PAGVIC 10

Language, Information and Computation



Proceedings of
The 10th Pacific Asia Conference

Edited by Benjamin K. T'sou and Tom B. Y. Lai

Language, Information and Computation

Proceedings of the 10th Pacific Asia Conference

27-28 December 1995, Hong Kong

Edited by Benjamin K T'sou and Tom B Y Lai

Language Information Sciences Research Centre City University of Hong Kong Hong Kong

Contents

	Foreword		•
	Acknowledgment		vi
	Programme Committee and List of Reviewers		iλ
	Conference Programme		>
1.	Distances and Trees in Linguistics	William SY. Wang	1
2.	Individuals and Modality	Hisashi Komatsu	7
3.	An Analysis of Generic Expressions in Situation Semantics	Ik-Hwan Lee	19
4.	Agreement Target Situations	Ishikawa Kiyoshi	29
5.	Comprehending Text: Achieving Coherence through a Connectionist Architecture	Samuel W K Chan	39
6.	Predication of Meaning of Bisyllabic Chinese Compound Words Using Back Propagation Neural Network	Lua Kim Teng	49
7.	Dative Shift in Chinese and English: A Lexical Mapping Account	One-Soon Her Hui-Ting Huang	57
8.	A Constraint-Based Lexical Approach to "Floating" Quantifiers	Byung-Soo Park	67
9.	Scrambling in German — Extraction into the Mittelfeld	Stefan Müller	79
10.	Crosslinguistic Notions of (In)definiteness	Ishikawa Kiyoshi	85
11.	Construction as a Theoretical Entity: An Argument Based on Mandarin Existential Sentences	Chao-ran Chen Chu-Ren Huang Kathleen Ahrens	91
12.	Clitic Analyses of Korean "Little Words"	Hee-Rahk Chae	97
13.	A Cognitive Account of the Lexical Polysemy of Chinese Kai	Flora Yu-Fang Wang	103

14.	Automatic Sense Disambiguation for Target Word Selection	Kwon Yang Kim Se Young Park Sang Jo Lee	109
15.	Structural Ambiguity and Conceptual Information Retrieval	Mathis Huey-chyun Chen Jason J.S. Chang	115
16.	Ambiguity Resolution in Chinese Word Segmentation	Sun Maosong Benjamin K T'sou	121
17.	The Postprocessing of Optical Character Recognition Based on Statistical Noisy Channel and Language Model	Jason J. S. Chang Shun-Der Chen	127
18.	A Quantitative Analysis of Word-Definition in a Machine-Readable Dictionary	Robert W.P. Luk Venus M.K. Chan	133
19.	Metaphorical Paradoxes: A Window on the Conceptual System	Kathleen Ahrens	139
20.	Predicate-Argument Structure of English Adjectives	Akira Ikeya	149
21.	Prototype Theory and Case Assignment	Yumiko Kinjo	157
22.	HMM Parameter Learning for Japanese Morphological Analyzer	Koichi Takeuchi Yuji Matsumoto	163
23.	Automatic Acquisition of Class-based Rules for Word Alignment	Sur-Jin Ker Jason J.S. Chang	173
24.	Automated Alignment in Multilingual Corpora	J.A. Campbell Alex Chengyu Fang	185
25.	Using Brackets to Improve Search for Statistical Machine Translation	Dekai Wu Cindy Ng	195
26.	Reference in Dialogues and Shared Belief Revision	Norihiro Ogata Tadayudi Matsuzawa Akira Ishikawa Hisashi Komatsu	205
27.	Recursion Problems in Concatenation: A Case of Korean Morphology	Kiyong Lee	215
28.	Interpretational Strategies and Semantic Identities	Ioannis Kanellos Victor Hugo Zaldivar-Carrillo	225

29.	Preferred Clause Structure in Mandarin Spoken and Written Discourse	Shu-Mei Liu	231
30.	A Corpus-Based Study of Adverbial Clauses in Mandarin Chinese Conversations: A Preliminary Analysis	Yu-Fang Wang	237
31.	A Network-based Writing System for French	Aesun Yoon Hyuk-Chul Kwon	243
32.	Web Access to a Lexical Database Using VB/Access CGI Programming	Jonathan J Webster	249
33.	Document Ranking Method for High Precision Rate	Mee-Sun Jeon Se-Young Park	255
34.	Natural Languages Analysis in Machine Translation (MT) based on the STCG (String-Tree Correspondence Grammar)	Tang Enya Kong Zaharin Yusoff	261
35.	An Implementation of a Multilingual Regular Expression Segmentor for Ordinary and Morphologically Rich Lexical Tokens	Paul Wu Horng Jyh Kevin Cheong	267
36.	Sorting by Sound — Arbitrary Lexical Ordering for Transcribed Thai Text	Doug Cooper	273
37.	A Unified Account of Polarity Phenomena	Chungmin Lee	281

Author Index

©1995 City University of Hong Kong

All rights 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.

