

CoNLL 2008

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

Conference chairs:
Alexander Clark and Kristina Toutanova

16–17 August 2008
Manchester, UK

©2008 The Coling 2008 Organizing Committee

Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Nonported* license
<http://creativecommons.org/licenses/by-nc-sa/3.0/>
Some rights reserved

Order copies of this and other Coling 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 978-1-905593-48-4

Design by Chimney Design, Brighton, UK
Production and manufacture by One Digital, Brighton, UK

Introduction

The 2008 Conference on Computational Natural Language Learning is the twelfth in the series of yearly meetings organized by SIGNLL, the ACL special interest group on natural language learning. CoNLL 2008 will be held in Manchester, UK, August 16-17, 2008, in conjunction with Coling 2008.

We are delighted to report that CoNLL's main session received a large number of submissions. A total of 85 papers were under consideration for the main session after several withdrawals, and of them only 20 were accepted. This makes this year's CoNLL especially competitive and contributes to an interesting program. We are grateful to the program committee members for their service in evaluating the submissions. Special thanks to the program committee members who joined on a short notice to help with the larger than expected number of submissions.

This year CoNLL had two special themes of interest, both of which solicited papers on models that explain natural phenomena relating to human language. The first concerned the central scientific problem addressed by CoNLL: the study of first language acquisition. The second theme was the central engineering problem: how to build systems that do something useful, especially complete systems that solve real problems.

The first theme contributed to an increased number of high-quality submissions in the first language acquisition area. Two sessions of the conference will be devoted to papers on this topic. The second theme led to submissions in diverse traditional NLP application areas.

As in previous years, CoNLL 2008 has a shared task. This year, the conference shared task proposed to merge the shared task topics from the last four years (2004-2007) into a unique task called "Joint Learning of Syntactic and Semantic Dependencies". Both syntactic dependencies (extracted from the Penn Treebank) and semantic dependencies (extracted from PropBank and NomBank) were jointly addressed under a unique unified representation.

The shared task was organized by Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre.

The call was very successful attracting the interest of more than 50 teams from all over the world, which represented a wide variety of universities, research institutions, and companies. At the end of the evaluation period, 22 teams submitted results (with 19 and 5 contributions to the closed and open challenges, respectively). All this work will be presented in the conference in the form of 5 selected oral talks and 14 posters.

In our opinion, the current shared task constitutes a qualitative step ahead and we hope that the resources created and the body of work presented will both serve as a benchmark and have a substantial impact on future research on syntactic-semantic parsing.

We are excited that the invited speakers at CoNLL 2008 will be Regina Barzilay and Nick Chater.

Finally, we would like to thank the SIGNLL board members for useful discussion, Erik Tjong Kim Sang, who acted as the information officer, and especially Lluís Màrquez and Joakim Nivre, who helped us greatly with advice around the conference organization, as well as to the organizers of COLING

2008, Harold Somers, Mark Stevenson and Roger Evans. Many thanks also to Microsoft Research for sponsoring CoNLL this year and to Priscilla Rasmussen for help with the finances.

Enjoy this year's conference!

Alex Clark and Kristina Toutanova

CoNLL 2008 Conference Chairs

Conference Chairs:

Alex Clark, Royal Holloway, University of London
Kristina Toutanova, Microsoft Research

Programme Committee:

Steve Abney, University of Michigan
Eneko Agirre, University of the Basque Country
Galen Andrew, Microsoft Research
Tim Baldwin, University of Melbourne
Regina Barzilay, MIT
Roberto Basili, University of Roma, Tor Vergata
Phil Blunsom, University of Edinburgh
Thorsten Brants, Google Research
Paula Buttery, Cambridge University
Xavier Carreras, MIT
Nick Chater, University College London
Ciprian Chelba, Google Research
Colin Cherry, Microsoft Research
Alexander Clark, Royal Holloway, University of London
Stephen Clark, Oxford University
James Cussens, University of York
Walter Daelemans, CNTS, Antwerp
Hal Daumé III, University of Utah
Jenny Finkel, Stanford University
Radu Florian, IBM research
Dayne Freitag, HNC Software
Michel Galley, Stanford University
Jianfeng Gao, Microsoft Research
Daniel Gildea, Rochester
Sharon Goldwater, University of Edinburgh
Jan Hajic, Institute of Formal and Applied Linguistics, Charles University in Prague
Marti Hearst, UC Berkeley
James Henderson, University of Geneva
Liang Huang, University of Pennsylvania
Rie Johnson (formerly, Ando), RJ Research Consulting
Rohit Kate, University of Texas at Austin
Philipp Koehn, University of Edinburgh
Rob Koeling, Sussex University
Anna Korhonen, Cambridge University
Shalom Lappin, King's College, London

