

CoNLL-2005

**Proceedings of the
Ninth Conference on
Computational Natural
Language Learning**

29-30 June 2005
University of Michigan
Ann Arbor, Michigan, USA

Production and Manufacturing by
Omnipress Inc.
Post Office Box 7214
Madison, WI 53707-7214

©2005 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
75 Paterson Street, Suite 9
New Brunswick, NJ 08901
USA
Tel: +1-732-342-9100
Fax: +1-732-342-9339
acl@aclweb.org

PREFACE

The 2005 Conference on Computational Natural Language Learning (CoNLL-2005) is the ninth in a series of meetings organized by SIGNLL, the ACL special interest group on natural language learning. This year's CoNLL will be held in Ann Arbor, Michigan, on June 29 and 30, in conjunction with the ACL 2005 conference.

This year we encouraged submissions addressing “deeper linguistic phenomena than have typically been covered in the past”, a theme reflected in the shared task, and one which will be addressed by our invited speakers, Mark Johnson and Mark Steedman. The latter talk will be part of what should be a very interesting joint session with the Workshop on Psychocomputational Models of Human Language Acquisition organized by William Sakas.

A total of 70 papers were submitted to CoNLL's main session, of which we were able to accept only 19, making this the most competitive CoNLL meeting to date. We are particularly grateful to our program committee for their work reviewing this large number of papers on a tight schedule.

In keeping with the unique tradition of CoNLL, we also have a shared task session this year, on semantic role labeling. This year's task included use of full syntactic parses, a step beyond the chunk-level information used in 2004. The shared task was coordinated by Xavier Carreras and Lluís Màrquez. Common training and test data were made available to all participants, who had to build their learning system and test it on this task. The shared task also achieved a record number of submissions this year, and these proceedings include system descriptions from each of the 19 participants.

In addition to the program committee and shared task organizers, we are very indebted to the 2005 ACL conference organizers, in particular Dragomir Radev, Mirella Lapata, Jason Eisner, and Philipp Koehn, for their help with local arrangements and the publication of the proceedings.

We hope you enjoy this year's meeting!

Ido Dagan and Dan Gildea
May 2005

Organizers:

Ido Dagan, Bar-Ilan University (Israel)
Daniel Gildea, University of Rochester (USA)

Shared Task Organizers:

Xavier Carreras and Lluís Màrquez
Technical University of Catalonia (Spain)

Program Committee:

Steven Abney, University of Michigan (USA)
Eneko Agirre, University of the Basque Country (Spain)
Regina Barzilay, Massachusetts Institute of Technology (USA)
Claire Cardie, Cornell University (USA)
John Carroll, University of Sussex (UK)
Eugene Charniak, Brown University (USA)
James Cussens, University of York (UK)
Walter Daelemans, University of Antwerp (Belgium)
Radu Florian, IBM (USA)
Dayne Freitag, Fair Isaac (USA)
Rebecca Hwa, University of Pittsburgh (USA)
Hang Li, Microsoft (China)
Dekang Lin, University of Alberta (Canada)
Diane Litman, University of Pittsburgh (USA)
Yuji Matsumoto, Nara Institute of Science and Technology (Japan)
Diana McCarthy, University of Sussex (UK)
Rada Mihalcea, University of North Texas (USA)
John Nerbonne, University of Groningen (Netherlands)
Hwee-Tou Ng, National University of Singapore (Singapore)
Grace Ngai, The Hong Kong Polytechnic University (Hong Kong)
Miles Osborne, University of Edinburgh (UK)
Patrick Pantel, Information Sciences Institute (USA)
David Powers, Flinders University (Australia)
Dragomir Radev, University of Michigan (USA)
Ellen Riloff, University of Utah (USA)
Dan Roth, University of Illinois at Urbana-Champaign (USA)
Anoop Sarkar, Simon Fraser University (Canada)
Suzanne Stevenson, University of Toronto (Canada)
Keh Yih Su, Behavior Design Corporation (ROC)
Erik Tjong Kim Sang, University of Antwerp (Belgium)
Antal van den Bosch, Tilburg University (Netherlands)
Janyce Wiebe, University of Pittsburgh (USA)

Additional Reviewers:

Iñaki Alegria, Toine Bogers, Sander Canisius, Yunbo Cao, Hal Daume III, Guy De Pauw, Alex Fraser, Véronique Hoste, David Martinez, Art Munson, Vivi Năstase, Liang Zhou

Invited Speakers:

