

ACL-IJCNLP 2009

**Joint Conference of the
47th Annual Meeting of the
Association for Computational Linguistics
and
4th International Joint Conference on
Natural Language Processing
of the AFNLP**

Proceedings of the Conference

2-7 August 2009
Suntec, Singapore

Production and Manufacturing by
World Scientific Publishing Co Pte Ltd
5 Toh Tuck Link
Singapore 596224

©2009 The Association for Computational Linguistics
and The Asian Federation of Natural Language Processing

Order copies of this and other ACL proceedings from:

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

ISBN 978-1-932432-45-9 / 1-932432-45-0 (Volume 1)
ISBN 978-1-932432-46-6 / 1-932432-46-9 (Volume 2)

Preface: General Chair

Welcome to the ACL-IJCNLP 2009, the first joint conference sponsored by the ACL (Association for Computational Linguistics) and the AFNLP (Asian Federation of Natural Language Processing). The idea to have a joint conference for ACL and AFNLP was first discussed at ACL-05 (Ann Arbor, Michigan) among Martha Palmer (ACL President), Benjamin T'sou (AFNLP President), Jun'ichi Tsujii (AFNLP Vice President) and Keh-Yih Su (AFNLP Conference Coordinating Committee Chair, also the Secretary General). We are glad that the original idea has come true four years later, and even the affiliation relationship between these two organizations has been built up now.

In this joint conference, we have tried to mix the spirit from both ACL and AFNLP; and, Singapore, which itself has a mixture of diversified cultures from eastern and western regions, is certainly a wonderful place to see how different languages meet each other. We hope you will enjoy this big event held in this garden city, which is brought to you via the efforts from each member of the conference organization team.

Among our hard working organizers, I would like to thank the Program Chairs, Jan Wiebe and Jian Su, who has carefully selected papers from our record high submissions, and the Local Arrangements Chair, Haizhou Li, who has shown his excellent capability in smoothly organizing various events and details. My thanks will also go to other chairs for their competent and hard work: The Webmaster, Minghui Dong; the Demo Chairs, Gary Geunbae Lee and Sabine Schulte im Walde; the Exhibits Chairs, Timothy Baldwin and Philipp Koehn; the Mentoring Service Chairs, Hwee Tou Ng and Florence Reeder; the Publication Chairs, Jing-Shin Chang and Regina Barzilay; the Publicity Chairs, Min-Yen Kan and Andy Way; the Sponsorship Chairs, Hitoshi Isahara and Kim-Teng Lua; the Student Research Workshop Chairs, Davis Dimalen, Jenny Rose Finkel, and Blaise Thomson; also the Faculty Advisors, Grace Ngai and Brian Roark; the Tutorial Chairs, Diana McCarthy and Chengqing Zong; the Workshop Chairs, Jimmy Lin and Yuji Matsumoto; last, the ACL Business Manager, Priscilla Rasmussen, who not only provides useful advice but also helps to contact more sponsors and get their support.

Besides, I need to express my gratitude to the Conference Coordination Committee for their valuable advice and support: in which Bonnie Dorr (chair), Steven Bird, Graeme Hirst, Kathleen McCoy, Martha Palmer, Dragomir Radev, Priscilla Rasmussen, Mark Steedman are from ACL; and Yuji Matsumoto, Keh-Yih Su, Jun'ichi Tsujii, Benjamin T'sou, Kam-Fai Wong are from AFNLP.

Last, I sincerely thank all the authors, reviewers, presenters, invited speakers, sponsors, exhibitors, local supporting staff, and all the conference attendants. It is you that make this conference possible. Wish you all enjoy the program that we provide.

Keh-Yih Su
ACL-IJCNLP 2009 General Chair
August 2009

Preface: Program Committee Co-Chairs

For the first time, the flagship conferences of the Association for Computational Linguistics (ACL) and the Asian Federation of Natural Language Processing (AFNLP) – the ACL and IJCNLP – are jointly organized as a single event. ACL-IJCNLP 2009 covers a broad spectrum of technical areas related to natural language and computation, representing a rich array of the state of the art. The conference includes full papers, short papers, demonstrations, a student research workshop, as well as pre- and post-conference tutorials and workshops.

This year, we again received a record number of submissions: 925 total valid paper submissions, a 24% increase over ACL-08: HLT. This includes 569 full-paper submissions and 356 short-paper submissions from more than 40 countries – approximately 51% from 20 countries in Asia Pacific, 27% from Canada, Cuba and the United States, 22% from 15 countries in Europe, fewer than 1% from Argentina, and one paper submitted anonymously. We thank all of the authors for submitting papers describing their recent work. The significant submission increase is a trend extending over multiple years, and shows how vigorous our field is. We also thank Hwee Tou Ng and Florence Reeder, the Mentoring Service Co-Chairs, for organizing a 19-mentor team who provided English scientific paper writing support.

20 Area Chairs worked with 489 Program Committee members and 85 additional reviewers to come up with 2551 reviews, in total, for the final paper selection. 21% of the full-paper submissions were accepted; all will be presented orally. 26% of the short-paper submissions were accepted; some will be presented orally and some as poster presentations. While short papers are distinguished from full papers in the proceedings, there are no distinctions in the proceedings between short papers presented orally and those presented as posters. We are absolutely indebted to the Area Chairs, Program Committee members, and additional reviewers for their intensive efforts.

We are delighted to have two keynote speakers: Qiang Yang, who will talk about heterogeneous transfer learning, and Bonnie Webber, who will address discourse and genre. Best (student) paper awards and the ACL Lifetime Achievement Award will be announced in the last session of the conference as well.

We thank General Conference Chair Keh Yih Su, the Local Arrangements Committee headed by Haizhou Li, and the ACL-AFNLP Conference Coordination Committee chaired by Bonnie Dorr, for their help and advice, as well as last years PC Co-Chairs, Johanna Moore and Simone Teufel, for sharing their experiences, Jason Eisner for his How to Serve as Program Chair of a Conference website and corresponding emails, Jing-Shin Chang and Regina Barzilay, the Publication Co-Chairs for putting the proceedings together, and all the other committee chairs for their work. Our thanks go to our assistant Chen Bin, who worked tirelessly throughout the entire process, and who made our work with START much easier. Together, everyone made such a wonderful event possible.

We hope that you enjoy the conference!

Jian Su, Institute for Infocomm Research
Jan Wiebe, University of Pittsburgh

Organizing Committee

General Conference Chair:

Su, Keh-Yih (Behavior Design Corp., Taiwan; kysu@bdc.com.tw)

Program Chairs:

Su, Jian (Institute for Infocomm Research, Singapore; sujian@i2r.a-star.edu.sg)

Wiebe, Janyce (University of Pittsburgh, USA; janycewiebe@gmail.com)

Local Organizing Chair:

Li, Haizhou (Institute for Infocomm Research, Singapore; hli@i2r.a-star.edu.sg)

Demo Chairs:

Lee, Gary Geunbae (POSTECH, Korea; gblee@postech.ac.kr)

Schulte im Walde, Sabine (University of Stuttgart, Germany; schulte@ims.uni-stuttgart.de)

Exhibits Chairs:

Baldwin, Timothy (University of Melbourne, Australia; tim@csse.unimelb.edu.au)

Koehn, Philipp (University of Edinburgh, UK; pkoehn@inf.ed.ac.uk)

Mentoring Service Chairs:

Ng, Hwee Tou (National University of Singapore, Singapore; nght@comp.nus.edu.sg)

Reeder, Florence (Mitre, USA; freeder@mitre.org)

Publication Chairs:

Barzilay, Regina (MIT, USA; regina@csail.mit.edu)

Chang, Jing-Shin (National Chi Nan University, Taiwan; jshin@csie.ncnu.edu.tw)

Publicity Chairs:

Kan, Min-Yen (National University of Singapore, Singapore; kanmy@comp.nus.edu.sg)

Way, Andy (Dublin City University, Ireland; away@computing.dcu.ie)

Sponsorship Chairs:

Americas Sponsorship Co-Chairs:

Bangalore, Srinivas

Doran, Christine

European Sponsorship Co-Chairs:

Koehn, Philipp

Genabith, Josef van

Asia Sponsorship Co-Chairs:

Isahara, Hitoshi (NICT, Japan; isahara@nict.go.jp)

Lua, Kim-Teng (COLIPS, Singapore; luakt@colips.org)

Student Chairs:

Dimalen, Davis (Academia Sinica, Taiwan; d.dimalen@yahoo.com)
Finkel, Jenny Rose (Stanford University, USA; jrfinkel@cs.stanford.edu)
Thomson, Blaise (Cambridge University, UK; brmt2@cam.ac.uk)

Student Workshop Faculty Advisors:

Ngai, Grace (Polytechnic University, Hong Kong; csgngai@polyu.edu.hk)
Roark, Brian (Oregon Health & Science University, USA; roark@cslu.ogi.edu)

