Proceedings of the
26th Pacific Asia Conference on
Language, Information and Computation
(PACLIC 26)

7 - 10 November 2012
Bali,Indonesia

© 2012 The PACLIC 26 Organizing Committee and PACLIC Steering Committee

All rights reserved. Except as otherwise expressly permitted under copyright law, no part of this publication may be reproduced, digitized, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, Internet or otherwise, without the prior permission of the publisher.

Copyright of contributed papers reserved by respective authors

ISBN: 978-979-1421-17-1
Published by Faculty of Computer Science, Universitas Indonesia

Welcome Message from Honorary Chairs

On behalf of the Organizing Committee of the 26th Pacific Asia Conference on Language, Information and Computation (PACLIC 26), we would like to extend our warm welcome to all of the participants and speakers, and in particular, we would like to express our sincere gratitude to our invited speakers.

This international conference is organized by the Faculty of Computer Science, Universitas Indonesia and is supported by the I-MHERE DIKTI project. We are very keen to host a conference about language processing fields which involves many researchers in this Asia Pacific region. We believe that this international conference will open up the opportunities for sharing and exchanging original research ideas and opinions, getting inspiration for future research, and broadening knowledge about various new topics and approaches in language study. We hope that in this conference, the attendees would have the opportunity to meet with new people and discuss the opportunity to collaborate together.

We chose to organize PACLIC 26 in Bali so that aside from attending this interesting conference, you can also enjoy the scenery and the culture of Bali. We realize that there might not be enough time to see all the nice places in Bali, but we hope that you can bring home some good memories.

We would like to express our sincere appreciation to the members of the Program Commitee for a fruitful reviews of the submitted papers, as well as the Organizing Commitee for the time and energy they have devoted to editing the proceedings and arranging the logistics of holding this conference. We would like to give an appreciation to the authors who have submitted their excellent works to this conference. Last but not least, we would like to extend our gratitude to the Ministry of Education and Culture of the Republic of Indonesia and the Dean of the Faculty of Computer Science at Universitas Indonesia for their continued support towards the PACLIC 26 conference.

Have a nice time in Bali and enjoy the conference.

Honorary Chairs:

Mirna Adriani (Universitas Indonesia)
I Wayan Arka (ANU / Universitas Udayana)

Welcome Message from Program Co-Chairs

Welcome to Bali! This is the first time that the PACLIC conference is being held in Indonesia, and we are very excited about this fact. By all accounts, Indonesia is a linguistic treasure trove, with over 700 living languages today according to the Ethnologue report. Moreover, with an increasing number of its 240 million population active on the Internet via the Web and social networks, clearly these are exciting times to be engaging in computational approaches towards the languages of Indonesia.

However, this PACLIC conference in 2012 is special for other reasons, most notably the commemoration of 25 years of the conference series. Over the years, the conference has developed into one of the leading conferences in the fields of theoretical and computational linguistics, extending beyond the Asia-Pacific region. This year, the specific research topics that the papers focus on can be classified into the following: discourse & pragmatics, grammar & syntax, information extraction, information retrieval, lexical semantics, machine translation, parsing, sentiment analysis, text summarization & paraphrasing, and word sense disambiguation & distributional semantics. Moreover, there is an interesting mix of both theoretical and computational approaches to almost all of the aforementioned topics.

We received paper submissions representing immense diversity, with authors from 29 countries or regions, namely Australia, Bahrain, Belgium, Canada, China, Czech Republic, Denmark, Egypt, France, Germany, Hong Kong, India, Indonesia, Japan, Korea, Macau, Malaysia, Pakistan, Philippines, Qatar, Singapore, Slovakia, Sri Lanka, Taiwan, Thailand, Tunisia, United Kingdom, United States, and Vietnam. To ensure that all accepted papers met the high quality standard of the PACLIC conference, all papers were sent to three reviewers. Of the 117 submissions that we received, 39 papers (33%) were accepted for oral presentation, and another 18 papers (15%) were accepted for poster presentation. We believe this has yielded an interesting, diverse, and high-quality collection of papers, and are confident that the conference will be successful as a result.

