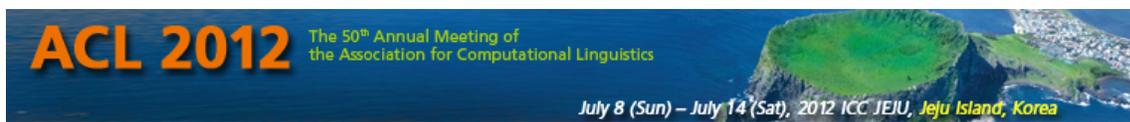


50th Annual Meeting of the Association for Computational Linguistics



Proceedings of the Conference

Volume 1: Long Papers

July 8 - 14, 2012

Jeju Island, Korea

PLATINUM SPONSOR



GOLD SPONSORS



SILVER SPONSORS



BRONZE SPONSORS



SUPPORTERS



SPONSOR FOR BEST PAPER AWARD

IBM Research

©2012 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

ISBN 978-1-937284-24-4 (Volume 1: Long Papers)
ISBN 978-1-937284-25-1 (Volume 2: Short Papers)

Preface: General Chair

Welcome to Jeju Island — where ACL makes a return to Asia!

As General Chair, I am indeed honored to pen the first words of ACL 2012 proceedings. In the past year, research in computational linguistics has continued to thrive across Asia and all over the world. On this occasion, I share with you the excitement of our community as we gather again at our annual meeting. On behalf of the organizing team, it is my great pleasure to welcome you to Jeju Island and ACL 2012.

In 2012, ACL turns 50. I feel privileged to chair the conference that marks such an important milestone for our community. We have prepared special programs to commemorate the 50th anniversary, including ‘Rediscovering 50 Years of Discovery’, a main conference workshop chaired by **Rafael Banchs** with a program on ‘the People, the Contents, and the Anthology’, which recollects some of the great moments in ACL history, and ‘ACL 50th Anniversary Lectures’ by **Mark Johnson**, **Aravind K. Joshi** and a Lifetime Achievement Award Recipient.

A large number of people have worked hard to bring this annual meeting to fruition. It has been an unforgettable experience for everyone involved. My deepest thanks go to the authors, reviewers, volunteers, participants, and all members and chairs of the organizing committees. It is your participation that makes a difference.

Program Chairs, **Chin-Yew Lin** and **Miles Osborne**, deserve our gratitude for putting an immense amount of work to ensure that each of the 940 submissions was taken care of. They put together a superb technical program like nobody else. Publication Chairs, **Maggie Li** and **Michael White**, extended the publishing tools to take care of every detail and compiled all the books within an impossible schedule. Tutorial Chair, **Michael Strube**, put together six tutorials that you can never miss. Workshop Chairs, **Massimo Poesio** and **Satoshi Sekine**, working with their EACL and NAACL counterparts, selected 11 quality workshops, many of which are new editions in their popular workshop series. Demo Chair, **Min Zhang**, started a novel review process and selected 29 quality system demos. Faculty Advisors, **Kentaro Inui**, **Greg Kondrak**, and **Yang Liu**, and Student Chairs, **Jackie Cheung**, **Jun Hatori**, **Carlos Henriquez** and **Ann Irvine**, assembled an excellent program for the Student Research Workshop with 12 accepted papers. Mentoring Chair, **Joyce Chai**, coordinated the mentorship of 13 papers. Publicity Chairs, **Jung-jae Kim** and **Youngjoong Ko**, developed the website, newsletters, and conference handbook that kept us updated all the time. Exhibition Chair, **Byeongchang Kim**, coordinated more than 10 exhibitors with a strong industry presence. All the events are now brought to us on Jeju Island by the Local Arrangements Chairs, **Gary Lee** and **Jong Park**, and their team. I can never thank them enough for all the preparations they have made to host us in such a spectacular place!

I would like to express my gratitude and appreciation to **Kevin Knight**, Chair of the ACL Conference Coordination Committee, **Dragomir Radev**, ACL Secretary, and **Priscilla Rasmussen**, ACL Business Manager, for their advice and guidance throughout the process.

The financial sponsors generously supported ACL 2012 in a meaningful way despite a challenging

economic outlook. We are honored to have Baidu as the Platinum Sponsor, Elsevier and Google as Gold Sponsors, Microsoft, KAIST and SK as Silver Sponsors, 7 Bronze Sponsors, and 3 Supporters. The Donald and Betty Walker Student Scholarship Fund and Asian Federation of Natural Language Processing have supported our student travel grants. The sponsorship program was made possible by the ACL sponsorship committee: **Eiichiro Sumita, Haifeng Wang, Michael Gamon, Patrick Pantel, Massimiliano Ciaramita, and Idan Szpektor.**

Finally, I do hope that you have an enjoyable and productive time on Jeju Island, and that you will leave with fond memories of ACL's 50th Anniversary. With my best wishes for a successful conference!

Haizhou Li
ACL 2012 General Chair
July 2012

Preface: Programme Committee Co-Chairs

This year we received 571 valid long paper submissions and 369 short paper submissions. 19% of the long papers and 20% of the short papers were accepted. As usual, some are presented orally and some as posters. Taking unigram counts from accepted long paper titles, and ignoring function words, the most popular word were:

entity 5
evaluation 5
hierarchical 5
information 5
joint 5
syntactic 5
topic 5
discriminative 6
lexical 6
statistical 6
chinese 7
dependency 7
machine 8
modeling 8
models 8
language 10
word 10
parsing 11
model 12
learning 14
translation 15

Some areas have grown over time and some have diminished. The most popular area for submissions (as expected) was Machine Translation. We promoted Social Media as a new area.

Twenty nine Area Chairs worked with 665 reviewers, producing 1830 long paper reviews and 1187 short paper reviews. Everything ran to a tight schedule and there were no slippages. This would not have been possible without our wonderful and diligent Area Chairs and Reviewers. Thanks!

We are delighted to have two keynote speakers, both of whom are very well known to the language community: Aravind Joshi and Mark Johnson. They will give coordinated talks addressing the 50th ACL anniversary: “Remembrance of ACLs past” and “Computational linguistics: Where do we go from here?” The ACL Lifetime Achievement Award will be announced on the last day of the conference.

Of the many papers, we selected two as being outstanding:

Bayesian Symbol-Refined Tree Substitution Grammars for Syntactic Parsing
Hiroyuki Shindo, Yusuke Miyao, Akinori Fujino, Masaaki Nagata

String Re-writing Kernel
Fan Bu, Hang Li, Xiaoyan Zhu

They will be presented as best papers in a dedicated session.

We thank the General Conference Chair Haizhou Li, the Local Arrangements Committee headed by Gary Geunbae Lee, Michael White and Maggie Li, the Publication Co-Chairs for coordinating and putting the proceedings together and all other committee chairs for their work. MO is especially thankful to Steve Clark for helpful tips on how to manage and run the whole process.

We hope you enjoy the conference!

Chin-Yew Lin, Microsoft Research Asia
Miles Osborne, University of Edinburgh

Organizing Committee

General Chair

Haizhou Li, Institute for Infocomm Research

Program Co-Chairs

Chin-Yew Lin, Microsoft Research Asia

Miles Osborne, University of Edinburgh

Local Arrangement Co-Chairs

Gary Geunbae Lee, Pohang University of Science and Technology (POSTECH)

Jong C. Park, Korea Advanced Institute of Science and Technology (KAIST)

Workshop Co-Chairs

Massimo Poesio, University of Essex

Satoshi Sekine, New York University

Publication Co-Chairs

Maggie Li, The Hong Kong Polytechnic University

Michael White, The Ohio State University

Publicity Chairs

Jung-jae Kim, Nanyang Technological University

Youngjoong Ko, Dong-A University

Tutorial Chair

Michael Strube, HITS gmbH

Demo Chair

Min Zhang, Institute for Infocomm Research

Special Session Chair

Rafael E. Banchs, Institute for Infocomm Research

Mentoring Service Chair

Joyce Chai, Michigan State University

Exhibit Chair

Byeongchang Kim, Catholic University of Daegu

Faculty Advisors (Student Research Workshop)

Kentaro Inui, Tohoku University

Greg Kondrak, University of Alberta

Yang Liu, University of Texas at Dallas

Student Chairs (Student Research Workshop)

Jackie Cheung, University of Toronto
Jun Hatori, University of Tokyo
Carlos Henriquez, Technical University of Catalonia (UPC)
Ann Irvine, Johns Hopkins University

Sponsorship Chairs

Eiichiro Sumita, NICT
Haifeng Wang, Baidu
Michael Gamon, Microsoft
Patrick Pantel, Microsoft
Massimiliano Ciaramita, Google
Idan Szpektor, Yahoo!

