ACL 2007

PRAGUE

ACL 2007

Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics

June 23–30, 2007
Prague, Czech Republic

The Association for Computational Linguistics

It is with great sadness that we record the death on 4 April 2007 of Karen Spärck Jones, who among many other honors and awards, was the President of ACL in 1994 and recipient of the ACL Lifetime Achievement Award in 2004.

These proceedings are dedicated to her memory.

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

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

Table of Contents

Preface: General Chair

On behalf of the organizing committee I am delighted to welcome you to the 45th Annual Meeting of the Association for Computational Linguistics, in Prague.

Setting up and running the ACL conference involves a lot of work by many people. Some of them are officially identified as being responsible for various aspects of the conference, while the contributions of others are less visible. I would like to say a warm thank you to the people named below, with apologies to anyone I have overlooked.

The Program Chairs, **Antal van den Bosch** and **Annie Zaenen** have done a great job in managing the almost 600 submissions for the main conference and putting together a high quality program. Through this process Antal has become a 'grandmaster' of the START paper management system and has given a lot of help to other chairs who have had to deal with their own sets of submissions. Many thanks also to their Area Chairs and the program committee of reviewers, and to **Florence Reeder** for coordinating the pre-submission mentoring service. (Antal and Annie reflect on their PC experience overleaf).

Sophia Ananiadou is Chair of the Demo/Poster part of the conference, and has overseen a separate review process to select a high quality set of presentations.

The Student Research Workshop Chairs, **Chris Biemann**, **Violeta Seretan** and **Ellen Riloff** have assembled an excellent program of papers and posters. I encourage everyone to attend the student workshop to hear about the exciting work being carried out by researchers just starting out on their careers in computational linguistics.

Workshops Chair **Simone Teufel** is overseeing 15 workshops – the most ever at an ACL conference – chosen (with the help of **Beth Ann Hockey**, **Katja Markert** and **Dekai Wu**) from a total of 27 proposals. The scale of the workshop program can be gauged by the fact that they received an aggregate total of 470 submissions (without even counting IWPT and EMNLP-CoNLL). **Joakim Nivre**, as Tutorials Chair, has assembled a program of 5 attractive and complementary tutorials, selected from 20 proposals with advice from **Walter Daelemans**, **Robert Dale**, **Nancy Ide**, **Diane Litman** and **Chris Manning**.

One of the most demanding yet least noticed organizational roles is that of Publications Chair. **Su Jian** has done a fantastic job in producing the hardcopy and electronic record of the conference, supported by his team—with notable contributions from **Upali Kohomban** who has cheerfully helped at all stages and whenever needed, day and night, weekdays and weekends.

Sponsorship is another success story, thanks to the Sponsorship Chairs **Martha Palmer**, **Gabor Proszeky**, **Jan Hajic** and **Jun'ichi Tsujii**, who have recruited 12 corporate sponsors. We are very grateful for their financial support.

Eva Hajicova, the Local Arrangements Chair, assisted by **Jan Hajic** and **Anna Kotesovcova**, have put in an enormous amount of detailed work to make this conference a success, ably supported by the local team of **Milan Fucik**, **Jaroslava Hlavacova**, **Marketa Lopatkova**, **Jiri Mirovsky**, **Pavel Pecina**,

Pavel Schlesinger, **Juraj Simlovic**, **Miroslav Spousta**, **Pavel Stranak**, **Zlatka Subrova**, **Jan Votrubec**, **Zdenek Zabokrtsky** and **Daniel Zeman**. From the ACL itself, **Kathy McCoy**, **Dragomir Radev**, **Priscilla Rasmussen**, **Mark Steedman** and **Jun'ichi Tsujii** have played strong roles in making decisions and giving advice, and kept everything on track while I was out of action due to ill health last year.

And finally, many thanks to the authors and presenters in the main conference, workshops and co-located events... and to all the participants. I hope you enjoy the formally organized aspects of the conference, take advantage of the opportunity to network with colleagues old and new, and that you also have a chance to appreciate the history and sights of Prague while you are here.

John Carroll
ACL 2007 General Chair

Preface: Program Chairs

The number of submissions for this ACL broke a new record: the program committee's selection of 131 papers was based on 588 submissions (after withdrawals). An updated program design with four parallel sessions and 25-minute papers allowed an acceptance rate of 22.3%, and an acceptance of all submissions that were recommended with priority by the area chairs.

