Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation

Edited by Ryo Otoguro, Kiyoshi Ishikawa, Hiroshi Umemoto, Kei Yoshimoto and Yasunari Harada

4 — 7 November 2010
Tohoku University

© 2010 The PACLIC 24 Organizing Committee and PACLIC Steering Committee

All rights reserved. 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 978-4-905166-00-9

Published by Institute for Digital Enhancement of Cognitive Development, Waseda University

Acknowledgments

PACLIC 24 is hosted by the Center for the Advancement of Higher Education, Tohoku University in conjunction with The Logico-Linguistic Society of Japan and supported by JSPS International Scientific Meetings in Japan and Research Center for Language, Brain and Cognition, Graduate School of International Cultural Studies, Tohoku University

Symposium: Linguistics of 'Ba' on Day 2: November 6th (Sat) is supported in part and organized in conjunction with JSPS Kakenhi, or Grant-inAid for Challenging Exploratory Research Number 21652041, entitled Construction of Linguistics of 'Ba', Semantics and Pragmatics of 'Ba'.

Foreword

Welcome to the beautiful city of Sendai and to PACLIC 24. Tohoku University thank you for your participation, the Logico-Linguistic Society of Japan thank you for your contribution, and I thank you for coming to join us.

We are gathered here for the 24th Pacific Asia Conference on Language, Information and Computation, to be held for the first time in Sendai. To our call for papers distributed earlier this year, we received a total of 102 submissions, out of which 28 are selected for oral presentations, 25 for poster presentations and 20 for the student session. This year, we witnessed a sudden increase of submissions from South Asia and on languages in the region, augmenting diversity of languages discussed and countries and regions of participants. With increasing number of submissions, we are no longer able to follow our earlier PACLIC tradition to organize the program into one series of single sessions, but we still hope to keep our spirit in which researchers in theoretical and computational linguistics from diverse research backgrounds on various issues of different languages gather together, share results and go home with fresh ideas and new friendships. Following the practice of PACLIC 23 held in 2009 at the City University of Hong Kong, we set up multiple sessions to accommodate more oral presentations, but the two invited open symposia and three plenary talks do not have counterpart parallel sessions. We have two pre-workshops, one on corpus and the other on meaning, on Day 0, which is Thursday November 4th, and on Day 3, which is Sunday November 7th, we have a special workshop, proposed and organized by Dr. Chungmin Lee, in honor of Professor Akira Ikeya, who is one of the "founding fathers" of this PACLIC series of conferences.

We introduced two new "innovations" in our program: the two organized symposia on Day 1 and Day 2 are open to the general public, namely general audience who are not registered for PACLIC 24 can attend the presentations for free. Another new attempt is to set up a Student Session and Student Awards. The session is for graduate students and authors of several best papers are awarded partial travel expense support. The former is intended to attract more audience, mostly from within Japan and especially from the area around the hosting institution, and the latter to motivate increased interests in submission, especially from students outside Japan.

All these new attempts were possible by the dedicated efforts by our Organizing Chair, Professor Kei Yoshimoto, Program Chair, Ryo Otoguro, and other members of both the Organizing and Program Committees and additional reviewers, with financial support from the Japan Society for the Promotion of Science and Tohoku University.

Now, please enjoy the presentations, get connected and get ready to start a new collaboration!

October, 2010
Yasunari Harada
Honorary Program Chair of PACLIC 24
President of the Logico-Linguistic Society of Japan

Organizers

Steering Committee

Jae-Woong Choe, Korea University
Yasunari Harada, Waseda University
Chu-Ren Huang, Hong Kong Polytechnic University
Kim Teng Lua, Chinese and Oriental Languages Information Processing Society
Rachel Edita Roxas, De La Salle University-Manila
Maosong Sun, Tsinghua University
Benjamin T'sou, The Hong Kong Institute of Education

Program Committee

Chairs

Ryo Otoguro, Waseda University
Kiyoshi Ishikawa, Hosei University
Hiroshi Umemoto, Fuji Xerox Co.
Yasunari Harada, Waseda University (Honorary Program Chair)

Co-Chairs

Olivia Kwong, City University of Hong Kong
Shu-Kai Hsieh, National Taiwan Normal University
Donghong Ji, Wuhan University
Yae-Sheik Lee, Kyungpook National University
Min Zhang, Institute for Infocomm Research
Rachel Edita Roxas, De La Salle University-Manila