Local Arrangements Committee

Gary Geunbae Lee (co-chair)
Jong C. Park (co-chair)
Jeong-Won Cha (social activities)
Hanmin Jung (local sponsorship)
Seungshik Kang (local finance)
Byeongchang Kim (local exhibit)
Harksoo Kim (government sponsorship)
Jung-jae Kim (conference handbook)
Youngjoong Ko (web site and flyer)
Heuseok Lim (student volunteer management)
Seong-Bae Park (internet, wifi and equipments)

Business Manager

Priscilla Rasmussen

Program Committee

Program Co-chairs

Chin-Yew Lin, Microsoft Research Asia
Miles Osborne, University of Edinburgh

Area Chairs

Hongyuan Zha, School of Computational Science and Engineering, College of Computing, Georgia Institute of Technology
Hsin-Hsi Chen, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan
Kemal Oflazer, Carnegie Mellon University - Qatar
Mikio Nakano, Honda Research Institute, Japan
Andrei Popescu-Belis, Idiap Research Institute, Switzerland
Eneko Agirre, University of the Basque Country, Spain
Jason Baldridge, The University of Texas at Austin, USA
David Weir, University of Sussex, UK
Trevor Cohn, University of Sheffield, UK
Mark Dredze, Johns Hopkins University, USA
Noah Smith, Carnegie Mellon University, USA
Jan Wiebe, University of Pittsburgh, USA
Daniel M. Bikel, Google Research, USA
Mona T. Diab, Center for Computational Learning Systems, Columbia University, USA
Fei Xia, Univ. of Washington, Seattle, USA
Marcello Federico, Fondazione Bruno Kessler, Trento, Italy
Adam Lopez, Human Language Technology Center of Excellence, Johns Hopkins University, USA
David Talbot, Google Research, USA
Hua Wu, Baidu, China
Evgeniy Gabrilovich, Google, USA
Naoaki Okazaki, Graduate School of Information Sciences, Tohoku University, Japan
Stephen Clark, University of Cambridge, UK
Liang Huang, University Southern California, USA
Anja Belz, School of Computing, Engineering and Maths, University of Brighton, UK
Ani Nenkova, University of Pennsylvania, USA
Chung-Hsien Wu, Department of Computer Science and Information Engineering, National Cheng Kung University, TAIWAN
Jian Su, Institute for Infocomm Research, Singapore
Jianfeng Gao, Microsoft Research, Redmond, USA
Vincent Ng, University of Texas at Dallas, USA

Program Committee

Apoorv Agarwal, Lars Ahrenberg, Hua Ai, Cem Akkaya, Inaki Alegria, Jan Alexandersson, Enrique Alfonseca, Ben Allison, Sophia Ananiadou, Abhishek Arun, Giuseppe Attardi, Necip Fazil Ayan, Cecilia Ovesdotter Alm

Jing Bai, Collin Baker, Kirk Baker, Jason Baldrige, Breck Baldwin, Tim Baldwin, Carmen Banea, Marco Baroni, Regina Barzilay, Roberto Basili, Tilman Becker, Michael Bendersky, Taylor Berg-Kirkpatrick, Sabine Bergler, Shane Bergsma, Indrajit Bhattacharya, Pushpak Bhat-tacharyya, Archana Bhattarai, Jiang Bian, Gann Bierner, Ann Bies, Arianna Bisazza, John Blitzer, Michael Bloodgood, Phil Blunsom, Bernd Bohnet, Ondrej Bojar, Danushka Bollegala, Francis Bond, Kalina Bontcheva, Stefano Borgo, Alexandre Bouchard, Jordan Boyd-Graber, Kristy Boyer, S.R.K. Branavan, Thorsten Brants, Chris Brew, Ted Briscoe, Samuel Brody, Sabine Buchholz, Paul Buitelaar, Paula Buttery, William Byrne, Donna Byron, Olga babko-malaya, Antal van den Bosch

Aoife Cahill, Mary Elaine Califf, Chris Callison-Burch, Nicoletta Calzolari Zamorani, Nicola Cancedda, Yunbo Cao, Claire Cardie, Michael Carl, Xavier Carreras, John Carroll, Francisco Casacuberta, Vittorio Castelli, Asli Celikyilmaz, Mauro Cettolo, Joyce Chai, Nate Chambers, Yee Seng Chan, Ming-Wei Chang, Pi-Chuan Chang, Berlin Chen, Chia-Ping Chen, John Chen, Keh-Jiann Chen, Xueqi Cheng, Colin Cherry, David Chiang, Christian Chiercos, Hai Leong Chieu, Laura Chiticariu, Yejin Choi, Jennifer Chu-Carroll, Tat-Seng Chua, Ken Church, Alexander Clark, Shay Cohen, Trevor Cohn, Kevyn Collins-Thompson, Marta Ruiz Costa Jussa, Dan Cristea, Hang Cui, Aron Culotta, James Cussens, Chaoliu Chaoliu

Walter Daelemans, Ido Dagan, R. I. Damper, Cristian Danescu-Niculescu-Mizil, Van Dang, Lawrence Danlos, Dipanjan Das, Hal Daume, Gerard De Melo, Guy De Pauw, Steve DeNeefe, John DeNero, Vera Demberg, Yasuharu Den, Pascal Denis, Markus Dickinson, Mike Dillinger, Xiaowen Ding, Kohji Dohsaka, John Dowding, Doug Downey, Markus Dreyer, Rebecca Dridan, Jinhua Du, Kevin Duh, Chris Dyer, Marc Dymetman

Judith Ecker-Kohler, Koji Eguchi, Andreas Eisele, Jacob Eisenstein, Jason Eisner, Michael Elhadad

Benoit Favre, Anna Feldman, Christiane Fellbaum, Raquel Fernandez, Margaret Fleck, Dan Flickinger, Radu Florian, Corina Forascu, Kate Forbes-Riley, George Foster, Anette Frank, Bob Frank, Dayne Freitag, Kotaro Funakoshi

Michel Galley, Michael Gamon, Kavita Ganesan, Juri Ganitkevitch, Wei Gao, Claire Gardent, Nikesh Garera, Michael Gasser, Albert Gatt, Dmitriy Genzel, Kallirroi Georgila, Sean Gerrish, Daniel Gildea, Jennifer Gillenwater, Dan Gillick, Kevin Gimpel, Roxana Girju, Claudio Giuliano, Amir Globerson, Yoav Goldberg, Sharon Goldwater, Julio Gonzalo, Cyril Goutte, Joao Graca, Spence Green, Charlie Greenbacker, Stephan Greene, Ralph Grishman, Jiafeng Guo, Iryna Gurevych, Adria de Gispert, Josef van Genabith

