

The 48th Annual Meeting of the Association for Computational Linguistics

# Conference Proceedings





UPPSALA UNIVERSITET

# ACL 2010

# 48th Annual Meeting of the Association for Computational Linguistics

**Proceedings of the Conference** 

11-16 July 2010 Uppsala University Uppsala, Sweden Production and Manufacturing by Taberg Media Group AB Box 94, 562 02 Taberg Sweden

©2010 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

ISBN 978-1-932432-66-4 / 1-932432-66-3 (Volume 1) ISBN 978-1-932432-67-1 / 1-932432-67-1 (Volume 2)

# **Preface: From the General Chair**

Welcome back to Europe at ACL 2010! After three years, the ACL crowd is meeting again in Europe, this time at the very north, to escape from the Central European heat it experienced in 2007.

This year, some significant changes can be found under the hood. The call for papers was formulated much more broadly than usual, and this idea brought up by the ACL membership and the Exec and then developed in detail by this year's program chairs, Sandra Carberry and Stephen Clark, really caught on - the number of submissions has been the highest of all times, forcing us to put some activities, such as the SRW, as the fifth track on Tuesday morning. The number of reviewers is hard to compute exactly - but a glimpse into their lists in this year's and previous years' proceedings reveals that we almost certainly set a new record here, too (thank you all!). Also, the proceedings have switched to electronic-only for all events, and adaptation of the START conference automation software has begun towards a fully automated workflow from submission to the production of the final proceedings in pdf format. It has been made possible thanks to Philipp Koehn's and Jing-Shin Chang's willingness to serve as Publication Chairs two years in a row in order to ensure a smooth transition from the semi-manual process employed in the past. However, there was one thing that overshadowed it all: the enthusiastic, meticulously precise and absolutely professional yet in every situation very polite approach of the local arrangements committee headed by Joakim Nivre. His efforts have made my job, as the General Chair, a piece of cake, limited essentially to watching the tons of emails exchanged between the local and other committees and to answering emails like "why wasn't I asked to be an invited speaker?" (obviously, from people no one would consider for this honor anyway).

Joakim has been helped by Beáta Megyesi, Rolf Carlson, Mats Dahllöf, Marco Kuhlmann, Mattias Nilsson, Markus Saers, Anna Sågvall Hein, Per Starbäck, Oscar Täckström, Jörg Tiedemann, Reut Tsarfaty and by the Akademikonferens team affiliated to Uppsala University headed by Ulla Conti; from her team, I would like to thank specifically Maria Carlson, Maria Bäckström and Johanna Thyselius Nilsson for taking care of the website.

There are traditional ACL conference features as well - the workshops (with CoNLL-2010 as the big one), tutorials and the Student Research Workshop, the banquet (at the Uppsala Castle), the invited talks (albeit not-so-traditional this year, please come and see yourself), the Lifetime Achievement award, the business meeting and the closing session where the "conference torch" will be handed over to the Americas, as planned.

The Workshop Chairs (Pushpak Bhattacharyya and David Weir) had a hard time deciding which workshops to turn down, and tutorials had to be kept to a reasonable number, too: quite an uneasy job for the Tutorials Chairs, Lluís Màrquez and Haifeng Wang. Demos have been selected by Sandra Kübler, and exhibitions handled by Jörg Tiedemann. Publicity has been the responsibility of Koenraad de Smedt and Beata Megyesi, the local ararngements vice-chair. Students had again the opportunity to submit papers to the Student Research Workshop, organized by the SRW Chairs Nils Reiter, Seniz Demir, and Jan Raab, helped by their Faculty Advisor Tomek Strzalkowski, who also handled the application for the usual NSF grant supporting the SRW. Markéta Lopatková, the other Faculty Advisor, then centrally handled the student travel grants. Mentoring was the responsibility of Björn Gambäck and Diana McCarthy. Taking about money and the budget, the sponsoring committee has been quite successful this year by securing grants both locally and internationally: Mats Wirén, Hercules Dalianis, Christy Doran, Srinivas Bangalore, Frédérique Segond, and Steven Pulman assembled and impressive lineup of sponsors. Thanks to them, all of you can benefit from low registration fees, subsidized banquet, the conference bag, and student scholarships and prizes.

No thank-you would be complete without mentioning Priscilla Rasmussen - her experience, insight, and ability to predict the numbers and other things was extremely helpful, to say the least. The secretary-

treasurer, Graeme Hirst, has helped to reassure us whenever there was doubt or an open budgetary question. And Steven Bird, who chaired the coordinating committee (a subcommittee of the ACL and EACL executive boards) which selected the conference venue and appointed the general chair and program chairs, has been with us throughout almost two years of preparations, helping to make sure we all (read and) follow the Conference organization handbook, and address all possible problems.

Finally, a conference without your contributions (and you as participants, of course) would not happen at all. Thank you for working hard, for submitting solid work and for preparing interesting talks and posters!

Enjoy the conference.

Jan Hajič ACL 2010 General Chair July 2010

# **Preface: Program Committee Co-Chairs**

Our goal this year has been to provide a broad conference program that acknowledges the very diverse field of computational linguistics and that recognizes the importance of both theoretical and empirical, data-driven research. In addition, we explicitly sought non-traditional conference papers such as surveys of important topics or emerging areas and papers that pose a challenge to the research community. A novel feature of the review process was the construction of different review forms for long and short papers and for the different types of papers, leading to a total of 20 different review forms. Not only were the review criteria for survey and challenge papers very different from the review criteria for research papers, but also the review criteria for theoretical research papers differed significantly from the review criteria for empirical research papers. Consequently, the program for the ACL 2010 conference includes a diverse set of papers, including papers in areas such as psycholinguistics and multimodal communication, as well as three survey papers and three challenge papers.

Both the number of paper submissions and attendance at ACL conferences continues to grow. Once again, the number of paper submissions to ACL 2010 broke the record set by the preceding year's conference. Discounting papers that were withdrawn or not reviewed due to failure to adhere to the specifications in the Call for Papers, there were 638 long paper submissions and 318 short paper submissions. Each submission was categorized according to topic and assigned to one of our paper tracks which were overseen by our 27 Area Chairs who selected expert reviewers to evaluate the submissions. Long paper submissions received at least three reviews and short paper submissions received at least two reviews. If there was a lack of consensus among the reviews for a paper, the reviewers then engaged in a discussion period in order to resolve disagreements. 25% of the long paper submissions and 22% of the short paper submissions were accepted for presentation at the conference. Unfortunately, even with adding a fifth parallel session, space prevented many very good papers from being accepted. As an experiment this year, authors were allowed to add an extra page of content to their final paper in order to enable them to address the comments and suggestions of their reviewers.