Tutorial Chairs:

McCarthy, Diana (University of Sussex, UK; dianam@sussex.ac.uk)
Zong, Chengqing (Chinese Academy of Sciences, China; cqzong@gmail.com)

Workshop Chairs:

Lin, Jimmy (University of Maryland, USA; jimmylin@umd.edu)
Matsumoto, Yuji (NAIST, Japan; matsu@is.naist.jp)

Webmaster:

Dong, Minghui (Institute for Infocomm Research, Singapore; mhdong@i2r.a-star.edu.sg)

Registration:

Rasmussen, Priscilla (ACL; rasmusse@ptd.net)

Program Committee

Program Chairs:

Su, Jian (Institute for Infocomm Research, Singapore; sujian@i2r.a-star.edu.sg)
Wiebe, Janyce (University of Pittsburgh, USA; janycewiebe@gmail.com)

Area Chairs:

Agirre, Eneko (University of Basque Country, Spain; e.agirre@ehu.es)
Ananiadou, Sophia (University of Manchester, UK; sophia.ananiadou@manchester.ac.uk)
Belz, Anja (University of Brighton, UK; a.s.belz@itri.brighton.ac.uk)
Carenini, Giuseppe (University of British Columbia, Canada; carenini@cs.ubc.ca)
Chen, Hsin-Hsi (National Taiwan University, Taiwan, hh_chen@csie.ntu.edu.tw)
Chen, Keh-Jiann (Sinica, Taiwan, kchen@iis.sinica.edu.tw)
Curran, James (University of Sydney, Australia; james@it.usyd.edu.au)
Gao, Jian Feng (MSR, USA; jfgao@microsoft.com)
Harabagiu, Sanda (University of Texas at Dallas, USA, sanda@hlt.utdallas.edu)
Koehn, Philipp (University of Edinburgh, UK; pkoeHN@inf.ed.ac.uk)
Kondrak, Grzegorz (University of Alberta, Canada; kondrak@cs.ualberta.ca)
Meng, Helen Mei-Ling (Chinese University of Hong Kong, HK; hmmeng@se.cuhk.edu.hk)
Mihalcea, Rada (University of North Texas, USA; rada@cs.unt.edu)
Poesio, Massimo (University of Trento, Italy; poesio@disi.unitn.it)
Riloff, Ellen (University of Utah, USA; riloff@cs.utah.edu)
Sekine, Satoshi (New York University, USA; sekine@cs.nyu.edu)
Smith, Noah (CMU, USA; nasmith@cs.cmu.edu)
Strube, Michael (EML Research, Germany; strube@eml-research.de)
Suzuki, Jun (NTT, Japan; jun@cslab.kecl.ntt.co.jp)
Wang, Hai Feng (Toshiba, China; wanghaifeng@rdc.toshiba.com.cn)

Program Committee Members:

Eugene Agichtein, Gregory Aist, Salah Ait-Mokhtar, Enrique Alfonseca, Yaser Al-Onaizan, Sophia Ananiadou, Alina Andreevskaia, Ion Androutsopoulos, Doug Appelt, Xabier Arregi, Abhishek Arun, Masayuki Asahara, Jordi Atserias, Michaela Atterer

Jason Baldridge, Srinivas Bangalore, Michele Banko, Marco Baroni, Regina Barzilay, Roberto Basili, Sabine Bergler, Shane Bergsma, Steven Bethard, Pushpak Bhattacharyya, Mikhail Bilenko, Philippe Blache, Alan Black, Sasha Blair-Goldensohn, John Blitzer, Phil Blunsom, Philip Blunsom, Bernd Bohnet, Johan Bos, Pierre Boullier, Thorsten Brants, Eric Breck, Chris Brew, Ted Briscoe, Chris Brockett, Paul Buitelaar, Razvan Bunescu, Harry Bunt, Stephan Busemann, Donna Byron

Aoife Cahill, Lynne Cahill Cahill, Nicoletta Calzolari, Nick Campbell, Claire Cardie, Giuseppe Carenini, Michael Carl, Xavier Carreras, John Carroll, Francisco Casacuberta, Jon Chamberlain, Ciprian Chelba, John Chen, Keh-Jiann Chen, Pu-Jen Cheng, Colin Cherry, David Chiang, Key-Sun Choi, Yejin Choi, Monojit Choudhury, Tat-Seng Chua, Jennifer Chu-Carroll, Philip Cimiano, Stephen Clark, James Clarke, Kevin Cohen, Shay Cohen, Trevor Cohn, Nigel Collier, John Conroy, Mark Craven, Dan Cristea, Andras Csomai, Silviu-PetrCucerzan, Hang Cui, Aron Culotta, James Curran

Robert Dale, Hal Daume, Eric de la Clergerie, Maarten de Rijke, Vera Demberg, Dina Demner, Dina Demner-Fushman, John DeNero, Li Deng, Yonggang Deng, Pascal Denis, Ann Devitt, Mona Diab, Arantza Diaz, Bill Dolan, John Dowding, Mark Dras, Mark Dredze, Jasha Droppo, Amit Dubey, Kevin Duh, Kenneth Dwyer, Chris Dyer

Phil Edmonds, Andreas Eisele, Jacob Eisenstein, Jason Eisner, Michael Elhadad, Mark Ellison, Micha Elsner, Katrin Erk, Andrea Esuli, Marc Ettliger, Stefan Evert

Ronen Feldman, Christiane Fellbaum, Raquel Fernandez, Elena Filatova, Katja Filippova, Jenny Finkel, Dan Flickinger, Radu Florian, Juliane Fluck, Eric Fosler-Lussier, George Foster, Alex Fraser, Marjorie Freedman, Carol Friedman, Qiang Fu

Evgeniy Gabrilovich, Rob Gaizauskas, Michael Gamon, Albert Gatt, Ulrich Germann, Daniel Gildea, Jesus Gimenez, Jonathan Ginzburg, Roxana Girju, Amir Globerson, Andrew Goldberg, John Goldsmith, Sharon Goldwater, Julio Gonzalo, Ralph Grishman, Rodrigo Guido, Tunga Gungor, Iryna Gurevych

Stephanie Haas, Aria Haghighi, Dilek Hakkani-Tur, Keith Hall, John Hansen, Sanda Harabagiu, Donna Harman, Saša Hasan, Samer Hassan, Timothy Hazen, Xiaodong He, Jeff Heinz, James Henderson, John Henderson, Andrew Hickl, Erhard Hinrichs, Keikichi Hirose, Julia Hirschberg, Graeme Hirst, Hieu Hoang, Julia Hockenmaier, Beth Ann Hockey, Mark Hopkins, Veronique Hoste, Churen Huang, Liang Huang, Sarmad Hussain, Rebecca Hwa, Mei-Yuh Hwang

Nancy Ide, Ryu Iida, Diana Inkpen, Kentaro Inui, Hitoshi Isahara, Abe Ittycheriah, Tatsuya Izuha

Heng Ji, Sittichai Jiampojarn, Jing Jiang, Mark Johnson, Doug Jones, Gareth Jones

Mijail Kabadjov, Laura Kallmeyer, Nanda Kambhatla, Hiroshi Kanayama, Nikiforos Karamanis, Tsuneaki Kato, Hisashi Kawai, Junichi Kazama, Frank Keller, Andre Kempe, Brett Kessler, Mitesh Khapra, Bernd Kiefer, Adam Kilgarriff, Brian Kingsbury, James Kirby, Ewan Klein, Alexandre Klementiev, Kevin Knight, Rob Koeling, Terry Koo, Moshe Koppel, Wessel Kraaij, Emiel Krahmer, Ivana Kruijff-Korabayova, Lun-Wei Ku, Sandra Kuebler, Marco Kuhlmann, Jonas Kuhn, Seth Kulick, Akira Kumano, Shankar Kumar, A. Kumaran, June-Jei Kuo, Sadao Kurohashi, Kui-Lam Kwok

Philippe Langlais, Mirella Lapata, Alberto Lavelli, Gary Lee, Lillian Lee, Yoong Keok Lee, Yue-Shi Lee, Haizhou Li, Hang Li, Xiaolong Li, Percy Liang, Elizabeth Liddy, Dekang Lin, Lucian Lita, Ken Litkowski, Diane Litman, Bing Liu, Qun Liu, Tie-Yan Liu, Yang Liu, Zhanyi Liu, Adam Lopez, Xiaoqiang Luo, Yajuan Lv