Nizar Habash, Ben Hachey, Barry Haddow, Patrick Haffner, Dilek Hakkani-Tur, David Hall, Keith Hall, Xianpei Han, Jirka Hana, Sanda Harabagiu, Christian Hardmeier, Kazi Saidul Hasan, Sasa Hasan, Chikara Hashimoto, Koiti Hasida, Ahmed Hassan, Helen Hastie, Katsuhiko Hayashi, Xiaodong He, Zhongjun He, Jeffrey Heinz, John Henderson, Iris Hendrickx, Graeme Hirst, Hieu Hoang, Julia Hockenmaier, Matt Honnibal, Mark Hopkins, Veronique Hoste, Eduard Hovy, Paul Hsu, Fei Huang, Liang Huang, Minlie Huang, Ruihong Huang, Zhongqiang Huang, Mans Hulden

Nancy Ide, Gonzalo Iglesias, Ryu Iida, Nitin Indurkha, Grant Ingersoll, Diana Inkpen, Pierre Isabelle, Mitsuru Ishizuka, Tatsuya Izuha, Aranta Diaz de Ilarraza, Gary Lee

Jagadeesh Jagarlamudi, Heng Ji, Sittichai Jiampojarn, Jing Jiang, Wenbin Jiang, Richard Johansson, Howard Johnson, Mark Johnson, Rie Johnson, Doug Jones, Vanja Josifovski

Min-Yen Kan, Damianos Karakos, Daisuke Kawahara, Tatsuya Kawahara, Jun'ichi Kazama, Frank Keller, Andre Kempe, Shahram Khadivi, Emre Kiciman, Bernd Kiefer, Jungi Kim, Su Nam Kim, Katrin Kirchhoff, Ioannis Klapaftis, Thomas Kleinbauer, Alexandre Klementiev, Kevin Knight, Philipp Koehn, Rob Koeling, Oskar Kohonen, Kazunori Komatani, Greg Kondrak, Moshe Koppel, Anna Korhonen, Andras Kornai, Zornitsa Kozareva, Emiel Kraemer, Lun-Wei Ku, Sandra Kuebler, Marco Kuhlmann, Roland Kuhn, Seth Kulick, Tom Kwiatkowski, Oi Yee Kwong

Mikel L. Forcada, Wai Lam, Mathias Lambert, Felipe Langlais, Mirella Lapata, Alberto Lavelli, Alon Lavie, Yoong Keok Lee, Oliver Lemon, James Lester, Gregor Leusch, Gina-Anne Levow, Roger Levy, William Lewis, Fangtao Li, Haizhou Li, Hang Li, Shoushan Li, Xiao Li, Zhifei Li, Percy Liang, Shasha Liao, Chuan-Jie Lin, Ken Litkowski, Marina Litvak, Bing Liu, Fei Liu, Qun Liu, Yan Liu, Yang Liu, Yi Liu, Elena Lloret, Adam Lopez, Annie Louis, Xiaofei Lu, Yue Lu, Yajuan Lv, Keith vander Linden

Bin Ma, Yanjun Ma, Klaus Macherey, Wolfgang Macherey, Nitin Madnani, Suresh Manandhar, Gideon Mann, Christopher Manning, Daniel Marcu, Katja Markert, Konstantin Markov, Erwin Marsi, Andre Martins, Yuval Marton, Spyros Matsoukas, Yuichiroh Matsubayashi, Yuji Matsumoto, Takuya Matsuzaki, Evgeny Matusov, Arne Mauser, Jon May, Diana Maynard, Andrew McCallum, Diana McCarthy, David McClosky, Kathy McCoy, Ryan McDonald, Tara McIntosh, Kathy McKeown, Paul McNamee, Arul Menezes, Paola Merlo, Donald Metzler, Adam Meyers, Haitao Mi, Jeff Mielke, Rada Mihalcea, Yusuke Miyao, Saif Mohammad, Emad Mohammed, Behrang Mohit, Karo Moilanen, Dan Moldovan, Christian Monson, Christof Monz, Taesun Moon, Robert Moore, Roser Morante, Hamish Morgan, Alessandro Moschitti, Smaranda Muresan, Gabriel Murray, Markos Mylonakis

Toshiaki Nakazawa, Preslav Nakov, Tahira Nasseem, Vivi Nastase, Roberto Navigli, Mark-Jan Nederhof, Graham Neubig, Günter Neumann, Hwee Tou Ng, Vincent Ng, Jian-Yun Nie, Zaiqing Nie, Joakim Nivre, Gertjan van Noord

Stephan Oepen, Jong-Hoon Oh, Manabu Okumura, Constantin Orasan

Martha Palmer, Sinno Pan, Bo Pang, Ivandre Paraboni, Christopher Parisien, Kristen P. Parton, Becky Passonneau, Siddharth Patwardhan, Michael Paul, Matthias Paulik, Adam Pauls, Adam Pease, Fuchun Peng, Jing Peng, Gerald Penn, Marco Pennacchiotti, Slav Petrov, Sasa Petrovic, Daniele Pighin, Manfred Pinkal, Elias Ponvert, Simone Paolo Ponzetto, Hoifung Poon, Andrei Popescu-Belis, Maja Popovic, Fred Popowich, Matt Post, Christopher Potts, David Powers, Sameer Pradhan, Rashmi Prasad, Daniel Preotiuc, Adam Przepiórkowski, Matthew Purver, James Pustejovsky, Sampo Pyysalo

Ariadna Quattoni

Rob Gaizauskas, Drago Radev, Dragomir Radev, Kira Radinsky, Hema Raghavan, Altaf Rahman, Daniel Ramage, Ganesh Ramakrishnan, Owen Rambow, Delip Rao, Ari Rappoport, Antoine Raux, Sujith Ravi, Emmanuel Rayner, Roi Reichart, Ehud Reiter, Norbert Reithinger, Sebastian Riedel, Jason Riesa, Stefan Riezler, German Rigau, Laura Rimell, Eric Ringger, Alan Ritter, Brian Roark, Horacio Rodríguez, Carolyn Rose, Andrew Rosenberg, Dan Roth, Vasile Rus, Alexander Rush, Graham Russell, Anton Rytting

Soo Ngee Koh, Kenji Sagae, Hassan Sajjad, Hasim Sak, Tapio Salakoski, Murat Saraclar, Anoop Sarkar, Sudeshna Sarkar, Christina Sauper, David Schlangen, Helmut Schmid, Nathan Schneider, William Schuler, Sabine Schulte im Walde, Hinrich Schütze, Satoshi Sekine, Jean Senellart, Violeta Seretan, Hendra Setiawan, Dipti Sharma, Libin Shen, Wade Shen, Shuming Shi, Eyal Shnarch, Luo Si, Candy Sidner, Michel Simard, Khalil Sima'an, Gabriel Skantze, Otakar Smrz, Rion Snow, Ben Snyder, Stephen Soderland, Yang Song, Youngin Song, Lucia Specia, Caroline Sporleder, Rohini Srihari, Manfred Stede, Mark Steedman, Josef Steinberger, Amanda Stent, Svetlana Stoyanchev, Veselin Stoyanov, Michael Strube, Keh-Yih Su, Fabian Suchanek, Weiwei Sun, Mihai Surdeanu, Hisami Suzuki, Jun Suzuki, Stan Szpakowicz, Idan Szpektor, Sien Moens, Diarmuid ó Séaghdha

Hiroya Takamura, Koichi Takeda, David Talbot, Partha Talukdar, Kumiko Tanaka-Ishii, Stefan Thater, Mariet Theune, Joerg Tiedemann, Christoph Tillmann, Ivan Titov, Cigdem Toprak, Kristina Toutanova, Roy Tromble, Junichi Tsujii, Yoshimasa Tsuruoka, Dan Tufis

Jakob Uszkoreit, Masao Utiyama

Benjamin Van Durme, Lucy Vanderwende, Vasudeva Varma, Tony Veale, Paola Velardi, Marc Vilain, David Vilar, Aline Villavicencio, Sami Virpioja, Andreas Vlachos, Piek Vossen

