Welcome Message from Local Chair, Dr Lua Kim Teng

Dear Friends and Colleagues,

A very warm welcome to the beautiful Sentosa Resort Island of Singapore!

The O-COCOSDA2003 and PACLIC17 are major events in the Asian natural language research arena. This is the seventh in the series of O-COCOSDA or Oriental International Coordinating Committee on Speech Databases and Speech I/O System Assessment and the seventeenth PACLIC or Pacific Asia Conference on Language, Information and Computation. We are very honored that Singapore is chosen to be the venue to host both meetings and co-lips is the organizer.

COCOSDA is an international workshop held annually by the oriental chapter of The International Committee for the Coordination and Standardization of Speech Databases and Assessment Techniques for Speech Input/Output. The first meeting was held in Hong Kong and then the past five conferences were held in Japan, Taiwan, China Mainland, Korea and Thailand. After the Singapore meeting, I understand that the next meeting, O-COCOSDA2004 will be held in New Delhi, India. The meeting will be organized under the leadership of Dr Shyam S. Agrawal, Emeritus Scientist, Central Scientific Instruments, Delhi. I wish that every one of us will support the meeting and make it a success!

The PACLIC series of conference is an annual meeting of scholars in theoretical and computational linguistics from the Pacific Asia region. PACLIC aims to cover all aspects of both theoretical and computational linguistics, including morphology, phonology, syntax, semantics, pragmatics, discourse analysis, typology, corpus linguistics, formal grammar theory, natural language processing, natural language systems and related computer applications.

We find that by holding the two conferences at the same venue side-by-side will allow direct interaction between members of the two communities, the speech and the text and NLP community. Perhaps, this synergy will create sparks that will be most beneficial to both research areas.

We received more than 100 submissions from research in Asia and many other parts of the world for the current joint meetings of O-COCOSDA2003 and PACLIC17. Only about 50% of the papers are selected for oral presentations. From the papers presented in the two volumes of proceedings, you will find that we are working hard to maintain high academic standards for the conferences.

I wish to thank the steering committee members: Dr Shyam Agrawal, Nick Campbell, Qiang Huo, Dawa Idomuso, Shuichi Itahashi, Aijun Li, Lin Shan Lee, Yongju Lee, Yung-Hwan Oh, Virach Sornlertlamanich, Hsiao-Chuan Wang, Jialu Zhang (above for O-COCOSDA2003), Huang Chu-Ren, Akira Ikeya, Ik-Hwan Lee, Byung-Soo Park, Benjamin T'sou (above are co-chairs for PACLIC17) for their assistance, guidance and
promoting the conferences. I also wish to thank all members of programs committees for reviewing the papers. The two program chairs, Dr Li Haizhou (under the assistance of Dr Zhang Min) and Dr Ji Dong Hong worked particularly hard in preparing the programs and the proceedings and Ms Lua Tse Min for assisting in the final typesetting and preparation of the proceedings. I also wish to thank the council members of colips for their supports and funding of the conferences. The travel grants received by the young researchers come from the colips council and Lee Foundation of Singapore jointly. Lee Foundation has been always very supportive to our academic activities. Last but not least, we also wish to thank staff members of the Sijori Hotel who help organize the conference.

Wish all of you the very best!

By

Lua Kim-Teng, Local Chair, O-COCOSDA2003 and PACLIC17
15-08-2003
# PACLIC17 Conference Committees

**International Co-Chairs**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huang Chu-Ren,</td>
<td>Academia Sinica, Taipei, Taiwan</td>
</tr>
<tr>
<td>Akira Ikeya,</td>
<td>Toyo Gakuen University, Tokyo</td>
</tr>
<tr>
<td>Ik-Hwan Lee</td>
<td>Yonsei University, Seoul</td>
</tr>
<tr>
<td>Kawamori Masahito</td>
<td>NTT, Japan</td>
</tr>
<tr>
<td>Byung-Soo Park</td>
<td>Kyung Hee University, Seoul</td>
</tr>
<tr>
<td>Benjamin Tsou</td>
<td>City University of Hong Kong, Hong Kong</td>
</tr>
</tbody>
</table>

**Local Chair**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim Teng LUA</td>
<td>National University of Singapore</td>
</tr>
</tbody>
</table>

