

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

Hwee Tou Ng and Ellen Riloff (editors)

Held in cooperation with HLT-NAACL 2004

**May 6–7, 2004
Boston, Massachusetts**

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

Hwee Tou Ng and Ellen Riloff (editors)

Held in cooperation with HLT-NAACL 2004

**May 6–7, 2004
Boston, Massachusetts**

Copyright © 2004 The Association for Computational Linguistics

ISBN: 1-932432-30-2

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics

3 Landmark Center

East Stroudsburg, PA 18301 USA

Phone: +1-570-476-8006

Fax: +1-570-476-0860

acl@aclweb.org

www.aclweb.org

SIGNLL

Computational Natural Language Learning (CoNLL-2004)

May 6-7, 2004

INVITED SPEAKERS:

Christopher Manning, Stanford University (USA)
Martha Palmer, University of Pennsylvania (USA)

PROGRAM COMMITTEE:

Hwee Tou Ng, National University of Singapore (Singapore), Co-Chair
Ellen Riloff, University of Utah (USA), Co-Chair
Xavier Carreras, UPC (Spain), Shared Task Co-Chair
Lluís Màrquez, UPC (Spain), Shared Task Co-Chair

Steven Abney, University of Michigan (USA)
Regina Barzilay, Massachusetts Institute of Technology (USA)
Thorsten Brants, Google Inc (USA)
Claire Cardie, Cornell University (USA)
Eugene Charniak, Brown University (USA)
James Cussens, University of York (UK)
Walter Daelemans, University of Antwerp (Belgium)
Ido Dagan, Bar Ilan University / LingoMotors (Israel)
Radu Florian, IBM (USA)
Ulf Hermjakob, USC Information Sciences Institute (USA)
Hang Li, Microsoft Research Asia (China)
Dekang Lin, University of Alberta (Canada)
Diane Litman, University of Pittsburgh (USA)
Yuji Matsumoto, Nara Institute of Science and Technology (Japan)
Rada Mihalcea, University of North Texas (USA)
Raymond Mooney, University of Texas at Austin (USA)
John Nerbonne, University of Groningen (Netherlands)
Grace Ngai, Hong Kong Polytechnic University (Hong Kong)
Franz Josef Och, USC Information Sciences Institute (USA)
Miles Osborne, University of Edinburgh (UK)
Ted Pedersen, University of Minnesota, Duluth (USA)
David Pierce, SUNY Buffalo (USA)
David Powers, Flinders University (Australia)
Dan Roth, University of Illinois at Urbana-Champaign (USA)
Erik Tjong Kim Sang, University of Antwerp (Belgium)
Anoop Sarkar, Simon Fraser University (Canada)
Suzanne Stevenson, University of Toronto (Canada)
Cynthia Thompson, University of Utah (USA)
Antal van den Bosch, Tilburg University (Netherlands)
Janyce Wiebe, University of Pittsburgh (USA)

Additional Reviewers:

Yunbo Cao, Microsoft Research Asia (China)
Ruifang Ge, University of Texas at Austin (USA)
Lluís Padró, Technical University of Catalonia (Spain)
Mihai Surdeanu, Technical University of Catalonia (Spain)
Yuk Wah Wong, University of Texas at Austin (USA)

SHARED TASK COORDINATORS:

Xavier Carreras, UPC (Spain)

Lluís Màrquez, UPC (Spain)

CONFERENCE WEBSITE: <http://cnts.uia.ac.be/conll2004/>

PREFACE

The 2004 Conference on Computational Natural Language Learning (CoNLL-2004) is the eighth 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 Boston, Massachusetts, USA on May 6 and 7, in conjunction with the HLT-NAACL 2004 Conference. This is the first time that CoNLL will be held in the USA.

The annual CoNLL meeting is intended to address all aspects of computational natural language learning. We have an exciting program this year, thanks to the many good submissions that we received. In all, 23 papers were submitted to the main session, of which 11 were selected by the program committee to be presented at the conference. They represent cutting-edge research on computational natural language learning.

In keeping with the unique tradition of CoNLL, we also have a shared task session this year, on semantic role labeling, which is a topic of growing interest in the field. 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. We received enthusiastic responses from many research groups, and the shared task papers contain descriptions of the ten participating systems and their results.

This year's CoNLL also features two invited presentations, one by Christopher Manning on what the future holds for computational language learning, and one by Martha Palmer on semantic role labeling, the topic of this year's shared task.

Our program committee and additional reviewers have spent much time carefully reviewing the submitted papers, for which we are most grateful. We would also like to thank the HLT-NAACL 2004 conference organizers for helping us with the local organization. And a special thanks for the enthusiastic and capable work of Xavier Carreras and Lluís Màrquez in organizing and running the semantic role labeling shared task.

We wish you an enjoyable time attending CoNLL-2004!

