for Word Sense Representation and Disambiguation</i>	
Xinxiong Chen, Zhiyuan Liu and Maosong Sun	1025
<i>Reducing Dimensions of Tensors in Type-Driven Distributional Semantics</i>	
Tamara Polajnar, Luana Fagarasan and Stephen Clark	1036
<i>An Etymological Approach to Cross-Language Orthographic Similarity. Application on Romanian</i>	
Alina Maria Ciobanu and Liviu P. Dinu	1047
<i>Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space</i>	
Arvind Neelakantan, Jeevan Shankar, Alexandre Passos and Andrew McCallum	1059
<i>Tailor knowledge graph for query understanding: linking intent topics by propagation</i>	
Shi Zhao and Yan Zhang	1070
<i>Queries as a Source of Lexicalized Commonsense Knowledge</i>	
Marius Pasca	1081
<i>Question Answering over Linked Data Using First-order Logic</i>	
Shizhu He, Kang Liu, Yuanzhe Zhang, Liheng Xu and Jun Zhao	1092
<i>Knowledge Graph and Corpus Driven Segmentation and Answer Inference for Telegraphic Entity-seeking Queries</i>	
Mandar Joshi, Uma Sawant and Soumen Chakrabarti	1104
<i>A Regularized Competition Model for Question Difficulty Estimation in Community Question Answering Services</i>	
Quan Wang, Jing Liu, Bin Wang and Li Guo	1115
<i>Vote Prediction on Comments in Social Polls</i>	
Isaac Persing and Vincent Ng	1127
<i>Exploiting Social Relations and Sentiment for Stock Prediction</i>	
Jianfeng Si, Arjun Mukherjee, Bing Liu, Sinno Jialin Pan, Qing Li and Huayi Li	1139
<i>Developing Age and Gender Predictive Lexica over Social Media</i>	
Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar and Hansen Andrew Schwartz	1146
<i>Dependency Parsing for Weibo: An Efficient Probabilistic Logic Programming Approach</i>	
William Yang Wang, Lingpeng Kong, Kathryn Mazaitis and William W Cohen	1152
<i>Exploiting Community Emotion for Microblog Event Detection</i>	
Gaoyan Ou, Wei Chen, Tengjiao Wang, Zhongyu Wei, Binyang LI, Dongqing Yang and Kam-Fai Wong	1159
<i>Detecting Disagreement in Conversations using Pseudo-Monologic Rhetorical Structure</i>	
Kelsey Allen, Giuseppe Carenini and Raymond Ng	1169

<i>+/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference</i> Yoonjung Choi and Janyce Wiebe	1181
<i>A Sentiment-aligned Topic Model for Product Aspect Rating Prediction</i> Hao Wang and Martin Ester	1192
<i>Learning Emotion Indicators from Tweets: Hashtags, Hashtag Patterns, and Phrases</i> Ashequl Qadir and Ellen Riloff	1203
<i>Fine-Grained Contextual Predictions for Hard Sentiment Words</i> Sebastian Ebert and Hinrich Schütze	1210
<i>An Iterative Link-based Method for Parallel Web Page Mining</i> Le Liu, Yu Hong, Jun Lu, Jun Lang, Heng Ji and Jianmin Yao.....	1216
<i>Human Effort and Machine Learnability in Computer Aided Translation</i> Spence Green, Sida I. Wang, Jason Chuang, Jeffrey Heer, Sebastian Schuster and Christopher D. Manning	1225
<i>Exact Decoding for Phrase-Based Statistical Machine Translation</i> Wilker Aziz, Marc Dymetman and Lucia Specia	1237
<i>Large-scale Expected BLEU Training of Phrase-based Reordering Models</i> Michael Auli, Michel Galley and Jianfeng Gao	1250
<i>Confidence-based Rewriting of Machine Translation Output</i> Benjamin Marie and Aurélien Max	1261
<i>Learning Compact Lexicons for CCG Semantic Parsing</i> Yoav Artzi, Dipanjan Das and Slav Petrov	1273
<i>Morpho-syntactic Lexical Generalization for CCG Semantic Parsing</i> Adrienne Wang, Tom Kwiatkowski and Luke Zettlemoyer	1284
<i>Semantic Parsing Using Content and Context: A Case Study from Requirements Elicitation</i> Reut Tsarfay, Ilia Pogrebezky, Guy Weiss, Yaarit Natan, Smadar Szekely and David Harel ..	1296
<i>Semantic Parsing with Relaxed Hybrid Trees</i> Wei Lu	1308
<i>Low-dimensional Embeddings for Interpretable Anchor-based Topic Inference</i> David Mimno and Moontae Lee	1319
<i>Weakly-Supervised Learning with Cost-Augmented Contrastive Estimation</i> Kevin Gimpel and Mohit Bansal	1329
<i>Don't Until the Final Verb Wait: Reinforcement Learning for Simultaneous Machine Translation</i> Alvin Grissom II, He He, Jordan Boyd-Graber, John Morgan and Hal Daumé III.....	1342
<i>PCFG Induction for Unsupervised Parsing and Language Modelling</i> James Scicluna and Colin de la Higuera.....	1353
<i>Can characters reveal your native language? A language-independent approach to native language identification</i> Radu Tudor Ionescu, Marius Popescu and Aoife Cahill.....	1363

<i>Formalizing Word Sampling for Vocabulary Prediction as Graph-based Active Learning</i> Yo Ehara, Yusuke Miyao, Hidekazu Oiwa, Issei Sato and Hiroshi Nakagawa	1374
<i>Language Transfer Hypotheses with Linear SVM Weights</i> Shervin Malmasi and Mark Dras	1385
<i>Predicting Dialect Variation in Immigrant Contexts Using Light Verb Constructions</i> A. Seza Dogruoz and Preslav Nakov	1391
<i>Device-Dependent Readability for Improved Text Understanding</i> A-Yeong Kim, Hyun-Je Song, Seong-Bae Park and Sang-Jo Lee	1396
<i>Predicting Chinese Abbreviations with Minimum Semantic Unit and Global Constraints</i> Longkai Zhang, li li, Houfeng WANG and Xu Sun	1405
<i>Using Structured Events to Predict Stock Price Movement: An Empirical Investigation</i> Xiao Ding, Yue Zhang, Ting Liu and Junwen Duan	1415
<i>Extracting Clusters of Specialist Terms from Unstructured Text</i> Aaron Gerow	1426
<i>Citation-Enhanced Keyphrase Extraction from Research Papers: A Supervised Approach</i> Cornelia Caragea, Florin Adrian Bulgarov, Andreea Godea and Sujatha Das Gollapalli	1435
<i>Using Mined Coreference Chains as a Resource for a Semantic Task</i> Heike Adel and Hinrich Schütze	1447
<i>Financial Keyword Expansion via Continuous Word Vector Representations</i> Ming-Feng Tsai and Chuan-Ju Wang	1453
<i>Intrinsic Plagiarism Detection using N-gram Classes</i> Imene Bensalem, Paolo Rosso and Salim Chikhi	1459
<i>Verifiably Effective Arabic Dialect Identification</i> Kareem Darwish, Hassan Sajjad and Hamdy Mubarak	1465
<i>Keystroke Patterns as Prosody in Digital Writings: A Case Study with Deceptive Reviews and Essays</i> Ritwik Banerjee, Song Feng, Jun Seok Kang and Yejin Choi	1469
<i>Leveraging Effective Query Modeling Techniques for Speech Recognition and Summarization</i> Kuan-Yu Chen, Shih-Hung Liu, Berlin Chen, Ea-Ee Jan, Hsin-Min Wang, Wen-Lian Hsu and Hsin-Hsi Chen	1474
<i>Staying on Topic: An Indicator of Power in Political Debates</i> Vinodkumar Prabhakaran, Ashima Arora and Owen Rambow	1481
<i>Language Modeling with Power Low Rank Ensembles</i> Ankur P. Parikh, Avneesh Saluja, Chris Dyer and Eric Xing	1487
<i>Modeling Biological Processes for Reading Comprehension</i> Jonathan Berant, Vivek Srikumar, Pei-Chun Chen, Abby Vander Linden, Brittany Harding, Brad Huang and Christopher D. Manning	1499
<i>Sensicon: An Automatically Constructed Sensorial Lexicon</i> Serra Sinem Tekiroglu, Gözde Özbal and Carlo Strapparava	1511