Members

Doug Arnold, University of Essex
Wirote Aroonmanakun, Chulalongkorn University
Timothy Baldwin, University of Melbourne
Ma. Lourdes Bautista, De La Salle University
Francis Bond, Nanyang Technological University
Hee-Rahk Chae, Hankuk University of Foreign Studies
Jae-Woong Choe, Korea University
Jing-Shin Chang, National Chi-Nan University
Eng Siong Chng, Nanyang Technological University
Siaw-Fong Chung, National Chengchi University
Beatrice Daille, University of Nantes

Danilo Dayag, De La Salle University
Shirley Dita, De La Salle University
Alex Fang, City University of Hong Kong
Maria Flouraki, School of Oriental and African Studies, University of London
Guohong Fu, Heilongjiang University
Helena Hong Gao, Nanyang Technological University
Zhao-Ming Gao, National Taiwan University
Josef van Genabith, Dublin City University
Yi Guan, Harbin Institute of Technology
Takao Gunji, Kobe Shoin University
Koji Kawahara, Fuji Women's University
Jong-Bok Kim, Kyung Hee University
Young-Gil Kim, Electronics and Telecommunications Research Institute
Chunyu Kit, City University of Hong Kong
Valia Kordoni, Saarland University
Paul Law, City University of Hong Kong
Chungmin Lee, Seoul National University
Hae-yun Lee, Hankuk University of Foreign Studies
Alessandro Lenci, University of Pisa
Gina-Anne Levow, University of Manchester
Haizhou Li, Institute for Infocomm Research
Qun Liu, Chinese Academy of Sciences
Qin Lu, Hong Kong Polytechnic University
Qing Ma, Ryukoku University
Takafumi Maekawa, Hokusei Gakuen Junior College
Ruli Manurung, University of Indonesia
Michinao Matsui, Osaka Health Science University
Yoshiki Mori, University of Tokyo
Vincent Ng, University of Texas at Dallas
Toshiyuki Ogihara, University of Washington
David Y. Oshima, Nagoya University
Jong C. Park, Korea Advanced Institute of Science and Technology
Prashant Pardeshi, National Institute for Japanese Language and Linguistics
Laurent Prevot, Universite de Provence
Hong-Mwun Pryo, Sunkyunkwan University
Long Qiu, Institute for Infocomm Research
Manabu Sassano, Yahoo Japan Corporation
Sachiko Shudo, Waseda University
Pornsiri Singhapreecha, Thammasat University

Virach Sornlertlamvanich, Thai Computational Linguistics Laboratory, NICT
Andrew Spencer, University of Essex
I-Wen Su, National Taiwan University
Keh-Yih Su, Behavior Design Corporation
Zhifang Sui, Peking University
Le Sun, Chinese Academy of Sciences
Takenobu Tokunaga, Tokyo Institute of Technology
Shu-Chuan Tseng, Academia Sinica
Aline Villavicencio, Federal University of Rio Grande do Sul
Haifeng Wang, Baidu
Houfeng Wang, Peking University
Hui Wang, National University of Singapore
Andy Way, Dublin City University
Kam-Fai Wong, Chinese University of Hong Kong
Jiun-Shiung Wu, National Chiayi University
Fei Xia, University of Washington
Jae Il Yeom, Hongik University
Satoru Yokoyama, Tohoku University
Kei Yoshimoto, Tohoku University
Hao Yu, Fujitsu (China) Research Center
Yujie Zhang, National Institute of Information and Communications Technology
Guo Dong Zhou, Soochow University
Qiang Zhou, Tsinghua University
Jingbo Zhu, Northeastern University
Michael Zock, Laboratoire d'Informatique Fondamentale de Marseille, C.N.R.S.
Chengqing Zong, Chinese Academy of Sciences

Additional Reviewers

Chikara Hashimoto, National Institute of Information and Communication Technology
Masato Ishizaki, University of Tokyo
Reiko Mazuka, RIKEN Brain Science Institute
Yasuhide Miura, Fuji Xerox Co.
Hiroshi Nakagawa, University of Tokyo
Satoshi Sato, Nagoya University
Daigo Sugihara, Fuji Xerox Co.
Motoyuki Takaai, Fuji Xerox Co.
Satoshi Tojo, Japan Advanced Institute of Science and Technology
Masatsugu Tonoike, Fuji Xerox Co.

