



**Proceedings of the
ACL-2000 Workshop on
Word Senses and Multi-linguality**

**Held in conjunction with
The 38th Annual Meeting of the
Association for Computational Linguistics**

**Sponsored by
The ACL Special Interest Group for the Lexicon (SIGLEX)**

**Organized by
Nancy Ide
Charles Fillmore
Philip Resnik
David Yarowsky**

**7 October 2000
Hong Kong University of Science and Technology (HKUST)
Hong Kong**

**Proceedings of the
ACL-2000 Workshop on
Word Senses and Multi-linguality**

**Held in conjunction with
The 38th Annual Meeting of the
Association for Computational Linguistics**

**Sponsored by
The ACL Special Interest Group for the Lexicon (SIGLEX)**

**Organized by
Nancy Ide
Charles Fillmore
Philip Resnik
David Yarowsky**

**7 October 2000
Hong Kong University of Science and Technology (HKUST)
Hong Kong**

©2000 The Association for Computational Linguistics

Order copies of this and other ACL workshop proceedings from:

Association for Computational Linguistics (ACL)
75 Paterson Street, Suite 9
New Brunswick, NJ 08901
USA
Tel: +1-732-342-9100
Fax: +1-732-342-9339
acl@aclweb.org

PREFACE

With an increasingly global economy and the explosive growth of the "World" in "World Wide Web", the computational linguistics community is faced as never before with the challenges and opportunities of multi-linguality. At the same time, the community has returned with renewed enthusiasm to problems of word meaning, especially the delineation and discrimination of word senses. An intimate relationship between the two issues is becoming apparent – for example, in the consideration of translation equivalence in parallel corpora, the construction of multilingual ontologies, and the examination of senses in relation to specific natural language applications such as machine translation, information retrieval, summarization, etc. The issue of multi-lingual approaches to sense distinctions was also a central topic of discussion at the first SENSEVAL conference in 1998, and is one of the areas to be covered at SENSEVAL-2 (to be held in Spring 2001).

This workshop addresses problems of word sense disambiguation and delineation of appropriate sense distinctions, with specific emphasis on approaches that involve more than one language and the ways in which observations about cross-linguistic equivalence affect our consideration of sense divisions in the individual languages. More generally, we seek to foster discussion and exchanges of insight in any area of computational linguistics where a non-monolingual approach to word sense issues is being taken.

Many people are owed thanks for their contributions to setting up this workshop. We are especially grateful to David Yarowsky and the staff of the Human Language Technology Center at Hong Kong University of Science and Technology for their work in producing these proceedings under a very tight schedule. The SIGLEX'00 Program Committee enabled us to work within a very brief time frame, by quickly turning around reviews for the substantial number of submissions to the conference. Finally, the Department of Computer Science at Vassar College provided administrative and organizational support. All of them are responsible for the success of SIGLEX'00.

Nancy Ide, SIGLEX'00 Chair
Poughkeepsie, New York
September, 2000

SPONSORS:

The Association for Computational Linguistics (ACL)
SIGLEX (ACL's SIG for the Lexicon)

ORGANIZERS:

Nancy Ide, Chair
Charles Fillmore
Philip Resnik
David Yarowsky

PROGRAM COMMITTEE:

Helge Dyvik	University of Bergen
Nancy Ide	Vassar College
Christiane Fellbaum	Princeton University
Charles Fillmore	UC Berkeley and ICSI
Adam Kilgarriff	ITRI, University of Brighton
Martha Palmer	University of Pennsylvania
Philip Resnik	University of Maryland
Evelyne Viegas	Microsoft Corporation
David Yarowsky	Johns Hopkins University

FURTHER INFORMATION:

Nancy Ide
Department of Computer Science
Vassar College
124 Raymond Avenue
Poughkeepsie, New York 12604-0520 USA
email: ide@cs.vassar.edu

WORKSHOP PROGRAM

Saturday, October 7

- 9:00-9:15 **OPENING AND OVERVIEW**
Nancy Ide (Vassar College, USA)
Martha Palmer (Univ. of Pennsylvania, USA)
- 9:15-9:45 *An Unsupervised Method for Multilingual Word Sense Tagging Using Parallel Corpora*
Mona Diab (University of Maryland, USA)
- 9:45-10:15 *Sense Clusters for Information Retrieval: Evidence from SemCor and the EuroWordNet InterLingual Index*
Julio Gonzalo, Irina Chugur and Felisa Verdejo (UNED, Spain)
- 10:15-10:30 **COFFEE BREAK**
- 10:30-11:00 *Chinese-Japanese Cross Language Information Retrieval: A Han Character Based Approach*
Maruf Hasan and Yuji Matsumoto (NARA Inst., Japan)
- 11:00-11:30 *Experiments in Word Domain Disambiguation for Parallel Texts*
Bernardo Magnini and Carlo Strapparava (IRST, Italy)
- 11:30-12:00 **DISCUSSION AND SUMMARY**
Nancy Ide (Vassar College, USA)
Adam Kilgarriff (ITRI, UK)
Martha Palmer (Univ. of Pennsylvania, USA)
David Yarowsky (Johns Hopkins Univ., USA)
- 12:00-12:15 **SIGLEX Business Meeting**

TABLE OF CONTENTS

PREFACE	i
PROGRAM COMMITTEE	ii
WORKSHOP PROGRAM	iii
TABLE OF CONTENTS	iv
<i>An Unsupervised Method for Multilingual Word Sense Tagging Using Parallel Corpora</i> Mona Diab.....	1
<i>Sense Clusters for Information Retrieval: Evidence from SemCor and the EuroWordNet InterLingual Index</i> Julio Gonzalo, Irina Chugur and Felisa Verdejo.....	10
<i>Chinese-Japanese Cross Language Information Retrieval: A Han Character Based Approach</i> Maruf Hasan and Yuji Matsumoto	19
<i>Experiments in Word Domain Disambiguation for Parallel Texts</i> Bernardo Magnini and Carlo Strapparava	27
AUTHOR INDEX	34