

Proceedings of the
28th Pacific Asia Conference on
Language, Information and Computation
(PACLIC 28)

Edited by

Wirote Aroonmanakun, Prachya Boonkwan, and Thepchai Supnithi

12–14 December 2014
Phuket, Thailand

© The PACLIC 28 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 978-616-551-887-1

Published by the Department of Linguistics, Faculty of Arts, Chulalongkorn University.

Acknowledgments

PACLIC 28, the 28th Pacific Asia Conference on Language, Information and Computation, is organized by Department of Linguistics, Chulalongkorn University (CU), Language and Semantic Technology Laboratory, National Electronics and Computer Technology Center (NECTEC), and Sirindhorn International Institute of Technology (SIIT), under the auspices of PACLIC Steering Committee.

We gratefully acknowledge the support from the Chula Global Network, the Division of Research Development and Promotion, and the Research Affairs Division, Faculty of Arts.

Welcome Message

Distinguished scholars and colleagues,

It is my great pleasure and honour to be here today to open the 28th Pacific Asia Conference on Language, Information and Computing which is held in Thailand for the first time.

This year's PACLIC annual meeting maintains the long-standing mission of PACLIC conferences to emphasize the synergy of theoretical analysis and processing of language, and to serve as a venue for scholars working on issues pertaining to different languages in the Pacific-Asia region to share their findings and experiences. It provides wonderful opportunities for participants to build a strong academic network, be enlightened by new insights, get entertained intellectually, and return home ready to initiate new significant contribution.

For the past years since its establishment, the PACLIC conferences have gained more and more interests and participations from linguistic researchers, as evidenced by the increasing number of papers and by the wider range of topics. Likewise, the current PACLIC conference has received an overwhelming response of 151 papers from 27 countries or regions namely China, Hong Kong, Japan, Republic of Korea, Taiwan, Indonesia, Malaysia, Philippines, Thailand, Vietnam, India, Pakistan, Kazakhstan, Czech Republic, Finland, France, Germany, Ireland, Norway, Romania, Russian Federation, United Kingdom, United States, Egypt, Tunisia, New Zealand and Australia. (78.15% from 13 regions in Asia, 13.24% from 9 regions in Europe, 3.97% from the United States, 2.65% from Africa, 1.33% from New Zealand, and 0.66% from Australia). To ensure that all accepted papers meet the high quality standard of the PACLIC conference, each submission was reviewed by three reviewers. As a result, only approximately 37 % of top-notch academic papers were accepted for oral presentations and 13 % for poster sessions. From these accepted papers, 69 papers were presented and published in this proceedings.

Ladies and gentlemen, the successful conference is the result of tremendous efforts and contributions from several parties. We congratulate the Department of Linguistics, Chulalongkorn University, the National Electronics and Computer Technology Center (NECTEC) and Sirindhorn International Institute of Technology (SIIT) for their collaboration towards this significant achievement. We would like to take this opportunity to thank our keynote and invited speakers, namely Professor Mark Steedman and Professor Bonnie Webber from the University of Edinburgh, Professor Christian Matthiessen from Hong Kong Polytechnic University, Dr. Virach Sornlertlamvanich from the Technology Promotion Association (Thailand-Japan), Professor Jae-Woong Choe from Korea University, and Professor Min Zhang from Soochow University. We are overwhelmed with a sense of gratitude for the presenters and colleagues for donating your valuable time to attend and enrich this conference. We also wish to extend our sincere appreciation to the steering committee for their guidance and to the organizing

team for their essential supports and dedicated works. Last but not least, we remain profoundly grateful to the sponsors: Chula Global Network, the Division of Research Development and Promotion, and the Research Affairs Division, Faculty of Arts.

It is now time to declare the 28th Pacific Asia Conference on Language, Information and Computing open. May this conference succeed in all its aims and may it stimulate broader insights into theoretical and computational linguistics for all participants and related organizations.