YanJun Ma, Lluís Màrquez, Karin Müller, Bernardo Magnini, Brian Mak, Rob Malouf, Gideon Mann, Daniel Marcu, Katja Markert, David Martínez, Andre Martins, Mstislav Maslennikov, Tomoko Matsui, Yuji Matsumoto, Takuya Matsuzaki, Evgeny Matusov, Arne Mauser, Diana McCarthy, David McClosky, Mark McConnell, Ryan McDonald, Michael McTear, Qiaozhu Mei, Chris Mellish, Arul Menezes, Paola Merlo, Detmar Meurers, David Mimno, Mandar Mitra, Vibhu Mittal, Yusuke Miyao, Daichi Mochihashi, Saif Mohammad, Rajat Mohanty, Diego Molla-Aliod, Christian Monson, Simonetta Montemagni, Bob Moore, Alex Morgan, Glyn Morrill, Alessandro Moschitti, Dragos Munteanu, Gabriel Murray, Sung Hyon Myaeng

Vivi Nastase, Tetsuya Nasukawa, Roberto Navigli, Mark-Jan Nederhof, Ani Nenkova, John Nerbonne, Hwee Tou Ng, Vincent Ng, Patrick Nguyen, Jian-Yun Nie, Zaiqing Nie, Takashi Ninomiya, Joakim Nivre, Chikashi Nobata, David Novick, Adrian Novischi

Franz Och, Stephan Oepen, Kemal Oflazer, Alice Oh, Jong-Hoon Oh, Daisuke Okanohara, Naoaki Okazaki, Fredrik Olsson, Constantin Orasan, Miles Osborne, Jahna Otterbacher

Tim Paek, Martha Palmer, Bo Pang, Patrick Pantel, Cecile Paris, Rebecca Passonneau, Michael Paul, Matthias Paulik, Anselmo Peñas, Adam Pease, Ted Pedersen, Catherine Pelachaud, Slav Petrov, Christopher Pinchak, Maja Popovic, Sameer Pradhan, John Prager, Rashmi Prasad, Detlef Prescher, Stephen Pulman, Sampo Pyysalo

Long Qiu, Silvia Quarteroni, Chris Quirk

Bhuvana Ramabhadran, Ganesh Ramakrishnan, Lance Ramshaw, Giuseppe Riccardi, Verena Rieser, German Rigau, Hae-Chang Rim, Brian Roark, James Rogers, Barbara Rosario, Carolyn Rose, Antti-Veikk Rosti, Patrick Ruch, Marta Ruiz, Andrey Rzhetsky

Kenji Sagae, Horacio Saggion, Benoit Sagot, Tetsuya Sakai, Baskaran Sankaran, Murat Saraclar, Ruhi Sarikaya, Yutaka Sasaki, Giorgio Satta, Anne Schiller, Helmut Schmid, Marc Schroeder, Holger Schwenk, Yohei Seki, Satoshi Sekine, Mike Seltzer, Jungyun Seo, Vijay Shanker, Hagit Shatkay, Libin Shen, Nobuyuki Shimizu, Luo Si, Advait Siddharthan, Candy Sidner, Khalil Simaan, Michel Simard, Gabriel Skantze, David Smith, Noah Smith, Rion Snow, Ian Soboroff, Swapna Somasundaran, Radu Soricut, Caroline Sporleder, Rohini Srihari, Mark Steedman, Josef Steinberger, Amanda Stent, Mark Stevenson, Veselin Stoyanov, Carlo Strapparava, Michael Strube, Tomek Strzalkowski, Jana Sukkarieh, Maoson Sun, Mihai Surdeanu, Richard Sutcliffe, Stan Szpakowicz, Idan Szpektor

Maite Taboada, John Tait, Hiroya Takamura, David Talbot, Ben Taskar, Joel Tetreault, Simone Teufel, Joerg Tiedemann, Christoph Tillmann, Ivan Titov, Roberto Togneri, Keiichi Tokuda, Kristina Toutanova, Roy Tromble, Yuen-Hsien Tseng, Jun'ichi Tsujii, Yoshimasa Tsuruoka, Dan Tufis, Gokhan Tur

Antal van den Bosch, Josef van Genabith, Keith Vander Linden, Sebastian Varges, Tony Veale, Ashish Venugopal, Paola Verlardi, Yannick Versley, David Vilar, Piek Vossen

Michael Walsh, Stephen Wan, Xiaojun Wan, Qin Wang, Shaojun Wang, Wei Wang, Xinglong Wang, Taro Watanabe, Andy Way, Bonnie Webber, David Weir, Fuiliang Weng, Janyce Wiebe, Theresa Wilson, Shuly Wintner, Yuk Wah Wong, Johan Wouters, Dekai Wu, Hua Wu

Zhuli Xie, Deyi Xiong, Jun Xu, Peng Xu, Jian Xue

Kazuhide Yamamoto, Christopher Yang, Jianwu Yang, Muyun Yang, Kaisheng Yao, Umit Yapanel, Scott Wen-tau Yih, Deniz Yuret

Fabio Zanzotto, Dmitry Zelenko, Heiga Zen, Richard Zens, Luke Zettlemoyer, Hao Zhang, Min Zhang, Rong Zhang, Tong Zhang, Yue Zhang, Tiejun Zhao, Bowen Zhou, Denny Zhou, Ming Zhou, Jerry Zhu, Andreas Zollmann, Chengqing Zong, Pierre Zweigenbaum

Additional Reviewers:

Shlomo Argamon, Javier Artiles, Giuseppe Attardi, S.R.K. Branavan, Julian Brooke, David Burkett, Yong Cao, Yufeng Chen, Yu Chen, Ying Chen, Bonaventura Coppola, Sajib Dasgupta, Anirban Dasgupta, Diego De Cao, Oier Lopez de Lacalle, Adi Eyal, Jung-wei Fan, Benoit Favre, Moshe Fresko, Oana Frunza, Bin Gao, Xi-Wu Han, Zhongjun He, Carmen Heger, Zhongjun He, Wenbin Jiang, Richard Johansson, Anna Kazantseva, Alistair Kennedy, Fazel Keshtkar, Gerhard Kremer, Patrik Lambert, Greg Langmead, Oliver Lemon, Gregor Leusch, Xiao Li, Shoushan Li, Wei Li, Sujian Li, Xiaojiang Liu, Ting Liu, Yang Liu, Yue Lu, Jia Lu, Weihua Luo, Gang Luo, Yong-Liang Ma, Saab Mansour, Haitao Mi, Peter Nabenda, Peter Nabende, Ramesh Nallapati, Dipasree Pal, Adam Pauls, Emanuele Pianta, Daniele Pighin, Guilin Qi, Wei Qiao, Tao Qin, Jason Riesa, Felipe Sanchez-Martinez, Steven Schockaert, Dou Shen, Kathrin Spreyer, Jun Sun, Milan Tofiloski, Lav Varshney, Ye-Yi

Wang, Yu-Chieh Wu, Gu Xu, Sibel Yaman, Shiren Ye, Huang Yun, Hui Zhang, Yi Zhang,
Bing Zhao, Yu Zhou, Zhemin Zhu, Conghui Zhu

Mentoring Service Committee

Chairs:

Ng, Hwee Tou (National University of Singapore, Singapore; ngh@comp.nus.edu.sg)
Reeder, Florence (Mitre, USA; freeder@mitre.org)

Members:

Razvan Bunescu, Yee Seng Chan, Chrys Chrystello, Ken Church, Walter Daelemans, Deborah Dahl, Robert Daland, Janet Hitzeman, Eduard Hovy, Marilyn Kupetz, Preslav Nakov, Jian-Yun Nie, Kemal Oflazer, Bea Oshika, Dan Roth, Kenneth Samuel, Antal van den Bosch, John White, Yuk Wah Wong

Invited Talk:

Heterogeneous Transfer Learning with Real-world Applications

Qiang Yang

Hong Kong University of Science and Technology

qyang@cse.ust.hk

Abstract

In many real-world machine learning and data mining applications, we often face the problem where the training data are scarce in the feature space of interest, but much data are available in other feature spaces. Many existing learning techniques cannot make use of these auxiliary data, because these algorithms are based on the assumption that the training and test data must come from the same distribution and feature spaces. When this assumption does not hold, we have to seek novel techniques for ‘transferring’ the knowledge from one feature space to another. In this talk, I will present our recent works on heterogeneous transfer learning. I will describe how to identify the common parts of different feature spaces and learn a bridge between them to improve the learning performance in target task domains. I will also present several interesting applications of heterogeneous transfer learning, such as image clustering and classification, cross-domain classification and collaborative filtering.

Biography

Qiang Yang is a professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. His research interests are artificial intelligence, including automated planning, machine learning and data mining. He graduated from Peking University in 1982 with BSc. in Astrophysics, and obtained his MSc. degrees in Astrophysics and Computer Science from the University of Maryland, College Park in 1985 and 1987, respectively. He obtained his PhD in Computer Science from the University of Maryland, College Park in 1989. He was an assistant/associate professor at the University of Waterloo between 1989 and 1995, and a professor and NSERC Industrial Research Chair at Simon Fraser University in Canada from 1995 to 2001.

