

Second Chinese Language Processing Workshop 2000

**Held in conjunction with the 38th Annual Meeting of the
Association for Computational Linguistics (ACL 2000)**

**Hong Kong
8 October 2000**

Editors:

**Martha Palmer
Aravind Joshi**

**Mitch Marcus
Fei Xia**

ISBN: 978-1-62748-371-1

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2000) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

PRESENTATIONS

<i>Two Statistical Parsing Models Applied to the Chinese Treebank</i> Daniel M. Bikel and David Chiang	1
<i>Sense-Tagging Chinese Corpus</i> Hsin-Hsi Chen and Chi-Ching Lin	7
<i>Knowledge Extraction for Identification of Chinese Organization Names</i> Keh-Jiann Chen and Chao-jan Chen	15
<i>Using Co-occurrence Statistics as an Information Source for Partial Parsing of Chinese</i> Elliott Franco Drabek and Qiang Zhou	22
<i>Sinica Treebank: Design Criteria, Annotation Guidelines, and On-line Interface</i> Chu-Ren Huang, Feng-Yi Chen, Keh-Jiann Chen, Zhao-ming Gao and Kuang-Yu Chen	29
<i>Enhancement of a Chinese Discourse Marker Tagger with C4.5</i> Benjamin K. T'sou, Tom B.Y. Lai, Samuel W.K. Chan, Weijun Gao and Xuegang Zhan	38
<i>Statistically-Enhanced New Word Identification in a Rule-Based Chinese System</i> Andi Wu and Zixin Jiang	46
<i>Comparing Lexicalized Treebank Grammars Extracted from Chinese, Korean, and English Corpora</i> Fei Xia, Chunghye Han, Martha Palmer and Aravind Joshi	52
<i>The Research of Word Sense Disambiguation Method Based on Co-occurrence Frequency of Hownet</i> Erhong Yang, Guoqing Zhang, and Yongkui Zhang	60
<i>A Trainable Method for Extracting Chinese Entity Names and Their Relations</i> Yimin Zhang and Joe F Zhou	66
<i>Statistics Based Hybrid Approach to Chinese Base Phrase Identification</i> Tie-jun Zhao, Mu-yun Yang, Fang Liu, Jian-min Yao and Hao Yu	73
<i>A Block-Based Robust Dependency Parser for Unrestricted Chinese Text</i> Ming Zhou	78

POSTERS

<i>Annotating Information Structures in Chinese Texts Using HowNet</i> Kok Wee Gan and Ping Wai Wong	85
<i>Machine Learning Methods for Chinese Web Page Categorization</i> Ji He, Ah-Hwee Tan and Chew-Lim Tan	93

<i>Semantic Annotation of Chinese Phrases Using Recursive Graph</i> Donghong Ji	101
<i>Text Meaning Representation for Chinese</i> Wanying Jin	109
<i>How Should a Large Corpus Be Built?—A Comparative Study of Closure in Annotated Newspaper Corpora from Two Chinese Sources, Towards Building a Larger Representative Corpus Merged from Representative Sublanguage Collections</i> John J. Kovarik	116
<i>A Clustering Algorithm for Chinese Adjectives and Nouns</i> Yang Wen, Chunfa Yuan and Changning Huang	124
<i>Extraction of Chinese Compound Words - An Experimental Study on a Very Large Corpus</i> Jian Zhang, Jianfeng Gao and Ming Zhou	132
<i>An Algorithm for Situation Classification of Chinese Verbs</i> Xiaodan Zhu, Chunfa Yuan, K.F. Wong and Wenjie Li	140
INVITED TALK	
<i>Zero Anaphors in Chinese Discourse Processing</i> Chin-Chuan Cheng	146