**Program Committee**

<table>
<thead>
<tr>
<th>Name</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dong Hong Ji (Chair)</td>
<td>Laboratories for Information Technology, Singapore.</td>
</tr>
<tr>
<td>Jing-Shin Chang</td>
<td>Dept. of CSIE, National Chi-Nan Univ., Puli, Nantou, Taiwan</td>
</tr>
<tr>
<td>Hsin-Hsi CHEN</td>
<td>National Taiwan University, Taipei, Taiwan</td>
</tr>
<tr>
<td>Keh-Jiann CHEN</td>
<td>Research Fellow, Institute of Information Science, Academia Sinica, Taipei, Taiwan</td>
</tr>
<tr>
<td>Key-Sun CHOI</td>
<td>KAIST Korea Terminology Research Center for Language and Knowledge Engineering, Daejeon, Korea</td>
</tr>
<tr>
<td>Zhiwei FENG</td>
<td>Institute of Applied Linguistics, Beijing, China</td>
</tr>
<tr>
<td>Yasunari HARADA</td>
<td>Waseda University, Tokyo, Japan</td>
</tr>
<tr>
<td>One-Soon Her</td>
<td>National Chengchi University, Taipei, Taiwan</td>
</tr>
<tr>
<td>Kaoru HORIE</td>
<td>Tohoku University, Japan</td>
</tr>
<tr>
<td>Kiyoshi, ISHIKAWA</td>
<td>Hosei University, Tokyo, Japan</td>
</tr>
<tr>
<td>Beom-mo KANG</td>
<td>Korea University, Korea</td>
</tr>
<tr>
<td>Sue-Jin, KER</td>
<td>Soochow University, Taipei, Taiwan</td>
</tr>
<tr>
<td>Jong-Bok KIM</td>
<td>Kyung Hee University, Seoul, Korea</td>
</tr>
<tr>
<td>Tom Bong Yeung LAI</td>
<td>City University of Hong Kong, Hong Kong</td>
</tr>
<tr>
<td>Chungmin LEE</td>
<td>Seoul National University, Seoul, Korea</td>
</tr>
<tr>
<td>Kiyong LEE</td>
<td>Korea University, Seoul, Korea</td>
</tr>
<tr>
<td>Wei LI</td>
<td>Cymfony Inc., Buffalo, USA</td>
</tr>
<tr>
<td>Robert Wing Pong LUK</td>
<td>Hong Kong Polytechnic University, Hong Kong, China</td>
</tr>
<tr>
<td>Kang Kwong LUKE</td>
<td>University of Hong Kong, Hong Kong</td>
</tr>
<tr>
<td>Yuji MATSUMOTO</td>
<td>Nara Institute of Science and Technology, Nara, Japan</td>
</tr>
<tr>
<td>Tetsuharu MORIYA</td>
<td>Kanazawa University, Kanazawa, Japan</td>
</tr>
<tr>
<td>Jian-Yun NIE</td>
<td>University of Montreal, Montreal, Canada</td>
</tr>
<tr>
<td>Fuji Ren</td>
<td>The University of Tokushima, Tokushima, Japan</td>
</tr>
<tr>
<td>Name</td>
<td>Institution</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Richard SPROAT</td>
<td>AT&amp;T Labs -- Research, Florham Park, NJ, USA</td>
</tr>
<tr>
<td>Lily I-wen SU</td>
<td>National Taiwan University, Taipei, Taiwan</td>
</tr>
<tr>
<td>Maosong SUN</td>
<td>Tsinghua University, Beijing, China</td>
</tr>
<tr>
<td>Shu-Chuan TSENG</td>
<td>Institute of Linguistics, Academia Sinica, Taipei, Taiwan</td>
</tr>
<tr>
<td>Jie XU</td>
<td>National University of Singapore, Singapore</td>
</tr>
<tr>
<td>Jianqi WANG</td>
<td>Ohio State University, Columbus, USA</td>
</tr>
<tr>
<td>Chung-Hsien WU</td>
<td>National Cheng Kung University, Tainan, Taiwan</td>
</tr>
<tr>
<td>Shiwen YU</td>
<td>Peking University, Beijing, China</td>
</tr>
</tbody>
</table>
MESSAGE FROM PACLIC 17 & O-COCOSDA 2003 PROGRAM COMMITTEE

It is with great pleasure that we welcome you to Singapore for PACLIC 17 and Oriental COCOSDA 2003. It provides an excellent forum for scientists and engineers from various research areas, including linguistics, natural language processing, and speech processing, etc. to meet and to exchange ideas, to share information and to discuss regional matters on any related research issues.