First and foremost, we thank all the authors for submitting papers describing their recent work; the sheer amount of submissions reflects how active our field is. We thank **Florence Reeder** for provided mentoring to 17 author teams who felt they needed some writing support. For the selection, we are indebted to the 332 program committee members, who produced one to eleven reviews per reviewer, for a total of close to 1,800 reviews, and to the ten area chairs on whose shoulders rested most of the work of organizing the review process. We decided to work without an area chairs meeting: the two program co-chairs met for two days at Tilburg University, and interacted during that time vigorously with the area chairs by email and sometimes by phone.

As usual the main program will run for three days: there will be four parallel sessions of main session presentations, a demo/poster session organized by **Sophia Ananiadou**, nameMiroslav Spousta and **Zdenek Zabokrtsky**, and a Student Research Workshop – thanks to **Ellen Riloff**, **Violeta Seretan** and **Chris Biemann** for organizing it. Also as usual the conference is flanked by tutorial sessions and workshops; our thanks go to **Joakim Nivre** and **Simone Teufel** for organizing and compiling an excellent package.

The announcements of the ACL Lifetime Award and of the Best Paper Award will provide the customary suspense. They will take place in plenary sessions. Other plenary sessions will be devoted to the business meeting and the two invited talks, which this year will be delivered by **Tom Mitchell** and **Barney Pell**. We are grateful for their kind acceptation of our invitation.

We thank **John Carroll**, General Conference Chair, the Local Arrangements Committee headed by **Eva Hajicova**, and the ACL executive, especially **Dragomir Radev**, for their help and advice, and last year's co-chairs, **Claire Cardie** and **Pierre Isabelle**, for sharing their experience. Our sincere thanks go to **Su Jian** for putting together the proceedings.

Enjoy the conference,

Antal van den Bosch and Annie Zaenen
ACL-2007 Program Chairs

Organizers

General Chair:

John Carroll, University of Sussex, UK

Local Arrangements Chair:

Eva Hajicova, Charles University, Czech Republic

Program Chairs:

Antal van den Bosch, Tilburg University, The Netherlands
Annie Zaenen, Palo Alto Research Center (PARC), USA

Student Research Workshop:

Violeta Seretan, University of Geneva, Switzerland
Chris Biemann, University of Leipzig, Germany
Ellen Riloff (Faculty Advisor), University of Utah, USA

Workshop Chair:

Simone Teufel, University of Cambridge, UK

Tutorial Chair:

Joakim Nivre, Växjö University, Sweden

Demo/Poster Chair:

Sophia Ananiadou, The University of Manchester, UK

Exhibits Chairs:

Jaroslava Hlavacova, Charles University, Czech Republic
Pavel Pecina, Charles University, Czech Republic

Sponsorship Chairs:

Martha Palmer, University of Colorado, USA
Gabor Proszeky, Morphologic, Hungary
Jan Hajic, Charles University, Czech Republic
Jun'ichi Tsujii, University of Tokyo, Japan

Publicity Chairs:

Pavel Stranak, Charles University, Czech Republic
Jiri Mirovsky, Charles University, Czech Republic

Publication Chair:

Su Jian, Institute for Infocomm Research (I2R), Singapore

Mentoring Service:

Florence Reeder, The MITRE Corporation, USA

Student Volunteers:

Marketa Lopatkova, Charles University, Czech Republic

Webmasters:

Zlatka Subrova, Czech Republic
Juraj Simlovic, Czech Republic

Secretariat:

Anna Kotesovcova, Charles University, Czech Republic

Registration:

Priscilla Rasmussen, Association for Computational Linguistics (ACL)

ACL Executive Committee:

Mark Steedman, University of Edinburgh, UK
Bonnie Dorr, University of Maryland, USA
Steven Bird, University of Melbourne, Australia
Jun'ichi Tsujii, The University of Tokyo, Japan
Kathleen F. McCoy, University of Delaware, USA
Dragomir R. Radev, University of Michigan, USA
Owen Rambow, Columbia University, USA
Alex Lascarides, University of Edinburgh, UK
Keh-Yih Su, Behavior Design Corporation, Taiwan
Claire Cardie, Cornell University, USA
Nicoletta Calzolari, Istituto di Linguistica Computazionale del CNR, Italy

Program Committee

Program Chairs:

