

NAACL HLT 2009

**Unsupervised and Minimally Supervised
Learning of Lexical Semantics**

Proceedings of the Workshop

June 5, 2009
Boulder, Colorado

Production and Manufacturing by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53707
USA

©2009 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

ISBN 978-1-932432-34-3

Introduction

Lexical semantics (the semantics of words) has become an important part of Natural Language Processing due to its practical application in a number of areas such as machine translation, web & enterprise search, ontology learning etc.

This fact can be observed by looking at the increasing interest in the field of learning of lexical semantics e.g. the last Semantic Evaluation Workshop (SemEval-2007) consisted of 18 tasks ranging from the traditional Word Sense Disambiguation (WSD) task to the most recent of Word Sense Induction (WSI), web people search, metonymy resolution and others.

Given the wide variety of applications exploiting lexical semantics it is significant to focus on methods and techniques, which can overcome the "Knowledge Acquisition Bottleneck" and deal with the cost-prohibitive, error-prone and labor-intensive processes of creating hand-tagged training data.

The emphasis of this workshop is on unsupervised and minimally supervised methods relevant to learning of lexical semantics. The goal of this workshop is to provide a venue for researchers to obtain a better understanding on the issues and challenges that need to be tackled in order to overcome a number of significant problems within lexical semantics, such as data sparsity, unsupervised and minimally supervised parameter estimation, efficient and effective use of the web, and applications of distributional similarity.

We are very happy that the workshop has accepted a set of seven high quality papers tackling the above problems, and hope that their contribution will have an impact on the field. We are grateful to the program committee for their effort to thoroughly review the submissions. We would also like to thank Martha Palmer for presenting her noteworthy work in the workshop.

Suresh Manandhar & Ioannis P. Klapaftis
Co-chairs

Organizers

Suresh Manandhar, University of York, UK
Ioannis P. Klapaftis, University of York, UK

Program Committee

Eneko Agirre, University of the Basque Country, Spain
Enrique Alfonseca, Google, Switzerland
Chris Biemann, Powerset, USA
Philipp Cimiano, TU Delft, The Netherlands
James Cussens, University of York, UK
Cristian Danescu Niculescu-Mizil, Cornell University, USA
Aria Haghighi, University of California, Berkeley, USA
Nancy Ide, Vassar College, USA
Kyo Kaguera, University of Tokyo, Japan
Ioannis Klapaftis, University of York, UK
Lillian Lee, Cornell University, USA
Suresh Manandhar, University of York, UK
Ted Pedersen, University of Minnesota, USA
German Rigua, University of the Basque Country, Spain
Carlo Strapparava, ITC-Irst, Italy
Fangzhong Su, University of Leeds, UK
David Weir, University of Sussex, UK

Invited speakers

Martha Palmer, University of Colorado at Boulder, USA

Invited Talk

Knowing a Word(sense) by its company

Martha Palmer
University of Colorado at Boulder, USA

Abstract Supervised word sense disambiguation requires training corpora that have been tagged with word senses, and these word senses typically come from a pre-existing sense inventory. Space limitations imposed by dictionary publishers have biased the field towards lists of discrete senses for an individual lexeme. This approach does not capture information about relatedness of individual senses. How important is this information to knowing which sense distinctions are critical for particular types of NLP applications? How much does sense relatedness affect automatic word sense disambiguation performance? Recent psycholinguistic evidence seems to indicate that closely related word senses may be represented in the mental lexicon much like a single sense, whereas distantly related senses may be represented more like discrete entities. These results suggest that, for the purposes of WSD, closely related word senses can be clustered together into a more general sense with little meaning loss. This talk will describe the relatedness of verb senses and its impact on NLP applications and WSD components as well as recent psycholinguistic research results.

Table of Contents

<i>Acquiring Applicable Common Sense Knowledge from the Web</i> Hansen A. Schwartz and Fernando Gomez	1
<i>Utilizing Contextually Relevant Terms in Bilingual Lexicon Extraction</i> Azniyah Ismail and Suresh Manandhar	10
<i>Corpus-based Semantic Lexicon Induction with Web-based Corroboration</i> Sean Igo and Ellen Riloff	18
<i>Cross-lingual Predicate Cluster Acquisition to Improve Bilingual Event Extraction by Inductive Learning</i> Heng Ji	27
<i>Graph Connectivity Measures for Unsupervised Parameter Tuning of Graph-Based Sense Induction Systems.</i> Ioannis Korkontzelos, Ioannis Klapaftis and Suresh Manandhar	36
<i>Combining Syntactic Co-occurrences and Nearest Neighbours in Distributional Methods to Remedy Data Sparseness.</i> Lonneke van der Plas	45
<i>Using DEDICOM for Completely Unsupervised Part-of-Speech Tagging</i> Peter Chew, Brett Bader and Alla Rozovskaya	54

Conference Program

Friday, June 5, 2009

- 9:15–9:30 Opening remarks
- 9:30–10:00 *Acquiring Applicable Common Sense Knowledge from the Web*
Hansen A. Schwartz and Fernando Gomez
- 10:00–10:30 *Utilizing Contextually Relevant Terms in Bilingual Lexicon Extraction*
Azniah Ismail and Suresh Manandhar
- 10:30–11:00 Morning break
- 11:00–12:00 Invited talk: Martha Palmer, Knowing a Word(sense) by its company
- 12:00–12:30 *Corpus-based Semantic Lexicon Induction with Web-based Corroboration*
Sean Igo and Ellen Riloff
- 12:30–14:00 Lunch break
- 14:00–14:30 *Cross-lingual Predicate Cluster Acquisition to Improve Bilingual Event Extraction
by Inductive Learning*
Heng Ji
- 14:30–15:00 *Graph Connectivity Measures for Unsupervised Parameter Tuning of Graph-Based
Sense Induction Systems.*
Ioannis Korkontzelos, Ioannis Klapaftis and Suresh Manandhar
- 15:00–15:30 *Combining Syntactic Co-occurrences and Nearest Neighbours in Distributional
Methods to Remedy Data Sparseness.*
Lonneke van der Plas
- 15:30–16:00 Afternoon break
- 16:00–16:30 *Using DEDICOM for Completely Unsupervised Part-of-Speech Tagging*
Peter Chew, Brett Bader and Alla Rozovskaya
- 16:30–17:00 Closing remarks & discussion