Mark Johnson, Brown University
Mark Steedman, University of Edinburgh

Table of Contents

Main Session

<i>Effective use of WordNet Semantics via Kernel-Based Learning</i> Roberto Basili, Marco Cammisa and Alessandro Moschitti	1
<i>A Statistical Semantic Parser that Integrates Syntax and Semantics</i> Ruifang Ge and Raymond Mooney	9
<i>Search Engine Statistics Beyond the n-gram: Application to Noun Compound Bracketing</i> Preslav Nakov and Marti Hearst	17
<i>New Experiments in Distributional Representations of Synonymy</i> Dayne Freitag, Matthias Blume, John Byrnes, Edmond Chow, Sadik Kapadia, Richard Rohwer and Zhiqiang Wang	25
<i>Word Independent Context Pair Classification Model for Word Sense Disambiguation</i> Cheng Niu, Wei Li, Rohini K. Srihari and Huifeng Li	33
<i>Computing Word Similarity and Identifying Cognates with Pair Hidden Markov Models</i> Wesley Mackay and Grzegorz Kondrak	40
<i>A Bayesian Mixture Model for Term Re-occurrence and Burstiness</i> Avik Sarkar, Paul H. Garthwaite and Anne De Roeck	48
<i>Domain Kernels for Text Categorization</i> Alfio Gliozzo and Carlo Strapparava	56
<i>Discriminative Training of Clustering Functions: Theory and Experiments with Entity Identification</i> Xin Li and Dan Roth	64
<i>Using Uneven Margins SVM and Perceptron for Information Extraction</i> Yaoyong Li, Kalina Bontcheva and Hamish Cunningham	72
<i>Improving Sequence Segmentation Learning by Predicting Trigrams</i> Antal van den Bosch and Walter Daelemans	80
<i>An Expectation Maximization Approach to Pronoun Resolution</i> Colin Cherry and Shane Bergsma	88
<i>Probabilistic Head-Driven Parsing for Discourse Structure</i> Jason Baldridge and Alex Lascarides	96
<i>Intentional Context in Situated Natural Language Learning</i> Michael Fleischman and Deb Roy	104
<i>Representational Bias in Unsupervised Learning of Syllable Structure</i> Sharon Goldwater and Mark Johnson	112

<i>An Analogical Learner for Morphological Analysis</i>	
Nicolas Stroppa and François Yvon	120
<i>Morphology Induction from Term Clusters</i>	
Dayne Freitag	128
<i>Beyond the Pipeline: Discrete Optimization in NLP</i>	
Tomasz Marciniak and Michael Strube	136
<i>Investigating the Effects of Selective Sampling on the Annotation Task</i>	
Ben Hachey, Beatrice Alex and Markus Becker	144
Shared Task	
<i>Introduction to the CoNLL-2005 Shared Task: Semantic Role Labeling</i>	
Xavier Carreras and Lluís Màrquez	152
<i>Inferring Semantic Roles Using Sub-categorization Frames and Maximum Entropy Model</i>	
Akshar Bharati, Sriram Venkatapathy and Prashanth Reddy	165
<i>Semantic Role Labelling with Tree Conditional Random Fields</i>	
Trevor Cohn and Philip Blunsom	169
<i>A Joint Model for Semantic Role Labeling</i>	
Aria Haghighi, Kristina Toutanova and Christopher Manning	173
<i>Sparse Bayesian Classification of Predicate Arguments</i>	
Richard Johansson and Pierre Nugues	177
<i>Generalized Inference with Multiple Semantic Role Labeling Systems</i>	
Peter Koomen, Vasin Punyakanok, Dan Roth and Wen-tau Yih	181
<i>Semantic Role Labeling via Consensus in Pattern-Matching</i>	
Chi-San Lin and Tony C. Smith	185
<i>Semantic Role Labeling System Using Maximum Entropy Classifier</i>	
Ting Liu, Wanxiang Che, Sheng Li, Yuxuan Hu and Huaijun Liu	189
<i>Semantic Role Labeling as Sequential Tagging</i>	
Lluís Màrquez, Pere Comas, Jesús Giménez and Neus Català	193
<i>Semantic Role Labeling Using Support Vector Machines</i>	
Tomohiro Mitsumori, Masaki Murata, Yasushi Fukuda, Kouichi Doi and Hirohumi Doi	197
<i>Hierarchical Semantic Role Labeling</i>	
Alessandro Moschitti, Ana-Maria Giuglea, Bonaventura Coppola and Roberto Basili	201
<i>Semantic Role Labeling Using libSVM</i>	
Necati Ercan Ozgencil and Nancy McCracken	205