Conference Chair and Co-Chair

Wirote Aroonmanakun (Chulalongkorn University)

Thepchai Supnithi (National Electronics and Computer Technology Center)

PACLIC 28 Organizers

Steering Committee: Standing Members (in alphabetic order of surnames)

Hee-Rahk Chae, Hankuk University of Foreign Studies, Seoul

Chu-Ren Huang, The Hong Kong Polytechnic University, Hong Kong

Rachel Edita O. Roxas, National University, Manila

Maosong Sun, Tsinghua University, Beijing

Benjamin T'sou, City University of Hong Kong, Hong Kong

Kei Yoshimoto, Tohoku University, Sendai

Min Zhang, Soochow University, Suzhou

Organizer:

Department of Linguistics, Chulalongkorn University

Co-Organizers:

National Electronics and Computer Technology Center

Sirindhorn International Institute of Technology

Program Committee:

Chairs:

Aroonmanakun, Wirote (Chulalongkorn University)

Chae, Hee-Rahk (Hankuk University of Foreign Studies, Seoul)

Lai, Huei-Ling (National Chengchi University)

Supnithi, Thepchai (National Electronics and Computer Technology Center)

Yoshimoto, Kei (Tohoku University)

PC Members

Bond, Francis (Nanyang Technological University)

Boonkwan, Prachya (National Electronics and Computer Technology Center)

Chen, Doris (National Taiwan Normal University)
Chen, Kuang-Hua (National Taiwan University)
Chng, Eng-Siong (Nanyang Technological University)
Chui, Kawai (National Chengchi University, Taiwan)
Daille, Beatrice (Laboratoire d'Informatique de Nantes Atlantique)
Dellwo, Volker (University of Zurich)
Dita, Shirley (De La Salle University-Manila)
Fang, Alex Chengyu (The City University of Hong Kong)
Fu, Guohong (Heilongjiang University)
Gao, Wei (Qatar Computing Research Institute)
Gao, Helena (Nanyang Technological University)
Harada, Yasunari (Waseda University)
Haruechaiyasak, Choochart (National Electronics and Computer Technology Center)
Hayashibe, Yuta (Kyoto University)
Hong, Munpyo (Sungkyunkwan University)
Hsieh, Shu-Kai (National Taiwan Normal University)
Jenks, Peter (UC Berkeley)
Ji, Donghong (Wuhan University)
Jiang, Wen Bin (Institute of Computing, CAS)
Kim, Jong-Bok (Kyung Hee University)
Kordoni, Valia (Humboldt University Berlin)
Kosawat, Krit (National Electronics and Computer Technology Center)
Kwong, Oi Yee (City University of Hong Kong)
Lai, Bong Yeung Tom (City University of Hong Kong)
Law, Paul (City University of Hong Kong)
Levow, Gina-Anne (University of Washington)
Li, Haizhou (Institute for Infocomm Research)
Liu, Jyi-Shane (National Chengchi University)
Ma, Qing (Ryukoku University)
Maekawa, Takafumi (Faculty of Sociology, Ryukoku University)
Matsumoto, Yuji (Nara Institute of Science and Technology)

Morey, Mathieu (LPL, Université d'Aix-Marseille & LMS, Nanyang Technological University)
Natpratan, Natchanan (Kasetsart University)
Netisopakul, Ponrudee (KMAKE LAB)
Ogihara, Toshiyuki (University of Washington)
Onsuwan, Chutamanee (Thammasat University)
Okada, Makoto (Osaka Prefecture University)
Otoguro, Ryo (Faculty of Law, Waseda University)
Parinyawuttichai, Tanyaporn (Chulalongkorn University)
Park, Jong C. (KAIST)
Pittayaporn, Pittayawat (Chulalongkorn University)
Pongpairoj, Nattama (Chulalongkorn University)
Prévot, Laurent (Laboratoire Parole et Langage)
Qi, Haoliang (Heilongjiang Institute of Technology)
Qiu, Long (Institute for Infocomm Research)
Ranaivo-Malançon, Bali (MALINDO)
Ratitamkul, Theeraporn (Chulalongkorn University)
Shaikh, Samira (State University of New York - University at Albany)
Shyu, Shu-Ing (National Sun Yat-sen University)
Siegel, Melanie (Hochschule Darmstadt)
Singhapreecha, Pornsiri (Thammasat University)
Smith, Simon (Coventry University)
Sornlertlamvanich, Virach (National Electronics and Computer Technology Center)
Srioutai, Jiranthara (Chulalongkorn University)
Su, Keh-Yih (Institute of Information Science, Academia Sinica)
Suchato, Atiwong (Chulalongkorn University)
Tasanawan, Soonklang (Sirindhorn International Institute of Technology)
Thepkanjana, Kingkarn (Chulalongkorn University)
Uehara, Satoshi (Tohoku University)
Van Genabith, Josef (Dublin City University)
Villavicencio, Aline (Universidade Federal do Rio Grande do Sul)
Wang, Hsu (Yuan Ze University, Taiwan)