Qiang Yang has been active in research on artificial intelligence planning, machine learning and data mining. His research teams won the 2004 and 2005 ACM KDDCUP international competitions on data mining. He has been on several editorial boards of international journals, including IEEE Intelligent Systems, IEEE Transactions on Knowledge and Data Engineering and Web Intelligence. He has been an organizer for several international conferences in AI and data mining, including being the conference co-chair for ACM IUI 2010 and ICCBR 2001, program co-chair for PRICAI 2006 and PAKDD 2007, workshop chair for ACM KDD 2007, AAAI tutorial chair for AAAI 2005 and 2006, data mining contest chair for IEEE ICDM 2007 and 2009, and vice chair for ICDM 2006 and CIKM 2009. He is a fellow of IEEE and a member of AAAI and ACM. His home page is at <http://www.cse.ust.hk/~qyang>

Invited Talk:

Discourse - Early Problems, Current Successes, Future Challenges

Bonnie Webber

University of Edinburgh, UK

`bonnie.webber@ed.ac.uk`

Abstract

I will look back through nearly forty years of computational research on discourse, noting some problems (such as context-dependence and inference) that were identified early on as a hindrance to further progress, some admirable successes that we have achieved so far in the development of algorithms and resources, and some challenges that we may want to (or that we may have to!) take up in the future, with particular attention to problems of data annotation and genre dependence.

Biography

Bonnie Webber was a researcher at Bolt Beranek and Newman while working on the PhD she received from Harvard University in 1978. She then taught in the Department of Computer and Information Science at the University of Pennsylvania for 20 years before joining the School of Informatics at the University of Edinburgh. Known for research on discourse and on question answering, she is a Past President of the Association for Computational Linguistics, co-developer (with Aravind Joshi, Rashmi Prasad, Alan Lee and Eleni Miltsakaki) of the Penn Discourse TreeBank, and co-editor (with Annie Zaenen and Martha Palmer) of the journal, *Linguistic Issues in Language Technology*.

Table of Contents

<i>Heterogeneous Transfer Learning for Image Clustering via the SocialWeb</i> Qiang Yang, Yuqiang Chen, Gui-Rong Xue, Wenyuan Dai and Yong Yu	1
<i>Investigations on Word Senses and Word Usages</i> Katrin Erk, Diana McCarthy and Nicholas Gaylord	10
<i>A Comparative Study on Generalization of Semantic Roles in FrameNet</i> Yuichiroh Matsubayashi, Naoaki Okazaki and Jun'ichi Tsujii	19
<i>Unsupervised Argument Identification for Semantic Role Labeling</i> Omri Abend, Roi Reichart and Ari Rappoport	28
<i>Brutus: A Semantic Role Labeling System Incorporating CCG, CFG, and Dependency Features</i> Stephen Boxwell, Dennis Mehay and Chris Brew	37
<i>Exploiting Heterogeneous Treebanks for Parsing</i> Zheng-Yu Niu, Haifeng Wang and Hua Wu	46
<i>Cross Language Dependency Parsing using a Bilingual Lexicon</i> Hai Zhao, Yan Song, Chunyu Kit and Guodong Zhou	55
<i>Topological Field Parsing of German</i> Jackie Chi Kit Cheung and Gerald Penn	64
<i>Unsupervised Multilingual Grammar Induction</i> Benjamin Snyder, Tahira Naseem and Regina Barzilay	73
<i>Reinforcement Learning for Mapping Instructions to Actions</i> S.R.K. Branavan, Harr Chen, Luke Zettlemoyer and Regina Barzilay	82
<i>Learning Semantic Correspondences with Less Supervision</i> Percy Liang, Michael Jordan and Dan Klein	91
<i>Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling</i> Daichi Mochihashi, Takeshi Yamada and Naonori Ueda	100
<i>Knowing the Unseen: Estimating Vocabulary Size over Unseen Samples</i> Suma Bhat and Richard Sproat	109
<i>A Ranking Approach to Stress Prediction for Letter-to-Phoneme Conversion</i> Qing Dou, Shane Bergsma, Sittichai Jiampojarn and Grzegorz Kondrak	118
<i>Reducing the Annotation Effort for Letter-to-Phoneme Conversion</i> Kenneth Dwyer and Grzegorz Kondrak	127
<i>Transliteration Alignment</i> Vladimir Pervouchine, Haizhou Li and Bo Lin	136
<i>Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike</i> Bart Jongejan and Hercules Dalianis	145

<i>Revisiting Pivot Language Approach for Machine Translation</i> Hua Wu and Haifeng Wang	154
<i>Efficient Minimum Error Rate Training and Minimum Bayes-Risk Decoding for Translation Hypergraphs and Lattices</i> Shankar Kumar, Wolfgang Macherey, Chris Dyer and Franz Och	163
<i>Forest-based Tree Sequence to String Translation Model</i> Hui Zhang, Min Zhang, Haizhou Li, Aiti Aw and Chew Lim Tan	172
<i>Active Learning for Multilingual Statistical Machine Translation</i> Gholamreza Haffari and Anoop Sarkar	181
<i>DEPEVAL(summ): Dependency-based Evaluation for Automatic Summaries</i> Karolina Owczarzak	190
<i>Summarizing Definition from Wikipedia</i> Shiren Ye, Tat-Seng Chua and Jie LU	199
<i>Automatically Generating Wikipedia Articles: A Structure-Aware Approach</i> Christina Sauper and Regina Barzilay	208
<i>Learning to Tell Tales: A Data-driven Approach to Story Generation</i> Neil McIntyre and Mirella Lapata	217
<i>Recognizing Stances in Online Debates</i> Swapna Somasundaran and Janyce Wiebe	226
<i>Co-Training for Cross-Lingual Sentiment Classification</i> Xiaojun Wan	235
<i>A Non-negative Matrix Tri-factorization Approach to Sentiment Classification with Lexical Prior Knowledge</i> Tao Li, Yi Zhang and Vikas Sindhwani	244
<i>Discovering the Discriminative Views: Measuring Term Weights for Sentiment Analysis</i> Jungi Kim, Jin-Ji Li and Jong-Hyeok Lee	253
<i>Compiling a Massive, Multilingual Dictionary via Probabilistic Inference</i> Mausam, Stephen Soderland, Oren Etzioni, Daniel Weld, Michael Skinner and Jeff Bilmes ...	262
<i>A Metric-based Framework for Automatic Taxonomy Induction</i> Hui Yang and Jamie Callan	271
<i>Learning with Annotation Noise</i> Eyal Beigman and Beata Beigman Klebanov	280
<i>Abstraction and Generalisation in Semantic Role Labels: PropBank, VerbNet or both?</i> Paola Merlo and Lonneke van der Plas	288
<i>Robust Machine Translation Evaluation with Entailment Features</i> Sebastian Pado, Michel Galley, Dan Jurafsky and Christopher D. Manning	297
<i>The Contribution of Linguistic Features to Automatic Machine Translation Evaluation</i> Enrique Amigó, Jesús Giménez, Julio Gonzalo and Felisa Verdejo	306

<i>A Syntax-Driven Bracketing Model for Phrase-Based Translation</i> Deyi Xiong, Min Zhang, Aiti Aw and Haizhou Li	315
<i>Topological Ordering of Function Words in Hierarchical Phrase-based Translation</i> Hendra Setiawan, Min Yen Kan, Haizhou Li and Philip Resnik	324
<i>Phrase-Based Statistical Machine Translation as a Traveling Salesman Problem</i> Mikhail Zaslavskiy, Marc Dymetman and Nicola Cancedda	333
<i>Concise Integer Linear Programming Formulations for Dependency Parsing</i> Andre Martins, Noah Smith and Eric Xing	342
<i>Non-Projective Dependency Parsing in Expected Linear Time</i> Joakim Nivre	351
<i>Semi-supervised Learning of Dependency Parsers using Generalized Expectation Criteria</i> Gregory Druck, Gideon Mann and Andrew McCallum	360
<i>Dependency Grammar Induction via Bitext Projection Constraints</i> Kuzman Ganchev, Jennifer Gillenwater and Ben Taskar	369
<i>Cross-Domain Dependency Parsing Using a Deep Linguistic Grammar</i> Yi Zhang and Rui Wang	378
<i>A Chinese-English Organization Name Translation System Using Heuristic Web Mining and Asymmetric Alignment</i> Fan Yang, Jun Zhao and Kang Liu	387
<i>Reducing Semantic Drift with Bagging and Distributional Similarity</i> Tara McIntosh and James R. Curran	396
<i>Jointly Identifying Temporal Relations with Markov Logic</i> Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara and Yuji Matsumoto	405
<i>Profile Based Cross-Document Coreference Using Kernelized Fuzzy Relational Clustering</i> Jian Huang, Sarah M. Taylor, Jonathan L. Smith, Konstantinos A. Fotiadis and C. Lee Giles ..	414
<i>Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task</i> Kristen Parton, Kathleen R. McKeown, Bob Coyne, Mona T. Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Wei Yun Ma, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu and Sibel Yaman	423
<i>Bilingual Co-Training for Monolingual Hyponymy-Relation Acquisition</i> Jong-Hoon Oh, Kiyotaka Uchimoto and Kentaro Torisawa	432
<i>Automatic Set Instance Extraction using the Web</i> Richard C. Wang and William W. Cohen	441
<i>Extracting Lexical Reference Rules from Wikipedia</i> Eyal Shnarch, Libby Barak and Ido Dagan	450
<i>Employing Topic Models for Pattern-based Semantic Class Discovery</i> Huibin Zhang, Mingjie Zhu, Shuming Shi and Ji-Rong Wen	459
<i>Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition</i> Dipanjan Das and Noah A. Smith	468