Marilyn Walker, Hanna Wallach, Michael Walsh, Stephen Wan, Xiaojun Wan, Haifeng Wang, Hsin-Min Wang, Leo Wanner, Taro Watanabe, Yotaro Watanabe, Bonnie Webber, Julie Weeds, Daniel S. Weld, Ben Wellner, Ji-Rong Wen, Michael Wiegand, Jason Williams, Theresa Wilson, Shuly Wintner, John Wong, Kam-Fai Wong, Tak-Lam Wong, Kristian Woodsend, Chung-Hsien Wu, Xianchao Wu, Fei Wu

Yunqing Xia, Lei Xie, Shasha Xie, Deyi Xiong, Gu Xu, Peng Xu, Nianwen Xue, Xiaobin Xue

Charles Yang, Muyun Yang, Shuang-Hong Yang, Roman Yangarber, Tae Yano, Alexander Yates, Xing Yi, Scott Wen-Tau Yih, Anssi Yli-Jyra, Dong Yu, Kai Yu, Liang-Chih Yu, Yisong Yue, Deniz Yuret, Yichang Yichang

Fabio Zanzotto, Jakub Zavrel, Klaus Zechner, Dmitry Zelenko, Richard Zens, Torsten Zesch, Luke Zettlemoyer, ChengXiang Zhai, Bing Zhang, Duo Zhang, Hui Zhang, Joy Zhang, Min Zhang, Qi Zhang, Yi Zhang, Yue Zhang, Liu Zhanyi, Bing Zhao, Jun Zhao, Shiqi Zhao, Tiejun Zhao, Jing Zheng, Guodong Zhou, Qiang Zhou, Michael Zock, Ingrid Zukerman

Invited Talk

Remembrance of ACLs past

Aravind K. Joshi

Henry Salvatori Professor of Computer and Cognitive Science
University of Pennsylvania

Abstract

Besides briefly covering some highlights of the past 50 years of ACL from my perspective, I will try to comment on (1) why some directions of research were pursued for a while and then dropped, sometimes for a good reason and sometimes apparently for no reason, (2) why the relationship to Linguistics, Psycholinguistics, and AI goes up and down, and (3) are there any leftovers that have the possibility of being turned into delicious contributions!

Short Bio

After completing his undergraduate work in Electrical and Communication Engineering in India, Aravind Joshi came to the University of Pennsylvania and obtained his Ph.D. in Electrical Engineering in 1960. At present, he is the Henry Salvatori Professor of Computer and Cognitive Science at the University of Pennsylvania.

Joshi has worked on several problems that overlap computer science and linguistics. More specifically, he has worked on topics in mathematical linguistics as they relate to formal and linguistic adequacy of different formalisms and their processing implications. He has also worked on several aspects of theories of representation and inference in natural language, especially as they relate to discourse.

Joshi was the President of ACL in 1975 and was appointed as a Founding Fellow of ACL in 2011. He was awarded the Lifetime Achievement Award of ACL in 2002, the David Rumelhart Prize of the Cognitive Science Society in 2003 and the Franklin Medal for Computer and Cognitive Science, Franklin Institute, Philadelphia, in 2005.

Invited Talk

Computational linguistics: Where do we go from here?

Mark Johnson

Professor of Language Sciences (CORE)
Director, Centre for Language Sciences (CLaS)
Department of Computing Faculty of Science
Macquarie University
Sydney, Australia

“Prediction is very difficult, especially about the future” —Niels Bohr

Abstract

The very fact that we’re having a 50th annual meeting means that our field hasn’t been a complete failure, but will there still be computational linguistics meetings in 50 years time? How do we fit into the larger intellectual picture, and what would it take to make computational linguistics into a real engineering discipline, or, for that matter, a scientific one? Prognosticating fearlessly (or perhaps just foolishly) I’ll draw some lessons from the last 50 years about what the next few might hold.

Short Bio

Mark Johnson is a Professor of Language Science (CORE) in the Department of Computing at Macquarie University. He was awarded a BSc (Hons) in 1979 from the University of Sydney, an MA in 1984 from the University of California, San Diego and a PhD in 1987 from Stanford University. He held a postdoctoral fellowship at MIT from 1987 until 1988, and has been a visiting researcher at the University of Stuttgart, the Xerox Research Centre in Grenoble, CSAIL at MIT and the Natural Language group at Microsoft Research. He has worked on a wide range of topics in computational linguistics, but his main research area is parsing and its applications to text and speech processing. He was President of the Association for Computational Linguistics in 2003, and was a professor from 1989 until 2009 in the Departments of Cognitive and Linguistic Sciences and Computer Science at Brown University.

Table of Contents

<i>Learning to Translate with Multiple Objectives</i>	
Kevin Duh, Katsuhito Sudoh, Xianchao Wu, Hajime Tsukada and Masaaki Nagata	1
<i>Joint Feature Selection in Distributed Stochastic Learning for Large-Scale Discriminative Training in SMT</i>	
Patrick Simianer, Stefan Riezler and Chris Dyer	11
<i>Prediction of Learning Curves in Machine Translation</i>	
Prasanth Kolachina, Nicola Cancedda, Marc Dymetman and Sriram Venkatapathy	22
<i>Probabilistic Integration of Partial Lexical Information for Noise Robust Haptic Voice Recognition</i>	
Khe Chai Sim	31
<i>A Nonparametric Bayesian Approach to Acoustic Model Discovery</i>	
Chia-ying Lee and James Glass	40
<i>Automated Essay Scoring Based on Finite State Transducer: towards ASR Transcription of Oral English Speech</i>	
Xingyuan Peng, Dengfeng Ke and Bo Xu	50
<i>Text-level Discourse Parsing with Rich Linguistic Features</i>	
Vanessa Wei Feng and Graeme Hirst	60
<i>PDTB-style Discourse Annotation of Chinese Text</i>	
Yuping Zhou and Nianwen Xue	69
<i>SITS: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multi-party Conversations</i>	
Viet-An Nguyen, Jordan Boyd-Graber and Philip Resnik	78
<i>Extracting Narrative Timelines as Temporal Dependency Structures</i>	
Oleksandr Kolomiyets, Steven Bethard and Marie-Francine Moens	88
<i>Labeling Documents with Timestamps: Learning from their Time Expressions</i>	
Nathanael Chambers	98
<i>Temporally Anchored Relation Extraction</i>	
Guillermo Garrido, Anselmo Peñas, Bernardo Cabaleiro and Álvaro Rodrigo	107
<i>Efficient Tree-based Approximation for Entailment Graph Learning</i>	
Jonathan Berant, Ido Dagan, Meni Adler and Jacob Goldberger	117
<i>Learning High-Level Planning from Text</i>	
S.R.K. Branavan, Nate Kushman, Tao Lei and Regina Barzilay	126