Roger Levy, UC San Diego
Percy Liang, UC Berkeley
Rob Malouf, San Diego State University
Lluís Màrquez, Technical University of Catalonia
Yuji Matsumoto, Nara Institute of Science and Technology
Diana McCarthy, Sussex University
Rada Mihalcea, University of North Texas
Alessandro Moschitti, DISI, University of Trento
John Nerbonne, University of Groningen
Hwee Tou Ng, National University of Singapore
Vincent Ng, University of Texas at Dallas
Joakim Nivre, Uppsala University
Franz Och, Google Research
Miles Osborne, University of Edinburgh
Slav Petrov, UC Berkeley
David Powers, Flinders University
Chris Quirk, Microsoft Research
Ari Rappoport, Hebrew University
Ellen Riloff, University of Utah
Dan Roth, University of Illinois, Urbana-Champaign
William Sakas, City University of New York
Anoop Sarkar, Simon Fraser University
Richard Sproat, University of Illinois, Urbana-Champaign
Suzanne Stevenson, University of Toronto
Mihai Surdeanu, Barcelona Media Information Center
Charles Sutton, UC Berkeley
Kristina Toutanova, Microsoft Research
Antal van den Bosch, Tilburg University
Charles Yang, University of Pennsylvania

Invited Speakers:

Regina Barzilay, MIT Computer Science & Artificial Intelligence Lab
Nick Chater, Department of Psychology, University College London

Table of Contents

<i>Semantic Parsing for High-Precision Semantic Role Labelling</i> Paola Merlo and Gabriele Musillo	1
<i>TAG, Dynamic Programming, and the Perceptron for Efficient, Feature-Rich Parsing</i> Xavier Carreras, Michael Collins and Terry Koo	9
<i>A Fast Boosting-based Learner for Feature-Rich Tagging and Chunking</i> Tomoya Iwakura and Seishi Okamoto	17
<i>Linguistic features in data-driven dependency parsing</i> Lilja Ovreid	25
<i>Transforming Meaning Representation Grammars to Improve Semantic Parsing</i> Rohit Kate	33
<i>Using LDA to detect semantically incoherent documents</i> Hemant Misra, Olivier Cappe and Francois Yvon	41
<i>Picking them up and Figuring them out: Verb-Particle Constructions, Noise and Idiomaticity</i> Carlos Ramisch, Aline Villavicencio, Leonardo Moura and Marco Idiart	49
<i>Fast Mapping in Word Learning: What Probabilities Tell Us</i> Afra Alishahi, Afsaneh Fazly and Suzanne Stevenson	57
<i>Improving Word Segmentation by Simultaneously Learning Phonotactics</i> Daniel Blanchard and Jeffrey Heinz	65
<i>A MDL-based Model of Gender Knowledge Acquisition</i> Harmony Marchal, Benoît Lemaire, Maryse Bianco and Philippe Dessus	73
<i>Baby SRL: Modeling Early Language Acquisition.</i> Michael Connor, Yael Gertner, Cynthia Fisher and Dan Roth	81
<i>An Incremental Bayesian Model for Learning Syntactic Categories</i> Christopher Parisien, Afsaneh Fazly and Suzanne Stevenson	89
<i>Fully Unsupervised Graph-Based Discovery of General-Specific Noun Relationships from Web Corpora Frequency Counts</i> Gaël Dias, Raycho Mukelov and Guillaume Cleuziou	97
<i>Acquiring Knowledge from the Web to be used as Selectors for Noun Sense Disambiguation</i> Hansen A. Schwartz and Fernando Gomez	105
<i>Automatic Chinese Catchword Extraction Based on Time Series Analysis</i> Han Ren, Donghong Ji, Jing Wan and Lei Han	113
<i>Easy as ABC? Facilitating Pictorial Communication via Semantically Enhanced Layout</i> Andrew B. Goldberg, Xiaojin Zhu, Charles R. Dyer, Mohamed Eldawy and Lijie Heng	119
<i>A Nearest-Neighbor Approach to the Automatic Analysis of Ancient Greek Morphology</i> John Lee	127