<i>Word Semantic Representations using Bayesian Probabilistic Tensor Factorization</i> Jingwei Zhang, Jeremy Salwen, Michael Glass and Alfio Gliozzo	1522
<i>Glove: Global Vectors for Word Representation</i> Jeffrey Pennington, Richard Socher and Christopher Manning	1532
<i>Jointly Learning Word Representations and Composition Functions Using Predicate-Argument Structures</i> Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa and Yoshimasa Tsuruoka.....	1544
<i>Combining Distant and Partial Supervision for Relation Extraction</i> Gabor Angeli, Julie Tibshirani, Jean Wu and Christopher D. Manning	1556
<i>Typed Tensor Decomposition of Knowledge Bases for Relation Extraction</i> Kai-Wei Chang, Wen-tau Yih, Bishan Yang and Christopher Meek	1568
<i>A convex relaxation for weakly supervised relation extraction</i> Edouard Grave	1580
<i>Knowledge Graph and Text Jointly Embedding</i> Zhen Wang, Jianwen Zhang, Jianlin Feng and Zheng Chen	1591
<i>Abstractive Summarization of Product Reviews Using Discourse Structure</i> Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng and Bitu Nejat	1602
<i>Clustering Aspect-related Phrases by Leveraging Sentiment Distribution Consistency</i> Li Zhao, Minlie Huang, Haiqiang Chen, Junjun Cheng and Xiaoyan Zhu.....	1614
<i>Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach</i> Yue Hu and Xiaojun Wan	1624
<i>Fast and Accurate Misspelling Correction in Large Corpora</i> Octavian Popescu and Ngoc Phuoc An Vo	1634
<i>Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models</i> Marcello Federico, Matteo Negri, Luisa Bentivogli and Marco Turchi.....	1643
<i>Refining Word Segmentation Using a Manually Aligned Corpus for Statistical Machine Translation</i> Xiaolin Wang, Masao Utiyama, Andrew Finch and Eiichiro Sumita.....	1654
<i>Improving Pivot-Based Statistical Machine Translation by Pivoting the Co-occurrence Count of Phrase Pairs</i> Xiaoning Zhu, Zhongjun He, Hua Wu, Conghui Zhu, Haifeng Wang and Tiejun Zhao	1665
<i>Word Translation Prediction for Morphologically Rich Languages with Bilingual Neural Networks</i> Ke M. Tran, Arianna Bisazza and Christof Monz	1676
<i>Dependency-Based Bilingual Language Models for Reordering in Statistical Machine Translation</i> Ekaterina Garmash and Christof Monz.....	1689
<i>Combining String and Context Similarity for Bilingual Term Alignment from Comparable Corpora</i> Georgios Kontonatsios, Ioannis Korkontzelos, Jun'ichi Tsujii and Sophia Ananiadou	1701
<i>Random Manhattan Integer Indexing: Incremental L1 Normed Vector Space Construction</i> Behrang Q. Zadeh and Siegfried Handschuh.....	1713

<i>Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation</i> Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk and Yoshua Bengio	1724
<i>Type-based MCMC for Sampling Tree Fragments from Forests</i> Xiaochang Peng and Daniel Gildea	1735
<i>Convolutional Neural Networks for Sentence Classification</i> Yoon Kim	1746
<i>Sometimes Average is Best: The Importance of Averaging for Prediction using MCMC Inference in Topic Modeling</i> Viet-An Nguyen, Jordan Boyd-Graber and Philip Resnik	1752
<i>Large-scale Reordering Model for Statistical Machine Translation using Dual Multinomial Logistic Regression</i> Abdullah Alrajeh and Mahesan Niranjana	1758
<i>Improved Decipherment of Homophonic Ciphers</i> Malte Nuhn, Julian Schamper and Hermann Ney	1764
<i>Cipher Type Detection</i> Malte Nuhn and Kevin Knight	1769
<i>Joint Learning of Chinese Words, Terms and Keywords</i> Ziqiang Cao, Sujian Li and Heng Ji	1774
<i>Cross-Lingual Part-of-Speech Tagging through Ambiguous Learning</i> Guillaume Wisniewski, Nicolas Pécheux, Souhir Gahbiche-Braham and François Yvon	1779
<i>Comparing Representations of Semantic Roles for String-To-Tree Decoding</i> Marzieh Bazrafshan and Daniel Gildea	1786
<i>Detecting Non-compositional MWE Components using Wiktionary</i> Bahar Salehi, Paul Cook and Timothy Baldwin	1792
<i>Joint Emotion Analysis via Multi-task Gaussian Processes</i> Daniel Beck, Trevor Cohn and Lucia Specia	1798
<i>Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics</i> Zubin Jelveh, Bruce Kogut and Suresh Naidu	1804
<i>A Model of Individual Differences in Gaze Control During Reading</i> Niels Landwehr, Sebastian Arzt, Tobias Scheffer and Reinhold Kliegl	1810
<i>Multi-label Text Categorization with Hidden Components</i> li li, Longkai Zhang and Houfeng WANG	1816
<i>#TagSpace: Semantic Embeddings from Hashtags</i> Jason Weston, Sumit Chopra and Keith Adams	1822
<i>Joint Decoding of Tree Transduction Models for Sentence Compression</i> Jin-ge Yao, Xiaojun Wan and Jianguo Xiao	1828
<i>Dependency-based Discourse Parser for Single-Document Summarization</i> Yasuhisa Yoshida, Jun Suzuki, Tsutomu Hirao and Masaaki Nagata	1834