The conference program consists of three kinds of presentations: long 25 minute oral presentations, short 10 minute oral talks along with a subsequent poster presentation, and solely poster presentations. While the conference proceedings differentiates papers according to whether they are long or short papers, there is no distinction in the proceedings with respect to the mode of presentation.

We are delighted to have two invited speakers who will give what promise to be exciting plenary lectures. Andrei Broder (Vice-president, Yahoo! Research) will discuss the emerging field of computational advertising. Zenzi Griffin (Professor, University of Texas) will discuss language processing in interpersonal interactions. In addition, the recipient of the ACL Lifetime Achievement Award will present a plenary lecture on the second day of the conference.

As is traditional, there will be an award for the best long paper, the best long paper by a student, and the best short paper. Reviewers were asked to indicate whether a paper might merit a best paper prize. Area chairs then reviewed suggested papers and nominated deserving papers for further consideration. Two Best Paper Prize Committees were formed, one for long papers and one for short papers; each committee consisted of 5-6 senior researchers. The committees, along with the Program Chairs, then selected the award recipients. The Best Long Paper Prize will be presented at the plenary final session of the conference on Wednesday afternoon, and the recipient will present his or her paper during this session. The prize for Best Long Paper by a Student and the prize for the Best Short Paper will be presented at the oral talks given by the recipients during one of the regular paper sessions.

As usual, there are many individuals to thank for their contributions to the conference program. Most notably, we thank our 27 Area Chairs who did a superb job of overseeing the review of papers in their domain, the members of the Best Paper Committees, and the over 600 reviewers who in almost all cases

provided detailed, comprehensive reviews. We also owe a tremendous debt of gratitude to Rich Gerber; his START system made the review and scheduling process manageable. Moreover, his quick response to questions and requested modifications were very much appreciated. We want to thank Jason Eisner for his web site that provides invaluable advice on "How to Serve as a Program Chair of a Conference". We were very fortunate to work with wonderful organizers: Jan Hajic as General Chair, Joakim Nivre as Local Arrangements Chair, Jing-Shin Chang and Philipp Koehn as Publications Co-Chairs, and the ACL 2010 Coordinating Committee. We also thank last year's program chairs, Jian Su and Jan Wiebe, for their advice and responses to our questions.

But the ACL conference is more than just the main technical presentations. We would like to acknowledge the efforts of Seniz Demir, Jan Raab, and Nils Reiter, along with the faculty advisers Marketa Lopatkova and Tomek Strzalkowski, for organizing the Student Research Workshop. We would also like to acknowledge the work of Lluis Marquez and Haifeng Wang for soliciting an excellent set of pre-conference tutorials, Pushpak Bhattacharyia and David Weir for compiling a set of very interesting post-conference workshops, Sandra Kubler for assembling informative system demonstrations, and Jörg Tiedemann for handling exhibits.

We hope that you enjoy the conference!

ACL 2010 Program Co-Chairs Sandra Carberry, University of Delaware Stephen Clark, University of Cambridge

# **Organizing Committee**

# **General Chair:**

Jan Hajič, Charles University in Prague, Czech Republic

## **Local Arrangements Chair:**

Joakim Nivre, Uppsala University, Sweden

# **Program Chairs:**

Sandra Carberry, University of Delaware, USA Stephen Clark, University of Cambridge, United Kingdom

# **Local Arrangements Committee:**

Beata Megyesi, Uppsala University, Sweden (vice-chair) Anna Sågvall Hein, Uppsala University, Sweden Rolf Carlson, KTH, Stockholm, Sweden Mats Dahllöf, Uppsala University, Sweden Marco Kuhlmann, Uppsala University, Sweden Mattias Nilsson, Uppsala University, Sweden Markus Saers, Uppsala University, Sweden Per Starbäck, Uppsala University, Sweden Oscar Täckström, Uppsala University, Sweden Jörg Tiedemann, Uppsala University, Sweden Reut Tsarfaty, Uppsala University, Sweden

# Local Organizing Secretariat (Akademikonferens, Uppsala, Sweden):

Ulla Conti (head) Maria Carlson, Maria Bäckström, Johanna Tyselius Nilsson (webmasters)

# **Publicity Chairs:**

Koenraad de Smedt, University of Bergen, Norway Beata Megyesi, Uppsala University, Sweden

# **Exhibits Chair:**

Jörg Tiedemann, Uppsala University, Sweden

## **Sponsorship Chairs:**

Stephen Pulman, University of Oxford, United Kingdom Frédérique Segond, Xerox Research Centre Europe, France Srinivas Bangalore, AT&T Labs-Research, USA Christy Doran, MITRE, USA Hercules Dalianis, Stockholm University, Sweden Mats Wirén, Stockholm University, Sweden

## **Workshop Chairs:**

David Weir, University of Sussex, United Kingdom Pushpak Bhattacharyya, IIT Mumbai, India

# **Tutorial Chairs:**

Lluís Màrquez, Technical University of Catalonia, Spain Haifeng Wang, Baidu, China

# **Publications Chairs:**

Jing-Shin Chang, National Chi Nan University, Taiwan Philipp Koehn, University of Edinburgh, United Kingdom

# **System Demonstrations Chair:**

Sandra Kübler, Indiana University, USA

# **Student Research Workshop Chairs:**

Nils Reiter, University of Heidelberg, Germany Seniz Demir, University of Delaware, USA Jan Raab, Charles University in Prague, Czech Republic Tomek Strzalkowski, University at Albany – SUNY, USA (Faculty advisor) Markéta Lopatková, Charles University in Prague (Faculty advisor)

# **Mentoring Service Chairs:**

Björn Gambäck, SICS, Sweden Diana McCarthy, Lexical Computing Ltd., United Kingdom