<i>Context-based Arabic Morphological Analysis for Machine Translation</i> Thuy Linh Nguyen and Stephan Vogel	135
<i>A Tree-to-String Phrase-based Model for Statistical Machine Translation</i> Thai Phuong Nguyen, Akira Shimazu, Tu Bao Ho, Minh Le Nguyen and Vinh Van Nguyen ...	143
<i>Trainable Speaker-Based Referring Expression Generation</i> Giuseppe Di Fabbrizio, Amanda Stent and Srinivas Bangalore	151
<i>The CoNLL 2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies</i> Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez and Joakim Nivre	159
<i>A Latent Variable Model of Synchronous Parsing for Syntactic and Semantic Dependencies</i> James Henderson, Paola Merlo, Gabriele Musillo and Ivan Titov	178
<i>Dependency-based Syntactic–Semantic Analysis with PropBank and NomBank</i> Richard Johansson and Pierre Nugues	183
<i>A Joint Model for Parsing Syntactic and Semantic Dependencies</i> Xavier Lluís and Lluís Màrquez	188
<i>Collective Semantic Role Labelling with Markov Logic</i> Sebastian Riedel and Ivan Meza-Ruiz	193
<i>Hybrid Learning of Dependency Structures from Heterogeneous Linguistic Resources</i> Yi Zhang, Rui Wang and Hans Uszkoreit	198
<i>Parsing Syntactic and Semantic Dependencies with Two Single-Stage Maximum Entropy Models</i> Hai Zhao and Chunyu Kit	203
<i>A Combined Memory-Based Semantic Role Labeler of English</i> Roser Morante, Walter Daelemans and Vincent Van Asch	208
<i>A Puristic Approach for Joint Dependency Parsing and Semantic Role Labeling</i> Alexander Volokh and Günter Neumann	213
<i>Discriminative Learning of Syntactic and Semantic Dependencies</i> Lu Li, Shixi Fan, Xuan Wang and Xiaolong Wang	218
<i>Discriminative vs. Generative Approaches in Semantic Role Labeling</i> Deniz Yuret, Mehmet Ali Yatbaz and Ahmet Engin Ural	223
<i>A Pipeline Approach for Syntactic and Semantic Dependency Parsing</i> Yotaro Watanabe, Masakazu Iwatate, Masayuki Asahara and Yuji Matsumoto	228
<i>Semantic Dependency Parsing using N-best Semantic Role Sequences and Roleset Information</i> Joo-Young Lee, Han-Cheol Cho and Hae-Chang Rim	233
<i>A Cascaded Syntactic and Semantic Dependency Parsing System</i> Wanxiang Che, Zhenghua Li, Yuxuan Hu, Yongqiang Li, Bing Qin, Ting Liu and Sheng Li ...	238
<i>The Integration of Dependency Relation Classification and Semantic Role Labeling Using Bilayer Maximum Entropy Markov Models</i> Weiwei Sun, Hongzhan Li and Zhifang Sui	243