<i>Improving Word Alignment using Word Similarity</i>	
Theerawat Songyot and David Chiang	1840
<i>Constructing Information Networks Using One Single Model</i>	
Qi Li, Heng Ji, Yu HONG and Sujian Li	1846
<i>Event Role Extraction using Domain-Relevant Word Representations</i>	
Emanuela Boros, Romaric Besançon, Olivier Ferret and Brigitte Grau	1852
<i>Modeling Joint Entity and Relation Extraction with Table Representation</i>	
Makoto Miwa and Yutaka Sasaki	1858
<i>ZORE: A Syntax-based System for Chinese Open Relation Extraction</i>	
Likun Qiu and Yue Zhang	1870
<i>Coarse-grained Candidate Generation and Fine-grained Re-ranking for Chinese Abbreviation Prediction</i>	
Longkai Zhang, Houfeng WANG and Xu Sun	1881
<i>Type-Aware Distantly Supervised Relation Extraction with Linked Arguments</i>	
Mitchell Koch, John Gilmer, Stephen Soderland and Daniel S. Weld	1891
<i>Automatic Inference of the Tense of Chinese Events Using Implicit Linguistic Information</i>	
Yuchen Zhang and Nianwen Xue	1902
<i>Joint Inference for Knowledge Base Population</i>	
Liwei Chen, Yansong Feng, Jinghui Mo, Songfang Huang and Dongyan Zhao	1912
<i>Combining Visual and Textual Features for Information Extraction from Online Flyers</i>	
Emilia Apostolova and Noriko Tomuro	1924
<i>CTPs: Contextual Temporal Profiles for Time Scoping Facts using State Change Detection</i>	
Derry Tanti Wijaya, Ndapandula Nakashole and Tom M. Mitchell	1930
<i>Noisy Or-based model for Relation Extraction using Distant Supervision</i>	
Ajay Nagesh, Gholamreza Haffari and Ganesh Ramakrishnan	1937
<i>Search-Aware Tuning for Machine Translation</i>	
Lemao Liu and Liang Huang	1942
<i>Latent-Variable Synchronous CFGs for Hierarchical Translation</i>	
Avneesh Saluja, Chris Dyer and Shay B. Cohen	1953
<i>Gender and Power: How Gender and Gender Environment Affect Manifestations of Power</i>	
Vinodkumar Prabhakaran, Emily E. Reid and Owen Rambow	1965
<i>Online topic model for Twitter considering dynamics of user interests and topic trends</i>	
Kentaro Sasaki, Tomohiro Yoshikawa and Takeshi Furuhashi	1977
<i>Self-disclosure topic model for classifying and analyzing Twitter conversations</i>	
JinYeong Bak, Chin-Yew Lin and Alice Oh	1986
<i>Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts</i>	
Jiwei Li, Alan Ritter, Claire Cardie and Eduard Hovy	1997

<i>Brighter than Gold: Figurative Language in User Generated Comparisons</i> Vlad Niculae and Cristian Danescu-Niculescu-Mizil	2008
<i>Classifying Idiomatic and Literal Expressions Using Topic Models and Intensity of Emotions</i> Jing Peng, Anna Feldman and Ekaterina Vylomova	2019
<i>Learning Spatial Knowledge for Text to 3D Scene Generation</i> Angel Chang, Manolis Savva and Christopher D. Manning	2028
<i>A Model of Coherence Based on Distributed Sentence Representation</i> Jiwei Li and Eduard Hovy	2039
<i>Discriminative Reranking of Discourse Parses Using Tree Kernels</i> Shafiq Joty and Alessandro Moschitti	2049
<i>Recursive Deep Models for Discourse Parsing</i> Jiwei Li, Rumeng Li and Eduard Hovy	2061
<i>Recall Error Analysis for Coreference Resolution</i> Sebastian Martschat and Michael Strube	2070
<i>A Rule-Based System for Unrestricted Bridging Resolution: Recognizing Bridging Anaphora and Finding Links to Antecedents</i> Yufang Hou, Katja Markert and Michael Strube	2082
<i>Resolving Referring Expressions in Conversational Dialogs for Natural User Interfaces</i> Asli Celikyilmaz, Zhaleh Feizollahi, Dilek Hakkani-Tur and Ruhi Sarikaya	2094
<i>Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure</i> Yancui Li, wenhe feng, jing sun, Fang Kong and Guodong Zhou	2105
<i>Prune-and-Score: Learning for Greedy Coreference Resolution</i> Chao Ma, Janardhan Rao Doppa, J. Walker Orr, Prashanth Mannem, Xiaoli Fern, Tom Dietterich and Prasad Tadepalli	2115
<i>Summarizing Online Forum Discussions – Can Dialog Acts of Individual Messages Help?</i> Sumit Bhatia, Prakhar Biyani and Prasenjit Mitra	2127

Conference Program

Sunday, October 26, 2014

08:00–17:00 *Registration*

08:00–09:00 *Refreshments*

Plenary Session

08:40–09:00 *Opening Remarks, QNRF Presentation, Conference Logistics, Paper Selection Process and Statistics*

Alessandro Moschitti, Rekha Pilla, Kareem M. Darwish, and Walter Daelemans

09:00–10:00 *Invited Talk: IBM Cognitive Computing - An NLP Renaissance!*

Salim Roukos

10:00–10:30 *Coffee Break*

Session 1a: Neural Net Mixer I

10:30–10:55 *Modeling Interestingness with Deep Neural Networks*

Jianfeng Gao, Patrick Pantel, Michael Gamon, Xiaodong He and Li Deng

10:55–11:20 *Translation Modeling with Bidirectional Recurrent Neural Networks*

Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker and Hermann Ney

11:20–11:45 *A Neural Network Approach to Selectional Preference Acquisition*

Tim Van de Cruys

11:45–12:10 *Learning Image Embeddings using Convolutional Neural Networks for Improved Multi-Modal Semantics*

Douwe Kiela and Léon Bottou

Sunday, October 26, 2014 (continued)

Session 1b: Discourse, Dialogue and Pragmatics

- 10:30–10:55 *Identifying Argumentative Discourse Structures in Persuasive Essays*
Christian Stab and Iryna Gurevych
- 10:55–11:20 *Policy Learning for Domain Selection in an Extensible Multi-domain Spoken Dialogue System*
Zhuoran Wang, Hongliang Chen, Guanchun Wang, Hao Tian, Hua Wu and Haifeng Wang
- 11:20–11:45 *A Constituent-Based Approach to Argument Labeling with Joint Inference in Discourse Parsing*
Fang Kong, Hwee Tou Ng and Guodong Zhou
- 11:45–12:10 *Strongly Incremental Repair Detection*
Julian Hough and Matthew Purver

Session 1c: Segmentation / Spoken Language

- 10:30–10:55 *Semi-Supervised Chinese Word Segmentation Using Partial-Label Learning With Conditional Random Fields*
Fan Yang and Paul Vozila
- 10:55–11:20 *Accurate Word Segmentation and POS Tagging for Japanese Microblogs: Corpus Annotation and Joint Modeling with Lexical Normalization*
Nobuhiro Kaji and Masaru Kitsuregawa
- 11:20–11:45 *Revisiting Embedding Features for Simple Semi-supervised Learning*
Jiang Guo, Wanxiang Che, Haifeng Wang and Ting Liu
- 11:45–12:10 *Combining Punctuation and Disfluency Prediction: An Empirical Study*
Xuancong Wang, Khe Chai Sim and Hwee Tou Ng

Sunday, October 26, 2014 (continued)

Session 1p: Machine Translation

10:30–12:10 *Poster Session*
Multiple presenters

Submodularity for Data Selection in Machine Translation
Katrin Kirchhoff and Jeff Bilmes

Improve Statistical Machine Translation with Context-Sensitive Bilingual Semantic Embedding Model
Haiyang Wu, Daxiang Dong, Xiaoguang Hu, Dianhai Yu, Wei He, Hua Wu, Haifeng Wang and Ting Liu

Transformation from Discontinuous to Continuous Word Alignment Improves Translation Quality
Zhongjun He, Hua Wu, Haifeng Wang and Ting Liu

Unsupervised Word Alignment Using Frequency Constraint in Posterior Regularized EM
Hidetaka Kamigaito, Taro Watanabe, Hiroya Takamura and Manabu Okumura

Asymmetric Features Of Human Generated Translation
Sauleh Eetemadi and Kristina Toutanova