# **Registration:**

Priscilla Rasmussen, Association for Computational Linguistics (ACL), USA

#### ACL Conference Coordinating Committee Chair for 2010:

Steven Bird, University of Melbourne, Australia

# **Mentoring Service Chairs:**

Björn Gambäck, SICS, Sweden Diana McCarthy, Lexical Computing Ltd., United Kingdom

## **Mentoring Committee:**

Chris Brew, The Ohio State University, USA John Carroll, University of Sussex, United Kingdom Ann Copestake, University of Cambridge, United Kingdom Robert Daland, University of California at Los Angeles, USA Roger Evans, University of Brighton, United Kingdom Debora Field, University of Sheffield, United Kingdom Louise Guthrie, Oxford Internet Institute, United Kingdom Graeme Hirst, University of Toronto, Canada Nancy Ide, Vassar College, USA Kemal Oflazer, Carnegie Mellon University - Qatar Dan Roth, University of Illinois at Urbana-Champaign, USA Donia Scott, University of Sussex, United Kingdom Mark Steedman, University of Edinburgh, United Kingdom Mark Stevenson, University of Sheffield, United Kingdom Antal van den Bosch, Tilburg University, Germany Nick Webb, University at Albany - SUNY, USA Bonnie Webber, University of Edinburgh, United Kingdom David Weir, University of Sussex, United Kingdom Richard Wicentowski, Swarthmore College, USA Graham Wilcock, University of Helsinki, Finland

# **Program Committee**

# **Program Chairs:**

Sandra Carberry, University of Delaware, USA Stephen Clark, University of Cambridge, United Kingdom

## Area Chairs:

Tim Baldwin (University of Melbourne, Australia) Phil Blunsom (University of Oxford, UK) Kalina Bontcheva (University of Sheffield, UK) Johan Bos (University of Rome - La Sapienza, Italy) Claire Cardie (Cornell University, United States) Walter Daelemans (University of Antwerp, Belgium) Rob Gaizauskas (University of Sheffield, UK) Keith Hall (Google Research - Zurich, Switzerland) Julia Hirschberg (Columbia University, United States) Nancy Ide (Vassar College, United States) Michael Johnston (AT&T Labs, United States) Roger Levy (University of California - San Diego, United States) Hang Li (Microsoft Research Asia, China) Chin-Yew Lin (Microsoft Research Asia, China) Yusuke Miyao (University of Tokyo, Japan) Roberto Navigli (University of Rome - La Sapienza, Italy) Ani Nenkova (University of Pennsylvania, United States) Jon Oberlander (University of Edinburgh, UK) Chris Quirk (Microsoft Research, United States) Stuart Shieber (Harvard University, United States) Khalil Sima'an (University of Amsterdam, The Netherlands) Richard Sproat (Oregon Health and Science University, United States) Matthew Stone (Rutgers University, United States) Jun'ichi Tsujii (University of Tokyo and University of Manchester, Japan and UK) Bonnie Webber (University of Edinburgh, UK) Theresa Wilson (University of Edinburgh, UK) ChengXiang Zhai (University of Illinois at Urbana-Champaign, United States)

#### **Best Paper Committee:**

David Chiang (Information Sciences Institute, United States) James R. Curran (University of Sydney, Australia) Walter Daelemans (University of Antwerp, Belgium) Nancy Ide (Vassar College, United States) Ann Copestake (University of Cambridge, UK) Julia Hirschberg (Columbia University, United States) Mark-Jan Nederhof (University of St Andrews, UK) Stefan Riezler (University of Heidelberg, Germany) Brian Roark (Oregon Health and Science University, United States) Rada Mihalcea (University of North Texas, United States)

# **Program Committee:**

Ahmed Abbasi, Anne Abeille, Eugene Agichtein, Eneko Agirre, David Ahn, Beatrice Alex, Enrique Alfonseca, Afra Alishahi, Alina Andreevskaia, Doug Appelt, Abhishek Arun, Nicholas Asher, Aiti Aw

Bogdan Babych, Nguyen Bach, Collin Baker, Tim Baldwin, Carmen Banea, Srinivas Bangalore, Colin Bannard, Ken Barker, Regina Barzilay, Roberto Basili, Anton Batliner, Tilman Becker, Nuria Bel, Anja Belz, Luisa Bentivogli, Sabine Bergler, Raffaella Bernardi, Tim Bickmore, Klinton Bicknell, Dan Bikel, Steven Bird, Rens Bod, Francis Bond, Stefan Bordag, Antal van den Bosch, Gosse Bouma, Jordan Boyd-Graber, S.R.K. Branavan, Eric Breck, Chris Brew, Ted Briscoe, Samuel Brody, Sabine Buchholz, Paul Buitelaar, Razvan Bunescu, Harry Bunt, Aljoscha Burchardt, Stephan Busemann, Miriam Butt, Bill Byrne

Michael Cafarella, Aoife Cahill, Chris Callison-Burch, Diego Calvanese, Nicoletta Calzolari, Nicola Cancedda, Yunbo Cao, Giuseppe Carenini, Michael Carl, Bob Carpenter, Marine Carpuat, Xavier Carreras, John Carroll, Ben Carterette, Diamantino Caseiro, Dan Cer, Joyce Chai, Yee Seng Chan, Jason Chang, Pi-Chuan Chang, Ciprian Chelba, Aitao Chen, Hsin-Hsi Chen, Stanley Chen, Wenliang Chen, Colin Cherry, David Chiang, Key-Sun Choi, Yejin Choi, Tat-Seng Chua, Kenneth Church, Massimiliano Ciaramita, Christopher Cieri, Philipp Cimiano, Alex Clark, Peter Clark, James Clarke, Andrew Clegg, Eric de la Clergerie, Aaron Cohen, Kevin Cohen, Shay Cohen, Trevor Cohn, Nigel Collier, Michael Collins, Kevyn Collins-Thompson, Gao Cong, Paul Cook, Anna Corazza, Marta Ruiz Costa-Jussa, Josep Crego, Mathias Creutz, Dan Cristea, Montse Cuadros, Hang Cui, James Curran

Robert Damper, Dipanjan Das, Hal Daume, Thierry Declerck, Vera Demberg, Steve DeNeefe, John DeNero, Tejaswini Deoskar, Ann Devitt, Mona Diab, Bill Dolan, Christy Doran, Doug Downey, Mark Dredze, Markus Dreyer, Greg Druck, Amit Dubey, Kevin Duh, Benjamin van Durme, Chris Dyer