A successful conference is the result of many peoples efforts and contributions. Aside from the efforts of the authors who will be presenting their current work, thanks must be given to the tremendous efforts made by the program committee members in their paper reviews. Besides the oral and poster paper presentations, the conference is enriched by several invited speakers. Firstly there is a Special Session commemorating 25 years of PACLIC, which brings together Prof Kiyong Lee from Korea University, Prof Yuji Matsumoto from the Nara Institute of Science and Technology, and Prof Benjamin T'sou from the Hong Kong Institute of Education, three figures who have been instrumental in the formation of the PACLIC tradition. We have also scheduled invited talks from Prof I Wayan Arka from ANU & Universitas Udayana and Prof Tim Baldwin from the University of Melbourne. The expertise in the respective fields of all five speakers will undoubtedly provide us with new insights for research. On behalf of the program committee, we express our heartfelt thanks to them all. We would also like to thank the steering committee for their guidance, and the local organizing committee at Universitas Indonesia for their dedicated efforts and their excellent coordination with all parties, which has ensured that this conference will be a successful event.

Finally, we wish that you will all enjoy the conference presentations and resulting discussions between old and new friends, and also have some time to enjoy the wondrous setting that is the island of Bali.

Program Co-Chairs:
Ruli Manurung (Universitas Indonesia)
Francis Bond (Nanyang Technological University)

PACLIC 26 Organizers

Steering Committee:

Jae-Woong Choe, Korea University
Yasunari Harada, Waseda University
Chu-Ren Huang, Hong Kong Polytechnic University
Rachel Edita Roxas, De La Salle University-Manila
Maosong Sun, Tsinghua University
Benjamin T'sou, City University of Hong Kong
Min Zhang, Institute for Infocomm Research

Honorary Chairs:

Mirna Adriani, Universitas Indonesia
I Wayan Arka, ANU/Universitas Udayana

Program Committee:

Co-Chairs:

Ruli Manurung, Universitas Indonesia
Francis Bond, Nanyang Technological University
Shu-Kai Hsieh, National Taiwan University
Donghong Ji, Wuhan University
Olivia Kwong, City University of Hong Kong
Seungho Nam, Seoul National University
Ryo Otoguro, Waseda University
Rachel Edita Roxas, De La Salle University-Manila

Members:

Wirote Aroonmanakun, Chulalongkorn University
Timothy Baldwin, University of Melbourne
Stephane Bressan, National University of Singapore
Hee-Rahk Chae, Hankuk University of Foreign Studies
Hsin-Hsi Chen, National Taiwan University
Eng Siong Chng, Nanyang Technological University
Siaw-Fong Chung, National Chengchi University
Beatrice Daille, University of Nantes
Mary Dalrymple, Oxford University
Danilo Dayag, De La Salle University

Minghui Dong, Institute for Infocomm Research
Rebecca Dridan, University of Oslo
Maria Flouraki, SOAS, University London
Guohong Fu, Heilongjiang University
Wei Gao, Chinese University of Hong Kong
Yasunari Harada, Waseda University
Munpyo Hong, Sungkyunkwan University
Shu-Kai Hsieh, National Taiwan Normal University
Xuanjing Huang, Fudan University
Kentoro Inui, Nara Institute of Science and Technology
Donghong Ji, Wuhan University
Nikiforos Karamanis, TouchType
Jong-Bok Kim, Kyung Hee University
Valia Kordoni, Saarland University
Sadao Kurohashi, Kyoto University
Oi Yee Kwong, City University of Hong Kong
Bong Yeung Tom Lai, City University of Hong Kong
Paul Law, City University of Hong Kong
Alessandro Lenci, University of Pisa
Gina-Anne Levow, University of Manchester
Haizhou Li, Institute for Infocomm Research
Qun Liu, Chinese Academy of Sciences
Qing Ma, Ryukoku University
Yanjun Ma, Baidu
Takafumi Maekawa, Hokusei Gakuen University Junior College
Yuji Matsumoto, Nara Institute of Science and Technology
Mathieu Morey, Universite d'Aix-Marseille & Nanyang Technological University
Yoshiki Mori, University of Tokyo
Seungho Nam, Seoul National University
Vincent Ng, University of Texas at Dallas
Jian-Yun Nie, Universite de Montreal
Toshiyuki Ogihara, University of Washington
David Yoshikazu Oshima, Nagoya University
Ryo Otoguro, Waseda University
Ceile Paris, Commonwealth Scientific and Industrial Research Organisation
Jong C. Park, Korea Advanced Institute of Science and Technology
Laurent Prevot, Universite de Provence
Long Qiu, Institute for Infocomm Research
Bali Ranaivo-Malancon, Universiti Malaysia Sarawak
Graeme Ritchie, University of Aberdeen
Rachel Edita Roxas, De La Salle University-Manila
Samira Shaikh, State University of New York - University at Albany
Sachiko Shudo, Waseda University
Melanie Siegel, Hochschule Darmstadt
Pornsiri Singhapreecha, Thammasat University
Virach Sornlertlamvanich, Thai Computational Linguistics Laboratory, NICT