Syntax-Augmented Machine Translation using Syntax-Label Clustering
Hideya MINO, Taro WATANABE and Eiichiro SUMITA

Testing for Significance of Increased Correlation with Human Judgment
Yvette Graham and Timothy Baldwin

Syntactic SMT Using a Discriminative Text Generation Model
Yue Zhang, Kai Song, Linfeng Song, Jingbo Zhu and Qun Liu

Learning Hierarchical Translation Spans
jingyi zhang, Masao Utiyama, Eiichiro Sumita and Hai Zhao

Neural Network Based Bilingual Language Model Growing for Statistical Machine Translation
Rui Wang, Hai Zhao, Bao-Liang Lu, Masao Utiyama and Eiichiro Sumita

Better Statistical Machine Translation through Linguistic Treatment of Phrasal Verbs
Kostadin Cholakov and Valia Kordoni

Sunday, October 26, 2014 (continued)

Fitting Sentence Level Translation Evaluation with Many Dense Features

Miloš Stanojević and Khalil Sima'an

A Human Judgement Corpus and a Metric for Arabic MT Evaluation

Houda Bouamor, Hanan Alshikhabobakr, Behrang Mohit and Kemal Oflazer

Learning to Differentiate Better from Worse Translations

Francisco Guzmán, Shafiq Joty, Lluís Màrquez, Alessandro Moschitti, Preslav Nakov and Massimo Nicosia

Two Improvements to Left-to-Right Decoding for Hierarchical Phrase-based Machine Translation

Maryam Siahbani and Anoop Sarkar

Reordering Model for Forest-to-String Machine Translation

Martin Cmejrek

12:10–13:30 Lunch Break

Session 2a: Computational Psycholinguistics

13:30–13:55 *Aligning context-based statistical models of language with brain activity during reading*

Leila Wehbe, Ashish Vaswani, Kevin Knight and Tom Mitchell

13:55–14:20 *A Cognitive Model of Semantic Network Learning*

Aida Nematzadeh, Afsaneh Fazly and Suzanne Stevenson

14:20–14:45 *The Benefits of a Model of Annotation*

Rebecca J. Passonneau and Bob Carpenter

14:45–15:10 *Learning Abstract Concept Embeddings from Multi-Modal Data: Since You Probably Can't See What I Mean*

Felix Hill and Anna Korhonen

Sunday, October 26, 2014 (continued)

Session 2b: Tagging, Chunking, Parsing and Syntax

- 13:30–13:55 *Go Climb a Dependency Tree and Correct the Grammatical Errors*
Longkai Zhang and Houfeng WANG
- 13:55–14:20 *An Unsupervised Model for Instance Level Subcategorization Acquisition*
Simon Baker, Roi Reichart and Anna Korhonen
- 14:20–14:45 *Parsing low-resource languages using Gibbs sampling for PCFGs with latent annotations*
Liang Sun, Jason Mielens and Jason Baldridge
- 14:45–15:10 *Incremental Semantic Role Labeling with Tree Adjoining Grammar*
Ioannis Konstas, Frank Keller, Vera Demberg and Mirella Lapata

Session 2c: NLP for the Web and Social Media

- 13:30–13:55 *A Graph-based Approach for Contextual Text Normalization*
Cagil Sonmez and Arzucan Ozgur
- 13:55–14:20 *Entity Linking on Microblogs with Spatial and Temporal Signals*
Yuan Fang and Ming-Wei Chang
- 14:20–14:45 *ReNoun: Fact Extraction for Nominal Attributes*
Mohamed Yahya, Steven Whang, Rahul Gupta and Alon Halevy
- 14:45–15:10 *Hierarchical Discriminative Classification for Text-Based Geolocation*
Benjamin Wing and Jason Baldridge

Sunday, October 26, 2014 (continued)

Session 2p: Semantics

13:30–15:10 *Poster Session*
Multiple presenters

Probabilistic Models of Cross-Lingual Semantic Similarity in Context Based on Latent Cross-Lingual Concepts Induced from Comparable Data
Ivan Vulić and Marie-Francine Moens

Multi-Predicate Semantic Role Labeling
Haitong Yang and Chengqing Zong

Werdy: Recognition and Disambiguation of Verbs and Verb Phrases with Syntactic and Semantic Pruning
Luciano Del Corro, Rainer Gemulla and Gerhard Weikum

Multi-Resolution Language Grounding with Weak Supervision
R. Koncel-Kedziorski, Hannaneh Hajishirzi and Ali Farhadi

Incorporating Vector Space Similarity in Random Walk Inference over Knowledge Bases
Matt Gardner, Partha Talukdar, Jayant Krishnamurthy and Tom Mitchell

Composition of Word Representations Improves Semantic Role Labelling
Michael Roth and Kristian Woodsend

Automatic Domain Assignment for Word Sense Alignment
Tommaso Caselli and Carlo Strapparava

Nothing like Good Old Frequency: Studying Context Filters for Distributional Thesauri
Muntsa Padró, Marco Idiart, Aline Villavicencio and Carlos Ramisch

Aligning English Strings with Abstract Meaning Representation Graphs
Nima Pourdamghani, Yang Gao, Ulf Hermjakob and Kevin Knight

A Shortest-path Method for Arc-factored Semantic Role Labeling
Xavier Lluís, Xavier Carreras and Lluís Màrquez

Semantic Kernels for Semantic Parsing
Iman Saleh, Alessandro Moschitti, Preslav Nakov, Lluís Màrquez and Shafiq Joty

Sunday, October 26, 2014 (continued)

Multi-Modal Models for Concrete and Abstract Concept Meaning

Felix Hill, Roi Reichart, Anna Korhonen

An I-vector Based Approach to Compact Multi-Granularity Topic Spaces Representation of Textual Documents

Mohamed Morchid, Mohamed Bouallegue, Richard Dufour, Georges Linares, Driss Matrouf and Renato de Mori

15:10–15:40 *Coffee Break*

Session 3a: Sentiment Analysis and Opinion Mining

15:40–16:05 *Explaining the Stars: Weighted Multiple-Instance Learning for Aspect-Based Sentiment Analysis*

Nikolaos Pappas and Andrei Popescu-Belis

16:05–16:30 *Sentiment Analysis on the People's Daily*

Jiwei Li and Eduard Hovy

16:30–16:55 *A Joint Segmentation and Classification Framework for Sentiment Analysis*

Duyu Tang, Furu Wei, Bing Qin, Li Dong, Ting Liu and Ming Zhou

16:55–17:20 *Positive Unlabeled Learning for Deceptive Reviews Detection*

yafeng ren, donghong ji and hongbin zhang

Session 3b: Semantics

15:40–16:05 *Resolving Shell Nouns*

Varada Kolhatkar and Graeme Hirst

16:05–16:30 *A Comparison of Selectional Preference Models for Automatic Verb Classification*

Will Roberts and Markus Egg

16:30–16:55 *Learning to Solve Arithmetic Word Problems with Verb Categorization*

Mohammad Javad Hosseini, Hannaneh Hajishirzi, Oren Etzioni and Nate Kushman

16:55–17:20 *NaturalLI: Natural Logic Inference for Common Sense Reasoning*

Gabor Angeli and Christopher D. Manning

Sunday, October 26, 2014 (continued)