<i>Distributional Semantics in Technicolor</i>	
Elia Bruni, Gemma Boleda, Marco Baroni and Nam Khanh Tran	136
<i>A Class-Based Agreement Model for Generating Accurately Inflected Translations</i>	
Spence Green and John DeNero	146
<i>Deciphering Foreign Language by Combining Language Models and Context Vectors</i>	
Malte Nuhn, Arne Mauser and Hermann Ney	156
<i>Machine Translation without Words through Substring Alignment</i>	
Graham Neubig, Taro Watanabe, Shinsuke Mori and Tatsuya Kawahara	165
<i>Fast Syntactic Analysis for Statistical Language Modeling via Substructure Sharing and Uptraining</i>	
Ariya Rastrow, Mark Dredze and Sanjeev Khudanpur	175
<i>Bootstrapping a Unified Model of Lexical and Phonetic Acquisition</i>	
Micha Elsner, Sharon Goldwater and Jacob Eisenstein	184
<i>Discriminative Pronunciation Modeling: A Large-Margin, Feature-Rich Approach</i>	
Hao Tang, Joseph Keshet and Karen Livescu	194
<i>Discriminative Strategies to Integrate Multiword Expression Recognition and Parsing</i>	
Matthieu Constant, Anthony Sigogne and Patrick Watrin	204
<i>Utilizing Dependency Language Models for Graph-based Dependency Parsing Models</i>	
Wenliang Chen, Min Zhang and Haizhou Li	213
<i>Spectral Learning of Latent-Variable PCFGs</i>	
Shay B. Cohen, Karl Stratos, Michael Collins, Dean P. Foster and Lyle Ungar	223
<i>Reducing Approximation and Estimation Errors for Chinese Lexical Processing with Heterogeneous Annotations</i>	
Weiwei Sun and Xiaojun Wan	232
<i>Capturing Paradigmatic and Syntagmatic Lexical Relations: Towards Accurate Chinese Part-of-Speech Tagging</i>	
Weiwei Sun and Hans Uszkoreit	242
<i>Fast Online Training with Frequency-Adaptive Learning Rates for Chinese Word Segmentation and New Word Detection</i>	
Xu Sun, Houfeng Wang and Wenjie Li	253
<i>Verb Classification using Distributional Similarity in Syntactic and Semantic Structures</i>	
Danilo Croce, Alessandro Moschitti, Roberto Basili and Martha Palmer	263
<i>Word Sense Disambiguation Improves Information Retrieval</i>	
Zhi Zhong and Hwee Tou Ng	273
<i>Efficient Search for Transformation-based Inference</i>	
Asher Stern, Roni Stern, Ido Dagan and Ariel Felner	283

<i>Maximum Expected BLEU Training of Phrase and Lexicon Translation Models</i> Xiaodong He and Li Deng	292
<i>Learning Translation Consensus with Structured Label Propagation</i> Shujie Liu, Chi-Ho Li, Mu Li and Ming Zhou	302
<i>Smaller Alignment Models for Better Translations: Unsupervised Word Alignment with the l0-norm</i> Ashish Vaswani, Liang Huang and David Chiang	311
<i>Modeling Review Comments</i> Arjun Mukherjee and Bing Liu	320
<i>A Joint Model for Discovery of Aspects in Utterances</i> Asli Celikyilmaz and Dilek Hakkani-Tur	330
<i>Aspect Extraction through Semi-Supervised Modeling</i> Arjun Mukherjee and Bing Liu	339
<i>Learning to "Read Between the Lines" using Bayesian Logic Programs</i> Sindhu Raghavan, Raymond Mooney and Hyeonseo Ku	349
<i>Collective Generation of Natural Image Descriptions</i> Polina Kuznetsova, Vicente Ordonez, Alexander Berg, Tamara Berg and Yejin Choi	359
<i>Concept-to-text Generation via Discriminative Reranking</i> Ioannis Konstas and Mirella Lapata	369
<i>A Discriminative Hierarchical Model for Fast Coreference at Large Scale</i> Michael Wick, Sameer Singh and Andrew McCallum	379
<i>Coreference Semantics from Web Features</i> Mohit Bansal and Dan Klein	389
<i>Subgroup Detection in Ideological Discussions</i> Amjad Abu-Jbara, Pradeep Dasigi, Mona Diab and Dragomir Radev	399
<i>Cross-Domain Co-Extraction of Sentiment and Topic Lexicons</i> Fangtao Li, Sinno Jialin Pan, Ou Jin, Qiang Yang and Xiaoyan Zhu	410
<i>Learning Syntactic Verb Frames using Graphical Models</i> Thomas Lippincott, Anna Korhonen and Diarmuid Ó Séaghdha	420
<i>Fast Online Lexicon Learning for Grounded Language Acquisition</i> David Chen	430
<i>Bayesian Symbol-Refined Tree Substitution Grammars for Syntactic Parsing</i> Hiroyuki Shindo, Yusuke Miyao, Akinori Fujino and Masaaki Nagata	440
<i>String Re-writing Kernel</i> Fan Bu, Hang Li and Xiaoyan Zhu	449

<i>Translation Model Adaptation for Statistical Machine Translation with Monolingual Topic Information</i> Jinsong Su, Hua Wu, Haifeng Wang, Yidong Chen, Xiaodong Shi, Huailin Dong and Qun Liu	459
<i>A Statistical Model for Unsupervised and Semi-supervised Transliteration Mining</i> Hassan Sajjad, Alexander Fraser and Helmut Schmid	469
<i>Modified Distortion Matrices for Phrase-Based Statistical Machine Translation</i> Arianna Bisazza and Marcello Federico	478
<i>Semantic Parsing with Bayesian Tree Transducers</i> Bevan Jones, Mark Johnson and Sharon Goldwater	488
<i>Dependency Hashing for n-best CCG Parsing</i> Dominick Ng and James R. Curran	497
<i>Strong Lexicalization of Tree Adjoining Grammars</i> Andreas Maletti and Joost Engelfriet	506
<i>Tweet Recommendation with Graph Co-Ranking</i> Rui Yan, Mirella Lapata and Xiaoming Li	516
<i>Joint Inference of Named Entity Recognition and Normalization for Tweets</i> Xiaohua Liu, Ming Zhou, Xiangyang Zhou, Zhongyang Fu and Furu Wei	526
<i>Finding Bursty Topics from Microblogs</i> Qiming Diao, Jing Jiang, Feida Zhu and Ee-Peng Lim	536
<i>Spice it up? Mining Refinements to Online Instructions from User Generated Content</i> Gregory Druck and Bo Pang	545
<i>Sentence Dependency Tagging in Online Question Answering Forums</i> Zhonghua Qu and Yang Liu	554
<i>Mining Entity Types from Query Logs via User Intent Modeling</i> Patrick Pantel, Thomas Lin and Michael Gamon	563
<i>Cross-Lingual Mixture Model for Sentiment Classification</i> Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu and Houfeng Wang	572
<i>Community Answer Summarization for Multi-Sentence Question with Group L1 Regularization</i> Wen Chan, Xiangdong Zhou, Wei Wang and Tat-Seng Chua	582
<i>Error Mining on Dependency Trees</i> Claire Gardent and Shashi Narayan	592
<i>Computational Approaches to Sentence Completion</i> Geoffrey Zweig, John C. Platt, Christopher Meek, Christopher J.C. Burges, Ainur Yessenalina and Qiang Liu	601

<i>Iterative Viterbi A* Algorithm for K-Best Sequential Decoding</i>	
Zhiheng Huang, Yi Chang, Bo Long, Jean-Francois Crespo, Anlei Dong, Sathiya Keerthi and Su-Lin Wu	611
<i>Bootstrapping via Graph Propagation</i>	
Max Whitney and Anoop Sarkar	620
<i>Selective Sharing for Multilingual Dependency Parsing</i>	
Tahira Naseem, Regina Barzilay and Amir Globerson	629
<i>The Creation of a Corpus of English Metalanguage</i>	
Shomir Wilson	638
<i>Crosslingual Induction of Semantic Roles</i>	
Ivan Titov and Alexandre Klementiev	647
<i>Head-driven Transition-based Parsing with Top-down Prediction</i>	
Katsuhiko Hayashi, Taro Watanabe, Masayuki Asahara and Yuji Matsumoto	657
<i>MIX Is Not a Tree-Adjoining Language</i>	
Makoto Kanazawa and Sylvain Salvati	666
<i>Exploiting Multiple Treebanks for Parsing with Quasi-synchronous Grammars</i>	
Zhenghua Li, Ting Liu and Wanxiang Che	675
<i>A Probabilistic Model for Canonicalizing Named Entity Mentions</i>	
Dani Yogatama, Yanchuan Sim and Noah A. Smith	685
<i>Multilingual Named Entity Recognition using Parallel Data and Metadata from Wikipedia</i>	
Sungchul Kim, Kristina Toutanova and Hwanjo Yu	694
<i>A Computational Approach to the Automation of Creative Naming</i>	
Gozde Ozbal and Carlo Strapparava	703
<i>Unsupervised Relation Discovery with Sense Disambiguation</i>	
Limin Yao, Sebastian Riedel and Andrew McCallum	712
<i>Reducing Wrong Labels in Distant Supervision for Relation Extraction</i>	
Shingo Takamatsu, Issei Sato and Hiroshi Nakagawa	721
<i>Finding Salient Dates for Building Thematic Timelines</i>	
Rémy Kessler, Xavier Tannier, Caroline Hagège, Véronique Moriceau and André Bittar	730
<i>Historical Analysis of Legal Opinions with a Sparse Mixed-Effects Latent Variable Model</i>	
William Yang Wang, Elijah Mayfield, Suresh Naidu and Jeremiah Dittmar	740
<i>A Topic Similarity Model for Hierarchical Phrase-based Translation</i>	
Xinyan Xiao, Deyi Xiong, Min Zhang, Qun Liu and Shouxun Lin	750