Markus Egg, Koji Eguchi, Jason Eisner, Michael Elhadad, Ralf Engbert, Tomaz Erjavec, Katrin Erk, Gunes Erkan, Gülşen Eryiğit, Maxine Eskenazi, Andrea Esuli, Oren Etzioni

Hui Fang, Marcello Federico, Anna Feldman, Ronen Feldman, Christiane Fellbaum, Raquel Fernandez, Raul Fernandez, Tim Fernando, Jenny Finkel, Dan Flickinger, Radu Florian, Juliane Fluck, Kate Forbes-Riley, Mary Ellen Foster, Nissim Francez, Anette Frank, Stefan Frank, Stella Frank, Alex Fraser, Dayne Freitag, Pascale Fung, Sadaoki Furui, Dina Demner Fushman

Evgeniy Gabrilovich, Michael Galley, Michael Gamon, Jianfeng Gao, Claire Gardent, Nikesh Garera, Josef van Genabith, Dmitriy Genzel, Kallirroi Georgila, Daniel Gildea, Alastair Gill, Jesus Gimenez, Kevin Gimpel, Roxana Girju, Adria de Gispert, Claudio Giuliano, Sharon Goldwater, Gevieve Gorrell, Agustin Gravano, Stephan Greene, Mark Greenwood, Gregory Grefenstette, Ralph Grishman, Philippe de Groote, Claire Grover, Joakim Gustafson

Barry Haddow, Gholamreza Haffari, Udo Hahn, Dilek Hakkani-Tur, Harald Hammarstrom, Siegfried Handschuh, Sanda Harabagiu, Henk Harkema, Donna Harman, Saša Hasan, Chikara Hashimoto, Hany Hassan, Xiaodong He, Peter Heeman, James Henderson, Mark Hepple, Ulf Hermjakob, Andrew Hickl, Ryuichiro Higashinaka, Erhard Hinrichs, Lynette Hirschman, Graeme Hirst, Julia Hockenmaier, Beth Ann Hockey, Veronique Hoste, Yunnhua Hu, Chu-Ren Huang, Jimmy Huang, Liang Huang, Minlie Huang, Xuanjing Huang, Sarmad Hussain, Gerhard van Huyssteen, Rebecca Hwa

Nancy Ide, Diana Inkpen

Florian Jaeger, Gerhard Jaeger, Martin Jansche, Jing Jiang, Wenbin Jiang, Richard Johansson, Howard Johnson, Mark Johnson, Rie Johnson, Michael Johnston, Aravind Joshi

Damianos Karakos, Nikiforos Karamanis, Lauri Karttunen, Daisuke Kawahara, Andrew Kehler, Frank Keller, Rodger Kibble, Adam Kilgarriff, Jin-Dong Kim, Su Nam Kim, Tracy Holloway

King, Kevin Knight, Alistair Knott, Philipp Koehn, Rob Koeling, Oskar Kohonen, Alexander Koller, Greg Kondrak, Moshe Koppel, Valia Kordoni, Anna Korhonen, Andras Kornai, Wessel Kraaij, Emiel Krahmer, Ivana Kruijff-Korbayova, Udo Kruschwitz, Lun-Wei Ku, Sandra Kuebler, Marco Kuhlmann, Roland Kuhn, Shankar Kumar, Mikko Kurimo, Oren Kurland, Sadao Kurohashi, Tom Kwiatkowski, Kui-Lam Kwok, Olivia Kwong

Krista Lagus, Patrik Lambert, Philippe Langlais, Mirella Lapata, Alex Lascarides, Alberto Lavelli, Victor Lavrenko, Gary Geunbae Lee, Yoong Keok Lee, Gregor Leusch, Gina-Anne Levow, Roger Levy, David Lewis, Chi-Ho Li, Haizhou Li, Maggie Wenjie Li, Mu Li, Wei Li, Xiao Li, Zhifei Li, Percy Liang, Chuan-Jie Lin, Ken Litkowski, Diane Litman, Bing Liu, Qun Liu, Ting Liu, Yang Liu, Yupeng Liu, Adam Lopez, Ramon Lopez-Cozar, Daniel Lopresti, Annie Louis, Bernd Ludwig

Klaus Macherey, Wolfgang Macherey, Rob Malouf, Suresh Manandhar, Daniel Marcu, Mitch Marcus, Katja Markert, Erwin Marsi, Andre Martins, David Martínez, Sameer Maskey, Yuji Matsumoto, Takuya Matsuzaki, Evgeny Matusov, Arne Mauser, Mike Maxwell, Diana Maynard, Diana McCarthy, David McClosky, Ryan McDonald, Kathleen McKeown, Qiaozhu Mei, Dan Melamed, Arul Menezes, Helen Meng, Paola Merlo, Donald Metzler, Haitao Mi, Jens Michaelis, Rada Mihalcea, Eleni Miltsakaki, David Mimno, Michael Minock, Teruko Mitamura, Bernd Moebius, Marie-Francine Moens, Saif Mohammad, Christian Monson, Christof Monz, Bob Moore, Michael Moortgat, Roser Morante, Louis-Philippe Morency, Tatsunori Mori, Glyn Morrill, Alessandro Moschitti, Larry Moss, Lluís Màrquez, Tony Mullen, Gabriel Murray, Sung-Hyon Myaeng

Tetsuji Nakagawa, Hiromi Nakaiwa, Yukiko Nakano, Tahira Naseem, Vivi Nastase, Tetsuya Nasukawa, Mark-Jan Nederhof, Rani Nelken, Goran Nenadic, Ani Nenkova, Günter Neumann, Hermann Ney, Hwee Tou Ng, Vincent Ng, Patrick Nguyen, Jeremy Nicholson, Jian-Yun Nie, Zaiqing Nie, Takashi Ninomiya, Gertjan van Noord, Scott Nowson

Jon Oberlander, Franz Och, Stephan Oepen, Kemal Oflazer, Naoaki Okazaki, Manabu Okumura, Constantin Orasan, Miles Osborne, Diarmuid OSeaghdha