<i>Stochastic Gradient Descent Training for L1-regularized Log-linear Models with Cumulative Penalty</i> Yoshimasa Tsuruoka, Jun'ichi Tsujii and Sophia Ananiadou	477
<i>A global model for joint lemmatization and part-of-speech prediction</i> Kristina Toutanova and Colin Cherry	486
<i>Distributional Representations for Handling Sparsity in Supervised Sequence-Labeling</i> Fei Huang and Alexander Yates	495
<i>Minimized Models for Unsupervised Part-of-Speech Tagging</i> Sujith Ravi and Kevin Knight	504
<i>An Error-Driven Word-Character Hybrid Model for Joint Chinese Word Segmentation and POS Tagging</i> Canasai Kruengkrai, Kiyotaka Uchimoto, Jun'ichi Kazama, Yiou Wang, Kentaro Torisawa and Hitoshi Isahara	513
<i>Automatic Adaptation of Annotation Standards: Chinese Word Segmentation and POS Tagging – A Case Study</i> Wenbin Jiang, Liang Huang and Qun Liu	522
<i>Linefeed Insertion into Japanese Spoken Monologue for Captioning</i> Tomohiro Ohno, Masaki Murata and Shigeki Matsubara	531
<i>Semi-supervised Learning for Automatic Prosodic Event Detection Using Co-training Algorithm</i> Je Hun Jeon and Yang Liu	540
<i>Summarizing multiple spoken documents: finding evidence from untranscribed audio</i> Xiaodan Zhu, Gerald Penn and Frank Rudzicz	549
<i>Improving Tree-to-Tree Translation with Packed Forests</i> Yang Liu, Yajuan Lü and Qun Liu	558
<i>Fast Consensus Decoding over Translation Forests</i> John DeNero, David Chiang and Kevin Knight	567
<i>Joint Decoding with Multiple Translation Models</i> Yang Liu, Haitao Mi, Yang Feng and Qun Liu	576
<i>Collaborative Decoding: Partial Hypothesis Re-ranking Using Translation Consensus between Decoders</i> Mu Li, Nan Duan, Dongdong Zhang, Chi-Ho Li and Ming Zhou	585
<i>Variational Decoding for Statistical Machine Translation</i> Zhifei Li, Jason Eisner and Sanjeev Khudanpur	593
<i>Unsupervised Learning of Narrative Schemas and their Participants</i> Nathanael Chambers and Dan Jurafsky	602
<i>Learning a Compositional Semantic Parser using an Existing Syntactic Parser</i> Ruifang Ge and Raymond Mooney	611
<i>Latent Variable Models of Concept-Attribute Attachment</i> Joseph Reisinger and Marius Pasca	620
<i>The Chinese Aspect Generation Based on Aspect Selection Functions</i> Guowen Yang and John Bateman	629

<i>Quantitative modeling of the neural representation of adjective-noun phrases to account for fMRI activation</i>	
Kai-min K. Chang, Vladimir L. Cherkassky, Tom M. Mitchell and Marcel Adam Just	638
<i>Capturing Saliency with a Trainable Cache Model for Zero-anaphora Resolution</i>	
Ryu Iida, Kentaro Inui and Yuji Matsumoto	647
<i>Conundrums in Noun Phrase Coreference Resolution: Making Sense of the State-of-the-Art</i>	
Veselin Stoyanov, Nathan Gilbert, Claire Cardie and Ellen Riloff	656
<i>A Novel Discourse Parser Based on Support Vector Machine Classification</i>	
David duVerle and Helmut Prendinger	665
<i>Genre distinctions for discourse in the Penn TreeBank</i>	
Bonnie Webber	674
<i>Automatic sense prediction for implicit discourse relations in text</i>	
Emily Pitler, Annie Louis and Ani Nenkova	683
<i>A Framework of Feature Selection Methods for Text Categorization</i>	
Shoushan Li, Rui Xia, Chengqing Zong and Chu-Ren Huang	692
<i>Mine the Easy, Classify the Hard: A Semi-Supervised Approach to Automatic Sentiment Classification</i>	
Sajib Dasgupta and Vincent Ng	701
<i>Modeling Latent Biographic Attributes in Conversational Genres</i>	
Nikesh Garera and David Yarowsky	710
<i>A Graph-based Semi-Supervised Learning for Question-Answering</i>	
Asli Celikyilmaz, Marcus Thint and Zhiheng Huang	719
<i>Combining Lexical Semantic Resources with Question & Answer Archives for Translation-Based Answer Finding</i>	
Delphine Bernhard and Iryna Gurevych	728
<i>Answering Opinion Questions with Random Walks on Graphs</i>	
Fangtao Li, Yang Tang, Minlie Huang and Xiaoyan Zhu	737
<i>What lies beneath: Semantic and syntactic analysis of manually reconstructed spontaneous speech</i>	
Erin Fitzgerald, Frederick Jelinek and Robert Frank	746
<i>Discriminative Lexicon Adaptation for Improved Character Accuracy - A New Direction in Chinese Language Modeling</i>	
Yi-cheng Pan, Lin-shan Lee and Sadaoki Furui	755
<i>Improving Automatic Speech Recognition for Lectures through Transformation-based Rules Learned from Minimal Data</i>	
Cosmin Munteanu, Gerald Penn and Xiaodan Zhu	764
<i>Quadratic-Time Dependency Parsing for Machine Translation</i>	
Michel Galley and Christopher D. Manning	773
<i>A Gibbs Sampler for Phrasal Synchronous Grammar Induction</i>	
Phil Blunsom, Trevor Cohn, Chris Dyer and Miles Osborne	782

<i>Source-Language Entailment Modeling for Translating Unknown Terms</i> Shachar Mirkin, Lucia Specia, Nicola Cancedda, Ido Dagan, Marc Dymetman and Idan Szpektor	791
<i>Case markers and Morphology: Addressing the crux of the fluency problem in English-Hindi SMT</i> Ananthakrishnan Ramanathan, Hansraj Choudhary, Avishek Ghosh and Pushpak Bhattacharyya	800
<i>Dependency Based Chinese Sentence Realization</i> Wei He, Haifeng Wang, Yuqing Guo and Ting Liu	809
<i>Incorporating Information Status into Generation Ranking</i> Aoife Cahill and Arndt Riester	817
<i>A Syntax-Free Approach to Japanese Sentence Compression</i> Tsutomu Hirao, Jun Suzuki and Hideki Isozaki	826
<i>Application-driven Statistical Paraphrase Generation</i> Shiqi Zhao, Xiang Lan, Ting Liu and Sheng Li	834
<i>Semi-Supervised Cause Identification from Aviation Safety Reports</i> Isaac Persing and Vincent Ng	843
<i>SMS based Interface for FAQ Retrieval</i> Govind Kothari, Sumit Negi, Tanveer A. Faruque, Venkatesan T. Chakaravarthy and L. Venkata Subramaniam	852
<i>Semantic Tagging of Web Search Queries</i> Mehdi Manshadi and Xiao Li	861
<i>Mining Bilingual Data from the Web with Adaptively Learnt Patterns</i> Long Jiang, Shiquan Yang, Ming Zhou, Xiaohua Liu and Qingsheng Zhu	870
<i>Comparing Objective and Subjective Measures of Usability in a Human-Robot Dialogue System</i> Mary Ellen Foster, Manuel Giuliani and Alois Knoll	879
<i>Setting Up User Action Probabilities in User Simulations for Dialog System Development</i> Hua Ai and Diane Litman	888
<i>Dialogue Segmentation with Large Numbers of Volunteer Internet Annotators</i> T. Daniel Midgley	897
<i>Robust Approach to Abbreviating Terms: A Discriminative Latent Variable Model with Global Information</i> Xu Sun, Naoaki Okazaki and Jun'ichi Tsujii	905
<i>A non-contiguous Tree Sequence Alignment-based Model for Statistical Machine Translation</i> Jun Sun, Min Zhang and Chew Lim Tan	914
<i>Better Word Alignments with Supervised ITG Models</i> Aria Haghighi, John Blitzer, John DeNero and Dan Klein	923
<i>Confidence Measure for Word Alignment</i> Fei Huang	932