<i>Mixing and Blending Syntactic and Semantic Dependencies</i>	
Yvonne Samuelsson, Oscar Täckström, Sumithra Velupillai, Johan Eklund, Mark Fishel and Markus Saers	248
<i>Dependency Tree-based SRL with Proper Pruning and Extensive Feature Engineering</i>	
Hongling Wang, Honglin Wang, Guodong Zhou and Qiaoming Zhu	253
<i>DeSRL: A Linear-Time Semantic Role Labeling System</i>	
Massimiliano Ciaramita, Giuseppe Attardi, Felice Dell’Orletta and Mihai Surdeanu	258
<i>Probabilistic Model for Syntactic and Semantic Dependency Parsing</i>	
Enhong Chen, Liu Shi and Dawei Hu	263
<i>Applying Sentence Simplification to the CoNLL-2008 Shared Task</i>	
David Vickrey and Daphne Koller	268

Conference Programme

Saturday, August 16, 2008

8:30–8:50 Opening Remarks

Session 1: Parsing

8:50–9:15 *Semantic Parsing for High-Precision Semantic Role Labelling*
Paola Merlo and Gabriele Musillo

9:15–9:40 *TAG, Dynamic Programming, and the Perceptron for Efficient, Feature-Rich Parsing*
Xavier Carreras, Michael Collins and Terry Koo

9:40–10:05 *A Fast Boosting-based Learner for Feature-Rich Tagging and Chunking*
Tomoya Iwakura and Seishi Okamoto

10:05–10:30 *Linguistic features in data-driven dependency parsing*
Lilja Øvrelid

10:30–11:00 Coffee break

Session 2: Semantics

11:00–11:25 *Transforming Meaning Representation Grammars to Improve Semantic Parsing*
Rohit Kate

11:25–11:50 *Using LDA to detect semantically incoherent documents*
Hemant Misra, Olivier Cappe and Francois Yvon

11:50–12:40 Invited talk by Regina Barzilay

12:40–14:00 Lunch

Saturday, August 16, 2008 (continued)

Shared Task

- 14:00–14:30 *The CoNLL 2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies*
Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez and Joakim Nivre

Oral Presentations

- 14:30–14:50 *A Latent Variable Model of Synchronous Parsing for Syntactic and Semantic Dependencies*
James Henderson, Paola Merlo, Gabriele Musillo and Ivan Titov
- 14:50–15:10 *Dependency-based Syntactic–Semantic Analysis with PropBank and NomBank*
Richard Johansson and Pierre Nugues
- 15:10–15:30 *A Joint Model for Parsing Syntactic and Semantic Dependencies*
Xavier Lluís and Lluís Màrquez
- 15:30–15:50 *Collective Semantic Role Labelling with Markov Logic*
Sebastian Riedel and Ivan Meza-Ruiz
- 15:50–16:10 *Hybrid Learning of Dependency Structures from Heterogeneous Linguistic Resources*
Yi Zhang, Rui Wang and Hans Uszkoreit
- 16:10–16:20 Closing remarks
- 16:20–16:45 Coffee break

Saturday, August 16, 2008 (continued)

Poster session 16:45–18:00

Parsing Syntactic and Semantic Dependencies with Two Single-Stage Maximum Entropy Models

Hai Zhao and Chunyu Kit

A Combined Memory-Based Semantic Role Labeler of English

Roser Morante, Walter Daelemans and Vincent Van Asch

A Puristic Approach for Joint Dependency Parsing and Semantic Role Labeling

Alexander Volokh and Günter Neumann

Discriminative Learning of Syntactic and Semantic Dependencies

Lu Li, Shixi Fan, Xuan Wang and Xiaolong Wang

Discriminative vs. Generative Approaches in Semantic Role Labeling

Deniz Yuret, Mehmet Ali Yatbaz and Ahmet Engin Ural

A Pipeline Approach for Syntactic and Semantic Dependency Parsing

Yotaro Watanabe, Masakazu Iwatate, Masayuki Asahara and Yuji Matsumoto

Semantic Dependency Parsing using N-best Semantic Role Sequences and Roleset Information

Joo-Young Lee, Han-Cheol Cho and Hae-Chang Rim

A Cascaded Syntactic and Semantic Dependency Parsing System

Wanxiang Che, Zhenghua Li, Yuxuan Hu, Yongqiang Li, Bing Qin, Ting Liu and Sheng Li