Antal van den Bosch, Tilburg University, The Netherlands
Annie Zaenen, Palo Alto Research Center, USA

Area Chairs:

Tim Baldwin, University of Melbourne, Australia
Kees van Deemter, University of Aberdeen, UK
Barbara Di Eugenio, University of Illinois at Chicago, USA
Josef van Genabith, Dublin City University, Ireland
Claire Grover, University of Edinburgh, UK
Diana McCarthy, University of Sussex, UK
Dan Roth, University of Illinois at Urbana-Champaign, USA
Richard Sproat, University of Illinois at Urbana-Champaign, USA
Marc Swerts, Tilburg University, The Netherlands
Andy Way, Dublin City University, Ireland

Program Committee Members:

Takeshi Abekawa, Eneko Agirre, Gregory Aist, Cyril Allauzen, Elisabeth André, Shlomo Argamon, Ron Artstein,

Collin Baker, Srinivas Bangalore, Colin Bannard, Regina Barzilay, Roberto Basili, John Bateman, Kenneth Beesley, Anja Belz, Slaven Bilac, Steven Bird, Guido Boella, Francis Bond, Kalina Bontcheva, Johan Bos, Gosse Bouma, Susan Brennan, Chris Brew, Ted Briscoe, Ralf Brown, Paul Buitelaar, Razvan Bunescu, Harry Bunt,

Aoife Cahill, Janet Cahn, Mary Elaine Califf, Charles Callaway, Chris Callison-Burch, Nicoletta Calzolari, Giuseppe Carenini, Michael Carl, Jean Carletta, Xavier Carreras, Justine Cassell, Joyce Chai, Jing-Shin Chang, Ming-Wei Chang, Jinying Chen, Keh-Jiann Chen, Colin Cherry, David Chiang, Grzegorz Chrupala, Jennifer Chu-Carroll, Ilyas Cicekli, Philipp Cimiano, Stephen Clark, Michael Collins, Michael Connor, Mark Core, Mathias Creutz, James Curran,

Walter Daelemans, Ido Dagan, Hercules Dalianis, Hal Daume, Eric de la Clergerie, Maarten de Rijke, Barbara Di Eugenio, Mona Diab, Gael Diaz, Bonnie Dorr, Laila Dybkjær,

Jason Eisner, Noemie Elhadad, Katrin Erk, Dominique Estival, Stefan Evert, David Farwell, Afsaneh Fazly, Marcello Federico, Katja Filippova, Kate Forbes Riley, Mikel Forcada, George Foster, Mary Ellen Foster, Anette Frank, Reva Freedman, Guohong Fu, Pascale Fung,

Rob Gaizauskas, Jianfeng Gao, Kallirroi Georgila, Daniel Gildea, Roxana Girju, Alfio Gliozzo, Sharon Goldwater, Fernando Gomez, Nancy Green, Mark Greenwood, Ralph Grishman, Declan Groves,

Nizar Habash, Keith Hall, Susan Haller, Sanda Harabagiu, Henk Harkema, Tony Hartley, Sven Hartrumpf, Mary Hearne, Marti Hearst, Peter Heeman, James Henderson, Mark Hepple, Ryuichiro Higashinaka, Julia Hirschberg, Graeme Hirst, Barbora Hladka, Julia Hockenmaier, Deirdre Hogan, Baden Hughes,

Diana Inkpen, Kentaro Inui,

Martin Jansche, Valentin Jijkoun, Mark Johnson, Kristiina Jokinen, Pamela Jordan,

Laura Kallmeyer, Stephan Kanthak, Vangelis Karkaletsis, Daisuke Kawahara, John Kelleher, Frank Keller, Andre Kempe, Rodger Kibble, Adam Kilgarriff, Tracy King, Ewan Klein, Alexandre Klementiev, Kevin Knight, Alistair Knott, Philipp Koehn, Rob Koeling, Alexander Koller, Grzegorz Kondrak, Moshe Koppel, Valia Kordoni, Anna Korhonen, Emiel Krahmer, Sandra Kuebler, Jonas Kuhn, Sadao Kurohashi,

Philippe Langlais, Guy Lapalme, Mirella Lapata, Staffan Larsson, Alberto Lavelli, Alon Lavie, Lillian Lee, Jochen Leidner, Oliver Lemon, Yves Lepage, Leonardo Lesmo, Gina Levow, Ian Lewin, William Lewis, Mu Li, Dekang Lin, Jimmy Lin, Diane Litman, Ting Liu, Saturnino Luz,