Wijitsopon, Raksangob (Chulalongkorn University)
Wu, Jiun-Shiung (National Chung Cheng University)
Yang, Cheng-Zen (Yuan Ze University)
Yeom, Jae-II (Hongik University)
Yokoyama, Satoru (Tohoku University)
Yu, Liang-Chih (Yuan Ze University)
Zhang, Jiajun (Institute of Automation Chinese Academy of Sciences)
Zhang, Min (Institute for Infocomm Research)
Zhao, Hai (Shanghai Jiao Tong University)
Zhou, Yu (Chinese Academy of Sciences)
Zock, Michael (CNRS-LIF)

Table of Contents

Keynote Speeches

Robust Semantics for Semantic Parsing	1
<i>Mark Steedman</i>	
Social Media Understanding by Word Cloud Timeline	2
<i>Virach Sornlertlamvanich</i>	
Registerial Cartography: Context-based Mapping of Text Types and their Rhetorical-relational Organization	5
<i>Christian M.I.M. Matthiessen</i>	
Discourse for Machine Translation	27
<i>Bonnie Webber</i>	

Invited Talks

Relating Keywords to the Top Ten News of the Year in Korean Newspapers	28
<i>Jae-Woong Choe</i>	
Word Sense Induction for Machine Translation	38
<i>Min Zhang</i>	

Regular Papers

Setting Syntactic Parameters with Implicit Negative Evidence: The Case of Zero-derived Causatives in English	39
<i>Isaac Gould</i>	
Pseudo-Passives as Adjectival Passives	48
<i>Kwang-Sup Kim</i>	
Phonological Suppression of Anaphoric Wh-expressions in English and Korean	57
<i>Myung-Kwan Park</i>	
Finding The Best Model Among Representative Compositional Models	65
<i>Masayasu Muraoka, Sonse Shimaoka, Kazeto Yamamoto, Yotaro Watanabe, Naoaki Okazaki and Kentaro Inui</i>	
Semantic Frame-based Statistical Approach for Topic Detection	75
<i>Yung-Chun Chang, Yu-Lun Hsieh, Cen-Chieh Chen, Chad Liu, Chun-Hung Lu and Wen-Lian Hsu</i>	
Zero-Shot Learning of Language Models for Describing Human Actions Based on Semantic Compositionality of Actions	85
<i>Hideki Asoh and Ichiro Kobayashi</i>	

Partial Case-Marking in Japanese Stripping/Sluicing: A Dynamic Syntax Account	93
<i>Tohru Seraku</i>	
A Corpus-Based Quantitative Study of Nominalizations across Chinese and British Media English	101
<i>Ying Liu, Alex Chengyu Fang and Naixing Wei</i>	
Machine-guided Solution to Mathematical Word Problems	111
<i>Bussaba Amnueypornsakul and Suma Bhat</i>	
The So-called Person Restriction of Internal State Predicates in Japanese in Contrast with Thai	120
<i>Satoshi Uehara and Kingkarn Thepkanjana</i>	
Investigation Into Using the Unicode Standard for Primitives of Unified Han Characters.....	129
<i>Henry Larkin</i>	
Taking Antonymy Mask off in Vector Space	135
<i>Enrico Santus, Qin Lu, Alessandro Lenci and Chu-Ren Huang</i>	
Improvement of Statistical Machine Translation using Charater-Based Segmentation with Monolingual and Bilingual Information	145
<i>Vipas Sutantayawalee, Peerachet Porkaew, Prachya Boonkwa, Sitthaa Phaholphinyo, and Thepchai Supnithi</i>	
Topic-based Multi-document Summarization using Differential Evolution for Combinatorial Optimization of Sentences	152
<i>Haruka Shigematsu and Ichiro Kobayashi</i>	
Mapping between Lexical Tones and Musical Notes in Thai Pop Songs	160
<i>Chawadon Ketkaew and Pittayawat Pittayaporn</i>	
Emphasized Accent Phrase Prediction from Text for Advertisement Text-To-Speech Synthesis	170
<i>Hideharu Nakajima, Hideyuki Mizuno and Sumitaka Sakauchi</i>	
Using Tone Information in Thai Spelling Speech Recognition	178
<i>Natthawut Kertkeidkachorn, Proadpran Punyabukkana and Atiwong Suchato</i>	
Automatically Building a Corpus for Sentiment Analysis on Indonesian Tweets	185
<i>Alfan Farizki Wicaksono, Clara Vania, Bayu Distiawan and Mirna Adriani</i>	
Automatic News Source Detection in Twitter Based on Text Segmentation.....	195
<i>Takashi Inui, Masaki Saito and Mikio Yamamoto</i>	
Sentiment Lexicon Interpolation and Polarity Estimation of Objective and Out-Of-Vocabulary Words to Improve Sentiment Classification on Microblogging	204
<i>Yongyos Kaewpitakkun, Kiyooki Shirai and Masnizah Mohd</i>	
How Mutual Knowledge Constrains the Choice of Anaphoric Demonstratives in Japanese and English	214
<i>David Yoshikazu Oshima and Eric McCready</i>	
The Influence of Givenness and Heaviness on OSV in Japanese	224
<i>Satoshi Imamura</i>	