Andrew Spencer, University of Essex
Jian Su, Institute for Infocomm Research
I-Wen Su, National Taiwan University
Keh-Yih Su, Behavior Design Corporation
Henry S. Thompson, University of Edinburgh
Takenobu Tokunaga, Tokyo Institute of Technology
Josef van Genabith, Dublin City University
Aline Villavicencio, Universidade Federal do Rio Grande do Sul
Haifeng Wang, Baidu
Houfeng Wang, Peking University
Hui Wang, National University of Singapore
Jiun-Shiung Wu, National Chiayi University
Jae II Yeom, Hongik University
Satoru Yokoyama, Tohoku University
Min Zhang, Institute for Infocomm Research
Qiang Zhou, Tsinghua University
Michael Zock, Laboratoire d'Informatique Fondamentale de Marseille, C.N.R.S.
Chengqing Zong, Chinese Academy of Sciences

Local Organizing Committee:

Bayu Distiawan, Universitas Indonesia
Muhammad Hilman, Universitas Indonesia
Samuel Louvan, Universitas Indonesia
Lelya Rimadhiana, Universitas Indonesia
Clara Vania, Universitas Indonesia

The First Workshop on Generative Lexicon for Asian Languages (GLAL)

Organizers:

Shu-Kai Hsieh (Institute of Linguistics, National Taiwan University)
Zuoyan Song (School of Chinese Language and Literature, Beijing Normal University)
Kyoko Kanzaki (National Institute for Japanese Language and Linguistics)

Program Committee:

Toni Badia(Universitat Pompeu Fabra, spain)
Christian BASSAC (Universit de Lyon2, France)
Pierrette Bouillon (ETI/TIM/ISSCO, Switzerland)
Nicoletta Calzolari (CNR-ILC, Italy)
Ann Copestake (University of Cambridge, UK)
Christiane Fellbaum (Princeton University, USA)

Catherine Havasi(MIT,USA)
Chu-Ren Huang (The Hongkong Polytechnic University, China)
Hitoshi Isahara (NICT, Kyoto, Japan)
Chungmin Lee(Seoul National University, Seoul, Korea)
Alessandro Lenci (Universita di Pisa, Pisa, Italy)
Kentaro Nakatani(Kounan University, Japan)
Seungho Nam (Seoul National University, Seoul, Korea)
Fiammetta Namer (ATILF-CNRS, University of Nancy, France)
Naoyuki Ono (Tohoku University, Sendai, Japan)
Laurent Prvot (Aix-Marseille Universit & CNRS, France)
James Pustejovsky (Brandeis University, USA)
Anna Rumshisky (Brandeis University, USA)
Patrick Saint-Dizier (CNRS, Toulouse, France)
Koichi Takeuchi(Okayama University, Japan)
Hongjun Wang (Peking University, Beijing, China)
Nianwen Xue (Brandeis University, Waltham, MA USA)
Yulin Yuan (Peking University, Beijing, China)
Seohyun Im (Brandeis University, USA)