<i>A Comparative Study of Hypothesis Alignment and its Improvement for Machine Translation System Combination</i>	
Boxing Chen, Min Zhang, Haizhou Li and Aiti Aw	941
<i>Incremental HMM Alignment for MT System Combination</i>	
Chi-Ho Li, Xiaodong He, Yupeng Liu and Ning Xi	949
<i>K-Best A* Parsing</i>	
Adam Pauls and Dan Klein	958
<i>Coordinate Structure Analysis with Global Structural Constraints and Alignment-Based Local Features</i>	
Kazuo Hara, Masashi Shimbo, Hideharu Okuma and Yuji Matsumoto	967
<i>Learning Context-Dependent Mappings from Sentences to Logical Form</i>	
Luke Zettlemoyer and Michael Collins	976
<i>An Optimal-Time Binarization Algorithm for Linear Context-Free Rewriting Systems with Fan-Out Two</i>	
Carlos Gómez-Rodríguez and Giorgio Satta	985
<i>A Polynomial-Time Parsing Algorithm for TT-MCTAG</i>	
Laura Kallmeyer and Giorgio Satta	994
<i>Distant supervision for relation extraction without labeled data</i>	
Mike Mintz, Steven Bills, Rion Snow and Daniel Jurafsky	1003
<i>Multi-Task Transfer Learning for Weakly-Supervised Relation Extraction</i>	
Jing Jiang	1012
<i>Unsupervised Relation Extraction by Mining Wikipedia Texts Using Information from the Web</i>	
Yulan Yan, Naoaki Okazaki, Yutaka Matsuo, Zhenglu Yang and Mitsuru Ishizuka	1021
<i>Phrase Clustering for Discriminative Learning</i>	
Dekang Lin and Xiaoyun Wu	1030
<i>Semi-Supervised Active Learning for Sequence Labeling</i>	
Katrin Tomanek and Udo Hahn	1039
<i>Word or Phrase? Learning Which Unit to Stress for Information Retrieval</i>	
Young-In Song, Jung-Tae Lee and Hae-Chang Rim	1048
<i>A Generative Blog Post Retrieval Model that Uses Query Expansion based on External Collections</i>	
Wouter Weerkamp, Krisztian Balog and Maarten de Rijke	1057
<i>Language Identification of Search Engine Queries</i>	
Hakan Ceylan and Yookyung Kim	1066
<i>Exploiting Bilingual Information to Improve Web Search</i>	
Wei Gao, John Blitzer, Ming Zhou and Kam-Fai Wong	1075

Conference Program

Monday, August 3, 2009

08:30–08:40 Opening Session

08:40–09:40 Invited Talk: Qiang Yang, Heterogeneous Transfer Learning with Real-world Applications

Invited Talk

08:40–09:40 *Heterogeneous Transfer Learning for Image Clustering via the SocialWeb*
Qiang Yang, Yuqiang Chen, Gui-Rong Xue, Wenyuan Dai and Yong Yu

09:40–10:10 Break

Session 1A: Semantics 1

Chaired by Graeme Hirst

10:10–10:35 *Investigations on Word Senses and Word Usages*
Katrin Erk, Diana McCarthy and Nicholas Gaylord

10:35–11:00 *A Comparative Study on Generalization of Semantic Roles in FrameNet*
Yuichiroh Matsubayashi, Naoaki Okazaki and Jun'ichi Tsujii

11:00–11:25 *Unsupervised Argument Identification for Semantic Role Labeling*
Omri Abend, Roi Reichart and Ari Rappoport

11:25–11:50 *Brutus: A Semantic Role Labeling System Incorporating CCG, CFG, and Dependency Features*
Stephen Boxwell, Dennis Mehay and Chris Brew

Monday, August 3, 2009 (continued)

Session 1B: Syntax and Parsing 1

Chaired by Christopher Manning

- 10:10–10:35 *Exploiting Heterogeneous Treebanks for Parsing*
Zheng-Yu Niu, Haifeng Wang and Hua Wu
- 10:35–11:00 *Cross Language Dependency Parsing using a Bilingual Lexicon*
Hai Zhao, Yan Song, Chunyu Kit and Guodong Zhou
- 11:00–11:25 *Topological Field Parsing of German*
Jackie Chi Kit Cheung and Gerald Penn
- 11:25–11:50 *Unsupervised Multilingual Grammar Induction*
Benjamin Snyder, Tahira Naseem and Regina Barzilay

Session 1C: Statistical and Machine Learning Methods 1

Chaired by Jun Suzuki

- 10:10–10:35 *Reinforcement Learning for Mapping Instructions to Actions*
S.R.K. Branavan, Harr Chen, Luke Zettlemoyer and Regina Barzilay
- 10:35–11:00 *Learning Semantic Correspondences with Less Supervision*
Percy Liang, Michael Jordan and Dan Klein
- 11:00–11:25 *Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling*
Daichi Mochihashi, Takeshi Yamada and Naonori Ueda
- 11:25–11:50 *Knowing the Unseen: Estimating Vocabulary Size over Unseen Samples*
Suma Bhat and Richard Sproat

Monday, August 3, 2009 (continued)

Session 1D: Phonology and Morphology

Chaired by Jason Eisner

- 10:10–10:35 *A Ranking Approach to Stress Prediction for Letter-to-Phoneme Conversion*
Qing Dou, Shane Bergsma, Sittichai Jiampojarn and Grzegorz Kondrak
- 10:35–11:00 *Reducing the Annotation Effort for Letter-to-Phoneme Conversion*
Kenneth Dwyer and Grzegorz Kondrak
- 11:00–11:25 *Transliteration Alignment*
Vladimir Pervouchine, Haizhou Li and Bo Lin
- 11:25–11:50 *Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike*
Bart Jongejan and Hercules Dalianis
- 11:50–13:20 Lunch

Session 2A: Machine Translation 1

Chaired by Qun Liu

- 13:20–13:45 *Revisiting Pivot Language Approach for Machine Translation*
Hua Wu and Haifeng Wang
- 13:45–14:10 *Efficient Minimum Error Rate Training and Minimum Bayes-Risk Decoding for Translation Hypergraphs and Lattices*
Shankar Kumar, Wolfgang Macherey, Chris Dyer and Franz Och
- 14:10–14:35 *Forest-based Tree Sequence to String Translation Model*
Hui Zhang, Min Zhang, Haizhou Li, Aiti Aw and Chew Lim Tan
- 14:35–15:00 *Active Learning for Multilingual Statistical Machine Translation*
Gholamreza Haffari and Anoop Sarkar

Monday, August 3, 2009 (continued)

Session 2B: Generation and Summarization 1

Chaired by Anja Belz

- 13:20–13:45 *DEPEVAL(summ): Dependency-based Evaluation for Automatic Summaries*
Karolina Owczarzak
- 13:45–14:10 *Summarizing Definition from Wikipedia*
Shiren Ye, Tat-Seng Chua and Jie LU
- 14:10–14:35 *Automatically Generating Wikipedia Articles: A Structure-Aware Approach*
Christina Sauper and Regina Barzilay
- 14:35–15:00 *Learning to Tell Tales: A Data-driven Approach to Story Generation*
Neil McIntyre and Mirella Lapata

Session 2C: Sentiment Analysis and Text Categorization 1

Chaired by Katja Markert

- 13:20–13:45 *Recognizing Stances in Online Debates*
Swapna Somasundaran and Janyce Wiebe
- 13:45–14:10 *Co-Training for Cross-Lingual Sentiment Classification*
Xiaojun Wan
- 14:10–14:35 *A Non-negative Matrix Tri-factorization Approach to Sentiment Classification with Lexical Prior Knowledge*
Tao Li, Yi Zhang and Vikas Sindhwani
- 14:35–15:00 *Discovering the Discriminative Views: Measuring Term Weights for Sentiment Analysis*
Jungi Kim, Jin-Ji Li and Jong-Hyeok Lee

Monday, August 3, 2009 (continued)

Session 2D: Language Resources

Chaired by Nicoletta Calzolari

- 13:20–13:45 *Compiling a Massive, Multilingual Dictionary via Probabilistic Inference*
Mausam, Stephen Soderland, Oren Etzioni, Daniel Weld, Michael Skinner and Jeff Bilmes
- 13:45–14:10 *A Metric-based Framework for Automatic Taxonomy Induction*
Hui Yang and Jamie Callan
- 14:10–14:35 *Learning with Annotation Noise*
Eyal Beigman and Beata Beigman Klebanov
- 14:35–15:00 *Abstraction and Generalisation in Semantic Role Labels: PropBank, VerbNet or both?*
Paola Merlo and Lonneke van der Plas
- 15:00–15:30 Break