ISBN 962-442-083-1

Published by City University of Hong Kong

For the Language Information Sciences
Research Centre
City University of Hong Kong
Tat Chee Avenue
Kowloon
Hong Kong

Foreword

The link between rationalists and empirists, perhaps like that between Yin and Yang, has been at times a tenuous one, but in the study of language, the legitimacy of both is increasingly recognised and the need for mutual understanding, if not cross-fertilization, is readily seen in recent decades. This is especially so as advancement in technology has provided unprecedented opportunities for the testing of models by the former on the one hand, and for the exploration of newer and better means to obtain and utilize data by the latter on the other. Moreover, the emergence of the information super highway in the new era also necessitates the full exploitation of technology in minimizing language as a barrier to better communication.

The Pacific Asia Conference on Language, Information and Computation (PACLIC) combines two traditions: the Asian Conference on Language, Information and Computation (ACLIC) and the Pacific Asia Conference on Formal and Computational Linguistics (PACFoCoL). These two traditions converged at Kyoto in 1994 when a joint conference was held there. It was also there that the two overlapping groups of researchers decided to consolidate the two traditions as one.

The name of the merged conference would be PACLIC and the longer history of ACLIC makes it necessary for us to number the 'first' PACLIC held in 1995 in Hong Kong the tenth, hence PACLIC10. ACLIC began more than a decade ago as a small forum providing an intimate occasion primarily for scholars working on formal linguistics in Japan and Korea to share the results of their research. The subsequent participation of scholars from Taiwan led to the first PACFoCoL held in 1993 in Taipei, where computational linguistics became a prominent additional theme. The involvement of researchers from Hong Kong and other places in 1993 and 1994 further widened the circle of participants.

Submissions for presentation at ACLIC and PACFoCoL were in the form of extended abstracts. Usually around twenty papers were accepted for presentation and many papers were published in proceedings.

This year there was a very good response to our call for extended abstracts. We tried our best to find the most suitable reviewers (at least two) for each submitted abstract. In the case of a number of papers, we went out of the Programme Committee for help with the anonymous review process.

Following the prompt response of many reviewers, we were faced with the difficult task of deciding whether we should really accept only around twenty long papers for presentation in plenary sessions and turn away a number of promising papers on the basis of the extended abstracts submitted. After careful consideration and consultation the decision was taken to go beyond tradition and schedule two quarter-day parallel sessions for short papers. This has had the net effect of allowing more papers to be presented. As a result, there are now all together 36 presentations at PACLIC10, out of more than 60 initial submissions.

The contributors are mainly from the Pacific Asia region — Japan, Korea, Taiwan, Hong Kong, Singapore, Mainland China, Malaysia, Australia and Thailand. We have had submissions from other parts of the world — Europe and America. As a result we have researchers from England, France and Germany presenting their results at PACLIC10. Altogether the participants are drawn from twelve communities of scholars from all over the world.

A conference like PACLIC10 will of course have papers on formal syntax/semantics and language information engineering. We have papers on linguistic formalisms and model-theoretic-based semantics/pragmatics. We also have papers on corpus-based linguistic studies as well as works in various areas of general linguistics. On the computing side, we have papers on work involving stochastic and corpus-based methods. We have papers on various aspects of computational linguistics. We have also accepted a small number of papers on 'softer' areas such as computer-aided language learning, linguistic information processing and multilingual computing, which are to many people the real meeting points of language and the computer. It may be observed that papers presented at PACLIC10 reflect a trend towards using more numerical methods, but discrete approaches remain a main stay.

The keynote speaker at the conference, Professor William S.-Y. Wang, the first holder of a Chair in Language Engineering in Asia, and probably anywhere in the world, introduces yet a new approach to the computational modeling of language — in the computation of historical affinity of languages.

The organizers are pleased that the papers accepted on the basis of the extended abstracts have been made available as full papers in this volume prior to the conference. Soft copies of many of these papers have also been put on the WEB at http://ctwww.cityu.edu.hk/.

The order of papers in this volume generally follows that in the programme, and the sequence in the Table of Contents is indicated by the number within square brackets after each paper in the programme.

