Production and Manufacturing by
Omnipress, Inc.
Post Office Box 7214
Madison, WI 53707-7314

The ACL-2001 conference, its workshops and tutorials are sponsored by:

Centre National de la Recherche (CNRS)
Institut de Recherches en Informatique de Toulouse (IRIT)
Université des Sciences Sociales
Université Paul Sabatier (UPS)
Association pour le Traitement Automatique des Langues (ATALA)
ELSNET
Ministère de l'éducation nationale et de la recherche (MRT)
Centre de Lingüistica y Computacio (CliC)
Ville de Toulouse
Conseil Régional de Midi-Pyrénées

Air France
Electricité de France (EDF)
France Telecom
IBM
Microsoft
Synapse développement
Xerox Research Center

©2001 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
75 Paterson Street, Suite 4
New Brunswick, NJ 08901
USA
Tel: +1-732-342-9100
Fax: +1-732-342-9339
acl@aclweb.org
http://www.aclweb.org

Preface

On behalf of everyone who was involved in SENSEVAL-2, we are happy to provide this volume of papers by the teams who participated in SENSEVAL-2, the second organized evaluation of word sense disambiguation systems. This volume constitutes the proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems that was held on July 5-6, 2001 in Toulouse (in conjunction with the 39th Annual Meeting of the ACL).

Everyone who participated in SENSEVAL-2 by either submitting a system or organizing a task was invited to submit a 4-page paper to this proceedings. Every paper would be accepted, but each paper would be reviewed by at least two reviewers. If one submitted a paper, then they agreed to be a reviewer, thus the reviewing committee is the set of authors.

Many people and organisations contributed to the success of SENSEVAL-2 and the workshop. Foremost, of course, are all the teams who submitted systems to be evaluated and the other teams (sometimes the same ones) who designed and organized tasks for those systems to compete in. These researchers form the core of the SENSEVAL community, and their papers are included within this proceedings.

We thank the panel organisers and, of course, the many panelists, including Dan Flickinger (YY & CSLI), Anna Samiotou (ESTeam AB), Steve Finch (Infogistics), and Patrick Hanks (Lingomotors). Diana McCarthy and Michael Oakes coordinated and put together the page of system descriptions. Patrick Saint-Dizier and his crew provided the local arrangements, including a most memorable lunch.

A special thanks goes to Randee Tengi for providing an advance-release of WordNet 1.7 specially for SENSEVAL.

We'd also like to thank the many sponsors of this essentially volunteer effort. The workshop, an ACL-SIGLEX event, was supported by EURALEX, ELSNET, EPSRC (grant GR/RO2337/01 - MATS), ELRA, and Pearson Education. Sharp Laboratories of Europe provided the main SENSEVAL-2 website, and the University of Pennsylvania the download site. The English tasks were sponsored by EPSRC (grant GR/RO2337/01 – MATS), the National Science Foundation (NSF-9800658 – VerbNet, NSF-9910603 – ISLE), DARPA (TIDES – PennTools), and the University of Pennsylvania. The Swedish organisers would like to thank The Swedish Council for Research in the Humanities and Social Sciences (HSFR) for providing financial support for the coordination of the task. The Estonian project (task) was sponsored in part by the Estonian Science Foundation (Grant No. 4467).

Scott Cotton, Phil Edmonds, Adam Kilgarriff, Martha Palmer
SENSEVAL-2 Committee

Judita Preiss, David Yarowsky
Proceedings Editors

October 2001

Table of Contents

PREFACE iii

TABLE OF CONTENTS v

AUTHOR INDEX viii

SENSEVAL-2: Overview
Philip Edmonds and Scott Cotton 1

TASK DESCRIPTIONS AND PANEL REPORTS 7

The Basque Task: Did Systems Perform in the Upperbound?
Eneko Agirre, Elena Garcia, Mikel Lersundi, David Martinez and Eli Pociello 9

Dutch Word Sense Disambiguation: Data and Preliminary Results
Iris Hendrickx and Antal van den Bosch 13