<i>Modeling Topic Dependencies in Hierarchical Text Categorization</i>	
Alessandro Moschitti, Qi Ju and Richard Johansson	759
<i>Attacking Parsing Bottlenecks with Unlabeled Data and Relevant Factorizations</i>	
Emily Pitler	768
<i>Semi-supervised Dependency Parsing using Lexical Affinities</i>	
Seyed Abolghasem Mirroshandel, Alexis Nasr and Joseph Le Roux	777
<i>Chinese Comma Disambiguation for Discourse Analysis</i>	
Yaqin Yang and Nianwen Xue	786
<i>Collective Classification for Fine-grained Information Status</i>	
Katja Markert, Yufang Hou and Michael Strube	795
<i>Structuring E-Commerce Inventory</i>	
Karin Mauge, Khash Rohanimanesh and Jean-David Ruvini	805
<i>Named Entity Disambiguation in Streaming Data</i>	
Alexandre Davis, Adriano Veloso, Altigran Soares, Alberto Laender and Wagner Meira Jr.	815
<i>Big Data versus the Crowd: Looking for Relationships in All the Right Places</i>	
Ce Zhang, Feng Niu, Christopher Ré and Jude Shavlik	825
<i>Automatic Event Extraction with Structured Preference Modeling</i>	
Wei Lu and Dan Roth	835
<i>Discriminative Learning for Joint Template Filling</i>	
Einat Minkov and Luke Zettlemoyer	845
<i>Classifying French Verbs Using French and English Lexical Resources</i>	
Ingrid Falk, Claire Gardent and Jean-Charles Lamirel	854
<i>Modeling Sentences in the Latent Space</i>	
Weimei Guo and Mona Diab	864
<i>Improving Word Representations via Global Context and Multiple Word Prototypes</i>	
Eric Huang, Richard Socher, Christopher Manning and Andrew Ng	873
<i>Exploiting Social Information in Grounded Language Learning via Grammatical Reduction</i>	
Mark Johnson, Katherine Demuth and Michael Frank	883
<i>You Had Me at Hello: How Phrasing Affects Memorability</i>	
Cristian Danescu-Niculescu-Mizil, Justin Cheng, Jon Kleinberg and Lillian Lee	892
<i>Modeling the Translation of Predicate-Argument Structure for SMT</i>	
Deyi Xiong, Min Zhang and Haizhou Li	902
<i>A Ranking-based Approach to Word Reordering for Statistical Machine Translation</i>	
Nan Yang, Mu Li, Dongdong Zhang and Nenghai Yu	912

<i>Character-Level Machine Translation Evaluation for Languages with Ambiguous Word Boundaries</i> Chang Liu and Hwee Tou Ng	921
<i>PORT: a Precision-Order-Recall MT Evaluation Metric for Tuning</i> Boxing Chen, Roland Kuhn and Samuel Larkin	930
<i>Mixing Multiple Translation Models in Statistical Machine Translation</i> Majid Razmara, George Foster, Baskaran Sankaran and Anoop Sarkar	940
<i>Hierarchical Chunk-to-String Translation</i> Yang Feng, Dongdong Zhang, Mu Li and Qun Liu	950
<i>Large-Scale Syntactic Language Modeling with Treelets</i> Adam Pauls and Dan Klein	959
<i>Text Segmentation by Language Using Minimum Description Length</i> Hiroshi Yamaguchi and Kumiko Tanaka-Ishii	969
<i>Improve SMT Quality with Automatically Extracted Paraphrase Rules</i> Wei He, Hua Wu, Haifeng Wang and Ting Liu	979
<i>Ecological Evaluation of Persuasive Messages Using Google AdWords</i> Marco Guerini, Carlo Strapparava and Oliviero Stock	988
<i>Polarity Consistency Checking for Sentiment Dictionaries</i> Eduard Dragut, Hong Wang, Clement Yu, Prasad Sistla and Weiyi Meng	997
<i>Combining Coherence Models and Machine Translation Evaluation Metrics for Summarization Evaluation</i> Ziheng Lin, Chang Liu, Hwee Tou Ng and Min-Yen Kan	1006
<i>Sentence Simplification by Monolingual Machine Translation</i> Sander Wubben, Antal van den Bosch and Emiel Krahmer	1015
<i>A Cost Sensitive Part-of-Speech Tagging: Differentiating Serious Errors from Minor Errors</i> Hyun-Je Song, Jeong-Woo Son, Tae-Gil Noh, Seong-Bae Park and Sang-Jo Lee	1025
<i>A Broad-Coverage Normalization System for Social Media Language</i> Fei Liu, Fuliang Weng and Xiao Jiang	1035
<i>Incremental Joint Approach to Word Segmentation, POS Tagging, and Dependency Parsing in Chinese</i> Jun Hatori, Takuya Matsuzaki, Yusuke Miyao and Jun'ichi Tsujii	1045
<i>Exploring Deterministic Constraints: from a Constrained English POS Tagger to an Efficient ILP Solution to Chinese Word Segmentation</i> Qiuye Zhao and Mitch Marcus	1054

Conference Program

Monday July 9, 2012

(9:00 – 10:30) Invited Talk: Remembrance of ACLs past, by Aravind K. Joshi

Session 1a: (11:00 – 12:30) Machine Translation

Learning to Translate with Multiple Objectives

Kevin Duh, Katsuhito Sudoh, Xianchao Wu, Hajime Tsukada and Masaaki Nagata

Joint Feature Selection in Distributed Stochastic Learning for Large-Scale Discriminative Training in SMT

Patrick Simianer, Stefan Riezler and Chris Dyer

Prediction of Learning Curves in Machine Translation

Prasanth Kolachina, Nicola Cancedda, Marc Dymetman and Sriram Venkatapathy

Session 1b: (11:00 – 12:30) Speech

Probabilistic Integration of Partial Lexical Information for Noise Robust Haptic Voice Recognition

Khe Chai Sim

A Nonparametric Bayesian Approach to Acoustic Model Discovery

Chia-ying Lee and James Glass

Automated Essay Scoring Based on Finite State Transducer: towards ASR Transcription of Oral English Speech

Xingyuan Peng, Dengfeng Ke and Bo Xu

Monday July 9, 2012 (continued)

Session 1c: (11:00 – 12:30) Discourse

Text-level Discourse Parsing with Rich Linguistic Features

Vanessa Wei Feng and Graeme Hirst

PDTB-style Discourse Annotation of Chinese Text

Yuping Zhou and Nianwen Xue

SITS: A Hierarchical Nonparametric Model using Speaker Identity for Topic Segmentation in Multiparty Conversations