Benjamin K. T'sou Tom B. Y. Lai

Acknowledgment

A meaningful conference requires the concerted efforts of many individuals. I should like to take this opportunity to express my gratitude towards Chu-Ren Huang, Akira Ikeya and Byung-Soo Park, co-chairpersons of the conference, and to Akira Ishikawa and Paul Horng-Jyh Wu for their unfailing support and tremendous contributions towards PACLIC10. We are no less thankful to the other members of the Programme Committee, the invited reviewers and all others who have made it possible for this conference to be held. We are specially grateful for the financial support of Onflo Computer Co Ltd, and for the sponsorship of the Linguistic Society of Hong Kong and the Department of Chinese, Translation and Linguistics, City University of Hong Kong. Pro-Vice-Chancellor Y.W. Lam and staff at the Vice-Chancellor's Office, Research Office, and Language Information Sciences Research Centre deserve a special vote of thanks for their support, especially Helena Tsang, Stella Lo, Grace Ko, Shirley So, Kei Yu and Yvonne Chan, for the great efforts they have put in to actually bring the conference into being. As I was away for much of the period from March to August, 1995, I must register my heartfelt gratitude to Tom Lai, who deserves much of the credit for putting PACLIC10 on a sure footing.

Benjamin K T'sou Chairperson, Programme Committee

Programme Committee

Chairman: Benjamin K. T'sou, City University of Hong Kong

Co-chairmen: Chu-Ren Huang, Academia Sinica

Akira Ikeya, Tokyo Gakuen University Byung-Soo Park, Kyung Hee University

Secretary: Tom Lai, City University of Hong Kong

Lydia S.C. Chan, National University of Singapore Jason Jyun-Sheng Chang, National Tsing Hua University

Hsin-hsi Chen, National Taiwan University

Jaewoong Choe, Korea University

Zhendong Dong, National University of Singapore

One-Soon Her, National Chengchi University

Changning Huang, Tsinghua University

Akira Ishikawa, Sophia University

Masato Ishizaki, Advanced Telecommunications Research Inst. International

Yan Jiang, Hong Kong Polytechnic University

Hisashi Komatsu, Hiroshima City University

Chungmin Lee, Seoul National University

Ik-Hwan Lee, Yonsei University

Kiyong Lee, Korea University

Shuhung Leung, City University of Hong Kong

Kim-Teng Lua, National University of Singapore

Robert W.P. Luk, City University of Hong Kong

Yuji Matsumoto, Nara Institute of Science and Technology

Haihua Pan, City University of Hong Kong

Jhing-fa Wang, National Chengkung University

Jonathan J. Webster, City University of Hong Kong

Dekai Wu, Hong Kong University of Science and Technology

Paul Horng-Jyh Wu, National University of Singapore

Tianshun Yao, North-East University

List of Additional Reviewers

Yang Gu, Chinese University of Hong Kong

Hing-Cheung Ho, City University of Hong Kong

Horace Ip, City University of Hong Kong

Thomas Lee, Chinese University of Hong Kong

Hing-Lung Lin, City University of Hong Kong

Weera Ostapirat, University of California, Berkeley

Dingxu Shi, Hong Kong Polytechnic University

Hsu Loke Soo, National University of Singapore

Christopher Tancredi, Yokohama National University

Peter Tung, City University of Hong Kong

William Wang, City University of Hong Kong and University of California, Berkeley

Udom Warotamasikkhadit, Ramkhmaheng University, Thailand

Liejiong Xu, City University of Hong Kong

Conference Programme

27 December 1995

Keynote Speech Chairperson: Benjamin K. T'sou City University of Hong Kong

"Distances and Trees in Linguistics "[1]

William S.-Y. Wang City University of Hong Kong and University of California, Berkeley

Plenary A1 Chairperson: Byung-Soo Park Kyung Hee University

"Individuals and Modality" [2]

Hisashi Komatsu Hiroshima City University

"An Analysis of Generic Expressions in Situation Semantics" [3]

Ik-Hwan Lee Yonsei University
"Agreement Target Situations" [4]
Ishikawa Kiyoshi Hosei University

Plenary A2 Chairperson: Jason J.S. Chang National Tsing Hua University

"Comprehending Text: Achieving Coherence through a Connectionist Architecture" [5]

Samuel W.K. Chan University of New South Wales

"Predication of Meaning of Bisyllabic Chinese Compound Words Using Back Propagation Neural

Network" [6]

Lua Kim Teng National University of Singapore

Plenary A3 Chairperson: Akira Ikeya Toyo Gakuen University

"Dative Shift in Chinese and English: A Lexical Mapping Account" [7]

One-Soon Her National Chengchi University, Hui-Ting Huang Academia Sinica

"A Unified Account of Polarity Phenomena" [37]

Chungmin Lee Seoul National University

"A Constraint-Based Lexical Approach to 'Floating' Quantifiers" [8]

Byung-Soo Park Kyung Hee University

Parallel B1 Chairpersons: Ik-Hwan Lee Yonsei University, Hisashi Komatsu Hiroshima City University

"Scrambling in German — Extraction into the Mittelfeld" [9]

Stefan Müller Humboldt Universität zu Berlin

"Crosslinguistic Notions of (In)definiteness" [10]

Ishikawa Kiyoshi Hosei University

"Construction as a Theoretical Entity: An Argument Based on Mandarin Existential Sentences" [11]

