ACL-SIGLEX 2005 Workshop on Deep Lexical Acquisition

Proceedings of the Workshop

Workshop Chairs:
Timothy Baldwin
Anna Korhonen
Aline Villavicencio

30 June 2005
University of Michigan
Ann Arbor, Michigan, USA

Introduction

This volume contains the papers accepted for presentation at the ACL-SIGLEX 2005 Workshop on Deep Lexical Acquisition, held at the University of Michigan, Ann Arbor, USA, on the 30th of June, 2005.

This workshop is supported by SIGLEX, the Special Interest Group on the Lexicon of the Association for Computational Linguistics (http://www.clres.com/siglex.html). Its goal is to bring together researchers interested in different facets of the automatic acquisition of deep lexical information, e.g. in the areas of computational grammars, computational lexicography, machine translation, information retrieval, question-answering, and text mining.

Deep lexical resources include lexicons for linguistically-precise grammars, template sets for information extraction systems, and ontologies for word sense disambiguation. Such resources are critical for enhancing the performance of systems and for improving their portability between domains. Most deep lexical resources in current use have been developed manually by lexicographers at considerable cost, and yet have limited coverage and require labour-intensive porting to new tasks. Automatic lexical acquisition is a more promising and cost-effective approach to take, and is increasingly viable given recent advances in NLP and machine learning technology, and corpus availability. However, a number of important challenges still need addressing before benefits can be reaped in practical language engineering, such as the (multilingual) acquisition of deep lexical information from corpora and the implementation of accurate, large-scale, portable acquisition techniques.

In the call for papers we solicited papers describing aspects of deep lexical acquisition including:

- Automatic acquisition of deep lexical information: subcategorization, diathesis alternations, selectional preferences, lexical/semantic classes, qualia structure, lexical ontologies, semantic roles, word senses, etc.

- Methods for supervised, unsupervised and weakly supervised deep lexical acquisition: machine learning, statistical, example- or rule-based, hybrid etc.

- Large-scale, cross-domain, domain-specific and portable deep lexical acquisition

- Extending and refining existing lexical resources with automatically acquired information

- Evaluation of deep lexical acquisition

- Application of deep lexical acquisition to NLP applications (e.g. machine translation, information extraction, language generation, question-answering)

- Multilingual deep lexical acquisition

Of the 22 papers submitted, the programme committee selected 11 papers for publication, representative of the state of the art in this subject today. Each full-length submission was independently reviewed

by three members of the program committee, who then collectively faced the difficult task of selecting a subset of papers for publication from a very strong field. The accepted papers include proposals for automatic annotation and extension of deep lexical resources, and methods for automatically acquiring deep lexical information. Languages targeted in the papers include English, Chinese, Japanese and Catalan.

We would like to thank all the authors who submitted papers, as well as the members of the program committee for the time and effort they contributed in reviewing the papers, and Chris Brew for complementing the workshop expertly with his invited talk. Our thanks go also to the organisers of the main conference, the publication chairs (Jason Eisner and Philipp Köhn) and the conference workshop committee (Mark Dras, Mary Harper, Dan Klein, Mirella Lapata and Shuly Wintner).

Timothy Baldwin, Anna Korhonen, Aline Villavicencio

Organizers:

Timothy Baldwin, University of Melbourne, Australia
Anna Korhonen, University of Cambridge, UK
Aline Villavicencio, University of Essex, UK

Program Committee:

Collin Baker, University of California Berkeley (USA)
Roberto Basili, University of Rome Tor Vergata (Italy)
Francis Bond, NTT (Japan)
Chris Brew, Ohio State University (USA)
Ted Briscoe, University of Cambridge (UK)
John Carroll, University of Sussex (UK)
Stephen Clark, University of Oxford (UK)
Sonja Eisenbeiss, University of Essex (UK)
Christiane Fellbaum, University of Princeton (USA)
Frederick Fouvry, University of Saarland (Germany)
Sadao Kurohashi, University of Tokyo (Japan)
Diana McCarthy, University of Sussex (UK)
Rada Mihalcea, University of North Texas (USA)
Tom O'Hara, University of Maryland, Baltimore County (USA)
Martha Palmer, University of Pennsylvania (USA)
Massimo Poesio, University of Essex (UK)
Philip Resnik, University of Maryland (USA)
Patrick Saint-Dizier, IRIT-CNRS (France)
Sabine Schulte im Walde, University of Saarland (Germany)
Mark Steedman, University of Edinburgh (UK)
Mark Stevenson, University of Sheffield (UK)
Suzanne Stevenson, University of Toronto (Canada)
Dominic Widdows, MAYA Design, Inc. (USA)
Yorick Wilks, University of Sheffield (UK)
Dekai Wu, Hong Kong University of Science and Technology (Hong Kong)

Invited Speaker:

Cris Brew, Linguistics, Computer Science and Engineering and Cognitive Science Departments,
Ohio State University

Table of Contents

Conference Program

08:55-09:00   Opening remarks

09:00-09:30   *Data Homogeneity and Semantic Role Tagging in Chinese*
Oi Yee Kwong and Benjamin K. Tsou

09:30-10:00   *Verb Subcategorization Kernels for Automatic Semantic Labeling*
Alessandro Moschitti and Roberto Basili

10:00-10:30   *Identifying Concept Attributes Using a Classifier*
Massimo Poesio and Abdulrahman Almuhareb

10:30-11:00   Coffee Break

11:00-11:30   *Automatically Learning Qualia Structures from the Web*
Philipp Cimiano and Johanna Wenderoth

11:30-12:00   *Automatically Distinguishing Literal and Figurative usages of Highly Polysemous Verbs*
Afsaneh Fazly, Ryan North and Suzanne Stevenson

12:00-12:30   *Automatic Extraction of Idioms using Graph Analysis and Asymmetric Lexicosyntactic Patterns*
Dominic Widdows and Beate Dorow

12:30-14:00   Lunch

14:00-14:30   *Frame Semantic Enhancement of Lexical-Semantic Resources*
Rebecca Green and Bonnie J. Dorr

14:30-15:30   *It might be Deep Enough, but is it Broad Enough? Diversity in the Lexicon*
Invited Speaker - Chris Brew, Ohio State University

15:30-16:00   Coffee Break

16:00-16:30   *Bootstrapping Deep Lexical Resources: Resources for Courses*
Timothy Baldwin

16:30-17:00   *Morphology vs. Syntax in Adjective Class Acquisition*
Gemma Boleda, Toni Badia and Sabine Schulte im Walde

17:00-17:30   *Automatic Acquisition of Bilingual Rules for Extraction of Bilingual Word Pairs from Parallel Corpora*
Hiroshi Echizen-ya, Kenji Araki and Yoshio Momouchi

17:30-18:00   *Approximate Searching for Distributional Similarity*
James Gorman and James Curran

18:00-18:05   Closing remarks