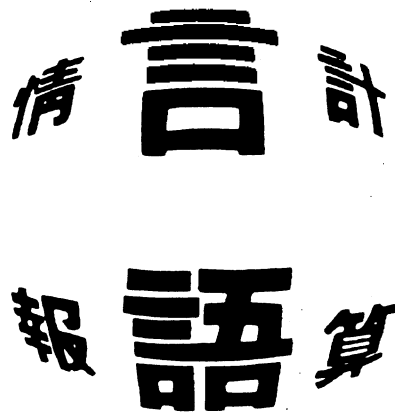


PACLIC12

Language, Information and Computation



18-20 Feb, 1998
Singapore

Edited by
Jin Guo, K T Lua and Jie Xu

Chinese and Oriental Languages Information Processing Society
Laboratory for Computational Linguistics,
Department of Information Systems and Computer Science, National University of Singapore

Language, Information and Computation

Proceedings of the
12th Pacific Asia Conference

18-20 February 1998, Singapore

Edited by
Jin Guo, Kim Teng Lua and Jie Xu

Chinese and Oriental Languages Information Processing Society
Laboratory for Computational Linguistics, Department of Information
Systems and Computer Science, National University of Singapore

©1998 Chinese and Oriental Languages Information
Society

All right reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Published by Chinese and Oriental Languages
Information Processing Society
c/o Department of Information Systems and
Computer Science, National University of Singapore,
Lower Kent Ridge Road, Singapore 119260.

Contents

Foreword

<i>Ettehkey</i> 'How' as a Small Clause Head, <i>Daeho Chung</i>	1
Three Kinds of Korean Reflexives: A Corpus Linguistic Investigation on Grammar and Usage, <i>Beom-mo Kang</i>	10
The Chinese Temporal Coverbs, Postpositions, Coverb-Postposition Pairs, and Their Temporal Logic, <i>York Chung-Ho Yang and June-Jei Kuo</i>	20
Grammatical Devices in the Processing of [+Wh] and [+Focus], <i>Jie Xu</i>	33
Mandarin Intransitive Reflexive Verbs and the Unaccusative Hypothesis (Mandarin Intransitive Reflexive Verbs), <i>Lian-Cheng Chief</i>	48
Analyzing Embedded Noun Phrase Structures Derived from Japanese Double-Nominal-Case Construction, <i>Masahiro Oku</i>	60
The Two Kinds of Japanese Negative Nai in Terms of Their NPI Licensing Condition, <i>Akira Ikeya and Masahito Kawamori</i>	69
A Multiple Inheritance Analysis of the Internally-Headed Relative Clause in Japanese, <i>Chiharu Uda Kikuta</i>	82
The Acquisition of Functional Categories: Data from Japanese, <i>Yoshie Yamashita</i>	94
Moras, Syllables, and Feet in Japanese, <i>Seiichiro Inaba</i>	106
English Nasal-Final Prefixes and Uniform Exponence, <i>Seok-keun Kang</i>	118
Common Grounds as Multiple Information States, <i>Jae-Il Yeom and Ik-Hwan Lee</i>	127
Co-Interpretation Network in English Discourse, <i>Yibin Ni</i>	139
Predictivity vs. Stipulativity in the Lexicon, <i>Cornelia Maria Verspoor</i>	152
Using Case Prototypicality as a Semantic Primitive, <i>Dan-Hee Yang, Ik-Hwan Lee and Mansuk Song</i>	163
Case Alternations in Potential Constructions in Japanese and Their Semantic Implications, <i>Hiroaki Nakamura and Takeshi Fujita</i>	172
On Plural Anaphora, <i>Kiyoshi Ishikawa</i>	184
MI-Trigger-Based Language Modelling, <i>Guodong Zhou and Kim-Teng Lua</i>	195

