

# SEMI-SUPERVISED LEARNING FOR NATURAL LANGUAGE PROCESSING

JUNE 4, 2009  
BOULDER, COLORADO (USA)

A WORKSHOP HELD IN CONJUNCTION WITH  
NAACL HUMAN LANGUAGE TECHNOLOGIES CONFERENCE

## Organizers:

Qin Wang (AT&T)  
Kevin Duh (U of Washington)  
Dekang Lin (Google Research)  
Contact: [ssl.nlp2009@gmail.com](mailto:ssl.nlp2009@gmail.com)

Keynote: Jason Eisner (Johns Hopkins)

## Program Committee:

Steven Abney (U of Michigan)  
Yasemin Altun (MPI-Tuebingen)  
Tim Baldwin (U of Melbourne)  
Shane Bergsma (U of Alberta)  
Antal van den Bosch (Tilburg Univ)  
John Blitzer (UC Berkeley)  
Ming-Wei Chang (UIUC)  
Walter Daelemans (U of Antwerp)  
Hal Daume III (U of Utah)  
Kevin Gimpel (Carnegie Mellon)  
Andrew Goldberg (U of Wisconsin)  
Liang Huang (Google Research)  
Rie Johnson (RJ Research Consulting)  
Katrín Kirchoff (U of Washington)  
Percy Liang (UC Berkeley)  
Gary Geunbae Lee (POSTECH)  
Gina-Anne Levow (U of Chicago)  
Gideon Mann (Google)  
David McClotsky (Brown Univ.)  
Ray Mooney (UT Austin)  
Hwee Tou Ng (Natl Univ of Singapore)  
Vincent Ng (UT Dallas)  
Miles Osborne (U of Edinburgh)  
Mari Ostendorf (U of Washington)  
Chris Pinchak (U of Alberta)  
Dragomir Radev (U of Michigan)  
Dan Roth (UIUC)  
Anoop Sarkar (Simon Fraser U.)  
Dale Schuurmans (U of Alberta)  
Akira Shimazu (JAIST)  
Jun Suzuki (NTT)  
Yee Whye Teh (Univ College London)  
Kristina Toutanova (Microsoft Research)  
Jason Weston (NEC)  
Tong Zhang (Rutgers Univ.)  
Ming Zhou (Microsoft Research Asia)  
Xiaojin Jerry Zhu (U of Wisconsin)

## MOTIVATION

Machine learning has become an indispensable tool for natural language processing (NLP) researchers. Unfortunately, given the limited availability of annotated data, progress on supervised learning often yields diminishing returns. Consequently, interest in semi-supervised learning has grown in the NLP community. Yet, while some promising results have been demonstrated to date, we suspect that consistently good results might not be easy to achieve across the board. In particular, many open questions remain:

1. Problem Structure: What are the different classes of NLP problem structures (e.g. sequences, trees, lists) and what algorithms are best suited for each class?
2. Background Knowledge: What kinds of NLP-specific background knowledge can we exploit to aid semi-supervised learning?
3. Scalability: NLP data are often large. What are the scalability solutions for applying existing semi-supervised learning algorithms to NLP data?
4. Evaluation and Negative Results: What can we learn from negative results? Can we make an educated guess as to when semi-supervised learning might outperform supervised or unsupervised learning?

## SUBMISSION

This workshop aims to bring together researchers dedicated to making semi-supervised learning work for NLP problems. Our goal is to help build a community of researchers and foster deep discussions about insights, speculations, and results (both positive and negative) that may otherwise not appear in a technical paper at a major conference.

We welcome submissions of 3 types: Technical papers (complete work, ~8 pages), Late-breaking papers (work-in-progress, ~4 pages), Position papers (perspectives/speculation, ~2 pages). For more information, see:

<http://sites.google.com/site/sslNlp/>

## IMPORTANT DATES

- March 6, 2009 – Submission Deadline  
March 30, 2009 – Notification of Acceptance  
April 12, 2009 – Camera-ready copies due  
June 4, 2009 – Workshop date