Table of Contents

1. Invited Talks

From All Possible Worlds to Small Worlds: A Story of How We Started and Where We Will Go Doing Semantics
Kiyong Lee 1

Developing a Deep Grammar of Indonesian within the ParGram Framework: Theoretical and Implementational Challenges
I Wayan Arka 19

Idiomaticity and Classical Traditions in Some East Asian Languages
Benjamin K Tsou 39

Things between Lexicon and Grammar
Yuji Matsumoto 56

Social Media: Friend or Foe of Natural Language Processing?
Timothy Baldwin 58

2. Regular Papers

Towards a Semantic Annotation of English Television News - Building and Evaluating a Constraint Grammar FrameNet
Eckhard Bick 60

Compositionality of NN Compounds: A Case Study on [N1+Artifactual-Type Event Nouns
Shan Wang, Chu-Ren Huang and Hongzhi Xu 70

Automatic Domain Adaptation for Word Sense Disambiguation Based on Comparison of Multiple Classifiers
Kanako Komiya and Manabu Okumura 80

Calculating Selectional Preferences of Transitive Verbs in Korean
Sanghoun Song and Jae-Woong Choe 89

Extracting and Visualizing Semantic Relationships from Chinese Biomedical Text
Qingliang Miao, Shu Zhang, Bo Zhang and Hao Yu 99

Entity Set Expansion using Interactive Topic Information
Kugatsu Sadamitsu Sadamitsu, Kuniko Saito, Kenji Imamura and Yoshihiro Matsuo 108

Improving Chinese-to-Japanese Patent Translation Using English as Pivot Language
Xianhua Li, Yao Meng and Yao Meng 117

Combining Social Cognitive Theories with Linguistic Features for Multi-genre Sentiment Analysis
Hao Li, Yu Chen, Heng Ji, Smaranda Muresan and Dequan Zheng 127

Indonesian Dependency Treebank: Annotation and Parsing
Nathan Green, Septina Dian Larasati and Zdenek Zabokrtsky 137

Handling Indonesian Clitics: A Dataset Comparison for an Indonesian-English Statistical Machine Translation System
Septina Dian Larasati 146

Two Types of Nominalization in Japanese as an Outcome of Semantic Tree Growth
Tohru Seraku 153

Semantic Distributions of the Color Terms, Black and White in Taiwanese Languages
Huei-ling Lai and Shu-chen Lu 163

Language Independent Sentence-Level Subjectivity Analysis with Feature Selection
Aditya Mogadala and Vasudeva Varma 171

Annotation Scheme for Constructing Sentiment Corpus in Korean
Hyopil Shin, Munhyong Kim, Hayeon Jang and Andrew Cattle 181

Lexical Gaps and Lexicalization: Implications for Word Segmentation Systems for Chinese NLP
Chan-Chia Hsu 191

Extracting Keywords from Multi-party Live Chats
Su Nam Kim and Timothy Baldwin 199

Extracting Networks of People and Places from Literary Texts
John Lee and Chak Yan Yeung 209

Pre- vs. Post-verbal Asymmetries and the Syntax of Korean RDC
Daeho Chung 219

Pattern Matching Refinements to Dictionary-Based Code-Switching Point Detection
Nathaniel Oco and Rachel Edita Roxas 229

An Adaptive Method for Organization Name Disambiguation with Feature Reinforcing
Shu Zhang, Jianwei Wu, Dequan Zheng, Yao Meng and Hao Yu 237

Predicting Answer Location Using Shallow Semantic Analogical Reasoning in a Factoid Question Answering System
Hapnes Toba, Mirna Adriani and Hisar Maruli Manurung 246

On the Alleged Condition on the Base Verb of the Indirect Passive in Japanese
Tomokazu Takehisa 254

Comparing Classifier use in Chinese and Japanese
Yue Hui Ting and Francis Bond 264

Nominative-marked Phrases in Japanese Tough Constructions
Akira Ohtani and Maria del Pilar Valverde Ibanez 272