Annotating Article Errors in Spanish Learner Texts: Design and Evaluation of an Annotation Scheme.....	234
<i>M. Pilar Valverde-Ibañez and Akira Ohtani</i>	
Needle in a Haystack: Reducing the Costs of Annotating Rare-Class Instances in Imbalanced Datasets.....	244
<i>Emily Jamison and Iryna Gurevych</i>	
Hybrid Approach to Zero Subject Resolution for multilingual MT - Spanish-to-Korean Cases -	254
<i>Arum Park and Munpyo Hong</i>	
Improving Statistical Machine Translation Accuracy Using Bilingual Lexicon Extraction with Paraphrases	262
<i>Chenhui Chu, Toshiaki Nakazawa and Sadao Kurohashi</i>	
Incrementally Updating the SMT Reordering Model.....	272
<i>Shachar Mirkin</i>	
TakeTwo: A Word Aligner based on Self Learning.....	282
<i>Jim Chang, Jian-Cheng Wu and Jason Chang</i>	
Frequency-influenced Choice of L2 Sound Realization and Perception: Evidence from Two Chinese Dialects.....	292
<i>Yizhou Lan</i>	
The L2 Acquisition of the Chinese Aspect Marking.....	299
<i>Suying Yang</i>	
Readability of Bangla News Articles for Children	309
<i>Zahrul Islam and Rashedur Rahman</i>	
Focusing on a Subset of Scripts Enhances the Learning Efficiency of Second Language Writing System.....	318
<i>Ching-Pong Au, Yuk-Man Cheung and Charles Chen, Jr.</i>	
Transition-based Knowledge Graph Embedding with Relational Mapping Properties.....	328
<i>Miao Fan, Qiang Zhou, Emily Chang and Thomas Fang Zheng</i>	
Retrieval Term Prediction Using Deep Belief Networks	338
<i>Qing Ma, Ibuki Tanigawa and Masaki Murata</i>	
Word-level Language Identification in Bi-lingual Code-switched Texts.....	348
<i>Harsh Jhamtani, Suleep Kumar Bhogi and Vaskar Raychoudhury</i>	
An Example-Based Approach to Difficult Pronoun Resolution	358
<i>Canasai Kruengkrai, Naoya Inoue, Jun Sugiura and Kentaro Inui</i>	
A Unified Analysis to Surpass Comparative and Experiential Aspect	368
<i>Charles Lam</i>	
A Quantitative View of Short Utterances in Daily Conversation: A Case Study of That’s right, That’s true and That’s correct	378
<i>Yanjiao Li, Alex Chengyu Fang and Jing Cao</i>	