Sebastian Pado, Tim Paek, Bo Pang, Patrick Pantel, Soo-Min Pantel, Cecile Paris, Marius Pasca, Rebecca Passonneau, Adam Pauls, Guy De Pauw, Gerald Penn, Fernando Pereira, John Pestian, Kay Peterson, Slav Petrov, Michael Picheny, Roberto Pieraccini, Manfred Pinkal, Stelios Piperidis, Emily Pitler, Paul Piwek, Massimo Poesio, Simone Paolo Ponzetto, Hoifung Poon, Ana-Maria Popescu, Andrei Popescu-Belis, Matt Post, David Powers, John Prager, Rashmi Prasad, Stephen Pulman, Vasin Punyakanok, Matthew Purver, James Pustejovsky, Sampo Pyysalo

Stephan Raaijmakers, Filip Radlinski, Bhuvana Ramabhadran, Owen Rambow, Allan Ramsay, Delip Rao, Reinhard Rapp, Ari Rappoport, Emmanuel Rayner, Dietrich Rebholz-Schuhmann, Flo Reeder, Erik Reichle, Ehud Reiter, David Reitter, Steve Renals, Christian Retore, Giuseppe Riccardi, Sebastian Riedel, German Rigau, Ellen Riloff, Hae-Chang Rim, Laura Rimell, Brian Roark, James Rogers, Doug Roland, Laurent Romary, Andrew Rosenberg, Paolo Rosso, Dan Roth, Shourya Roy, Patrick Ruch

Kenji Sagae, William Sakas, Tapio Salakoski, Greg Sanders, Rajeev Sangal, Murat Saraclar, Giorgio Satta, Roser Sauri, Frank Schilder, David Schlangen, Helmut Schmid, Hinrich Schuetze, William Schuler, Rolf Schwitter, Donia Scott, Yohei Seki, Satoshi Sekine, Violeta Seretan, Burr Settles, Chung-chieh Shan, Vijay Shanker, Libin Shen, Stuart Shieber, Kiyoaki Shirai, Mario Silva, David Smith, Jason Smith, Nathaniel Smith, Noah Smith, Ben Snyder, Stephen Soderland, Hagen Soltau, Swapna Somasundaran, Young-In Song, Abdelhadi Soudi, Lucia Specia, Sebastian Spiegler, Michael Spivey, Caroline Sporleder, Richard Sproat, Rohini Srihari, Padmini Srinivasan, Ed Stabler, Manfred Stede, Mark Steedman, Ralf Steinberger, Amanda Stent, Mark Stevenson, Matthew Stone, Veselin Stoyanov, Carlo Strapparava, Michael Strube, Patrick Sturt, Jian Su, L Venkata Subramaniam, Mihai Surdeanu, Marc Swerts, Stan Szpakowicz

Hiroya Takamura, David Talbot, Kumiko Tanaka-Ishii, Jie Tang, Ben Taskar, Michael Tepper, Stefan Thater, Mariet Theune, Christoph Tillmann, Ivan Titov, Take Tokunaga, Kentaro Torisawa,

Kristina Toutanova, Roy Tromble, Vivian Tsang, Reut Tsarfaty, Yoshimasa Tsuruoka, Gokhan Tur, Peter Turney

Lucy Vanderwende, Sebastian Varges, Vasudeva Varma, Shravan Vasishth, Tony Veale, Paola Velardi, Henriette Viethen, Marc Vilain, David Vilar, Sami Virpioja, Andreas Vlachos, Stephan Vogel, Piek Vossen

Sabine Schulte im Walde, Hanna Wallach, Stephen Wan, Xiaojun Wan, Haifeng Wang, Wei Wang, Taro Watanabe, Andy Way, Bonnie Webber, David Weir, Ralph Weischedel, Ben Wellner, Edward Whittaker, Janyce Wiebe, John Wilbur, Theresa Wilson, Shuly Wintner, Yuk Wah Wong, Dekai Wu

Fei Xia, Deyi Xiong, Peng Xu, Nianwen Xue

Mei Yang, Roman Yangarber, Patrick Ye, Lars Yencken, Zhang Yi, Scott Wen-tau Yih

Richard Zens, Luke Zettlemoyer, Cheng Zhai, Hao Zhang, Yi Zhang, Yue Zhang, Zhu Zhang, Jing Zheng, Jun Zhu, Andreas Zollmann, Willem (Jelle) Zuidema, Ingrid Zukerman

# **Table of Contents**

Efficient Third-Order Dependency Parsers Terry Koo and Michael Collins
Dependency Parsing and Projection Based on Word-Pair Classification         Wenbin Jiang and Qun Liu       12
<i>Bitext Dependency Parsing with Bilingual Subtree Constraints</i> Wenliang Chen, Jun'ichi Kazama and Kentaro Torisawa
Computing Weakest Readings Alexander Koller and Stefan Thater
Identifying Generic Noun Phrases           Nils Reiter and Anette Frank         40
Structural Semantic Relatedness: A Knowledge-Based Method to Named Entity Disambiguation Xianpei Han and Jun Zhao
Correcting Errors in Speech Recognition with Articulatory Dynamics Frank Rudzicz
Learning to Adapt to Unknown Users: Referring Expression Generation in Spoken Dialogue Systems Srinivasan Janarthanam and Oliver Lemon
A Risk Minimization Framework for Extractive Speech Summarization Shih-Hsiang Lin and Berlin Chen
The Human Language Project: Building a Universal Corpus of the World's Languages         Steven Abney and Steven Bird       88
Bilingual Lexicon Generation Using Non-Aligned Signatures         Daphna Shezaf and Ari Rappoport
Automatic Evaluation Method for Machine Translation Using Noun-Phrase Chunking Hiroshi Echizen-ya and Kenji Araki
<i>Open Information Extraction Using Wikipedia</i> Fei Wu and Daniel S. Weld
<i>SystemT: An Algebraic Approach to Declarative Information Extraction</i> Laura Chiticariu, Rajasekar Krishnamurthy, Yunyao Li, Sriram Raghavan, Frederick Reiss and Shivakumar Vaithyanathan
Extracting Social Networks from Literary Fiction David Elson, Nicholas Dames and Kathleen McKeown
Pseudo-Word for Phrase-Based Machine Translation Xiangyu Duan, Min Zhang and Haizhou Li
<i>Hierarchical Search for Word Alignment</i> Jason Riesa and Daniel Marcu