Emotional Tendency Identification for Micro-blog Topics Based on Multiple Characteristics
Quanchao Liu, Chong Feng and Heyan Huang 280

Product Name Classification for Product Instance Distinction
Hye-Jin Min and Jong C. Park 289

Automatic Detection of Gender and Number Agreement Errors in Spanish Texts Written by Japanese Learners
Maria del Pilar Valverde Ibanez and Akira Ohtani 299

A Reranking Approach for Dependency Parsing with Variable-sized Subtree Features
Mo Shen, Daisuke Kawahara and Sadao Kurohashi 308

Applying Statistical Post-Editing to English-to-Korean Rule-based Machine Translation System
Ki-Young Lee and Young-Gil Kim 318

A Model of Vietnamese Person Named Entity Question Answering System
Mai-Vu Tran, Duc-Trong Le, Xuan- Tu Tran and Tien-Tung Nguyen 325

Towards a Semantic Annotation of English Television News - Building and Evaluating a Constraint Grammar FrameNet
Shaohua Yang, Hai Zhao and Bao-liang Lu 333

Emotion Estimation from Sentence Using Relation between Japanese Slangs and Emotion Expressions
Kazuyuki Matsumoto, Kenji Kita and Fuji Ren 343

Can Word Segmentation be Considered Harmful for Statistical Machine Translation Tasks between Japanese and Chinese?
Jing Sun and Yves Lepage 351

Introduction of a Probabilistic Language Model to Non-Factoid Question Answering Using Example Q&A Pairs
Kosuke Yoshida, Taro Ueda, Madoka Ishioroshi, Hideyuki Shibuki and Tatsunori Mori 361

Answering Questions Requiring Cross-passage Evidence
Kisuh Ahn and Hee-Rahk Chae 371

Thai Sentence Paraphrasing from the Lexical Resource
Krittaporn Phucharasupa and Ponrudee Netisopakul 381

Anaphora Annotation in Hindi Dependency TreeBank
Praveen Dakwale, Himanshu Sharma and Dipti M Sharma 391

Improving Statistical Machine Translation with Processing Shallow Parsing
Hoai-Thu Vuong, Vinh Van Nguyen, Viet Hong Tran and Akira Shimazu 401

Psycholinguistics, Lexicography, and Word Sense Disambiguation
Oi Yee Kwong 408

Thought De se, first person indexicals and Chinese reflexive ziji
Yingying Wang and Haihua Pan 418

The Headedness of Mandarin Chinese Serial Verb Constructions: A Corpus-Based Study
Jingxia Lin, Chu-Ren Huang, Huarui Zhang and Hongzhi Xu 428

Japanese Pseudo-NPI Dare-mo as an Unrestricted Universal Quantifier
Katsuhiko Yabushita 436

Automatic Tripartite Classification of Intransitive Verbs
Nitesh Surtani and Soma Paul 446

The Transliteration from Alphabet Queries to Japanese Product Names
Rieko Tsuji, Yoshinori Nemoto, Wimvipa Luangpiensamut, Yuji Abe, Takeshi Kimura, Kanako Komiya, Koji Fujimoto and Yoshiyuki Kotani 456

Classifying Dialogue Acts in Multi-party Live Chats
Su Nam Kim, Lawrence Cavedon and Timothy Baldwin 463

Syntax-semantics mapping of locative arguments
Seungho Nam 473

Deep Lexical Acquisition of Type Properties in Low-resource Languages: A Case Study in Wambaya
Jeremy Nicholson, Rachel Nordlinger and Timothy Baldwin 481

Chinese Sentiments on the Clouds: A Preliminary Experiment on Corpus Processing and Exploration on Cloud Service
Shu-Kai Hsieh, Yu-Yun Chang and Meng-Xian Shih 491

Cross-Lingual Topic Alignment in Time Series Japanese / Chinese News
Shuo Hu, Yusuke Takahashi, Liyi Zheng, Takehito Utsuro, Masaharu Yoshioka, Noriko Kando, Tomohiro Fukuhara, Hiroshi Nakagawa and Yoji Kiyota 498