Chao-ran Chen, Chu-Ren Huang Academia Sinica, Kathleen Ahrens National Chung Cheng

University

"Clitic Analysis of Korean 'Little Words'" [12]

Hee-Rahk Chae Hankuk University of Foreign Studies

"A Cognitive Account of the Lexical Polysemy of Chinese Kai" [13]

Flora Yu-Fang Wang National Taiwan Normal University

Parallel C1 Chairpersons: Lua Kim Teng, Paul Wu Horng Jyh National University of Singapore

"Automatic Sense Disambiguation for Target Word Selection" [14]

Kwon Yang Kim Kyungpook Sanup University, Se Young Park Electronics and Telecommunications

Research Institute, Sang Jo Lee Kyungpook National University

"Structural Ambiguity and Conceptual Information Retrieval" [15]

Mathis Huey-chyun Chen, Jason J. S. Chang National Tsing Hua University

"Ambiguity Resolution in Chinese Word Segmentation" [16]

Sun Maosong Tsinghua University and City University of Hong Kong, Benjamin K T'sou City

University of Hong Kong

"The Postprocessing of Optical Character Recognition Based on Statistical Noisy Channel and

Language Model" [17]

Jason J. S. Chang, Shun-Der Chen National Tsing Hua University

"A Quantitative Analysis of Word-Definition in a Machine-Readable Dictionary" [18]

Robert W.P. Luk, Venus M.K. Chan City University of Hong Kong

28 December 1995

Plenary A4

Chairperson: Chu-Ren Huang Academia Sinica

"Metaphorical Paradoxes: A Window on the Conceptual System" [19]

Kathleen Ahrens National Chung Cheng University

"Predicate-Argument Structure of English Adjectives" [20]

Akira Ikeya Toyo Gakuen University

"Prototype Theory and Case Assignment" [21]

Yumiko Kinjo Kobe University

Plenary A5

Chairpersons: Yuji Matsumoto Nara Institute of Science and Technology, Dekai Wu Hong Kong University of Science and Technology

"HMM Parameter Learning for Japanese Morphological Analyzer" [22] Koichi Takeuchi, Yuji Matsumoto Nara Institute of Science and Technology

"Automatic Acquisition of Class-based Rules for Word Alignment" [23]

Sur-Jin Ker, Jason J. S. Chang National Tsing Hua University

"Automatic Alignment in Multilingual Corpora" [24]

J.A. Campbell, Alex Chengyu Fang University College London

"Using Brackets to Improve Search for Statistical Machine Translation" [25] Dekai Wu, Cindy Ng Hong Kong University of Science and Technology

Plenary A6

Chairperson: Chungmin Lee Seoul National University

"Reference in Dialogues and Shared Belief Revision" [26]

Norihiro Ogata University of Tsukuba, Tadayuki Matsuzawa DEC Japan Akira Ishikawa Sophia University, Hisashi Komatsu Hiroshima City University "Recursion Problems in Concatenation: A Case of Korean Morphology" [27]

Kiyong Lee Korea University

Parallel B2

Chairpersons: Kiyong Lee Korea University, Yan Jiang Hong Kong Polytechnic University

"Interpretational Strategies and Semantic Identities" [28]

Ioannis Kanellos, Victor Hugo Zaldivar-Carrillo Ecole Nationale Supérieure des Telecommunications de Bretagne

"Preferred Clause Structure in Mandarin Spoken and Written Discourse" [29]

Shu-Mei Liu National Chengchi University and Academia Sinica

"A Corpus-Based Study of Adverbial Clauses in Mandarin Chinese Conversations: A Preliminary Analysis" [30]

Yu-Fang Wang National Taiwan Normal University

"A Network-Based Writing System for French" [31]

Aesun Yoon, Hyuk-Chul Kwon Pusan National University

"Web Access to a Lexical Database Using VB/Access CGI Programming" [32]

Jonathan J Webster City University of Hong Kong

Parallel C2

Chairpersons: Yao Tianshun North-East University, Haihua Pan City University of Hong Kong

"Document Ranking Method for High Precision Rate" [33]

Mee-Sun Jeon, Se-Young Park Electronics and Telecommunications Research Institute

"Natural Languages Analysis in Machine Translation (MT) based on the STCG (String-Tree Correspondence Grammar)" [34]

Tang Enya Kong, Zaharin Yusoff Universiti Sains Malaysia

"An Implementation of a Multilingual Regular Expression Segmentor for Ordinary and Morphologically Rich Lexical Tokens" [35]

Paul Wu Horng Jyh, Kevin Cheong National University of Singapore

"Sorting by Sound — Arbitrary Lexical Ordering for Transcribed Thai Text" [36]

Doug Cooper Centre for Research in Computational Linguistics, Bangkok