Hwee Tou Ng & Ellen Riloff (editors)
May 2004

Table of Contents

<i>A Linear Programming Formulation for Global Inference in Natural Language Tasks</i> Dan Roth and Wen-tau Yih	1
<i>Semantic Lexicon Construction: Learning from Unlabeled Data via Spectral Analysis</i> Rie Kubota Ando.....	9
<i>A Semantic Kernel for Predicate Argument Classification</i> Alessandro Moschitti and Cosmin Adrian Bejan	17
<i>Combining Lexical and Syntactic Features for Supervised Word Sense Disambiguation</i> Saif Mohammad and Ted Pedersen	25
<i>Co-training and Self-training for Word Sense Disambiguation</i> Rada Mihalcea	33
<i>Word Sense Discrimination by Clustering Contexts in Vector and Similarity Spaces</i> Amruta Purandare and Ted Pedersen	41
<i>Memory-Based Dependency Parsing</i> Joakim Nivre, Johan Hall and Jens Nilsson	49
<i>Modeling Category Structures with a Kernel Function</i> Hiroya Takamura, Yuji Matsumoto and Hiroyasu Yamada.....	57
<i>A Comparison of Manual and Automatic Constructions of Category Hierarchy for Classifying Large Corpora</i> Fumiyo Fukumoto and Yoshimi Suzuki	65
<i>Thesauruses for Prepositional Phrase Attachment</i> Mark McLauchlan	73
<i>Calculating Semantic Distance between Word Sense Probability Distributions</i> Vivian Tsang and Suzanne Stevenson.....	81
<i>Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling</i> Xavier Carreras and Lluís Màrquez	89
<i>Semantic Role Labelling With Chunk Sequences</i> Ulrike Baldewein, Katrin Erk, Sebastian Padó and Detlef Prescher	98
<i>Memory-based semantic role labeling: Optimizing features, algorithm, and output</i> Antal van den Bosch, Sander Canisius, Walter Daelemans, Iris Hendrickx and Erik Tjong Kim Sang	102
<i>Hierarchical Recognition of Propositional Arguments with Perceptrons</i> Xavier Carreras, Lluís Màrquez and Grzegorz Chrupala.....	106

<i>Semantic Role Labeling by Tagging Syntactic Chunks</i> Kadri Hacioglu, Sameer Pradhan, Wayne Ward, James H. Martin and Daniel Jurafsky	110
<i>A transformation-based approach to argument labeling</i> Derrick Higgins	114
<i>A Memory-Based Approach for Semantic Role Labeling</i> Beata Kouchnir	118
<i>Semantic Role Labeling using Maximum Entropy Model</i> Joon-Ho Lim, Young-Sook Hwang, So-Young Park and Hae-Chang Rim	122
<i>Two-Phase Semantic Role Labeling based on Support Vector Machines</i> Kyung-Mi Park, Young-Sook Hwang and Hae-Chang Rim	126
<i>Semantic Role Labeling Via Generalized Inference Over Classifiers</i> Vasin Punyakanok, Dan Roth, Wen-tau Yih, Dav Zimak and Yuancheng Tu	130
<i>Learning Transformation Rules for Semantic Role Labeling</i> Ken Williams, Christopher Dozier and Andrew McCulloh	134
<i>Language Learning: Beyond Thunderdome</i> Christopher D. Manning	138
<i>Putting Meaning into Your Trees</i> Martha Palmer	139

CONFERENCE PROGRAM

Thursday, May 6

9:00-9:05 Welcome (Hwee Tou Ng and Ellen Riloff)

Session 1: Inference and Semantics

9:05-9:35 *A Linear Programming Formulation for Global Inference in Natural Language Tasks*
Dan Roth and Wen-tau Yih

9:35-10:05 *Semantic Lexicon Construction: Learning from Unlabeled Data via Spectral Analysis*
Rie Kubota Ando

10:05-10:35 *A Semantic Kernel for Predicate Argument Classification*
Alessandro Moschitti and Cosmin Adrian Bejan

10:35-11:00 Break

Session 2: Word Sense Disambiguation

11:00-11:30 *Combining Lexical and Syntactic Features for Supervised Word Sense Disambiguation*
Saif Mohammad and Ted Pedersen

11:30-12:00 *Co-training and Self-training for Word Sense Disambiguation*
Rada Mihalcea

12:00-12:30 *Word Sense Discrimination by Clustering Contexts in Vector and Similarity Spaces*
Amruta Purandare and Ted Pedersen

12:30-14:00 Lunch

Session 3: Parsing

14:00-14:30 *Memory-Based Dependency Parsing*
Joakim Nivre, Johan Hall and Jens Nilsson