A CRF Sequence Labeling Approach to Chinese Punctuation Prediction
Yanqing Zhao, Chaoyue Wang and Guohong Fu 508

Analysis of Social and Expressive Factors of Requests by Methods of Text Mining
Dasa Munkova, Michal Munk, Zuzana Fraterova and Beata Durackova 515

Set Expansion using Sibling Relations between Semantic Categories
Sho Takase, Naoaki Okazaki and Kentaro Inui 525

Building a Diverse Document Leads Corpus Annotated with Semantic Relations
Masatsugu Hangyo, Daisuke Kawahara and Sadao Kurohashi 535

Text Readability Classification of Textbooks of a Low-Resource Language
Zahurul Islam, Alexander Mehler and Rashedur Rahman 545

Hybrid Approach for the Interpretation of Nominal Compounds using Ontology
Sruti Rallapalli and Soma Paul 554

Improved Constituent Context Model with Features
Yun Huang, Min Zhang and Chew Lim Tan 564

Accuracy and robustness in measuring the lexical similarity of semantic role fillers for automatic semantic MT evaluation
Anand Karthik Tumuluru, Chi-kiu Lo and Dekai Wu 574

3. The First Workshop on Generative Lexicon for Asian Languages

Type Construction of Event Nouns in Mandarin Chinese
Shan Wang and Chu-Ren Huang 582

On Interpretation of Resultative Phrases in Japanese
Tsuneko Nakazawa 592

Event Coercion of Mandarin Chinese Temporal Connective hou after
Zuoyan Song 602

To Construct the Interpretation Templates for the Chinese Noun Compounds Based on Semantic Classes and Qualia Structures
Xue Wei and Yulin Yuan 609

Compositional Mechanisms of Japanese Numeral Classifiers
Miho Mano 620

Psych-Predicates: How They Are Different
Chungmin Lee 626

The Role of Qualia Structure in Mandarin Children Acquiring Noun-modifying Constructions
Zhaojing Liu and Angel Wing-shan Chan 632

Gap in Gapless Relative Clauses in Korean and Other Asian Languages
Jeong-Shik Lee and Chungmin Lee 640

Invited Talk 1

All Possible Worlds to Small Worlds: A Story of How We Started and Where We Will Go Doing Semantics

Kiyong Lee, Korea University Seoul

Bio

Kiyong Lee is Professor emeritus of linguistics, Korea University, Seoul. He has been convenor of an ISO working group for the development of semantic annotation schemes since June 2004. He was invited as Visiting Professor to Department of Korean, Tenri University, Nara, Japan, in 1999-2000 and also as Visiting Professor to the Department of Chinese, Translation and Linguistics, City University of Hong Kong, on three different occasions. He was a keynote speaker on formal semantics at the 18th Congress of Linguists (July 21-26, 2008) in Seoul. He was awarded a prize for academic excellence from the National Academy of Sciences, Korea, on the basis of a three-volume book on Semantics: Formal, Possible Worlds, and Situation Semantics, and also a book award for his Computational Morphology from the Ministry of Culture and Tourism, Korea, in 2002. Since he graduated with an A.B. degree from Saint Louis University, St. Louis, MO, USA, in 1963, Kiyong Lee has taught Latin, English, Philosophy, and Linguistics at four different universities full-time and at over 20 universities part-time. As a Fulbright student, he also received a Ph.D. in Linguistics from the University of Texas, Austin, TX, USA, in 1974 and did research work as a Fulbright scholar at CSLI, Stanford University, Palo Alto, CA, USA, and as a DAAD scholar at the Computational Linguistics Lab, University of Erlangen, Germany. Kiyong Lee has been president of the Linguistics Society of Korea (1990-1992) and that of the Korean Society of Cognitive Science (1989-1990). He was also one of the founding members of the Korean Society for Language and Information and the first representative of its precursor, named the Seoul Workshop on Formal Grammar Theory. He has thus helped organize and host several PACLICs in Korea and abroad since its inception in December 1981.