" <i>Was It Good? It Was Provocative.</i> " <i>Learning the Meaning of Scalar Adjectives</i> Marie-Catherine de Marneffe, Christopher D. Manning and Christopher Potts
Importance-Driven Turn-Bidding for Spoken Dialogue Systems           Ethan Selfridge and Peter Heeman         177
<i>Entity-Based Local Coherence Modelling Using Topological Fields</i> Jackie Chi Kit Cheung and Gerald Penn
Syntactic and Semantic Factors in Processing Difficulty: An Integrated Measure Jeff Mitchell, Mirella Lapata, Vera Demberg and Frank Keller
Rebanking CCGbank for Improved NP Interpretation           Matthew Honnibal, James R. Curran and Johan Bos         207
BabelNet: Building a Very Large Multilingual Semantic Network         Roberto Navigli and Simone Paolo Ponzetto         216
Fully Unsupervised Core-Adjunct Argument Classification         Omri Abend and Ari Rappoport       226
<i>Towards Open-Domain Semantic Role Labeling</i> Danilo Croce, Cristina Giannone, Paolo Annesi and Roberto Basili
A Bayesian Method for Robust Estimation of Distributional Similarities Jun'ichi Kazama, Stijn De Saeger, Kow Kuroda, Masaki Murata and Kentaro Torisawa 247
Recommendation in Internet Forums and Blogs Jia Wang, Qing Li, Yuanzhu Peter Chen and Zhangxi Lin
Learning Phrase-Based Spelling Error Models from Clickthrough Data Xu Sun, Jianfeng Gao, Daniel Micol and Chris Quirk
Inducing Domain-Specific Semantic Class Taggers from (Almost) Nothing Ruihong Huang and Ellen Riloff
<i>Learning 5000 Relational Extractors</i> Raphael Hoffmann, Congle Zhang and Daniel S. Weld
Unsupervised Ontology Induction from Text Hoifung Poon and Pedro Domingos
<i>Exploring Syntactic Structural Features for Sub-Tree Alignment Using Bilingual Tree Kernels</i> Jun Sun, Min Zhang and Chew Lim Tan
Discriminative Pruning for Discriminative ITG Alignment Shujie Liu, Chi-Ho Li and Ming Zhou
<i>Fine-Grained Tree-to-String Translation Rule Extraction</i> Xianchao Wu, Takuya Matsuzaki and Jun'ichi Tsujii
Accurate Context-Free Parsing with Combinatory Categorial Grammar Timothy A. D. Fowler and Gerald Penn

Faster Parsing by Supertagger Adaptation         Jonathan K. Kummerfeld, Jessika Roesner, Tim Dawborn, James Haggerty, James R. Curran and         Stephen Clark
Using Smaller Constituents Rather Than Sentences in Active Learning for Japanese Dependency Parsing Manabu Sassano and Sadao Kurohashi
Conditional Random Fields for Word Hyphenation Nikolaos Trogkanis and Charles Elkan
Enhanced Word Decomposition by Calibrating the Decision Threshold of Probabilistic Models and Us- ing a Model Ensemble Sebastian Spiegler and Peter A. Flach
Word Representations: A Simple and General Method for Semi-Supervised Learning Joseph Turian, Lev-Arie Ratinov and Yoshua Bengio
Identifying Text Polarity Using Random Walks Ahmed Hassan and Dragomir Radev
Sentiment Learning on Product Reviews via Sentiment Ontology Tree         Wei Wei and Jon Atle Gulla       404
<i>Employing Personal/Impersonal Views in Supervised and Semi-Supervised Sentiment Classification</i> Shoushan Li, Chu-Ren Huang, Guodong Zhou and Sophia Yat Mei Lee
A Latent Dirichlet Allocation Method for Selectional Preferences Alan Ritter, Mausam and Oren Etzioni
Latent Variable Models of Selectional Preference Diarmuid Ó Séaghdha
Improving the Use of Pseudo-Words for Evaluating Selectional Preferences           Nathanael Chambers and Dan Jurafsky         445
Syntax-to-Morphology Mapping in Factored Phrase-Based Statistical Machine Translation from English to Turkish Reyyan Yeniterzi and Kemal Oflazer
<i>Hindi-to-Urdu Machine Translation through Transliteration</i> Nadir Durrani, Hassan Sajjad, Alexander Fraser and Helmut Schmid
Training Phrase Translation Models with Leaving-One-Out         Joern Wuebker, Arne Mauser and Hermann Ney         475
<i>Efficient Staggered Decoding for Sequence Labeling</i> Nobuhiro Kaji, Yasuhiro Fujiwara, Naoki Yoshinaga and Masaru Kitsuregawa
Minimized Models and Grammar-Informed Initialization for Supertagging with Highly Ambiguous Lex-
<i>icons</i> Sujith Ravi, Jason Baldridge and Kevin Knight
Practical Very Large Scale CRFs Thomas Lavergne, Olivier Cappé and François Yvon