English Lexical Sample Task Desciption
Adam Kilgarriff 17

English Tasks: All-Words and Verb Lexical Sample
Martha Palmer, Christiane Fellbaum, Scott Cotton, Lauren Delfs
and Hoa Trang Dang 21

Sensiting Inflectionality: Estonian Task for SENSEVAL-2
Neeme Kahusk, Heili Orav and Haldur Õim 25

The Italian Lexical Sample Task
Francesca Bertagna, Claudia Soria and Nicoletta Calzolari 29

SENSEVAL-2 Japanese Dictionary Task
Kiyoaki Shirai 33

SENSEVAL-2 Japanese Translation Task
Sadao Kurohashi 37

Framework and Results for the Spanish SENSEVAL
German Rigau, Mariona Taulé, Ana Fernandez and Julio Gonzalo 41

SENSEVAL-2: The Swedish Framework
Dimitrios Kokkinakis, Jerker Järborg and Yvonne Cederholm 45

The SENSEVAL-2 Panel on Domains, Topics and Senses
Paul Buitelaar 49

SYSTEM DESCRIPTIONS 53

The Japanese Translation Task: Lexical and Structural Perspectives
Timothy Baldwin, Atsushi Okazaki, Takenobu Tokunaga and Hozumi Tanaka 55

Supervised Sense Tagging using Support Vector Machines
Clara Cabezas, Philip Resnik and Jessica Stevens 59

Probabilistic Network Models for Word Sense Disambiguation
Gerald Chao and Michael Dyer 63

Improving WSD with Multi-Level View of Context Monitored by Similarity Measure
Eric Crestan, Marc El-Bèze and Claude de Loupy 67

Using Lazy Boosting for Word Sense Disambiguation
Gerard Escudero, Lluís Màrquez and German Rigau 71

The UNED Systems at SENSEVAL-2
David Fernández-Amorós, Julio Gonzalo and Felisa Verdejo 75

Semantic Tagging Using WordNet Examples
Sherwood Haynes 79

Classifier Optimization and Combination in the English All Words Task
Véronique Hoste, Anne Kool and Walter Daelemans 83

Combining Heterogeneous Classifiers for Word-Sense Disambiguation
H. Tolga Ilhan, Sepandar Kamvar, Dan Klein, Christopher Manning
and Kristina Toutanova 87

The Språkdata-ML System as Used for SENSEVAL-2
Dimitrios Kokkinakis 91

ATR-SLT System for SENSEVAL-2 Japanese Translation Task
Tadashi Kumano, Hideki Kashioka and Hideki Tanaka 95

Sense and Deduction: The Power of Peewees Applied to the SENSEVAL-2 Swedish Lexical Sample Task
Torbjörn Lager and Natalia Zinovjeva 99

Primitive-Based Word Sense Disambiguation for SENSEVAL2
Lim Beng Tat, Zaharin Yusoff, Tang Enya Kong and Guo Cheng Ming 103

Use of Machine Readable Dictionaries for Word-Sense Disambiguation in SENSEVAL-2
Kenneth Litkowski 107

Using Domain Information for Word Sense Disambiguation
Bernardo Magnini, Carlo Strapparava, Giovanni Pezzulo and Alfio Gliozzo 111

Decision Lists for English and Basque
David Martinez and Eneko Agirre 115

Disambiguating Noun and Verb Senses Using Automatically Acquired Selectional Preferences
Diana McCarthy, John Carroll and Judita Preiss 119

Combination of Contextual Features for Word Sense Disambiguation: LIU-WSD
Magnus Merkel and Mikael Andersson 123

Pattern Learning and Active Feature Selection for Word Sense Disambiguation
Rada Mihalcea and Dan Moldovan 127

The University of Alicante Word Sense Disambiguation System
Andrés Montoyo and Armando Suárez 131

Japanese Word Sense Disambiguation using the Simple Bayes and Support Vector Machine Methods
Masaki Murata, Masao Utiyama, Kiyotaka Uchimoto, Qing Ma and Hitoshi Isahara ... 135

Machine Learning with Lexical Features: The Duluth Approach to SENSEVAL-2
Ted Pedersen 139