Bernardo Magnini, Gideon Mann, Chris Manning, Daniel Marcu, Katja Markert, Lluis Marquez, Erwin Marsi, David Martinez, Carlos Martin-Vide, Evgeny Matusov, John Maxwell, Mike Maxwell, Diana Maynard, Andrew McCallum, Ryan McDonald, Kathy McKeown, Susan McRoy, Mike McTear, Dan Melamed, Chris Mellish, Arul Menezes, Helen Meng, Paola Merlo, Detmar Meurers, Rada Mihalcea, Maria Milosavljevic, Eleni Miltsakaki, David Milward, Yusuke Miyao, Dan Moldovan, Diego Molla, Raymond Mooney, Roger Moore, Alessandro Moschitti, Karin Mueller, Reinhard Muskens,

Hiroshi Nakagawa, Hiromi Nakaiwa, Roberto Navigli, Mark-Jan Nederhof, Goran Nenadic, John Nerbonne, Hwee Tou Ng, Vincent Ng, Jian-Yun Nie, Malvina Nissim, Joakim Nivre, David Novick,

Jon Oberlander, Franz Och, Stephan Oepen, Kemal Oflazer, Ian O'Neill, Miles Osborne, Mari Ostendorf,

Tim Paek, Martha Palmer, Patrick Pantel, Ivandre Paraboni, Becky Passonneau, Michael Paul, Fernando Pereira, Scott Piao, Manfred Pinkal, Paul Piwek, Massimo Poesio, Joe Polifroni, Richard Power, John Prager, Stephen Pulman, Vasin Punyakanok,

Chris Quirk,

Reinhard Rapp, Lev Ratinov, Ehud Reiter, Martin Reynaert, Stefan Riezler, German Rigau, Ellen Riloff, Brian Roark, Laurent Romary,

Bogdan Sacaleanu, Horacio Saggion, Anna Sagvall Hein, Mark Sanderson, Anoop Sarkar, Avik Sarkar, Helmut Schmid, Gerold Schneider, Patrick Schone, Hinrich Schuetze, Sabine Schulte im Walde, Satoshi Sekine, Stephanie Seneff, Izhak Shafran, Advaith Siddharthan, Khalil Simaan, Kevin Small, Noah A. Smith, Harold Somers, Virach Sornlertlamvanich, Manfred Stede, Mark Steedman, Amanda Stent, Mark Stevenson, Carlo Strapparava, Oliver Streiter, Nicolas Stroppa, Michael Strube, Jian Su, Rajen Subba, Stan Szpakowicz,

Maite Taboada, Hiroya Takamura, Kumiko Tanaka-Ishii, Louis ten Bosch, Joel Tetreault, Simone Teufel, Mariet Theune, Joerg Tiedemann, Takenobu Tokunaga, Kristina Toutanova, Isabel Trancoso, Richard Tsai, Gokhan Tur,

Nicola Ueffing, Nathan Vaillette, Gertjan van Noord, Menno van Zaanen, Sebastian Varges, Paola Velardi, Renata Vieira, Begona Villada Moiron, Carl Vogel, Stephan Vogel, Piek Vossen,

Marilyn Walker, Haifeng Wang, Xinglong Wang, Bonnie Webber, David Weir, Ben Wellner, Richard Wicentowski, Janyce Wiebe, Ross Wilkinson, Yorick Wilks, Theresa Wilson, Shuly Wintner, Dekai Wu,

Fei Xia, Nianwen Xue, Endong Xun,

Muyun Yang, Scott Wen-tau Yih, Anssi Yli-Jyrä,

Dmitry Zelenko, Richard Zens, ChengXiang Zhai, Min Zhang, Tong Zhang, Guodong Zhou

Conference Program

Monday, June 25, 2007

8:45–9:00 Opening Session

9:00–10:00 Invited Talk by Tom Mitchell

10:00–10:30 Break

Session 1A: Machine Translation 1

Session 1B: Word Sense Disambiguation

Monday, June 25, 2007 (continued)

Session 1C: Language Modeling 1

10:30–10:55 *A Bayesian Model for Discovering Typological Implications*
Hal Daumé III and Lyle Campbell

10:55–11:20 *A Discriminative Language Model with Pseudo-negative Samples*
Daisuke Okanohara and Jun'ichi Tsujii