Session 3c: Machine Translation

- 15:40–16:05 *Modeling Term Translation for Document-informed Machine Translation*
Fandong Meng, Deyi Xiong, Wenbin Jiang and Qun Liu
- 16:05–16:30 *Beyond Parallel Data: Joint Word Alignment and Decipherment Improves Machine Translation*
Qing Dou, Ashish Vaswani and Kevin Knight
- 16:30–16:55 *Latent Domain Phrase-based Models for Adaptation*
Cuong Hoang and Khalil Sima'an
- 16:55–17:20 *Translation Rules with Right-Hand Side Lattices*
Fabien Cromieres and Sadao Kurohashi

Session 3p: Information Retrieval, Summarization and Question Answering

- 15:40–17:20 *Poster Session*
Multiple presenters
- Learning to Translate: A Query-Specific Combination Approach for Cross-Lingual Information Retrieval*
Ferhan Ture and Elizabeth Boschee
- Semantic-Based Multilingual Document Clustering via Tensor Modeling*
Salvatore Romeo, Andrea Tagarelli and Dino Ienco
- Lexical Substitution for the Medical Domain*
Martin Riedl, Michael Glass and Alfio Gliozzo
- Question Answering with Subgraph Embeddings*
Antoine Bordes, Sumit Chopra and Jason Weston
- Correcting Keyboard Layout Errors and Homoglyphs in Queries*
Derek Barnes, Mahesh Joshi and Hassan Sawaf
- Non-linear Mapping for Improved Identification of 1300+ Languages*
Ralf Brown

Sunday, October 26, 2014 (continued)

A Neural Network for Factoid Question Answering over Paragraphs

Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher and Hal Daumé III

Joint Relational Embeddings for Knowledge-based Question Answering

Min-Chul Yang, Nan Duan, Ming Zhou and Hae-Chang Rim

Adding High-Precision Links to Wikipedia

Thanapon Noraset, Chandra Bhagavatula and Doug Downey

Crosslingual and Multilingual Construction of Syntax-Based Vector Space Models

Jason Utt and Sebastian Padó

Finding Good Enough: A Task-Based Evaluation of Query Biased Summarization for Cross-Language Information Retrieval

Jennifer Williams, Sharon Tam and Wade Shen

Chinese Poetry Generation with Recurrent Neural Networks

Xingxing Zhang and Mirella Lapata

Fear the REAPER: A System for Automatic Multi-Document Summarization with Reinforcement Learning

Cody Rioux, Sadid A. Hasan and Yllias Chali

Improving Multi-documents Summarization by Sentence Compression based on Expanded Constituent Parse Trees

Chen Li, Yang Liu, Fei Liu, Lin Zhao and Fuliang Weng

Analyzing Stemming Approaches for Turkish Multi-Document Summarization

Muhammed Yavuz Nuzumlalı and Arzucan Özgür

Monday, October 27, 2014

08:00–17:00 *Registration*

08:00–09:00 *Refreshments*

Plenary Session

09:00–10:00 *Invited Talk: Learning from Rational Behavior*
Thorsten Joachims

10:00–10:30 *Coffee Break*

Session 4a: Neural Net Mixer II

10:30–10:55 *Evaluating Neural Word Representations in Tensor-Based Compositional Settings*
Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh and Matthew Purver

10:55–11:20 *Opinion Mining with Deep Recurrent Neural Networks*
Ozan Irsoy and Claire Cardie

11:20–11:45 *The Inside-Outside Recursive Neural Network model for Dependency Parsing*
Phong Le and Willem Zuidema

11:45–12:10 *A Fast and Accurate Dependency Parser using Neural Networks*
Danqi Chen and Christopher Manning

Monday, October 27, 2014 (continued)

Session 4b: Discourse, Dialogue and Pragmatics / Summarization and Generation

- 10:30–10:55 *Why are You Taking this Stance? Identifying and Classifying Reasons in Ideological Debates*
Kazi Saidul Hasan and Vincent Ng
- 10:55–11:20 *Chinese Zero Pronoun Resolution: An Unsupervised Probabilistic Model Rivaling Supervised Resolvers*
Chen Chen and Vincent Ng
- 11:20–11:45 *Unsupervised Sentence Enhancement for Automatic Summarization*
Jackie Chi Kit Cheung and Gerald Penn
- 11:45–12:10 *ReferItGame: Referring to Objects in Photographs of Natural Scenes*
Sahar Kazemzadeh, Vicente Ordonez, Mark Matten and Tamara Berg

Session 4c: Information Extraction

- 10:30–10:55 *Unsupervised Template Mining for Semantic Category Understanding*
Lei Shi, Shuming Shi, Chin-Yew Lin, Yi-Dong Shen and Yong Rui
- 10:55–11:20 *Taxonomy Construction Using Syntactic Contextual Evidence*
Tuan Luu Anh, Jung-jae Kim and See Kiong Ng
- 11:20–11:45 *Analysing recall loss in named entity slot filling*
Glen Pink, Joel Nothman and James R. Curran
- 11:45–12:10 *Relieving the Computational Bottleneck: Joint Inference for Event Extraction with High-Dimensional Features*
Deepak Venugopal, Chen Chen, Vibhav Gogate and Vincent Ng

Monday, October 27, 2014 (continued)

Session 4p: Segmentation, Tagging and Parsing

10:30–12:10 *Poster Session*
Multiple presenters

Syllable weight encodes mostly the same information for English word segmentation as dictionary stress

John K Pate and Mark Johnson

A Joint Model for Unsupervised Chinese Word Segmentation

Miaohong Chen, Baobao Chang and Wenzhe Pei

Domain Adaptation for CRF-based Chinese Word Segmentation using Free Annotations

Yijia Liu, Yue Zhang, Wanxiang Che, Ting Liu and Fan Wu

Balanced Korean Word Spacing with Structural SVM

Changki Lee, Edward Choi and Hyunki Kim

Morphological Segmentation for Keyword Spotting

Karthik Narasimhan, Damianos Karakos, Richard Schwartz, Stavros Tsakalidis and Regina Barzilay

What Can We Get From 1000 Tokens? A Case Study of Multilingual POS Tagging For Resource-Poor Languages

Long Duong, Trevor Cohn, Karin Verspoor, Steven Bird and Paul Cook

An Experimental Comparison of Active Learning Strategies for Partially Labeled Sequences

Diego Marcheggiani and Thierry Artières

Language Modeling with Functional Head Constraint for Code Switching Speech Recognition

Ying LI and Pascale Fung

A Polynomial-Time Dynamic Oracle for Non-Projective Dependency Parsing

Carlos Gómez-Rodríguez, Francesco Sartorio and Giorgio Satta

Ambiguity Resolution for Vt-N Structures in Chinese

Yu-Ming Hsieh, Jason S. Chang and Keh-Jiann Chen

Neural Networks Leverage Corpus-wide Information for Part-of-speech Tagging

Yuta Tsuboi

Monday, October 27, 2014 (continued)

System Combination for Grammatical Error Correction

Raymond Hendy Susanto, Peter Phandi and Hwee Tou Ng

Dependency parsing with latent refinements of part-of-speech tags

Thomas Mueller, Richárd Farkas, Alex Judea, Helmut Schmid and hinrich schuetze

Importance weighting and unsupervised domain adaptation of POS taggers: a negative result