A Listenability Measuring Method for an Adaptive Computer-assisted Language Learning and Teaching System.....	387
<i>Katsunori Kotani, Shota Ueda, Takehiko Yoshimi and Hiroaki Nanjo</i>	
Prosodic Differences Between Declaratives and Polar Questions In Fataluku	395
<i>Tyler Heston</i>	
Recognition of Sarcasms in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches	404
<i>Piyoros Tunghamthiti, Shirai Kiyooki and Masnizah Mohd</i>	
CHULA TTS: A Modularized Text-To-Speech Framework	414
<i>Natthawut Kertkeidkachorn, Supadaech Chanjaradwichai, Proadpran Punyabukkana and Atiwong Suchato</i>	
Modeling Structural Topic Transitions for Automatic Lyrics Generation	422
<i>Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui and Masataka Goto</i>	
Influence of Information Structure on Word Order Change and Topic Marker WA in Japanese .	432
<i>Satoshi Imamura, Yohei Sato and Masatoshi Koizumi</i>	
On the Functional Differences between the Discourse Particles Ne and Yone in Japanese.	442
<i>David Yoshikazu Oshima</i>	
The Centre and Periphery of Discourse Connectives	452
<i>Magdal'ena Rysov' a and Kateřina Rysov' a</i>	
Toward a Discourse Theory for Annotating Causal Relations in Japanese	460
<i>Kimi Kaneko and Daisuke Bekki</i>	
On-line Summarization of Time-series Documents using a Graph-based Algorithm	470
<i>Satoko Suzuki and Ichiro Kobayashi</i>	
Adjacency Pair Recognition in Wikipedia Discussions using Lexical Pairs	479
<i>Emily Jamison and Iryna Gurevych</i>	
A Hierarchical Word Sequence Language Model.....	489
<i>Xiaoyi Wu and Yuji Matsumoto</i>	
An Analysis of Radicals-based Features in Subjectivity Classification on Simplified Chinese Sentences.....	495
<i>Ge Xu and Churen Huang</i>	
A Semantics for Honorifics with Reference to Thai	503
<i>Eric McCready</i>	
Disunity in Cohesion: How Purpose Affects Methods and Results When Analyzing Lexical Cohesion	513
<i>Stuart G. Towns and Richard Watson Todd</i>	
Topics are Conditionals: A Case Study from Exhaustification Over Questions	522
<i>Yurie Hara</i>	
Detecting the Untranslatable Colloquial Expressions of Japanese Verbs in Cross-Language Instant Messaging	532
<i>Yuchang Cheng, Masaru Fuji, Tomoki Nagase, Minoru Uegaki and Isaac Okada</i>	

A Keyword-based Monolingual Sentence Aligner in Text Simplification.....	542
<i>Chung-Chi Huang</i>	
Automatic Detection of Comma Splices.....	551
<i>John Lee, Chak Yan Yeung and Martin Chodorow</i>	
Predicting the Use of BA construction in Mandarin Chinese Discourse: A Modeling Study with Two Verbs	561
<i>Yao Yao</i>	
Guo1 and Guo2 in Chinese Temporal System	568
<i>Qiu Zhuang and Su Qi</i>	
A Non-local Attachment Preference in the Production and Comprehension of Thai Relative Clauses.....	575
<i>Teeranoot Siriwittayakorn, Edson T. Miyamoto, Theeraporn Ratitamkul and Heeyoun Cho</i>	
Encoding Generalized Quantifiers in Dependency-based Compositional Semantics	585
<i>Yubing Dong, Ran Tian and Yusuke Miyao</i>	
On Common Ground, Context and Information Structure: The Case of Counter-Expectation in Thai	595
<i>Upsorn Tawilapakul</i>	
Semantics and Pragmatics of Cantonese Polar Questions: an inquisitive approach.....	605
<i>Yurie Hara</i>	
On the Argument Structures of the Transitive Verb 'annoy; be annoyed; bother to do': A Study Based on Two Comparable Corpora	615
<i>Jiajuan Xiong and Chu-Ren Huang</i>	
The Semantics of <i>khin3</i> and <i>loŋ1</i> in Thai Compared to up and down in English: A Corpus-based Study	624
<i>Junyawan Suwannarat and Theeraporn Ratitamkul</i>	
Constructions: a New Unit of Analysis for Corpus-based Discourse Analysis	634
<i>Samia Touileb and Andrew Salway</i>	
Noun Paraphrasing Based on a Variety of Contexts	644
<i>Tomoyuki Kajiwara and Kazuhide Yamamoto</i>	
Sentential Paraphrase Generation for Agglutinative Languages Using SVM with a String Kernel.....	650
<i>Hancheol Park, Gahgene Gweon, Ho-Jin Choi, Jeong Heo and Pum-Mo Ryu</i>	
K-repeating Substrings: a String-Algorithmic Approach to Privacy-Preserving Publishing of Textual Data	658
<i>Yusuke Matsubara and Koiti Hasida</i>	