Extracting Recurrent Phrases and Terms from Texts Using a Purely Statistical Method, <i>Zhao-Ming Gao and Harold Somers</i>	206
Chinese Word Segmentation, <i>Haizhou Li and Baosheng Yuan</i>	212
Error-Driven Learning of Chinese Word Segmentation. <i>Julia Hockenmaier and Chris Brew</i>	218
Automatic Bunsetsu Segmentation of Japanese Sentences Using a Classification Tree, <i>Yujie Zhang and Kazuhiko Ozeki</i>	230
Word-Sense Classification by Hierarchical Clustering, <i>Ken Y. K. Lau and Robert W. P. Luk</i>	236
Automatic Acquisition of a High-Precision Translation Lexicon from Parallel Chinese-English Corpora, <i>Zhao-Ming Gao</i>	248
Surrogater: A Simple Yet Efficient Document Condensation System, <i>Joe Zhou</i>	255
A Computational Method for Resolving Ambiguities in Coordinate Structures, <i>Haodong Wu and Teiji Furugori</i>	263
On Removing Ambiguity in Text Understanding, <i>Simin Li and Yukihiro Itoh</i>	271
An Automatic Chinese Document Revision System Using Bit and Character Mask Approach, <i>June-Jei Kuo</i>	283
Japanese Kana-to-Kanji Conversion Using Large Scale Collocation Data, <i>Yasuo Koyama, Masako Yasutake, Kenji Yoshimura and Kosho Shudo</i>	295
Syntactic Verifier as a Filter to Compound Unit Recognizer, <i>Hanmin Jung, Sanghwa Yuh, Taewan Kim and Dong-In Park</i>	303
Machine-Readable Dictionary Headwords, <i>Yasuhito Tanaka and Kenji Kita</i>	310
The Advantages of 3D-Trees in Modelling Human Sentence Processing, <i>Charles C. Lee</i>	316
Improvement of Korean Proofreading System Using Corpus and Collocation Rules, <i>Young-Soog Chae</i>	328
An Investigation into the Use of Argument Structure and Lexical Mapping Theory for Machine Translation, <i>Shun Ha Sylvia Wong and Peter Hancox</i>	334
Using A Semantic Classification in Parsing Chinese: Some Preliminary Results, <i>Kok Wee Gan</i>	340
Adjunct Roles and External Predication, <i>Yong-Beom Kim</i>	348

The Underlying Representation of the <i>Tough</i> Construction in English, <i>Mariko Saiki</i>	357
Tense and the Speaker's Attitude in English, <i>Minako Nakayasu</i>	363

Foreword

The Pacific Asia Conference on Language, Information and Computation (PACLIC) is an annual gathering of scholars with a wide range of interests in theoretical and computational linguistics. The Conference solicits papers treating all fields in theoretical and computational linguistics, including syntax, morphology, phonology, semantics, pragmatics, discourse analysis, corpus linguistics, formal grammar theory, natural language processing, and computer applications. PACLIC12 is the twelfth of this series of conferences held in the Asia Pacific region.

This year, PACLIC12 is hosted by the Singapore Chinese and Oriental Languages Information Society (COLIPS) and Computational Linguistic Lab, Department of Information Systems and Computer Science, National University of Singapore(NUS). It is to be held from 18-20 Feb, 1998 at the Science Auditorium of NUS.

For PACLIC12, we had requested full papers for reviews. Though it is common for conferences in computer information sciences to ask for full papers, we admit that this is not an usual practice in linguistics. We are now very pleased to conclude that such a requirement did not in any way affect the enthusiasm of our learned colleagues on submitting papers. We received a total of 62 submissions from 15 countries and regions. These include: Japan, Korea, Taiwan, United Kingdom, Singapore United States, Australia, France, Hong Kong, Russia, Brazil, China, Denmark, Greece and Slovenia. From this list, we can see that the influence of PACLIC has spread far and wide.