The Integration of Dependency Relation Classification and Semantic Role Labeling Using Bilayer Maximum Entropy Markov Models

Weiwei Sun, Hongzhan Li and Zhifang Sui

Mixing and Blending Syntactic and Semantic Dependencies

Yvonne Samuelsson, Oscar Täckström, Sumithra Velupillai, Johan Eklund, Mark Fishel and Markus Saers

Dependency Tree-based SRL with Proper Pruning and Extensive Feature Engineering

Hongling Wang, Honglin Wang, Guodong Zhou and Qiaoming Zhu

DeSRL: A Linear-Time Semantic Role Labeling System

Massimiliano Ciaramita, Giuseppe Attardi, Felice Dell'Orletta and Mihai Surdeanu

Saturday, August 16, 2008 (continued)

Probabilistic Model for Syntactic and Semantic Dependency Parsing

Enhong Chen, Liu Shi and Dawei Hu

Applying Sentence Simplification to the CoNLL-2008 Shared Task

David Vickrey and Daphne Koller

Sunday, August 17, 2008

Session 1: Language Acquisition I

8:50–9:15 *Picking them up and Figuring them out: Verb-Particle Constructions, Noise and Idiomatcity*

Carlos Ramisch, Aline Villavicencio, Leonardo Moura and Marco Idiart

9:15–9:40 *Fast Mapping in Word Learning: What Probabilities Tell Us*

Afra Alishahi, Afsaneh Fazly and Suzanne Stevenson

9:40–10:05 *Improving Word Segmentation by Simultaneously Learning Phonotactics*

Daniel Blanchard and Jeffrey Heinz

10:05–10:30 *A MDL-based Model of Gender Knowledge Acquisition*

Harmony Marchal, Benoît Lemaire, Maryse Bianco and Philippe Dessus

10:30–11:00 Coffee break

Session 2: Language Acquisition II

11:00–11:25 *Baby SRL: Modeling Early Language Acquisition.*

Michael Connor, Yael Gertner, Cynthia Fisher and Dan Roth

11:25–11:50 *An Incremental Bayesian Model for Learning Syntactic Categories*

Christopher Parisien, Afsaneh Fazly and Suzanne Stevenson

11:50–12:40 Invited talk by Nick Chater

12:40–14:00 Lunch

Sunday, August 17, 2008 (continued)

13:40–14:00 CoNLL Business Meeting

Session 3: Semantic extraction

14:00–14:25 *Fully Unsupervised Graph-Based Discovery of General-Specific Noun Relationships from Web Corpora Frequency Counts*

Gaël Dias, Raycho Mukelov and Guillaume Cleuziou

14:25–14:50 *Acquiring Knowledge from the Web to be used as Selectors for Noun Sense Disambiguation*

Hansen A. Schwartz and Fernando Gomez

14:50–15:15 *Automatic Chinese Catchword Extraction Based on Time Series Analysis*

Han Ren, Donghong Ji, Jing Wan and Lei Han

15:15–15:40 *Easy as ABC? Facilitating Pictorial Communication via Semantically Enhanced Layout*

Andrew B. Goldberg, Xiaojin Zhu, Charles R. Dyer, Mohamed Eldawy and Lijie Heng

15:40–16:00 Coffee break

Session 4: Morphology, MT and Generation

16:00–16:25 *A Nearest-Neighbor Approach to the Automatic Analysis of Ancient Greek Morphology*

John Lee

16:25–16:50 *Context-based Arabic Morphological Analysis for Machine Translation*

Thuy Linh Nguyen and Stephan Vogel

16:50–17:15 *A Tree-to-String Phrase-based Model for Statistical Machine Translation*

Thai Phuong Nguyen, Akira Shimazu, Tu Bao Ho, Minh Le Nguyen and Vinh Van Nguyen

17:15–17:40 *Trainable Speaker-Based Referring Expression Generation*

Giuseppe Di Fabbrizio, Amanda Stent and Srinivas Bangalore

17:40–18:00 Closing remarks and best paper award