The call-for-papers this year was met with such enthusiasm that we are able to prepare a technical programme that comprises a large number of high quality papers. For PACLIC 17, we received more than 80 submissions from 13 countries or regions: China, Hong Kong, Taiwan, Japan, Korea, Singapore, Thailand, Vietnam, Mongolia, United states, United Kingdom, France and Australia. For O-COCOSDA, we received over 50 submissions from 9 countries or regions: China, Hong Kong, India, Indonesia, Japan, Korea, Singapore, Thailand and Taiwan. In addition to the 47 papers for PACLIC (36 for oral presentation and 11 for poster) and 39 papers for COCOSDA accepted for presentation, we are honoured to have 4 keynotes speeches given by academic and industrial leaders in the field.

The excellent program offered by PACLIC 17 and Oriental COCOSDA 2003 is a result of the excellent review efforts undertaken by the two program committees and the local organizing committee here in Singapore. We would like to thank the steering committees for their guidance, the program committees for their great reviews, the organizing committees for their hard working and many people who contributed to PACLIC 17 and Oriental COCOSDA 2003 in one way or another. Especially, we would like to thank all the authors who submitted their papers.

On behalf of the program committees, we thank you very much for your tremendous supports in making today's event a success. October is the best to visit Singapore. With the intensive technical program, the open atmosphere and the multi-cultural surroundings, we hope that you can take this excellent opportunity to meet old friends and to make new ones. After SARS, the Lion City is roaring again, we are sure that you will enjoy the city as well as the conference.

Ji Donghong
PACLIC 17 program Committee Chair

LI Haizhou
Oriental COCOSDA Program Committee Chair
# Table of Contents

## Invited Speeches

Virtual Linked Lexical Knowledge Base for Causality Reasoning  
*Choi Key-Sun*  
1

Focus-Marking in Chinese and Malay: A Comparative Perspective  
*Jie Xu*  
2

## Paper Presentation

### Linguistics-1

A Constraint-Based Analysis of Association With Focus in Japanese  
*Yusuke Kubota*  
16

On the Event Structure of Indirect Passives in Japanese  
*Naoyuki Ono*  
28

An Event-Based Interpretation of Japanese Honorific Constructions Using RRG Operators  
*Akira Ishikawa*  
38

Multiple Nominative Constructions in Japanese and Their Theoretical Implications  
*Masahiro Akiyama*  
50

### NLP-1

Automatic Learning of Stemming Rules for the Indonesian Language  
*Lily Suryana Indradjaja and Stephane Bressan*  
62

A Simplified Latent Semantic Indexing Approach for Multi-Linguistic Information Retrieval  
*Liu Yi, Lu Haiming, Lu Zengxiang and Wang Pu*  
69
An integrated approach for Chinese word segmentation  
*Guohong Fu and K.K. Luke*  
80

Korean Phrase Structure Grammar and Its Implementations into the LKB system  
*Jong-Bok Kim and Jaehyung Yang*  
88

Porting Grammars between Typologically Similar Languages: Japanese to Korean  
*Roger Kim, Mary Dalrymple, Ronald M. Kaplan and Tracy Holloway King*  
98

**Linguistics-2**

Mandarin Chinese *Shenme* in Interaction  
*Fuhui Hsieh*  
106

The Semantics of Shapes: A Study based on Mandarin Quan1zi5 (圈子)  
*Cui-xia Weng and Chu-Ren Huang*  
115

Stock Markets as Ocean Water: A Corpus-based, Comparative Study of Mandarin Chinese, English and Spanish  
*Siaw-Fong Chung, Kathleen Ahrens and Ya-hui Sung*  
124

Subject Positions and Derivational Scope Calculation in Minimalist Syntax: A Phase-Based approach  
*Yukiko Ueda*  
134

**NLP-2**

Context-rule Model for POS tagging  
*Yu-Fang Tsai and Keh-Jiann Chen*  
146

Chinese Word Segmentation based on Contextual Entropy  
*Jin Hu Huang and David Powers*  
152