It is never an easy task to review papers from linguists and computer information scientists at the same time. For PACLIC12, each paper is reviewed by 3-5 reviewers. Comments from reviewers whose expertise does not fall into the area of the paper carry substantially less weightage than those who are specialized in the area. In the end, a total of 30 papers are carefully selected for oral presentations. This implies a stringent acceptance rate of only 44%. Another 6 papers are included for postal presentations. All 36 papers are included in this proceedings.

We have observed that there is a fairly equal distribution of papers in both linguistic and language information processing. Papers from linguistics cover grammar, syntax, phonology, lexicon and semantics, while papers from language information processing covers language modeling, parsing and processing.

The strength of PACLIC is its ability to attract, in a single conference, two completely different disciplines which traditionally do not have any interaction with each other at all. Yet, they cover mostly similar areas of research. In this conference, researchers from these two fields will sit down together and discuss face-to-face common topics from very different angles and perspective. This synergy will definitely benefit both fields tremendously.

In PACLIC12, we are fortunate enough to have been able to invite 2 key researchers to deliver speeches in their respective fields of expertise. Professor Jhing-fa Wang, who is well know in speech processing (and who will be chairing the PACLIC13 in Tawian), will speak on "Experience on the Development of Spoken Language System based on Continuous Speech Recognition". Professor Mohanan K. P. from National University of Singapore will deliver his keynote speech on " Non-Monotonicity in Linguistics"

We wish to take this opportunity to thank the National University of Singapore for providing the conference facilities and logistic support for the conference. We would also like to express our sincere thanks to the PACLIC12 co-chairs and technical committee members. Reviewing papers from unfamiliar areas is always a difficult experience. Special thanks go to Professor Jhing-fa Wang and Professor K. P. Mohanan for delivering the keynote speeches. Prof Wang has to make special arrangement to rush back from a conference in India. We also wish to thank all our learned colleagues who submitted papers and who come to Singapore to attend the conference. Without them, PACLIC12 will lose all its meaning. Last but not least, we wish to thank Ms Toh Siew Kek, our COLIPS secretary for her help in the booking of hotels, arrangement for the conference banquet and all other 'little-little' things.

Lua Kim Teng (PACLIC12 Chair)
Feb, 1998

PACLIC12 Committees

CONFERENCE CHAIRS

Kim Teng Lua	National University of Singapore, and President of the Chinese and Oriental Language Information Processing Society (COLIPS)	Chairman
Benjamin K. T'sou	City University of Hong Kong	Co-chairman
Chu-Ren Huang	Taiwan Academia Sinica	Co-chairman
Young-Hern Lee	Korea Chosun University, and President of the Korean Society for Language and Information (KSLI)	Co-chairman
Akira Ikeya	Tokyo Gakuen University	Co-chairman
Jhing-fa Wang	National Cheng Kung University	Co-chairman

PROGRAM COMMITTEE

Jin Guo	National University of Singapore (Chairman)
Akira Ishikawa	Sophia University
Baosheng Yuan	National University of Singapore
Byung-Soo Park	Kyung Hee University Korea
Changning Huang	Tsinghua University
Chris Tancredi	The Yokohama National University
Chungmin Lee	Seoul National University
Haizhou Li	National University of Singapore
Hongyin Tao	Cornell University
Ik-Hwan Lee	Yonsei University
Jerry Seligman	National Chung Cheng University
Jie Xu	National University of Singapore
Kathleen Ahrens	National Taiwan University
Jian Su	National University of Singapore
Keh-Jiann Chen	Academia Sinica
Kevin Knight	University of Southern California
Kiyoshi Ishikawa	Hosei University
Ling Cao	National University of Singapore
Maosong Sun	Tsinghua University
Masato Ishizaki	ATR
Martha Palmer	University of Pennsylvania
Paul Horng Jyh Wu	National University of Singapore
Richard Sproat	AT&T Bell Labs
Robert Luk	The Hong Kong Polytechnic University
Shiwen Yu	Beijing University
Shuichi Yatabe	University of Tokyo
Taiyi Huang	Institute of Automation, Academia Sinica
Von-Wun Soo	National Tsing Hua University