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

Session 1d: (11:00 – 12:30) Time

Extracting Narrative Timelines as Temporal Dependency Structures

Oleksandr Kolomiyets, Steven Bethard and Marie-Francine Moens

Labeling Documents with Timestamps: Learning from their Time Expressions

Nathanael Chambers

Temporally Anchored Relation Extraction

Guillermo Garrido, Anselmo Peñas, Bernardo Cabaleiro and Álvaro Rodrigo

Session 1e: (11:00 – 12:30) Meaning

Efficient Tree-based Approximation for Entailment Graph Learning

Jonathan Berant, Ido Dagan, Meni Adler and Jacob Goldberger

Learning High-Level Planning from Text

S.R.K. Branavan, Nate Kushman, Tao Lei and Regina Barzilay

Distributional Semantics in Technicolor

Elia Bruni, Gemma Boleda, Marco Baroni and Nam Khanh Tran

Monday July 9, 2012 (continued)

Session 2a: (14:00 – 15:30) Machine Translation

A Class-Based Agreement Model for Generating Accurately Inflected Translations

Spence Green and John DeNero

Deciphering Foreign Language by Combining Language Models and Context Vectors

Malte Nuhn, Arne Mauser and Hermann Ney

Machine Translation without Words through Substring Alignment

Graham Neubig, Taro Watanabe, Shinsuke Mori and Tatsuya Kawahara

Session 2b: (14:00 – 15:30) Speech

Fast Syntactic Analysis for Statistical Language Modeling via Substructure Sharing and Uptraining

Ariya Rastrow, Mark Dredze and Sanjeev Khudanpur

Bootstrapping a Unified Model of Lexical and Phonetic Acquisition

Micha Elsner, Sharon Goldwater and Jacob Eisenstein

Discriminative Pronunciation Modeling: A Large-Margin, Feature-Rich Approach

Hao Tang, Joseph Keshet and Karen Livescu

Session 2c: (14:00 – 15:30) Parsing

Discriminative Strategies to Integrate Multiword Expression Recognition and Parsing

Matthieu Constant, Anthony Sigogne and Patrick Watrin

Utilizing Dependency Language Models for Graph-based Dependency Parsing Models

Wenliang Chen, Min Zhang and Haizhou Li

Spectral Learning of Latent-Variable PCFGs

Shay B. Cohen, Karl Stratos, Michael Collins, Dean P. Foster and Lyle Ungar

Monday July 9, 2012 (continued)

Session 2d: (14:00 – 15:30) Chinese lexical processing

Reducing Approximation and Estimation Errors for Chinese Lexical Processing with Heterogeneous Annotations

Weiwei Sun and Xiaojun Wan

Capturing Paradigmatic and Syntagmatic Lexical Relations: Towards Accurate Chinese Part-of-Speech Tagging

Weiwei Sun and Hans Uszkoreit

Fast Online Training with Frequency-Adaptive Learning Rates for Chinese Word Segmentation and New Word Detection

Xu Sun, Houfeng Wang and Wenjie Li

Session 2e: (14:00 – 15:30) Lexical semantics

Verb Classification using Distributional Similarity in Syntactic and Semantic Structures

Danilo Croce, Alessandro Moschitti, Roberto Basili and Martha Palmer

Word Sense Disambiguation Improves Information Retrieval

Zhi Zhong and Hwee Tou Ng

Efficient Search for Transformation-based Inference

Asher Stern, Roni Stern, Ido Dagan and Ariel Felner

Session 3a: (16:00 – 17:30) Machine Translation

Maximum Expected BLEU Training of Phrase and Lexicon Translation Models

Xiaodong He and Li Deng

Learning Translation Consensus with Structured Label Propagation

Shujie Liu, Chi-Ho Li, Mu Li and Ming Zhou

Smaller Alignment Models for Better Translations: Unsupervised Word Alignment with the l0-norm

Ashish Vaswani, Liang Huang and David Chiang

Monday July 9, 2012 (continued)

Session 3b: (16:00 – 17:30) Aspect

Modeling Review Comments

Arjun Mukherjee and Bing Liu

A Joint Model for Discovery of Aspects in Utterances

Asli Celikyilmaz and Dilek Hakkani-Tur

Aspect Extraction through Semi-Supervised Modeling

Arjun Mukherjee and Bing Liu

Session 3c: (16:00 – 17:30) Generation and Machine Reading

Learning to "Read Between the Lines" using Bayesian Logic Programs

Sindhu Raghavan, Raymond Mooney and Hyeonsoo Ku

Collective Generation of Natural Image Descriptions

Polina Kuznetsova, Vicente Ordonez, Alexander Berg, Tamara Berg and Yejin Choi

Concept-to-text Generation via Discriminative Reranking

Ioannis Konstas and Mirella Lapata

Session 3d: (16:00 – 17:30) Dialogue and Discourse

A Discriminative Hierarchical Model for Fast Coreference at Large Scale

Michael Wick, Sameer Singh and Andrew McCallum

Coreference Semantics from Web Features

Mohit Bansal and Dan Klein

Subgroup Detection in Ideological Discussions

Amjad Abu-Jbara, Pradeep Dasigi, Mona Diab and Dragomir Radev

Monday July 9, 2012 (continued)

Session 3e: (16:00 – 17:30) Lexicon

Cross-Domain Co-Extraction of Sentiment and Topic Lexicons

Fangtao Li, Sinno Jialin Pan, Ou Jin, Qiang Yang and Xiaoyan Zhu

Learning Syntactic Verb Frames using Graphical Models

Thomas Lippincott, Anna Korhonen and Diarmuid Ó Séaghdha

Fast Online Lexicon Learning for Grounded Language Acquisition

David Chen

(18:00 – 20:30) Poster Session

Tuesday July 10, 2012

(9:00 – 10:30) Best Paper Session

Bayesian Symbol-Refined Tree Substitution Grammars for Syntactic Parsing

Hiroyuki Shindo, Yusuke Miyao, Akinori Fujino and Masaaki Nagata

String Re-writing Kernel

Fan Bu, Hang Li and Xiaoyan Zhu

Session 4b: (11:00 – 12:30) Machine Translation

Translation Model Adaptation for Statistical Machine Translation with Monolingual Topic Information

Jinsong Su, Hua Wu, Haifeng Wang, Yidong Chen, Xiaodong Shi, Huailin Dong and Qun Liu

A Statistical Model for Unsupervised and Semi-supervised Transliteration Mining

Hassan Sajjad, Alexander Fraser and Helmut Schmid

Modified Distortion Matrices for Phrase-Based Statistical Machine Translation

Arianna Bisazza and Marcello Federico

Tuesday July 10, 2012 (continued)

Session 4c: (11:00 – 12:30) Parsing

Semantic Parsing with Bayesian Tree Transducers

Bevan Jones, Mark Johnson and Sharon Goldwater

Dependency Hashing for n-best CCG Parsing

Dominick Ng and James R. Curran

Strong Lexicalization of Tree Adjoining Grammars

Andreas Maletti and Joost Engelfriet

Session 4d: (11:00 – 12:30) Social Media

Tweet Recommendation with Graph Co-Ranking

Rui Yan, Mirella Lapata and Xiaoming Li

Joint Inference of Named Entity Recognition and Normalization for Tweets

Xiaohua Liu, Ming Zhou, Xiangyang Zhou, Zhongyang Fu and Furu Wei

Finding Bursty Topics from Microblogs

Qiming Diao, Jing Jiang, Feida Zhu and Ee-Peng Lim

Session 4e: (11:00 – 12:30) User generated content

Spice it up? Mining Refinements to Online Instructions from User Generated Content

Gregory Druck and Bo Pang

Sentence Dependency Tagging in Online Question Answering Forums

Zhonghua Qu and Yang Liu

Mining Entity Types from Query Logs via User Intent Modeling

Patrick Pantel, Thomas Lin and Michael Gamon

Tuesday July 10, 2012 (continued)

Session 5b: (14:00 – 15:30) NLP Apps

Cross-Lingual Mixture Model for Sentiment Classification

Xinfan Meng, Furu Wei, Xiaohua Liu, Ming Zhou, Ge Xu and Houfeng Wang

Community Answer Summarization for Multi-Sentence Question with Group L1 Regularization