Barbara Plank, Anders Johannsen and Anders Søgaard

POS Tagging of English-Hindi Code-Mixed Social Media Content

Yogarshi Vyas, Spandana Gella, Jatin Sharma, Kalika Bali and Monojit Choudhury

Data Driven Grammatical Error Detection in Transcripts of Children's Speech

Eric Morley, Anna Eva Hallin and Brian Roark

12:10–13:30 *Lunch Break*

12:50–13:30 *SIGDAT Business Meeting*

Session 5a: Tagging, Chunking, Parsing and Syntax

13:30–13:55 *Improved CCG Parsing with Semi-supervised Supertagging*

Mike Lewis and Mark Steedman

13:55–14:20 *A* CCG Parsing with a Supertag-factored Model*

Mike Lewis and Mark Steedman

14:20–14:45 *A Dependency Parser for Tweets*

Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archana Bhatia, Chris Dyer and Noah A. Smith

14:45–15:10 *Greed is Good if Randomized: New Inference for Dependency Parsing*

Yuan Zhang, Tao Lei, Regina Barzilay and Tommi Jaakkola

Monday, October 27, 2014 (continued)

Session 5b: Semantics

- 13:30–13:55 *A Unified Model for Word Sense Representation and Disambiguation*
Xinxiong Chen, Zhiyuan Liu and Maosong Sun
- 13:55–14:20 *Reducing Dimensions of Tensors in Type-Driven Distributional Semantics*
Tamara Polajnar, Luana Fagarasan and Stephen Clark
- 14:20–14:45 *An Etymological Approach to Cross-Language Orthographic Similarity. Application on Romanian*
Alina Maria Ciobanu and Liviu P. Dinu
- 14:45–15:10 *Efficient Non-parametric Estimation of Multiple Embeddings per Word in Vector Space*
Arvind Neelakantan, Jeevan Shankar, Alexandre Passos and Andrew McCallum

Session 5c: Information Retrieval and Question Answering

- 13:30–13:55 *Tailor knowledge graph for query understanding: linking intent topics by propagation*
Shi Zhao and Yan Zhang
- 13:55–14:20 *Queries as a Source of Lexicalized Commonsense Knowledge*
Marius Pasca
- 14:20–14:45 *Question Answering over Linked Data Using First-order Logic*
Shizhu He, Kang Liu, Yuanzhe Zhang, Liheng Xu and Jun Zhao
- 14:45–15:10 *Knowledge Graph and Corpus Driven Segmentation and Answer Inference for Tele-graphic Entity-seeking Queries*
Mandar Joshi, Uma Sawant and Soumen Chakrabarti

Monday, October 27, 2014 (continued)

Session 5p: NLP for the Web, Social Media and Sentiment Analysis

13:30–15:10 *Poster Session*
Multiple presenters

A Regularized Competition Model for Question Difficulty Estimation in Community Question Answering Services

Quan Wang, Jing Liu, Bin Wang and Li Guo

Vote Prediction on Comments in Social Polls

Isaac Persing and Vincent Ng

Exploiting Social Relations and Sentiment for Stock Prediction

Jianfeng Si, Arjun Mukherjee, Bing Liu, Sinno Jialin Pan, Qing Li and Huayi Li

Developing Age and Gender Predictive Lexica over Social Media

Maarten Sap, Gregory Park, Johannes Eichstaedt, Margaret Kern, David Stillwell, Michal Kosinski, Lyle Ungar and Hansen Andrew Schwartz

Dependency Parsing for Weibo: An Efficient Probabilistic Logic Programming Approach

William Yang Wang, Lingpeng Kong, Kathryn Mazaitis and William W Cohen

Exploiting Community Emotion for Microblog Event Detection

Gaoyan Ou, Wei Chen, Tengjiao Wang, Zhongyu Wei, Binyang LI, Dongqing Yang and Kam-Fai Wong

Detecting Disagreement in Conversations using Pseudo-Monologic Rhetorical Structure

Kelsey Allen, Giuseppe Carenini and Raymond Ng

+/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference

Yoonjung Choi and Janyce Wiebe

A Sentiment-aligned Topic Model for Product Aspect Rating Prediction

Hao Wang and Martin Ester

Learning Emotion Indicators from Tweets: Hashtags, Hashtag Patterns, and Phrases

Ashequl Qadir and Ellen Riloff

Fine-Grained Contextual Predictions for Hard Sentiment Words

Sebastian Ebert and Hinrich Schütze

Monday, October 27, 2014 (continued)

An Iterative Link-based Method for Parallel Web Page Mining

Le Liu, Yu Hong, Jun Lu, Jun Lang, Heng Ji and Jianmin Yao

Exploiting Social Network Structure for Person-to-Person Sentiment Analysis

Robert West, Hristo Paskov, Jure Leskovec, Christopher Potts

15:10–15:40 *Coffee Break*

Session 6a: Machine Translation

15:40–16:05 *Human Effort and Machine Learnability in Computer Aided Translation*

Spence Green, Sida I. Wang, Jason Chuang, Jeffrey Heer, Sebastian Schuster and Christopher D. Manning

16:05–16:30 *Exact Decoding for Phrase-Based Statistical Machine Translation*

Wilker Aziz, Marc Dymetman and Lucia Specia

16:30–16:55 *Large-scale Expected BLEU Training of Phrase-based Reordering Models*

Michael Auli, Michel Galley and Jianfeng Gao

16:55–17:20 *Confidence-based Rewriting of Machine Translation Output*

Benjamin Marie and Aurélien Max

Session 6b: Semantic Parsing

15:40–16:05 *Learning Compact Lexicons for CCG Semantic Parsing*

Yoav Artzi, Dipanjan Das and Slav Petrov

16:05–16:30 *Morpho-syntactic Lexical Generalization for CCG Semantic Parsing*

Adrienne Wang, Tom Kwiatkowski and Luke Zettlemoyer

16:30–16:55 *Semantic Parsing Using Content and Context: A Case Study from Requirements Elicitation*

Reut Tsarfaty, Ilia Pogrebezky, Guy Weiss, Yaarit Natan, Smadar Szekely and David Harel

16:55–17:20 *Semantic Parsing with Relaxed Hybrid Trees*

Wei Lu

Monday, October 27, 2014 (continued)

Session 6c: NLP-Related Machine Learning

- 15:40–16:05 *Low-dimensional Embeddings for Interpretable Anchor-based Topic Inference*
David Mimno and Moontae Lee
- 16:05–16:30 *Weakly-Supervised Learning with Cost-Augmented Contrastive Estimation*
Kevin Gimpel and Mohit Bansal
- 16:30–16:55 *Don't Until the Final Verb Wait: Reinforcement Learning for Simultaneous Machine Translation*
Alvin Grissom II, He He, Jordan Boyd-Graber, John Morgan and Hal Daumé III
- 16:55–17:20 *PCFG Induction for Unsupervised Parsing and Language Modelling*
James Scicluna and Colin de la Higuera

Session 6p: Computational Psycholinguistics, Text Mining and NLP Applications

- 15:40–17:20 *Poster Session*
Multiple presenters
- Can characters reveal your native language? A language-independent approach to native language identification*
Radu Tudor Ionescu, Marius Popescu and Aoife Cahill
- Formalizing Word Sampling for Vocabulary Prediction as Graph-based Active Learning*
Yo Ehara, Yusuke Miyao, Hidekazu Oiwa, Issei Sato and Hiroshi Nakagawa
- Language Transfer Hypotheses with Linear SVM Weights*
Shervin Malmasi and Mark Dras
- Predicting Dialect Variation in Immigrant Contexts Using Light Verb Constructions*
A. Seza Dogruoz and Preslav Nakov
- Device-Dependent Readability for Improved Text Understanding*
A-Yeong Kim, Hyun-Je Song, Seong-Bae Park and Sang-Jo Lee
- Predicting Chinese Abbreviations with Minimum Semantic Unit and Global Constraints*
Longkai Zhang, li li, Houfeng WANG and Xu Sun