Topic Segmentation for Short Texts  
*Tao-Hsing Chang and Chia-Honag Lee*  
159

Cross-Lingual Text Filtering Based on Text Concepts and KNN  
*Li Shaozi, Su Weifeng, Li Tangqiu and Chen Huowang*  
166
Linguistics-3

The Semantics of Onomatopoeic Speech Act Verbs
I-Ni Tsai and Chu-Ren Huang

Mandarin Adverbial Jiu in Discourse
Fuhui Hsieh

A Synchronous Corpus-Based Study of Verb-Noun Fluidity in Chinese
Oi Yee Kwong and Benjamin K. Tsou

Non-Monotonic Negativity
Sumiyo Nishiguchi

NLP-3

The SVM With Uneven Margins and Chinese Document Categorization
Yaoyong Li and John Shawe-Taylor

The development of Tagged Uyghur Corpus
Yusup Aibaidula and Kim-Teng Lua

A Vector-Based Algorithm for Chinese Text Classification
Luo Changri and He Tingting

A Large-scale Lexical Semantic Knowledge-Base of Chinese
Wang Hui and Yu Shiwen

NLP-4

An Effective Combination of Different Order N-grams
Sen Zhang and Na Dong

Efficient Methods for Multigram Compound Discovery
Wu Horng Jyh Paul, Ng Hong I and Gong Ruibin

Translation Template Learning Based on Hidden Markov Modeling
Nguyen Minh Le, Akari Shimazu and Susumu Horiguchi

News-Oriented Keyword Indexing with Maximum Entropy Principle
Li Sujian, Wang Houfeng, Yu Shiwen and Xin Chengsheng
NLP-5

Extracting Chinese Multi-Word Units from Large-Scale Balanced Corpus

Liu Jianzhou, He Tingting and Liu Xiaohua

A New Sentence Reduction based on Decision Tree Model

Nguyen Minh Le and Susumu Horiguchi

Japanese Parser on the basis of the Lexical-FG Formalism and its Evaluation

Hiroshi Masuichi, Tomoko Ohkuma, Hiroki Yoshimura and Yasunari Harada

A Statistical Approach to Chinese-to-English Back-Transliteration

Chun-Jen Lee, Jason S. Chang and Jyh-Shing Roger Jang

On the Sentence Category Transfer of Action-effect Sentences in Chinese-English Machine Translation

Keliang Zhang

Linguistics-4

Modeling verb Order in Complex Multi-Verbal Predicate Constructions

Olivia S.-C. Lam and Adams B. Bodomo

The Structure of Spatial Expressions in Saisiyat

Pei-Shu Tsai

A Constraint-Based Grammar of Case: To Correctly Predict Case Phrases Occurring without Their Head Verb

Hiroki Koga

The Floating of Negative Factors and the Recognition of Semantic Patterns of Huaiyi Sentences in Mandrain

Xiao Guozheng and Guo Tingting

Poster

Using Mutual Information to Identify New Features for Text documents of Various Domains

Guo Zhi Li

Applicability Analysis of Corpus-derived Paraphrases toward Example-based Paraphrasing

Kiyonori Ohtake and Kazuhide Yamamoto
A Word Selection Model Based on Lexical Semantic Knowledge in English Generation
Chen Yi-Dong, Li Tang-Qiu and Zheng Xu-Ling

392

Corpus-Based Ontology Learning for Word Sense Disambiguation
Sin Jae Kang

399

On Intra-page and Inter-page Semantic Analysis of Web Pages
Jun Wang, Jicheng Wang, Gangshan Wu and Hiroshi Tsuda

408

Towards a Multi-Objective Corpus for Vietnamese Language
Vu Hai Quan, Pham Nam Trung, Nguyen Duc Hoang Ha, Huynh Bao Toan, Le Hoai Bac and Hoang Kiem

416

Using Zero Anaphora Resolution to Improve Text Categorization
Ching-Long Yeh and Yi-Chun Chen

423

Dependency of Long-Distance Reflexives
Hyeran Lee

431

Voicing Constraint and Segmental-Tonal Neighborhood Effects on Clusters in Thai
Rattima Nitisaroj

441

The treatment of Japanese Focus Particles based on Lexical-Functional Grammar
Tomoko Okkuma, Hiroshi Masuchi, Hiroki Yoshimura and Yasunari Harada

448

Empty Category and the effect of Teaching in Sentence Processing
Takako Kawasaki and Kiyoshi Ishikawa

456