Invited Talk

14:30-15:30 *Language Learning: Beyond Thunderdome*
Christopher D. Manning

15:30-16:00 Break

CONFERENCE PROGRAM

Thursday, May 6

Session 4: Text Classification

- 16:00-16:30 *Modeling Category Structures with a Kernel Function*
Hiroya Takamura, Yuji Matsumoto and Hiroyasu Yamada
- 16:30-17:00 *A Comparison of Manual and Automatic Constructions of Category Hierarchy for Classifying Large Corpora*
Fumiyo Fukumoto and Yoshimi Suzuki
- 17:00-17:30 SIGNLL Business Meeting

Friday, May 7

Session 5: Syntax and Semantics

- 9:30-10:00 *Thesauruses for Prepositional Phrase Attachment*
Mark McLauchlan
- 10:00-10:30 *Calculating Semantic Distance between Word Sense Probability Distributions*
Vivian Tsang and Suzanne Stevenson
- 10:30-11:00 Break
- Invited Talk
- 11:00-12:00 *Putting Meaning into Your Trees*
Martha Palmer
- 12:00-13:30 Lunch

Shared Task Session

- 13:30-14:00 *Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling*
Xavier Carreras and Lluís Màrquez
- 14:00-14:10 *Semantic Role Labelling With Chunk Sequences*
Ulrike Baldewein, Katrin Erk, Sebastian Padó and Detlef Prescher
- 14:10-14:20 *Memory-based semantic role labeling: Optimizing features, algorithm, and output*
Antal van den Bosch, Sander Canisius, Walter Daelemans, Iris Hendrickx and Erik Tjong Kim Sang
- 14:20-14:30 *Hierarchical Recognition of Propositional Arguments with Perceptrons*
Xavier Carreras, Lluís Màrquez and Grzegorz Chrupała
- 14:30-14:40 *Semantic Role Labeling by Tagging Syntactic Chunks*
Kadri Hacioglu, Sameer Pradhan, Wayne Ward, James H. Martin and Daniel Jurafsky

CONFERENCE PROGRAM

Friday, May 7

- 14:40-14:50 *A transformation-based approach to argument labeling*
Derrick Higgins
- 14:50-15:00 *A Memory-Based Approach for Semantic Role Labeling*
Beata Kouchnir
- 15:00-15:10 *Semantic Role Labeling using Maximum Entropy Model*
Joon-Ho Lim, Young-Sook Hwang, So-Young Park and Hae-Chang Rim
- 15:10-15:20 *Two-Phase Semantic Role Labeling based on Support Vector Machines*
Kyung-Mi Park, Young-Sook Hwang and Hae-Chang Rim
- 15:20-16:00 Break
- 16:00-16:10 *Semantic Role Labeling Via Generalized Inference Over Classifiers*
Vasin Punyakanok, Dan Roth, Wen-tau Yih, Dav Zimak and Yuancheng Tu
- 16:10-16:20 *Learning Transformation Rules for Semantic Role Labeling*
Ken Williams, Christopher Dozier and Andrew McCulloh
- 16:20-17:15 Summary and Discussion on Shared Task

Author Index

Ando, Rie Kubota	9	van den Bosch, Antal	102
Baldewein, Ulrike	98	Ward, Wayne	110
Bejan, Cosmin Adrian	17	Williams, Ken	134
Canisius, Sander	102	Yamada, Hiroyasu	57
Carreras, Xavier	89, 106	Yih, Wen-tau	1, 130
Chrupała, Grzegorz	106	Zimak, Dav	130
Daelemans, Walter	102		
Dozier, Christopher	134		
Erk, Katrin	98		
Fukumoto, Fumiyo	65		
Hacioglu, Kadri	110		
Hall, Johan	49		
Hendrickx, Iris	102		
Higgins, Derrick	114		
Hwang, Young-Sook	122, 126		
Jurafsky, Daniel	110		
Kouchnir, Beata	118		
Lim, Joon-Ho	122		
Manning, Christopher D.	138		
Màrquez, Lluís	89, 106		
Martin, James H.	110		
Matsumoto, Yuji	57		
McCulloh, Andrew	134		
McLauchlan, Mark	73		
Mihalcea, Rada	33		
Mohammad, Saif	25		
Moschitti, Alessandro	17		
Nilsson, Jens	49		
Nivre, Joakim	49		
Padó, Sebastian	98		
Palmer, Martha	139		
Park, Kyung-Mi	126		
Park, So-Young	122		
Pedersen, Ted	25, 41		
Pradhan, Sameer	110		
Prescher, Detlef	98		
Punyakankok, Vasin	130		
Purandare, Amruta	41		
Rim, Hae-Chang	122, 126		
Roth, Dan	1, 130		
Stevenson, Suzanne	81		
Suzuki, Yoshimi	65		
Takamura, Hiroya	57		
Tjong Kim Sang, Erik	102		
Tsang, Vivian	81		
Tu, Yuancheng	130		