<i>Maximum Entropy based Semantic Role Labeling</i> Kyung-Mi Park and Hae-Chang Rim	209
<i>Semantic Role Labeling Using Lexical Statistical Information</i> Simone Paolo Ponzetto and Michael Strube	213
<i>Semantic Role Chunking Combining Complementary Syntactic Views</i> Sameer Pradhan, Kadri Hacioglu, Wayne Ward, James H. Martin and Daniel Jurafsky	217
<i>Semantic Role Labeling Using Complete Syntactic Analysis</i> Mihai Surdeanu and Jordi Turmo	221
<i>Joint Parsing and Semantic Role Labeling</i> Charles Sutton and Andrew McCallum	225
<i>Applying Spelling Error Correction Techniques for Improving Semantic Role Labelling</i> Erik Tjong Kim Sang, Sander Canisius, Antal van den Bosch and Toine Bogers	229
<i>Exploiting Full Parsing Information to Label Semantic Roles Using an Ensemble of ME and SVM via Integer Linear Programming</i> Tzong-Han Tsai, Chia-Wei Wu, Yu-Chun Lin and Wen-Lian Hsu	233
<i>The Integration of Syntactic Parsing and Semantic Role Labeling</i> Szu-ting Yi and Martha Palmer	237

Conference Program

Wednesday June 29, 2005

8:45 Welcome

Session 1: Syntax and Semantics

8:50 *Effective use of WordNet Semantics via Kernel-Based Learning*
Roberto Basili, Marco Cammisa and Alessandro Moschitti

9:15 *A Statistical Semantic Parser that Integrates Syntax and Semantics*
Ruifang Ge and Raymond Mooney

9:40 *Search Engine Statistics Beyond the n-gram: Application to Noun Compound Bracketing*
Preslav Nakov and Marti Hearst

10:05 *New Experiments in Distributional Representations of Synonymy*
Dayne Freitag, Matthias Blume, John Byrnes, Edmond Chow, Sadik Kapadia, Richard Rohwer and Zhiqiang Wang

10:30 coffee break

Session 2: Semantics and Text Classification

11:00 *Word Independent Context Pair Classification Model for Word Sense Disambiguation*
Cheng Niu, Wei Li, Rohini K. Srihari and Huifeng Li

11:25 *Computing Word Similarity and Identifying Cognates with Pair Hidden Markov Models*
Wesley Mackay and Grzegorz Kondrak

11:50 *A Bayesian Mixture Model for Term Re-occurrence and Burstiness*
Avik Sarkar, Paul H. Garthwaite and Anne De Roeck

12:15 *Domain Kernels for Text Categorization*
Alfio Gliozzo and Carlo Strapparava

12:40 lunch

Wednesday June 29, 2005 (continued)

Session 3: Information Extraction and Learning Methods

- 14:00 *Discriminative Training of Clustering Functions: Theory and Experiments with Entity Identification*
Xin Li and Dan Roth
- 14:25 *Using Uneven Margins SVM and Perceptron for Information Extraction*
Yaoyong Li, Kalina Bontcheva and Hamish Cunningham
- 14:50 *Improving Sequence Segmentation Learning by Predicting Trigrams*
Antal van den Bosch and Walter Daelemans
- 15:15 SIGNLL business meeting
- 15:40 coffee break
- 16:10 Invited Speaker: Mark Johnson

Session 4: Discourse and Anaphora

- 17:10 *An Expectation Maximization Approach to Pronoun Resolution*
Colin Cherry and Shane Bergsma
- 17:35 *Probabilistic Head-Driven Parsing for Discourse Structure*
Jason Baldridge and Alex Lascarides

Thursday June 30, 2005

Session 5: Joint Session with Workshop on Psychocomputational Models of Human Language Acquisition

- 9:00 Invited Speaker: Mark Steedman
- 9:50 *Steps Toward Deep Lexical Acquisition*
Sourabh Niyogi (Workshop on Psychocomputational Models)
- 10:15 *Intentional Context in Situated Natural Language Learning*
Michael Fleischman and Deb Roy

10:40 coffee break

Session 6: Morphology

- 11:10 *Representational Bias in Unsupervised Learning of Syllable Structure*
Sharon Goldwater and Mark Johnson
- 11:35 *An Analogical Learner for Morphological Analysis*
Nicolas Stroppa and François Yvon
- 12:00 *Morphology Induction from Term Clusters*
Dayne Freitag

12:25 lunch

Session 7: Learning Methods and Architectures

- 14:00 *Beyond the Pipeline: Discrete Optimization in NLP*
Tomasz Marciniak and Michael Strube
- 14:25 *Investigating the Effects of Selective Sampling on the Annotation Task*
Ben Hachey, Beatrice Alex and Markus Becker
- 14:50 **Shared Task:** Introduction and System Descriptions
- 15:30 coffee break
- 16:00 **Shared Task:** System Descriptions
- 17:00 **Shared Task:** Discussion
- 18:00 Closing