On the Computational Complexity of Dominance Links in Grammatical Formalisms Sylvain Schmitz
Optimal Rank Reduction for Linear Context-Free Rewriting Systems with Fan-Out Two         Benoît Sagot and Giorgio Satta
<i>The Importance of Rule Restrictions in CCG</i> Marco Kuhlmann, Alexander Koller and Giorgio Satta
Automatic Evaluation of Linguistic Quality in Multi-Document Summarization Emily Pitler, Annie Louis and Ani Nenkova
Identifying Non-Explicit Citing Sentences for Citation-Based Summarization. Vahed Qazvinian and Dragomir R. Radev
Automatic Generation of Story Highlights Kristian Woodsend and Mirella Lapata
Sentence and Expression Level Annotation of Opinions in User-Generated Discourse Cigdem Toprak, Niklas Jakob and Iryna Gurevych
<i>Generating Focused Topic-Specific Sentiment Lexicons</i> Valentin Jijkoun, Maarten de Rijke and Wouter Weerkamp
<i>Evaluating Multilanguage-Comparability of Subjectivity Analysis Systems</i> Jungi Kim, Jin-Ji Li and Jong-Hyeok Lee
<i>Error Detection for Statistical Machine Translation Using Linguistic Features</i> Deyi Xiong, Min Zhang and Haizhou Li
TrustRank: Inducing Trust in Automatic Translations via Ranking         Radu Soricut and Abdessamad Echihabi         612
Bridging SMT and TM with Translation Recommendation Yifan He, Yanjun Ma, Josef van Genabith and Andy Way
<i>On Jointly Recognizing and Aligning Bilingual Named Entities</i> Yufeng Chen, Chengqing Zong and Keh-Yih Su
Generating Templates of Entity Summaries with an Entity-Aspect Model and Pattern Mining Peng Li, Jing Jiang and Yinglin Wang
Comparable Entity Mining from Comparative Questions Shasha Li, Chin-Yew Lin, Young-In Song and Zhoujun Li
Towards Robust Multi-Tool Tagging. An OWL/DL-Based Approach         Christian Chiarcos       659
<i>Temporal Information Processing of a New Language: Fast Porting with Minimal Resources</i> Francisco Costa and António Branco
A Taxonomy, Dataset, and Classifier for Automatic Noun Compound Interpretation Stephen Tratz and Eduard Hovy
<i>Models of Metaphor in NLP</i> Ekaterina Shutova

A Game-Theoretic Model of Metaphorical Bargaining Beata Beigman Klebanov and Eyal Beigman	. 698
Kernel Based Discourse Relation Recognition with Temporal Ordering Information WenTing Wang, Jian Su and Chew Lim Tan	. 710
Hierarchical Joint Learning: Improving Joint Parsing and Named Entity Recognition with Non-Jo Labeled Data Jenny Rose Finkel and Christopher D. Manning	
Detecting Errors in Automatically-Parsed Dependency Relations Markus Dickinson	. 729
Boosting-Based System Combination for Machine Translation Tong Xiao, Jingbo Zhu, Muhua Zhu and Huizhen Wang	. 739
<i>Fine-Grained Genre Classification Using Structural Learning Algorithms</i> Zhili Wu, Katja Markert and Serge Sharoff	. 749
Metadata-Aware Measures for Answer Summarization in Community Question Answering Mattia Tomasoni and Minlie Huang	. 760
A Hybrid Rule/Model-Based Finite-State Framework for Normalizing SMS Messages Richard Beaufort, Sophie Roekhaut, Louise-Amélie Cougnon and Cédrick Fairon	.770
<i>Letter-Phoneme Alignment: An Exploration</i> Sittichai Jiampojamarn and Grzegorz Kondrak	.780
Using Document Level Cross-Event Inference to Improve Event Extraction Shasha Liao and Ralph Grishman	. 789
Now, Where Was I? Resumption Strategies for an In-Vehicle Dialogue System Jessica Villing	.798
Learning to Follow Navigational Directions Adam Vogel and Dan Jurafsky	. 806
A Hybrid Hierarchical Model for Multi-Document Summarization Asli Celikyilmaz and Dilek Hakkani-Tur	.815
Improving Statistical Machine Translation with Monolingual Collocation Zhanyi Liu, Haifeng Wang, Hua Wu and Sheng Li	. 825
Bilingual Sense Similarity for Statistical Machine Translation Boxing Chen, George Foster and Roland Kuhn	. 834
Untangling the Cross-Lingual Link Structure of Wikipedia Gerard de Melo and Gerhard Weikum	. 844
Bucking the Trend: Large-Scale Cost-Focused Active Learning for Statistical Machine Translation Michael Bloodgood and Chris Callison-Burch	. 854
Creating Robust Supervised Classifiers via Web-Scale N-Gram Data Shane Bergsma, Emily Pitler and Dekang Lin	. 865

Convolution Kernel over Packed Parse Forest Min Zhang, Hui Zhang and Haizhou Li
Estimating Strictly Piecewise Distributions Jeffrey Heinz and James Rogers
String Extension Learning         Jeffrey Heinz
Compositional Matrix-Space Models of Language Sebastian Rudolph and Eugenie Giesbrecht
Cross-Language Document Summarization Based on Machine Translation Quality Prediction Xiaojun Wan, Huiying Li and Jianguo Xiao
A New Approach to Improving Multilingual Summarization Using a Genetic Algorithm Marina Litvak, Mark Last and Menahem Friedman
Bayesian Synchronous Tree-Substitution Grammar Induction and Its Application to Sentence Compression Elif Yamangil and Stuart M. Shieber
Contextualizing Semantic Representations Using Syntactically Enriched Vector Models Stefan Thater, Hagen Fürstenau and Manfred Pinkal
Bootstrapping Semantic Analyzers from Non-Contradictory Texts           Ivan Titov and Mikhail Kozhevnikov         958
Open-Domain Semantic Role Labeling by Modeling Word Spans           Fei Huang and Alexander Yates         968
<i>Learning Script Knowledge with Web Experiments</i> Michaela Regneri, Alexander Koller and Manfred Pinkal
Starting from Scratch in Semantic Role Labeling         Michael Connor, Yael Gertner, Cynthia Fisher and Dan Roth         989
Modeling Norms of Turn-Taking in Multi-Party Conversation Kornel Laskowski
Optimising Information Presentation for Spoken Dialogue Systems           Verena Rieser, Oliver Lemon and Xingkun Liu         1009
Combining Data and Mathematical Models of Language Change Morgan Sonderegger and Partha Niyogi
Finding Cognate Groups Using Phylogenies         David Hall and Dan Klein       1030
An Exact A* Method for Deciphering Letter-Substitution Ciphers Eric Corlett and Gerald Penn
A Statistical Model for Lost Language Decipherment Benjamin Snyder, Regina Barzilay and Kevin Knight

