Acquisition of Lexical Knowledge from Text

Proceedings of a Workshop
Sponsored by the
Special Interest Group on the Lexicon
of the
Association for Computational Linguistics

Edited by
Branimir Boguraev
and
James Pustejovsky

21 June 1993
Ohio State University
Columbus, Ohio, USA
Acquisition of Lexical Knowledge from Text

Proceedings of a Workshop
Sponsored by the
Special Interest Group on the Lexicon of the
Association for Computational Linguistics

Edited by
Branimir Boguraev
and
James Pustejovsky

21 June 1993
Ohio State University
Columbus, Ohio, USA

Published by the Association for Computational Linguistics
ACQUISITION OF LEXICAL KNOWLEDGE FROM TEXT

21, June 1993
Ohio State University
Columbus, Ohio, USA

A workshop sponsored by the
Special Interest Group on the Lexicon (SIGLEX)
of the
Association for Computational Linguistics

FINAL PROGRAM

Session 1: (chair: Branimir Boguraev)
8:45-9:15 Word Sense Disambiguation by Human Subjects: Computational and Psycholinguistic Applications
Thomas Ahlswede and David Lorand
9:15-9:45 Towards building Contextual Representations of Word Senses Using Statistical Models
Claudia Leacock, Geoffrey Towell, and Ellen Voorhees
9:45-10:15 Lexical Concept Acquisition from Collocation Map
Youn S. Han, Young Kyoong Han, and Key-Sun Choi

10:15-10:45 Coffee Break

Session 2 (chair: Gregory Grefenstette)
10:45-11:15 Internal and External Evidence in the Identification and Semantic Categorization of Proper Names
David McDonald
11:15-11:45 Identifying Unknown Proper Names in Newsire Text
Inderjeet Mani, T. Richard Macmillan, Susann Luperfoy, Elaine Lusher, and Sharon Laskowski
11:45-12:15 Customizing a Lexicon to Better Suit a Computational Task
Marti Hearst and Hinrich Schuetze

11:45-1:45 Lunch

Session 3 (chair: James Pustejovsky)
1:45-2:15 Hierarchical Clustering of Verbs
Roberto Basili, Maria Pazienza, and Paola Velardi
2:15-2:45 Detecting Dependencies between Semantic Verb Subclasses and Subcategorization Frames in Text Corpora
Victor Poznanski and Antonio Sanfilippo
2:45-3:15 The Automatic Acquisition of Frequencies of Verb Subcategorization Frames from Tagged Corpora
Akira Ushioda, David Evans, Ted Gibson, Alex Waibel
3:15-3:45 Acquiring Predicate-Argument Mapping Information from Multilingual Texts
Chinatsu Aone and Douglas McKee
3:45 - 4:15 Coffee Break

**Session 4** (chair: Susan Armstrong-Warwick)

4:15 - 4:45 *Experiments in Syntactic and Semantic Classification and Disambiguation using Bootstrapping*
Robert Futrelle and Susan Gauch

4:45 - 5:15 *Structural Methods for Lexical/Semantic Patterns*
Scott Waterman

5:15 - 5:45 *Evaluation Techniques for Automatic Semantic Extraction:*
Gregory Grefenstette

Alternate Papers:

*Categorization and Standardizing Proper Nouns for Efficient Information Retrieval*
Woojin Paik, Elizabeth Liddy, Edmund Yu, and Mary McKenna

*The Long Journey from the Core to the Real Size of Large LDBs*
Elena Paskaleva, Kiril Simov, Mariana Damova, and Milena Slavcheva

**PROGRAM CHAIR:**
Bran Boguraev
Apple Computer, Advanced Technologies Group
One Infinite Loop, MS 301-3C
Cupertino, CA 95014
email: bkb@apple.com

**PROGRAM COMMITTEE:**
Susan Armstrong-Warwick, ISSCO, Geneva
Branimir Boguraev, Apple Computer
Ann Copestake, University of Cambridge
Gregory Grefenstette, University of Pittsburgh
James Pustejovsky, Brandeis University
Mats Rooth, University of Stuttgart
David Waltz, Thinking Machines Corporation
Uri Zernik, GE-CRD
## TABLE OF CONTENTS

**Word Sense Disambiguation by Human Subjects: Computational and Psycholinguistic Applications**  
Thomas Ahlswede and David Lorand .................................................... 1

**Towards Building Contextual Representations of Word Senses Using Statistical Models**  
Claudia Leacock, Geoffrey Towell, and Ellen Voorhees .......................... 10

**Lexical Concept Acquisition from Collocation Map**  
Youn S. Han, Young Kyoon Han, and Key-Sun Choi .............................. 22

**Internal and External Evidence in the Identification and Semantic Categorization of Proper Names**  
David McDonald ........................................................................... 32

**Identifying Unknown Proper Names in Newswire Text**  
Inderjeet Mani, T. Richard Macmillan, Susann Luperføy, Elaine Lusher, and Sharon Laskowski .......................................................... 44

**Customizing a Lexicon to Better Suit a Computational Task**  
Marti Hearst and Hinrich Schuetze .................................................. 55

**Hierarchical Clustering of Verbs**  
Roberto Basili, Maria Pazienza, and Paola Velardi ............................... 70

**Detecting Dependencies between Semantic Verb Subclasses and Subcategorization Frames in Text Corpora**  
Victor Poznanski and Antonio Sanfilippo .......................................... 82

**The Automatic Acquisition of Frequencies of Verb Subcategorization Frames from Tagged Corpora**  
Akira Ushioda, David Evans, Ted Gibson, Alex Waibel .......................... 95

**Acquiring Predicate-Argument Mapping Information from Multilingual Texts**  
Chinatsu Aone and Douglas McKee .................................................. 107

**Experiments in Syntactic and Semantic Classification and Disambiguation using Bootstrapping**  
Robert Futrelle and Susan Gauch ..................................................... 117

**Structural Methods for Lexical/Semantic Patterns**  
Scott Waterman ............................................................................. 128

**Evaluation Techniques for Automatic Semantic Extraction: Comparing Syntactic and Window Based Approaches**  
Gregory Grefenstette ................................................................... 143

**Categorization and Standardizing Proper Nouns for Efficient Information Retrieval**  
Woowjin Paik, Elizabeth Liddy, Edmund Yu, and Mary McKeena ............ 154

**The Long Journey from the Core to the Real Size of Large LDBs**  
Elena Paskaleva, Kiril Simov, Mariana Damova, and Milena Slavcheva ..... 161