

HLT-NAACL-2006

**Computationally Hard  
Problems and  
Joint Inference in  
Speech and Language  
Processing**

**Proceedings of the Workshop**

9 June 2006

New York City, New York, USA

Production and Manufacturing by  
*Omnipress Inc.*  
2600 Anderson Street  
Madison, WI 53704

©2006 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](mailto:acl@aclweb.org)

## Introduction

We are pleased to present the proceedings of the Workshop on Computationally Hard Problems and Joint Inference in Speech and Language Processing, held at HLT/NAACL 2006 in New York City, New York.

Recent work on ranking, sampling and other approximate solutions to natural language processing problems indicate that researchers are coming back to the hard problems in speech and text, for which efficient algorithms are not known to exist. In addition, there has been increasing interest in moving away from systems that make chains of local decisions independently, and instead toward systems that make multiple decisions jointly using global information. The goal of this workshop is to bring together researchers working on NLP problems whose solutions are computationally hard—whether because the problem is not well modeled by only local features, or because the problem is best solved in a joint, rather than pipelined, manner.

We are grateful to the program committee for providing thoughtful and helpful reviews of the submitted papers. We also thank our invited speakers, Jeff Bilmes, Chris Manning, Dan Roth, and Giorgio Satta. Finally, we thank the organizers of the main HLT/NAACL 2006 conference, without which this workshop would not be possible.

Ryan McDonald, Charles Sutton, Hal Daumé III, Andrew McCallum, Jeff Bilmes, and Fernando Pereira  
organizers



**Invited Speakers:**

Jeff Bilmes, University of Washington  
Chris Manning, Stanford University  
Dan Roth, University of Illinois  
Giorgio Satta, University of Padua

**Organizers:**

Ryan McDonald, University of Pennsylvania  
Charles Sutton, University of Massachusetts  
Hal Daumé III, Information Sciences Institute, University of Southern California  
Andrew McCallum, University of Massachusetts  
Fernando Pereira, University of Pennsylvania  
Jeff Bilmes, University of Washington

**Program Committee:**

Razvan Bunescu, University of Texas  
Bill Byrne, University of Cambridge  
Xavier Carreras, Technical University of Catalonia  
Ozgur Cetin, University of California, Berkeley  
David Chiang, Information Sciences Institute, University of Southern California  
Michael Collins, Massachusetts Institute of Technology  
Jason Eisner, Johns Hopkins University  
Radu Florian, IBM TJ Watson Research Center  
Eric Fosler-Lussier, The Ohio State University  
Dan Gildea, University of Rochester  
Ralph Grishman, NYU  
Julia Hockenmaier, University of Pennsylvania  
Eric Horvitz, Microsoft  
Liang Huang, University of Pennsylvania  
Thorsten Joachims, Cornell University  
Katrin Kirchhoff, University of Washington  
Philipp Koehn, University of Edinburgh  
Shankar Kumar, Google  
Chris Manning, Stanford University  
Lluís Màrquez, Technical University of Catalonia  
Gideon Mann, University of Massachusetts  
Erik McDermott, NTT  
Ray Mooney, University of Texas  
Franz Och, Google

Kishore Papineni, IBM TJ Watson Research Center  
Chris Quirk, Microsoft  
Brian Roark, Oregon Graduate Institute  
Dan Roth, University of Illinois  
Salim Roukos, IBM TJ Watson Research Center  
Libin Shen, University of Pennsylvania  
Koichi Shinoda, Tokyo Institute of Technology  
Noah Smith, Johns Hopkins University  
Andreas Stolcke, SRI  
Ben Taskar, University of California, Berkeley

## Table of Contents

<i>A Syntax-Directed Translator with Extended Domain of Locality</i> Liang Huang, Kevin Knight and Aravind Joshi .....	1
<i>Efficient Dynamic Programming Search Algorithms for Phrase-Based SMT</i> Christoph Tillmann .....	9
<i>Computational Challenges in Parsing by Classification</i> Joseph Turian and I. Dan Melamed .....	17
<i>All-word Prediction as the Ultimate Confusable Disambiguation</i> Antal van den Bosch .....	25
<i>A Probabilistic Search for the Best Solution Among Partially Completed Candidates</i> Filip Ginter, Aleksandr Mylläri and Tapio Salakoski .....	33
<i>Practical Markov Logic Containing First-Order Quantifiers with Application to Identity Uncertainty</i> Aron Culotta and Andrew McCallum .....	41
<i>Re-Ranking Algorithms for Name Tagging</i> Heng Ji, Cynthia Rudin and Ralph Grishman .....	49





# Conference Program

## Friday, June 9, 2006

- 8:45–9:00      Opening Remarks
- 9:00–9:40      Invited Talk by Giorgio Satta
- 9:40–10:05     *A Syntax-Directed Translator with Extended Domain of Locality*  
Liang Huang, Kevin Knight and Aravind Joshi
- 10:05–10:30    *Efficient Dynamic Programming Search Algorithms for Phrase-Based SMT*  
Christoph Tillmann
- 10:30–11:00    Break
- 11:00–11:25    *Computational Challenges in Parsing by Classification*  
Joseph Turian and I. Dan Melamed
- 11:25–11:50    *All-word Prediction as the Ultimate Confusable Disambiguation*  
Antal van den Bosch
- 11:50–12:30    Invited Talk by Jeff Bilmes
- 12:30–14:00    Lunch
- 14:00–14:25    *A Probabilistic Search for the Best Solution Among Partially Completed Candidates*  
Filip Ginter, Aleksandr Mylläri and Tapio Salakoski
- 14:25–14:50    *Practical Markov Logic Containing First-Order Quantifiers with Application to Identity Uncertainty*  
Aron Culotta and Andrew McCallum
- 14:50–15:30    Invited Talk by Chris Manning
- 15:30–16:00    Break
- 16:00–16:25    *Re-Ranking Algorithms for Name Tagging*  
Heng Ji, Cynthia Rudin and Ralph Grishman

**Friday, June 9, 2006 (continued)**

16:25–17:05 Invited Talk by Dan Roth