Wen Chan, Xiangdong Zhou, Wei Wang and Tat-Seng Chua

Error Mining on Dependency Trees

Claire Gardent and Shashi Narayan

Session 5c: (14:00 – 15:30) Machine Learning

Computational Approaches to Sentence Completion

Geoffrey Zweig, John C. Platt, Christopher Meek, Christopher J.C. Burges, Ainur Yesse-
nalina and Qiang Liu

Iterative Viterbi A Algorithm for K-Best Sequential Decoding*

Zhiheng Huang, Yi Chang, Bo Long, Jean-Francois Crespo, Anlei Dong, Sathiya Keerthi
and Su-Lin Wu

Bootstrapping via Graph Propagation

Max Whitney and Anoop Sarkar

Session 5d: (14:00 – 15:30) Multilinguality

Selective Sharing for Multilingual Dependency Parsing

Tahira Naseem, Regina Barzilay and Amir Globerson

The Creation of a Corpus of English Metalanguage

Shomir Wilson

Crosslingual Induction of Semantic Roles

Ivan Titov and Alexandre Klementiev

Tuesday July 10, 2012 (continued)

Session 5e: (14:00 – 15:30) Parsing

Head-driven Transition-based Parsing with Top-down Prediction

Katsuhiko Hayashi, Taro Watanabe, Masayuki Asahara and Yuji Matsumoto

MIX Is Not a Tree-Adjoining Language

Makoto Kanazawa and Sylvain Salvati

Exploiting Multiple Treebanks for Parsing with Quasi-synchronous Grammars

Zhenghua Li, Ting Liu and Wanxiang Che

Session 6b: (16:00 – 17:30) Names

A Probabilistic Model for Canonicalizing Named Entity Mentions

Dani Yogatama, Yanchuan Sim and Noah A. Smith

Multilingual Named Entity Recognition using Parallel Data and Metadata from Wikipedia

Sungchul Kim, Kristina Toutanova and Hwanjo Yu

A Computational Approach to the Automation of Creative Naming

Gozde Ozbal and Carlo Strapparava

Session 6c: (16:00 – 17:30) Relations

Unsupervised Relation Discovery with Sense Disambiguation

Limin Yao, Sebastian Riedel and Andrew McCallum

Reducing Wrong Labels in Distant Supervision for Relation Extraction

Shingo Takamatsu, Issei Sato and Hiroshi Nakagawa

Finding Salient Dates for Building Thematic Timelines

Rémy Kessler, Xavier Tannier, Caroline Hagège, Véronique Moriceau and André Bittar

Tuesday July 10, 2012 (continued)

Session 6d: (16:00 – 17:30) Topics

Historical Analysis of Legal Opinions with a Sparse Mixed-Effects Latent Variable Model

William Yang Wang, Elijah Mayfield, Suresh Naidu and Jeremiah Dittmar

A Topic Similarity Model for Hierarchical Phrase-based Translation

Xinyan Xiao, Deyi Xiong, Min Zhang, Qun Liu and Shouxun Lin

Modeling Topic Dependencies in Hierarchical Text Categorization

Alessandro Moschitti, Qi Ju and Richard Johansson

Session 6e: (16:00 – 17:30) Parsing

Attacking Parsing Bottlenecks with Unlabeled Data and Relevant Factorizations

Emily Pitler

Semi-supervised Dependency Parsing using Lexical Affinities

Seyed Abolghasem Mirroshandel, Alexis Nasr and Joseph Le Roux

Wednesday July 11, 2012

**(9:00 – 10:30) Invited Talk: Computational linguistics: Where do we go from here?,
by Mark Johnson**

Monday July 9, 2012

(18:00 – 20:30) Poster Session

Chinese Comma Disambiguation for Discourse Analysis

Yaqin Yang and Nianwen Xue

Collective Classification for Fine-grained Information Status

Katja Markert, Yufang Hou and Michael Strube

Structuring E-Commerce Inventory

Karin Mauge, Khash Rohanimanesh and Jean-David Ruvini

Monday July 9, 2012 (continued)

Named Entity Disambiguation in Streaming Data

Alexandre Davis, Adriano Veloso, Altigran Soares, Alberto Laender and Wagner Meira Jr.

Big Data versus the Crowd: Looking for Relationships in All the Right Places

Ce Zhang, Feng Niu, Christopher Ré and Jude Shavlik

Automatic Event Extraction with Structured Preference Modeling

Wei Lu and Dan Roth

Discriminative Learning for Joint Template Filling

Einat Minkov and Luke Zettlemoyer

Classifying French Verbs Using French and English Lexical Resources

Ingrid Falk, Claire Gardent and Jean-Charles Lamirel

Modeling Sentences in the Latent Space

Weiwei Guo and Mona Diab

Improving Word Representations via Global Context and Multiple Word Prototypes

Eric Huang, Richard Socher, Christopher Manning and Andrew Ng

Exploiting Social Information in Grounded Language Learning via Grammatical Reduction

Mark Johnson, Katherine Demuth and Michael Frank

You Had Me at Hello: How Phrasing Affects Memorability

Cristian Danescu-Niculescu-Mizil, Justin Cheng, Jon Kleinberg and Lillian Lee

Modeling the Translation of Predicate-Argument Structure for SMT

Deyi Xiong, Min Zhang and Haizhou Li

A Ranking-based Approach to Word Reordering for Statistical Machine Translation

Nan Yang, Mu Li, Dongdong Zhang and Nenghai Yu

Character-Level Machine Translation Evaluation for Languages with Ambiguous Word Boundaries

Chang Liu and Hwee Tou Ng

Monday July 9, 2012 (continued)

PORT: a Precision-Order-Recall MT Evaluation Metric for Tuning

Boxing Chen, Roland Kuhn and Samuel Larkin

Mixing Multiple Translation Models in Statistical Machine Translation

Majid Razmara, George Foster, Baskaran Sankaran and Anoop Sarkar

Hierarchical Chunk-to-String Translation

Yang Feng, Dongdong Zhang, Mu Li and Qun Liu

Large-Scale Syntactic Language Modeling with Treelets

Adam Pauls and Dan Klein

Text Segmentation by Language Using Minimum Description Length

Hiroshi Yamaguchi and Kumiko Tanaka-Ishii

Improve SMT Quality with Automatically Extracted Paraphrase Rules

Wei He, Hua Wu, Haifeng Wang and Ting Liu

Ecological Evaluation of Persuasive Messages Using Google AdWords

Marco Guerini, Carlo Strapparava and Oliviero Stock

Polarity Consistency Checking for Sentiment Dictionaries

Eduard Dragut, Hong Wang, Clement Yu, Prasad Sistla and Weiyi Meng

Combining Coherence Models and Machine Translation Evaluation Metrics for Summarization Evaluation

Ziheng Lin, Chang Liu, Hwee Tou Ng and Min-Yen Kan

Sentence Simplification by Monolingual Machine Translation

Sander Wubben, Antal van den Bosch and Emiel Krahmer

A Cost Sensitive Part-of-Speech Tagging: Differentiating Serious Errors from Minor Errors

Hyun-Je Song, Jeong-Woo Son, Tae-Gil Noh, Seong-Bae Park and Sang-Jo Lee

A Broad-Coverage Normalization System for Social Media Language

Fei Liu, Fuliang Weng and Xiao Jiang

Monday July 9, 2012 (continued)

Incremental Joint Approach to Word Segmentation, POS Tagging, and Dependency Parsing in Chinese

Jun Hatori, Takuya Matsuzaki, Yusuke Miyao and Jun'ichi Tsujii

Exploring Deterministic Constraints: from a Constrained English POS Tagger to an Efficient ILP Solution to Chinese Word Segmentation

Qiuye Zhao and Mitch Marcus