11:20–11:45 *Detecting Erroneous Sentences using Automatically Mined Sequential Patterns*
Guihua Sun, Xiaohua Liu, Gao Cong, Ming Zhou, Zhongyang Xiong, John Lee and Chin-Yew Lin

11:45–12:10 *Vocabulary Decomposition for Estonian Open Vocabulary Speech Recognition*
Antti Puurula and Mikko Kurimo

Session 1D: Phonology and Morphology 1

10:30–10:55 *Phonological Constraints and Morphological Preprocessing for Grapheme-to-Phoneme Conversion*
Vera Demberg, Helmut Schmid and Gregor Möhler

10:55–11:20 *Redundancy Ratio: An Invariant Property of the Consonant Inventories of the World's Languages*
Animesh Mukherjee, Monojit Choudhury, Anupam Basu and Niloy Ganguly

11:20–11:45 *Multilingual Transliteration Using Feature based Phonetic Method*
Su-Youn Yoon, Kyoung-Young Kim and Richard Sproat

11:45–12:10 *Semantic Transliteration of Personal Names*
Haizhou Li, Khe Chai Sim, Jin-Shea Kuo and Minghui Dong

12:10–13:30 Lunch

Session 2A: Machine Translation 2

13:30–13:55 *Generating Complex Morphology for Machine Translation*
Einat Minkov, Kristina Toutanova and Hisami Suzuki

13:55–14:20 *Assisting Translators in Indirect Lexical Transfer*
Bogdan Babych, Anthony Hartley, Serge Sharoff and Olga Mudraya

Monday, June 25, 2007 (continued)

Session 2D: Language Resources

Session 3A, Machine Learning Methods 1

Session 3B: Machine Translation 3

Monday, June 25, 2007 (continued)

Session 3C: Generation

Session 3D: Multimodality 1

Tuesday, June 26, 2007

Session 4A, Parsing 1

09:00–09:25 *Fast Unsupervised Incremental Parsing*
Yoav Seginer

09:25–09:50 *K-best Spanning Tree Parsing*
Keith Hall

09:50–10:15 *Is the End of Supervised Parsing in Sight?*
Rens Bod

10:15–10:40 *An Ensemble Method for Selection of High Quality Parses*
Roi Reichart and Ari Rappoport

Session 4B: Sentiment 1

09:00–09:25 *Opinion Mining using Econometrics: A Case Study on Reputation Systems*
Anindya Ghose, Panagiotis Ipeirotis and Arun Sundararajan

09:25–09:50 *PageRanking WordNet Synsets: An Application to Opinion Mining*
Andrea Esuli and Fabrizio Sebastiani

09:50–10:15 *Structured Models for Fine-to-Coarse Sentiment Analysis*
Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells and Jeff Reynar

10:15–10:40 *Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification*
John Blitzer, Mark Dredze and Fernando Pereira

Session 4C: Paraphrasing, Textual Entailment

09:00–09:25 *Clustering Clauses for High-Level Relation Detection: An Information-theoretic Approach*
Samuel Brody

09:25–09:50 *Instance-based Evaluation of Entailment Rule Acquisition*
Idan Szpektor, Eyal Shnarch and Ido Dagan

09:50–10:15 *Statistical Machine Translation for Query Expansion in Answer Retrieval*
Stefan Riezler, Alexander Vasserman, Ioannis Tsochantaridis, Vibhu Mittal and Yi Liu

10:15–10:40 *A Computational Model of Text Reuse in Ancient Literary Texts*
John Lee

Tuesday, June 26, 2007 (continued)

Tuesday, June 26, 2007 (continued)

Session 6B: Parsing 2

Session 6C: Multilinguality 1

Session 6D: Language Modeling 3

Wednesday, June 27, 2007

Session 7A: Machine Translation 4

Session 7B: Sequence Processing

Session 7C: Question Answering

Wednesday, June 27, 2007 (continued)

Session 7D: Discourse and Dialog 2

Session 8A: Machine Learning Methods 2

Wednesday, June 27, 2007 (continued)

Session 8B: Machine Translation and Multilinguality

Session 8C: Lexicon and Lexical Semantics

Wednesday, June 27, 2007 (continued)

Session 8D: Phonology and Morphology 2

Session 9A: Parsing 3

Session 9B: Sentiment 2