Anaphora Resolution with Word Sense Disambiguation
Judita Preiss 143

KUNLP system using Classification Information Model at SENSEVAL-2
Hee-Cheol Seo, Sang-Zoo Lee, Hae-Chang Rim and Ho Lee 147

WASP-Bench: a Lexicographic Tool Supporting Word Sense Disambiguation
David Tugwell and Adam Kilgarriff 151

Word Translation Based on Machine Learning Models Using Translation Memory and Corpora
Kiyotaka Uchimoto, Satoshi Sekine, Masaki Murata and Hitoshi Isahara 155

Automatic WSD: Does it Make Sense of Estonian?
Kadri Vider and Kaarel Kaljurand 159

The Johns Hopkins SENSEVAL-2 System Descriptions
David Yarowsky, Silviu Cucerzan, Radu Florian, Charles Schafer and Richard Wicentowski 163

Author Index

Agirre, Eneko 9, 115
Andersson, Mikael 123
Baldwin, Timothy 55
Bertagna, Francesca 29
Buitelaar, Paul 49
Cabezas, Clara 59
Calzolari, Nicoletta 29
Carroll, John 119
Cederholm, Yvonne 45
Chao, Gerald 63
Cotton, Scott 1, 21
Crestan, Eric 67
Cucerzan, Silviu 163
Daelemans, Walter 83
Dang, Hoa Trang 21
De Loupy, Claude 67
Delfs, Lauren 21
Dyer, Michael 63
Edmonds, Philip 1
El-Bèze, Marc 67
Escudero, Gerard 71
Fellbaum, Christiane 21
Fernandez, Ana 41
Fernández-Amorós, David 75
Florian, Radu 163
Garcia, Elena 9
Gliozzo, Alfio 111
Gonzalo, Julio 41, 75
Guo, Cheng Ming 103
Haynes, Sherwood 79
Hendrickx, Iris 13
Hoste, Véronique 83
Ilhan, H. Tolga 87
Isahara, Hitoshi 135, 155
Järborg, Jerker 45
Kahusk, Neeme 25
Kaljurand, Kaarel 159
Kamvar, Sepandar 87
Kashioka, Hideki 95
Kilgarriff, Adam 17, 151
Klein, Dan 87
Kokkinakis, Dimitrios 45, 91
Kool, Anne 83
Kumano, Tadashi 95
Kurohashi, Sadao 37
Lager, Torbjörn 99
Lee, Ho 147
Lee, Sang-Zoo 147
Lersundi, Mikel 9
Lim, Beng Tat 103
Litkowski, Kenneth 107
Ma, Qing 135
Magnini, Bernardo 111
Manning, Christopher 87
Màrquez, Lluís 71
Martinez, David 9, 115
McCarthy, Diana 119
Merkel, Magnus 123
Mihalcea, Rada 127
Moldovan, Dan 127
Montoyo, Andrés 131
Murata, Masaki 135, 155
Õim, Haldur 25
Okazaki, Atsushi 55
Orav, Heili 25
Palmer, Martha 21
Pedersen, Ted 139
Pezzulo, Giovanni 111
Pociello, Eli 9
Preiss, Judita 119, 143
Resnik, Philip 59
Rigau, German 41, 71
Rim, Hae-Chang 147
Schafer, Charles 163
Sekine, Satoshi 155
Seo, Hee-Cheol 147
Shirai, Kiyoaki 33
Soria, Claudia 29
Stevens, Jessica 59
Strapparava, Carlo 111
Suárez, Armando 131
Tanaka, Hideki 95
Tanaka, Hozumi 55
Tang, Enya Kong 103
Taulé, Mariona 41
Tokunaga, Takenobu 55
Toutanova, Kristina 87
Tugwell, David 151
Uchimoto, Kiyotaka 135, 155
Utiyama, Masao 135
Van den Bosch, Antal 13
Verdejo, Felisa 75
Vider, Kadri 159
Wicentowski, Richard 163
Yarowsky, David 163
Yusoff, Zaharin 103
Zinovjeva, Natalia 99