Invited Talk 2

All Possible Worlds to Small Worlds: A Story of How We Started and Where We Will Go Doing Semantics

Yuji Matsumoto

Bio

Yuji Matsumoto is now a professor of Computational Linguistics in the Graduate School of Information Science, Nara Institute of Science and Technology. He got his PhD degree from Kyoto University in 1990. He has experienced a researcher at Electrotechnical Laboratory, a deputy chief of the first laboratory at New Generation Computer Technology Research Center, an Associate professor at Kyoto University, before getting the current position. He is now the Vice-President of the Asian Federation of Natural Language Processing, and the President of ACL SIGDAT, and a Advisory Board member of ACL SIGNLL. He is a Fellow of Information Processing Society of Japan, and the Association for Computational Linguistics.

Invited Talk 3

Developing a Deep Grammar of Indonesian within the ParGram Framework: theoretical and implementational challenges

I Wayan Arka, Australian National University/Udayana University

Bio

I Wayan Arka is affiliated with the Australian National University (as a Fellow in Linguistics at School of Culture, History and Language, College of Asia and the Pacific) and Udayana University Bali (English Department and Graduate Program in Linguistics). His interests are in descriptive, theoretical and typological aspects of Austronesian and Papuan languages of Indonesia. Wayan is currently working on a number of projects: NSF-funded research on voice in the Austronesian languages of eastern Indonesia (2008-2011), ARC-funded projects for the development of computational grammar for Indonesian (2008-2011) and the Languages of Southern New Guinea (2011-2014).

Invited Talk 4

Idiomaticity and Classical Traditions in Some East Asian Languages
Benjamin Tsou, The Hong Kong Institute of Education

Bio

Benjamin Tsou has been doing research on corpus linguistics and sociolinguistics via the on-going Linguistic Variation in Chinese Speech Communities project (http://livac.org) which focuses on the characteristics and evolving use of Chinese media language in Beijing, Hong Kong, Macau, Shanghai, Singapore and Taipei, involving the sophisticated processing and analysis of more than 450 million Chinese characters since 1995. His group has been tracking new and different neologistic developments as well as underlying sociolinguistic changes, and has also worked on the alignment and comparison of English-Chinese bilingual texts in the legal and technical domains. His research on the Language Atlas of China and his textbook on sociolinguistics have won awards from the Chinese Academy of Social Sciences and the Chinese Ministry of Education respectively.

Professor Tsou is the Chiang Chen Chair Professor of Linguistics and Language Sciences and the Director of the Research Centre on Linguistics and Language Information Sciences at The Hong Kong Institute of Education. He is a member of Acadmie Royale des Sciences dOutre-Mer of Belgium. He serves on the Standing Committee of the Executive Board of the Chinese Information Processing Society of China, and is the founding President of the Asian Federation of Natural Language Processing and of the Linguistic Society of Hong Kong. He publishes widely and is also a member of numerous editorial boards. Professor Tsou received his Ph.D from the University of California, Berkeley, and MA from Harvard University.

Invited Talk 5

Social Media: Friend or Foe of Natural Language Processing?
Tim Baldwin, University of Melbourne, Australia

Bio

Timothy Baldwin is an Associate Professor and Deputy Head of the Department of Computing and Information Systems, The University of Melbourne, and a contributed research staff member of the NICTA Victoria Research Laboratories. He has previously held visiting positions at the University of Washington, University of Tokyo, Saarland University, and NTT Communication Science Laboratories. His research interests cover topics including social media, deep linguistic processing, multiword expressions, computer-assisted language learning, information extraction, web mining and machine learning, with a particular interest in the interface between computational and theoretical linguistics. Current projects include web user forum mining, biomedical text mining, and intelligent interfaces for Japanese language learners. He is President of the Australasian Language Technology Association in 2011-2012. Tim completed a BSc(CS/Maths) and BA(Linguistics/Japanese) at the University of Melbourne in 1995, and an MEng(CS) and PhD(CS) at the Tokyo Institute of Technology in 1998 and 2001, respectively. Prior to commencing his current position at The University of Melbourne, he was a Senior Research Engineer at the Center for the Study of Language and Information, Stanford University (2001-2004).