Monday, October 27, 2014 (continued)

Using Structured Events to Predict Stock Price Movement: An Empirical Investigation

Xiao Ding, Yue Zhang, Ting Liu and Junwen Duan

Extracting Clusters of Specialist Terms from Unstructured Text

Aaron Gerow

Citation-Enhanced Keyphrase Extraction from Research Papers: A Supervised Approach

Cornelia Caragea, Florin Adrian Bulgarov, Andreea Godea and Sujatha Das Gollapalli

Using Mined Coreference Chains as a Resource for a Semantic Task

Heike Adel and Hinrich Schütze

Financial Keyword Expansion via Continuous Word Vector Representations

Ming-Feng Tsai and Chuan-Ju Wang

Intrinsic Plagiarism Detection using N-gram Classes

Imene Bensalem, Paolo Rosso and Salim Chikhi

Verifiably Effective Arabic Dialect Identification

Kareem Darwish, Hassan Sajjad and Hamdy Mubarak

Keystroke Patterns as Prosody in Digital Writings: A Case Study with Deceptive Reviews and Essays

Ritwik Banerjee, Song Feng, Jun Seok Kang and Yejin Choi

Leveraging Effective Query Modeling Techniques for Speech Recognition and Summarization

Kuan-Yu Chen, Shih-Hung Liu, Berlin Chen, Ea-Ee Jan, Hsin-Min Wang, Wen-Lian Hsu and Hsin-Hsi Chen

Staying on Topic: An Indicator of Power in Political Debates

Vinodkumar Prabhakaran, Ashima Arora and Owen Rambow

Tuesday, October 28, 2014

08:00–17:00 *Registration*

08:00–09:00 *Refreshments*

Plenary Session

09:00–09:05 *Best Paper Award*
Bo Pang and Walter Daelemans

09:05–09:30 *Language Modeling with Power Low Rank Ensembles*
Ankur P. Parikh, Avneesh Saluja, Chris Dyer and Eric Xing

09:30–09:55 *Modeling Biological Processes for Reading Comprehension*
Jonathan Berant, Vivek Srikumar, Pei-Chun Chen, Abby Vander Linden, Brittany Harding, Brad Huang and Christopher D. Manning

10:00–10:30 *Coffee Break*

Session 7a: Semantics

10:30–10:55 *Sensicon: An Automatically Constructed Sensorial Lexicon*
Serra Sinem Tekiroglu, Gözde Özbal and Carlo Strapparava

10:55–11:20 *Word Semantic Representations using Bayesian Probabilistic Tensor Factorization*
Jingwei Zhang, Jeremy Salwen, Michael Glass and Alfio Gliozzo

11:20–11:45 *Glove: Global Vectors for Word Representation*
Jeffrey Pennington, Richard Socher and Christopher Manning

11:45–12:10 *Jointly Learning Word Representations and Composition Functions Using Predicate-Argument Structures*
Kazuma Hashimoto, Pontus Stenetorp, Makoto Miwa and Yoshimasa Tsuruoka

Tuesday, October 28, 2014 (continued)

Session 7b: Information Extraction

- 10:30–10:55 *Combining Distant and Partial Supervision for Relation Extraction*
Gabor Angeli, Julie Tibshirani, Jean Wu and Christopher D. Manning
- 10:55–11:20 *Typed Tensor Decomposition of Knowledge Bases for Relation Extraction*
Kai-Wei Chang, Wen-tau Yih, Bishan Yang and Christopher Meek
- 11:20–11:45 *A convex relaxation for weakly supervised relation extraction*
Edouard Grave
- 11:45–12:10 *Knowledge Graph and Text Jointly Embedding*
Zhen Wang, Jianwen Zhang, Jianlin Feng and Zheng Chen

Session 7c: Sentiment Analysis and NLP Applications

- 10:30–10:55 *Abstractive Summarization of Product Reviews Using Discourse Structure*
Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng and Bitan Nejat
- 10:55–11:20 *Clustering Aspect-related Phrases by Leveraging Sentiment Distribution Consistency*
Li Zhao, Minlie Huang, Haiqiang Chen, Junjun Cheng and Xiaoyan Zhu
- 11:20–11:45 *Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach*
Yue Hu and Xiaojun Wan
- 11:45–12:10 *Fast and Accurate Misspelling Correction in Large Corpora*
Octavian Popescu and Ngoc Phuoc An Vo

Tuesday, October 28, 2014 (continued)

Session 7p: Machine Translation and Machine Learning

10:30–12:10 *Poster Session*
Multiple presenters

Assessing the Impact of Translation Errors on Machine Translation Quality with Mixed-effects Models

Marcello Federico, Matteo Negri, Luisa Bentivogli and Marco Turchi

Refining Word Segmentation Using a Manually Aligned Corpus for Statistical Machine Translation

Xiaolin Wang, Masao Utiyama, Andrew Finch and Eiichiro Sumita

Improving Pivot-Based Statistical Machine Translation by Pivoting the Co-occurrence Count of Phrase Pairs

Xiaoning Zhu, Zhongjun He, Hua Wu, Conghui Zhu, Haifeng Wang and Tiejun Zhao

Word Translation Prediction for Morphologically Rich Languages with Bilingual Neural Networks

Ke M. Tran, Arianna Bisazza and Christof Monz

Dependency-Based Bilingual Language Models for Reordering in Statistical Machine Translation

Ekaterina Garmash and Christof Monz

Combining String and Context Similarity for Bilingual Term Alignment from Comparable Corpora

Georgios Kontonatsios, Ioannis Korkontzelos, Jun'ichi Tsujii and Sophia Ananiadou

Random Manhattan Integer Indexing: Incremental L1 Normed Vector Space Construction

Behrang Q. Zadeh and Siegfried Handschuh

Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk and Yoshua Bengio

Type-based MCMC for Sampling Tree Fragments from Forests

Xiaochang Peng and Daniel Gildea

Convolutional Neural Networks for Sentence Classification

Yoon Kim

Sometimes Average is Best: The Importance of Averaging for Prediction using MCMC Inference in Topic Modeling

Viet-An Nguyen, Jordan Boyd-Graber and Philip Resnik

Tuesday, October 28, 2014 (continued)

Large-scale Reordering Model for Statistical Machine Translation using Dual Multinomial Logistic Regression

Abdullah Alrajeh and Mahesan Niranjan

Dynamic Language Models for Streaming Text

Dani Yogatama, Chong Wang, Bryan Routledge, Noah A. Smith, Eric P. Xing

Improved Decipherment of Homophonic Ciphers

Malte Nuhn, Julian Schamper and Hermann Ney

Cipher Type Detection

Malte Nuhn and Kevin Knight

12:10–13:30 Lunch Break

Session 8sa: Segmentation and Tagging / Spoken Language / Semantics

13:30–13:50 *Joint Learning of Chinese Words, Terms and Keywords*

Ziqiang Cao, Sujian Li and Heng Ji

13:50–14:10 *Cross-Lingual Part-of-Speech Tagging through Ambiguous Learning*

Guillaume Wisniewski, Nicolas Pécheux, Souhir Gahbiche-Braham and François Yvon

14:10–14:30 *Comparing Representations of Semantic Roles for String-To-Tree Decoding*