<i>Efficient Inference through Cascades of Weighted Tree Transducers</i> Jonathan May, Kevin Knight and Heiko Vogler
A Tree Transducer Model for Synchronous Tree-Adjoining Grammars Andreas Maletti
Dynamic Programming for Linear-Time Incremental Parsing Liang Huang and Kenji Sagae
Hard Constraints for Grammatical Function Labelling Wolfgang Seeker, Ines Rehbein, Jonas Kuhn and Josef Van Genabith
Simple, Accurate Parsing with an All-Fragments Grammar Mohit Bansal and Dan Klein
Joint Syntactic and Semantic Parsing of Chinese Junhui Li, Guodong Zhou and Hwee Tou Ng
Cross-Language Text Classification Using Structural Correspondence Learning Peter Prettenhofer and Benno Stein
Cross-Lingual Latent Topic Extraction Duo Zhang, Qiaozhu Mei and ChengXiang Zhai1128
Topic Models for Word Sense Disambiguation and Token-Based Idiom Detection         Linlin Li, Benjamin Roth and Caroline Sporleder
PCFGs, Topic Models, Adaptor Grammars and Learning Topical Collocations and the Structure of Proper Names Mark Johnson
A Cognitive Cost Model of Annotations Based on Eye-Tracking Data Katrin Tomanek, Udo Hahn, Steffen Lohmann and Jürgen Ziegler
A Rational Model of Eye Movement Control in Reading Klinton Bicknell and Roger Levy
The Influence of Discourse on Syntax: A Psycholinguistic Model of Sentence Processing         Amit Dubey       1179
Complexity Metrics in an Incremental Right-Corner Parser Stephen Wu, Asaf Bachrach, Carlos Cardenas and William Schuler
"Ask Not What Textual Entailment Can Do for You" Mark Sammons, V.G.Vinod Vydiswaran and Dan Roth
Assessing the Role of Discourse References in Entailment Inference Shachar Mirkin, Ido Dagan and Sebastian Pado
Global Learning of Focused Entailment Graphs Jonathan Berant, Ido Dagan and Jacob Goldberger
Modeling Semantic Relevance for Question-Answer Pairs in Web Social Communities Baoxun Wang, Xiaolong Wang, Chengjie Sun, Bingquan Liu and Lin Sun

How Many Words Is a Picture Worth? Automatic Caption Generation for News Images Yansong Feng and Mirella Lapata
Generating Image Descriptions Using Dependency Relational Patterns Ahmet Aker and Robert Gaizauskas
Incorporating Extra-Linguistic Information into Reference Resolution in Collaborative Task Dialogue Ryu Iida, Syumpei Kobayashi and Takenobu Tokunaga
Reading between the Lines: Learning to Map High-Level Instructions to Commands S.R.K. Branavan, Luke Zettlemoyer and Regina Barzilay
Profiting from Mark-Up: Hyper-Text Annotations for Guided Parsing Valentin I. Spitkovsky, Daniel Jurafsky and Hiyan Alshawi
Phylogenetic Grammar Induction           Taylor Berg-Kirkpatrick and Dan Klein           1288
<i>Improved Unsupervised POS Induction through Prototype Discovery</i> Omri Abend, Roi Reichart and Ari Rappoport
<i>Extraction and Approximation of Numerical Attributes from the Web</i> Dmitry Davidov and Ari Rappoport
Learning Word-Class Lattices for Definition and Hypernym Extraction Roberto Navigli and Paola Velardi
On Learning Subtypes of the Part-Whole Relation: Do Not Mix Your Seeds Ashwin Ittoo and Gosse Bouma
Understanding the Semantic Structure of Noun Phrase Queries Xiao Li
<i>Multilingual Pseudo-Relevance Feedback: Performance Study of Assisting Languages</i> Manoj Kumar Chinnakotla, Karthik Raman and Pushpak Bhattacharyya
<i>Wikipedia as Sense Inventory to Improve Diversity in Web Search Results</i> Celina Santamaría, Julio Gonzalo and Javier Artiles
A Unified Graph Model for Sentence-Based Opinion Retrieval Binyang Li, Lanjun Zhou, Shi Feng and Kam-Fai Wong
Generating Fine-Grained Reviews of Songs from Album Reviews Swati Tata and Barbara Di Eugenio
A Study of Information Retrieval Weighting Schemes for Sentiment Analysis Georgios Paltoglou and Mike Thelwall
Supervised Noun Phrase Coreference Research: The First Fifteen Years Vincent Ng
Unsupervised Event Coreference Resolution with Rich Linguistic Features Cosmin Bejan and Sanda Harabagiu
Coreference Resolution across Corpora: Languages, Coding Schemes, and Preprocessing Information Marta Recasens and Eduard Hovy

Constituency to Dependency Translation with Forests Haitao Mi and Qun Liu
Learning to Translate with Source and Target Syntax David Chiang
Discriminative Modeling of Extraction Sets for Machine Translation John DeNero and Dan Klein
Detecting Experiences from Weblogs           Keun Chan Park, Yoonjae Jeong and Sung Hyon Myaeng           1464
<i>Experiments in Graph-Based Semi-Supervised Learning Methods for Class-Instance Acquisition</i> Partha Pratim Talukdar and Fernando Pereira
Learning Arguments and Supertypes of Semantic Relations Using Recursive Patterns Zornitsa Kozareva and Eduard Hovy
A Transition-Based Parser for 2-Planar Dependency Structures Carlos Gómez-Rodríguez and Joakim Nivre
Viterbi Training for PCFGs: Hardness Results and Competitiveness of Uniform Initialization Shay Cohen and Noah A Smith
A Generalized-Zero-Preserving Method for Compact Encoding of Concept Lattices Matthew Skala, Victoria Krakovna, János Kramár and Gerald Penn
<i>Knowledge-Rich Word Sense Disambiguation Rivaling Supervised Systems</i> Simone Paolo Ponzetto and Roberto Navigli
All Words Domain Adapted WSD: Finding a Middle Ground between Supervision and Unsupervision Mitesh Khapra, Anup Kulkarni, Saurabh Sohoney and Pushpak Bhattacharyya 1532
Combining Orthogonal Monolingual and Multilingual Sources of Evidence for All Words WSD Weiwei Guo and Mona Diab1542
<i>Phrase-Based Statistical Language Generation Using Graphical Models and Active Learning</i> Francois Mairesse, Milica Gasic, Filip Jurcicek, Simon Keizer, Blaise Thomson, Kai Yu and Steve Young
Plot Induction and Evolutionary Search for Story Generation Neil McIntyre and Mirella Lapata
Automated Planning for Situated Natural Language Generation Konstantina Garoufi and Alexander Koller
Beyond NomBank: A Study of Implicit Arguments for Nominal Predicates Matthew Gerber and Joyce Chai