Session 3A: Machine Translation 2

Chaired by Haifeng Wang

- 15:30–15:55 *Robust Machine Translation Evaluation with Entailment Features*
Sebastian Pado, Michel Galley, Dan Jurafsky and Christopher D. Manning
- 15:55–16:20 *The Contribution of Linguistic Features to Automatic Machine Translation Evaluation*
Enrique Amigó, Jesús Giménez, Julio Gonzalo and Felisa Verdejo
- 16:20–16:45 *A Syntax-Driven Bracketing Model for Phrase-Based Translation*
Deyi Xiong, Min Zhang, Aiti Aw and Haizhou Li
- 16:45–17:10 *Topological Ordering of Function Words in Hierarchical Phrase-based Translation*
Hendra Setiawan, Min Yen Kan, Haizhou Li and Philip Resnik
- 17:10–17:35 *Phrase-Based Statistical Machine Translation as a Traveling Salesman Problem*
Mikhail Zaslavskiy, Marc Dymetman and Nicola Cancedda

Monday, August 3, 2009 (continued)

Session 3B: Syntax and Parsing 2

Chaired by Dan Klein

- 15:30–15:55 *Concise Integer Linear Programming Formulations for Dependency Parsing*
Andre Martins, Noah Smith and Eric Xing
- 15:55–16:20 *Non-Projective Dependency Parsing in Expected Linear Time*
Joakim Nivre
- 16:20–16:45 *Semi-supervised Learning of Dependency Parsers using Generalized Expectation Criteria*
Gregory Druck, Gideon Mann and Andrew McCallum
- 16:45–17:10 *Dependency Grammar Induction via Bitext Projection Constraints*
Kuzman Ganchev, Jennifer Gillenwater and Ben Taskar
- 17:10–17:35 *Cross-Domain Dependency Parsing Using a Deep Linguistic Grammar*
Yi Zhang and Rui Wang

Session 3C: Information Extraction 1

Chaired by Eduard Hovy

- 15:30–15:55 *A Chinese-English Organization Name Translation System Using Heuristic Web Mining and Asymmetric Alignment*
Fan Yang, Jun Zhao and Kang Liu
- 15:55–16:20 *Reducing Semantic Drift with Bagging and Distributional Similarity*
Tara McIntosh and James R. Curran
- 16:20–16:45 *Jointly Identifying Temporal Relations with Markov Logic*
Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara and Yuji Matsumoto
- 16:45–17:10 *Profile Based Cross-Document Coreference Using Kernelized Fuzzy Relational Clustering*
Jian Huang, Sarah M. Taylor, Jonathan L. Smith, Konstantinos A. Fotiadis and C. Lee Giles
- 17:10–17:35 *Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task*
Kristen Parton, Kathleen R. McKeown, Bob Coyne, Mona T. Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Wei Yun Ma, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu and Sibel Yaman

Monday, August 3, 2009 (continued)

Session 3D: Semantics 2

Chaired by Patrick Pantel

- 15:30–15:55 *Bilingual Co-Training for Monolingual Hyponymy-Relation Acquisition*
Jong-Hoon Oh, Kiyotaka Uchimoto and Kentaro Torisawa
- 15:55–16:20 *Automatic Set Instance Extraction using the Web*
Richard C. Wang and William W. Cohen
- 16:20–16:45 *Extracting Lexical Reference Rules from Wikipedia*
Eyal Shnarch, Libby Barak and Ido Dagan
- 16:45–17:10 *Employing Topic Models for Pattern-based Semantic Class Discovery*
Huibin Zhang, Mingjie Zhu, Shuming Shi and Ji-Rong Wen
- 17:10–17:35 *Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition*
Dipanjan Das and Noah A. Smith

Tuesday, August 4, 2009

Session 4A: Statistical and Machine Learning Methods 2

Chaired by Hal Daumé III

- 08:30–08:55 *Stochastic Gradient Descent Training for L1-regularized Log-linear Models with Cumulative Penalty*
Yoshimasa Tsuruoka, Jun'ichi Tsujii and Sophia Ananiadou
- 08:55–09:20 *A global model for joint lemmatization and part-of-speech prediction*
Kristina Toutanova and Colin Cherry
- 09:20–09:45 *Distributional Representations for Handling Sparsity in Supervised Sequence-Labeling*
Fei Huang and Alexander Yates

Tuesday, August 4, 2009 (continued)

Session 4B: Word Segmentation and POS Tagging

Chaired by Hwee Tou Ng

- 08:30–08:55 *Minimized Models for Unsupervised Part-of-Speech Tagging*
Sujith Ravi and Kevin Knight
- 08:55–09:20 *An Error-Driven Word-Character Hybrid Model for Joint Chinese Word Segmentation and POS Tagging*
Canasai Kruengkrai, Kiyotaka Uchimoto, Jun'ichi Kazama, Yiou Wang, Kentaro Torisawa and Hitoshi Isahara
- 09:20–09:45 *Automatic Adaptation of Annotation Standards: Chinese Word Segmentation and POS Tagging – A Case Study*
Wenbin Jiang, Liang Huang and Qun Liu

Session 4C: Spoken Language Processing 1

Chaired by Brian Roark

- 08:30–08:55 *Linefeed Insertion into Japanese Spoken Monologue for Captioning*
Tomohiro Ohno, Masaki Murata and Shigeki Matsubara
- 08:55–09:20 *Semi-supervised Learning for Automatic Prosodic Event Detection Using Co-training Algorithm*
Je Hun Jeon and Yang Liu
- 09:20–09:45 *Summarizing multiple spoken documents: finding evidence from untranscribed audio*
Xiaodan Zhu, Gerald Penn and Frank Rudzicz

Session 4DI: Short Paper 1 (Syntax and Parsing)

Session 4DII: Short Paper 2 (Discourse and Dialogue)

- 09:45–10:15 Break

Tuesday, August 4, 2009 (continued)

Session 5A: Machine Translation 3

Chaired by Dan Gildea

- 10:15–10:40 *Improving Tree-to-Tree Translation with Packed Forests*
Yang Liu, Yajuan Lü and Qun Liu
- 10:40–11:05 *Fast Consensus Decoding over Translation Forests*
John DeNero, David Chiang and Kevin Knight
- 11:05–11:30 *Joint Decoding with Multiple Translation Models*
Yang Liu, Haitao Mi, Yang Feng and Qun Liu
- 11:30–11:55 *Collaborative Decoding: Partial Hypothesis Re-ranking Using Translation Consensus between Decoders*
Mu Li, Nan Duan, Dongdong Zhang, Chi-Ho Li and Ming Zhou
- 11:55–12:20 *Variational Decoding for Statistical Machine Translation*
Zhifei Li, Jason Eisner and Sanjeev Khudanpur

Session 5B: Semantics 3

Chaired by Diana McCarthy

- 10:15–10:40 *Unsupervised Learning of Narrative Schemas and their Participants*
Nathanael Chambers and Dan Jurafsky
- 10:40–11:05 *Learning a Compositional Semantic Parser using an Existing Syntactic Parser*
Ruifang Ge and Raymond Mooney
- 11:05–11:30 *Latent Variable Models of Concept-Attribute Attachment*
Joseph Reisinger and Marius Pasca
- 11:30–11:55 *The Chinese Aspect Generation Based on Aspect Selection Functions*
Guowen Yang and John Bateman
- 11:55–12:20 *Quantitative modeling of the neural representation of adjective-noun phrases to account for fMRI activation*
Kai-min K. Chang, Vladimir L. Cherkassky, Tom M. Mitchell and Marcel Adam Just

Tuesday, August 4, 2009 (continued)

Session 5C: Discourse and Dialogue 1

Chaired by Kathy McKeown

- 10:15–10:40 *Capturing Saliency with a Trainable Cache Model for Zero-anaphora Resolution*
Ryu Iida, Kentaro Inui and Yuji Matsumoto
- 10:40–11:05 *Conundrums in Noun Phrase Coreference Resolution: Making Sense of the State-of-the-Art*
Veselin Stoyanov, Nathan Gilbert, Claire Cardie and Ellen Riloff
- 11:05–11:30 *A Novel Discourse Parser Based on Support Vector Machine Classification*
David duVerle and Helmut Prendinger
- 11:30–11:55 *Genre distinctions for discourse in the Penn TreeBank*
Bonnie Webber
- 11:55–12:20 *Automatic sense prediction for implicit discourse relations in text*
Emily Pitler, Annie Louis and Ani Nenkova

Session 5D: Student Research Workshop

- 12:20–14:20 Short Paper Poster / SRW Poster Session (Lunch)

Session 6A: Short Paper 3 (Machine Translation)

Session 6B: Short Paper 4 (Semantics)

Session 6C: Short Paper 5 (Spoken Language Processing)

Tuesday, August 4, 2009 (continued)

Session 6D: Short Paper 6 (Statistical and Machine Learning Methods 1)

Session 6E: Short Paper 7 (Summarization and Generation)

Session 6F: Short Paper 8 (Sentiment Analysis)

Session 6G: Short Paper 9 (Question Answering)