Marzieh Bazrafshan and Daniel Gildea

14:30–14:50 *Detecting Non-compositional MWE Components using Wiktionary*

Bahar Salehi, Paul Cook and Timothy Baldwin

14:50–15:10 *(Empty slot)*

(No presentation)

Tuesday, October 28, 2014 (continued)

Session 8sb: Sentiment Analysis / Social / Computational Psycholinguistics / Text Classification

- 13:30–13:50 *Joint Emotion Analysis via Multi-task Gaussian Processes*
Daniel Beck, Trevor Cohn and Lucia Specia
- 13:50–14:10 *Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics*
Zubin Jelveh, Bruce Kogut and Suresh Naidu
- 14:10–14:30 *A Model of Individual Differences in Gaze Control During Reading*
Niels Landwehr, Sebastian Arzt, Tobias Scheffer and Reinhold Kliegl
- 14:30–14:50 *Multi-label Text Categorization with Hidden Components*
li li, Longkai Zhang and Houfeng WANG
- 14:50–15:10 *#TagSpace: Semantic Embeddings from Hashtags*
Jason Weston, Sumit Chopra and Keith Adams

Session 8sc: Summarization / Machine Translation / Information Extraction

- 13:30–13:50 *Joint Decoding of Tree Transduction Models for Sentence Compression*
Jin-ge Yao, Xiaojun Wan and Jianguo Xiao
- 13:50–14:10 *Dependency-based Discourse Parser for Single-Document Summarization*
Yasuhisa Yoshida, Jun Suzuki, Tsutomu Hirao and Masaaki Nagata
- 14:10–14:30 *Improving Word Alignment using Word Similarity*
Theerawat Songyot and David Chiang
- 14:30–14:50 *Constructing Information Networks Using One Single Model*
Qi Li, Heng Ji, Yu HONG and Sujian Li
- 14:50–15:10 *Event Role Extraction using Domain-Relevant Word Representations*
Emanuela Boros, Romaric Besançon, Olivier Ferret and Brigitte Grau

Tuesday, October 28, 2014 (continued)

Session 8p: Information Extraction

13:30–15:10 *Poster Session*
Multiple presenters

Modeling Joint Entity and Relation Extraction with Table Representation
Makoto Miwa and Yutaka Sasaki

ZORE: A Syntax-based System for Chinese Open Relation Extraction
Likun Qiu and Yue Zhang

Coarse-grained Candidate Generation and Fine-grained Re-ranking for Chinese Abbreviation Prediction
Longkai Zhang, Houfeng WANG and Xu Sun

Type-Aware Distantly Supervised Relation Extraction with Linked Arguments
Mitchell Koch, John Gilmer, Stephen Soderland and Daniel S. Weld

Automatic Inference of the Tense of Chinese Events Using Implicit Linguistic Information
Yuchen Zhang and Nianwen Xue

Joint Inference for Knowledge Base Population
Liwei Chen, Yansong Feng, Jinghui Mo, Songfang Huang and Dongyan Zhao

Combining Visual and Textual Features for Information Extraction from Online Flyers
Emilia Apostolova and Noriko Tomuro

CTPs: Contextual Temporal Profiles for Time Scoping Facts using State Change Detection
Derry Tanti Wijaya, Ndapandula Nakashole and Tom M. Mitchell

Noisy Or-based model for Relation Extraction using Distant Supervision
Ajay Nagesh, Gholamreza Haffari and Ganesh Ramakrishnan

15:10–15:40 *Coffee Break*

Tuesday, October 28, 2014 (continued)

Session 9a: Machine Learning and Machine Translation

- 15:40–16:05 *Search-Aware Tuning for Machine Translation*
Lemao Liu and Liang Huang
- 16:05–16:30 *Latent-Variable Synchronous CFGs for Hierarchical Translation*
Avneesh Saluja, Chris Dyer and Shay B. Cohen
- 16:30–16:55 *Dynamically Shaping the Reordering Search Space of Phrase-Based Statistical Machine Translation*
Arianna Bisazza and Marcello Federico
- 16:55–17:20 *(Empty slot)*
(No presentation)

Session 9b: NLP for the Web and Social Media

- 15:40–16:05 *Gender and Power: How Gender and Gender Environment Affect Manifestations of Power*
Vinodkumar Prabhakaran, Emily E. Reid and Owen Rambow
- 16:05–16:30 *Online topic model for Twitter considering dynamics of user interests and topic trends*
Kentaro Sasaki, Tomohiro Yoshikawa and Takeshi Furuhashi
- 16:30–16:55 *Self-disclosure topic model for classifying and analyzing Twitter conversations*
JinYeong Bak, Chin-Yew Lin and Alice Oh
- 16:55–17:20 *Major Life Event Extraction from Twitter based on Congratulations/Condolences Speech Acts*
Jiwei Li, Alan Ritter, Claire Cardie and Eduard Hovy

Tuesday, October 28, 2014 (continued)

Session 9c: Semantics

- 15:40–16:05 *Brighter than Gold: Figurative Language in User Generated Comparisons*
Vlad Niculae and Cristian Danescu-Niculescu-Mizil
- 16:05–16:30 *Classifying Idiomatic and Literal Expressions Using Topic Models and Intensity of Emotions*
Jing Peng, Anna Feldman and Ekaterina Vylomova
- 16:30–16:55 *TREETALK: Composition and Compression of Trees for Image Descriptions*
Polina Kuznetsova, Vicente Ordonez, Tamara Berg, Yejin Choi
- 16:55–17:20 *Learning Spatial Knowledge for Text to 3D Scene Generation*
Angel Chang, Manolis Savva and Christopher D. Manning

Session 9p: Discourse, Dialogue and Pragmatics

- 15:40–17:20 *Poster Session*
Multiple presenters
- A Model of Coherence Based on Distributed Sentence Representation*
Jiwei Li and Eduard Hovy
- Discriminative Reranking of Discourse Parses Using Tree Kernels*
Shafiq Joty and Alessandro Moschitti
- Recursive Deep Models for Discourse Parsing*
Jiwei Li, Rumeng Li and Eduard Hovy
- Recall Error Analysis for Coreference Resolution*
Sebastian Martschat and Michael Strube
- A Rule-Based System for Unrestricted Bridging Resolution: Recognizing Bridging Anaphora and Finding Links to Antecedents*
Yufang Hou, Katja Markert and Michael Strube
- Resolving Referring Expressions in Conversational Dialogs for Natural User Interfaces*
Asli Celikyilmaz, Zhaleh Feizollahi, Dilek Hakkani-Tur and Ruhi Sarikaya

Tuesday, October 28, 2014 (continued)

Building Chinese Discourse Corpus with Connective-driven Dependency Tree Structure

Yancui Li, wenhe feng, jing sun, Fang Kong and Guodong Zhou

Prune-and-Score: Learning for Greedy Coreference Resolution

Chao Ma, Janardhan Rao Doppa, J. Walker Orr, Prashanth Mannem, Xiaoli Fern, Tom Dietterich and Prasad Tadepalli

Summarizing Online Forum Discussions – Can Dialog Acts of Individual Messages Help?

Sumit Bhatia, Prakhar Biyani and Prasenjit Mitra

Closing Session

17:20–17:40 *Final Thanks and EMNLP 2015 Preview*
Alessandro Moschitti