Session 6H: Short Paper 10 (Statistical and Machine Learning Methods 2)

16:25–16:50 Break

Session 7A: Sentiment Analysis and Text Categorization 2

Chaired by Bing Liu

16:50–17:15 *A Framework of Feature Selection Methods for Text Categorization*
Shoushan Li, Rui Xia, Chengqing Zong and Chu-Ren Huang

17:15–17:40 *Mine the Easy, Classify the Hard: A Semi-Supervised Approach to Automatic Sentiment Classification*
Sajib Dasgupta and Vincent Ng

17:40–18:05 *Modeling Latent Biographic Attributes in Conversational Genres*
Nikesh Garera and David Yarowsky

Session 7B: Question Answering

Chaired by Hae-Chang Rim

16:50–17:15 *A Graph-based Semi-Supervised Learning for Question-Answering*
Asli Celikyilmaz, Marcus Thint and Zhiheng Huang

17:15–17:40 *Combining Lexical Semantic Resources with Question & Answer Archives for Translation-Based Answer Finding*
Delphine Bernhard and Iryna Gurevych

17:40–18:05 *Answering Opinion Questions with Random Walks on Graphs*
Fangtao Li, Yang Tang, Minlie Huang and Xiaoyan Zhu

Tuesday, August 4, 2009 (continued)

Session 7C: Spoken Language Processing 2

Chaired by Yang Liu

- 16:50–17:15 *What lies beneath: Semantic and syntactic analysis of manually reconstructed spontaneous speech*
Erin Fitzgerald, Frederick Jelinek and Robert Frank
- 17:15–17:40 *Discriminative Lexicon Adaptation for Improved Character Accuracy - A New Direction in Chinese Language Modeling*
Yi-cheng Pan, Lin-shan Lee and Sadaoki Furui
- 17:40–18:05 *Improving Automatic Speech Recognition for Lectures through Transformation-based Rules Learned from Minimal Data*
Cosmin Munteanu, Gerald Penn and Xiaodan Zhu

Session 7D: Short Paper 11 (Information Extraction)

Wednesday, August 5, 2009

08:30–09:30 Invited Talk: Bonnie Webber, Discourse – Early Problems, Current Successes, Future Challenges

09:30–10:00 Break

Session 8A: Machine Translation 4

Chaired by Dekai Wu

- 09:55–10:20 *Quadratic-Time Dependency Parsing for Machine Translation*
Michel Galley and Christopher D. Manning
- 10:20–10:45 *A Gibbs Sampler for Phrasal Synchronous Grammar Induction*
Phil Blunsom, Trevor Cohn, Chris Dyer and Miles Osborne
- 10:45–11:10 *Source-Language Entailment Modeling for Translating Unknown Terms*
Shachar Mirkin, Lucia Specia, Nicola Cancedda, Ido Dagan, Marc Dymetman and Idan Szpektor
- 11:10–11:35 *Case markers and Morphology: Addressing the crux of the fluency problem in English-Hindi SMT*
Ananthakrishnan Ramanathan, Hansraj Choudhary, Avishek Ghosh and Pushpak Bhattacharyya

Wednesday, August 5, 2009 (continued)

Session 8B: Generation and Summarization 2

Chaired by Regina Barzilay

09:55–10:20 *Dependency Based Chinese Sentence Realization*
Wei He, Haifeng Wang, Yuqing Guo and Ting Liu

10:20–10:45 *Incorporating Information Status into Generation Ranking*
Aoife Cahill and Arndt Riester

10:45–11:10 *A Syntax-Free Approach to Japanese Sentence Compression*
Tsutomu Hirao, Jun Suzuki and Hideki Isozaki

11:10–11:35 *Application-driven Statistical Paraphrase Generation*
Shiqi Zhao, Xiang Lan, Ting Liu and Sheng Li

Session 8C: Text Mining and NLP applications

Chaired by Sophia Ananioudou

09:55–10:20 *Semi-Supervised Cause Identification from Aviation Safety Reports*
Isaac Persing and Vincent Ng

10:20–10:45 *SMS based Interface for FAQ Retrieval*
Govind Kothari, Sumit Negi, Tanveer A. Faruque, Venkatesan T. Chakaravarthy and L. Venkata Subramaniam

10:45–11:10 *Semantic Tagging of Web Search Queries*
Mehdi Manshadi and Xiao Li

11:10–11:35 *Mining Bilingual Data from the Web with Adaptively Learnt Patterns*
Long Jiang, Shiquan Yang, Ming Zhou, Xiaohua Liu and Qingsheng Zhu

Wednesday, August 5, 2009 (continued)

Session 8D: Discourse and Dialogue 2

Chaired by Gary Geunbae Lee

- 09:55–10:20 *Comparing Objective and Subjective Measures of Usability in a Human-Robot Dialogue System*
Mary Ellen Foster, Manuel Giuliani and Alois Knoll
- 10:20–10:45 *Setting Up User Action Probabilities in User Simulations for Dialog System Development*
Hua Ai and Diane Litman
- 10:45–11:10 *Dialogue Segmentation with Large Numbers of Volunteer Internet Annotators*
T. Daniel Midgley
- 11:10–11:35 *Robust Approach to Abbreviating Terms: A Discriminative Latent Variable Model with Global Information*
Xu Sun, Naoaki Okazaki and Jun'ichi Tsujii
- 11:35–12:30 Lunch
- 12:30–14:00 Business Meeting
- 14:00–14:25 Break

Session 9A: Machine Translation 5

Chaired by Kevin Knight

- 14:25–14:50 *A non-contiguous Tree Sequence Alignment-based Model for Statistical Machine Translation*
Jun Sun, Min Zhang and Chew Lim Tan
- 14:50–15:15 *Better Word Alignments with Supervised ITG Models*
Aria Haghighi, John Blitzer, John DeNero and Dan Klein
- 15:15–15:40 *Confidence Measure for Word Alignment*
Fei Huang
- 15:40–16:05 *A Comparative Study of Hypothesis Alignment and its Improvement for Machine Translation System Combination*
Boxing Chen, Min Zhang, Haizhou Li and Aiti Aw
- 16:05–16:30 *Incremental HMM Alignment for MT System Combination*
Chi-Ho Li, Xiaodong He, Yupeng Liu and Ning Xi

Wednesday, August 5, 2009 (continued)

Session 9B: Syntax and Parsing 3

Chaired by James Curran

- 14:25–14:50 *K-Best A* Parsing*
Adam Pauls and Dan Klein
- 14:50–15:15 *Coordinate Structure Analysis with Global Structural Constraints and Alignment-Based Local Features*
Kazuo Hara, Masashi Shimbo, Hideharu Okuma and Yuji Matsumoto
- 15:15–15:40 *Learning Context-Dependent Mappings from Sentences to Logical Form*
Luke Zettlemoyer and Michael Collins
- 15:40–16:05 *An Optimal-Time Binarization Algorithm for Linear Context-Free Rewriting Systems with Fan-Out Two*
Carlos Gómez-Rodríguez and Giorgio Satta
- 16:05–16:30 *A Polynomial-Time Parsing Algorithm for TT-MCTAG*
Laura Kallmeyer and Giorgio Satta

Session 9C: Information Extraction 2

Chaired by Jian Su

- 14:25–14:50 *Distant supervision for relation extraction without labeled data*
Mike Mintz, Steven Bills, Rion Snow and Daniel Jurafsky
- 14:50–15:15 *Multi-Task Transfer Learning for Weakly-Supervised Relation Extraction*
Jing Jiang
- 15:15–15:40 *Unsupervised Relation Extraction by Mining Wikipedia Texts Using Information from the Web*
Yulan Yan, Naoaki Okazaki, Yutaka Matsuo, Zhenglu Yang and Mitsuru Ishizuka
- 15:40–16:05 *Phrase Clustering for Discriminative Learning*
Dekang Lin and Xiaoyun Wu
- 16:05–16:30 *Semi-Supervised Active Learning for Sequence Labeling*
Katrin Tomanek and Udo Hahn

Wednesday, August 5, 2009 (continued)

Session 9D: Information Retrieval

Chaired by Kam-Fai Wong

- 14:25–14:50 *Word or Phrase? Learning Which Unit to Stress for Information Retrieval*
Young-In Song, Jung-Tae Lee and Hae-Chang Rim
- 14:50–15:15 *A Generative Blog Post Retrieval Model that Uses Query Expansion based on External Collections*
Wouter Weerkamp, Krisztian Balog and Maarten de Rijke
- 15:15–15:40 *Language Identification of Search Engine Queries*
Hakan Ceylan and Yookyung Kim
- 15:40–16:05 *Exploiting Bilingual Information to Improve Web Search*
Wei Gao, John Blitzer, Ming Zhou and Kam-Fai Wong
- 16:35–18:00 Best Paper Awards, Lifetime Achievement Award and Presentation
- 18:00–18:30 Closing Session