Local Organizing Committee

Kei Yoshimoto, Tohoku University
Natsue Sugaya, Tohoku University
Satoru Yokoyama, Tohoku University
Shigeru Sato, Juntendo University
Chidori Nakamura, Iwate University

Table of Contents

Foreword iii

Organizers iv

Invited and Plenary Talks

Symposium on Language Resources

PodCastle: A Spoken Document Retrieval Service Improved by Anonymous User Contributions
Masataka Goto and Jun Ogata 3

The Future Role of Language Resources for Natural Language Parsing (We Won't Be Able to Rely on Pierre Vinken Rorever. . . or Will We Have to?)
Julia Hockenmaier 13

The Acquisition of Word Order in a Topic-prominent Language: Corpus Findings and Experimental Investigation
Thomas Hun-Tak Lee 15

Symposium: Toward a Linguistic Theory of Ba: Semantics and Pragmatics of Ba

How and Why Two Strangers Can Co-create a Story: An Application of the 'Ba'-Theory Based Approach to the Discourse
Sachiko Ide 19

Situated Language: Case of Japanese
Hideyuki Nakashima 25

Listening In
Stanley Peters 31

Plenary Talks

Basic Units of Lexicons and Ontologies: Words, Senses and Concepts
Hee-Rahk Chae 35

A Morphosyntactic Analysis of the Pronominal System of Philippine Languages
Shirley N. Dita 45

Statistical Translation Model Based On Source Syntax Structure
Qun Liu, Yang Liu and Haitai Mi 61

Regular Papers

Lexical Semantics-Syntactic Model for Defining and Subcategorizing Attribute Noun Class
Xiaopeng Bai and Hui Wang 65

Benglish Verbs: A Case of Code-mixing in Bengali
Shishir Bhattacharja 75

Enhanced Genre Classification through Linguistically Fine-Grained POS Tags
Alex Chengyu Fang and Jing Cao 85

Identifying Emotional Expressions, Intensities and Sentence Level Emotion Tags Using a Supervised Framework
Dipankar Das and Sivaji Bandyopadhyay 95

Decision Theory and Discourse Particles: A Case Study from a Large Japanese Sentiment Corpus
Christopher Davis 105

Finding Appropriate Subset of Votes Per Classifier Using Multiobjective Optimization: Application to Named Entity Recognition
Asif Ekbal, Sriparna Saha and Md. Hasanuzzaman 115

Experimental Approach to Reference in Discourse: Working Memory Capacity and Language Comprehension in Russian
Olga Fedorova, Ekaterina Delikishkina and Anna Uspenskaya 125

A Computational Model of Language Generation Applied to Japanese Wh-*questions*
Jason Ginsburg 133

Chains in Syntax and Morphology
Thomas Gross 143

Feature Subset Selection Using Genetic Algorithm for Named Entity Recognition
Md. Hasanuzzaman, Sriparna Saha and Asif Ekbal 153

Developing Punjabi Morphology, Corpus and Lexicon
Muhammad Humayoun and Aarne Ranta 163

Effective Use of Linguistic Features for Sentiment Analysis of Korean
Hayeon Jang and Hyopil Shin 173

Word Order and NP Structure in Korean: A Constraint Based Approach
Jong-Bok Kim, Nam-Guen Lee and Yae-shiek Lee 183

Evidentials and Epistemic Modal in Korean: Evidence from Their Intractions
Chungmin Lee 193

Implementation of Korean Syllable Structures in the Typed Feature Structure Formalism
Gyu-hyung Lee, Ye-seul Park and Yong-hun Lee 203

Focus Types and Subject-Object Asymmetry in Korean Case Ellipsis: A New Look at Focus Effects
Hanjung Lee and Haejeong Choi 213

eSpatialML: An Event-Driven Spatial Annotation Framework
Kiyong Lee, Jonathan Webster and Alex Chengyu Fang 223

Domain-Independent Novel Event Discovery and Semi-Automatic Event Annotation
Hao Li, Xiang Li, Heng Ji and Yuval Marton 233

Developing an Online Indonesian Corpora Repository
Ruli Manurung, Bayu Distiawan and Desmond Darma Putra 243

Topicalization and Truth Conditions: A Categorial Account
Hiroaki Nakamura 251

Unsupervised Classification of Biomedical Abstracts Using Lexical Association
Jonathon Read, Jonathan Webster and Alex Chengyu Fang 261

Licensing Nominals in the Multiple Nominative Constructions in Korean: A Mereological Perspective
Byong-Rae Ryu 271

Combination of 3 Types of Speech Recognizers for Anaphora Resolution
Kazutaka Shimada, Noriko Tanamachi and Tsutomu Endo 281

Syntactic and Semantic Discrepancies among the Verbs for 'kill' in English, Chinese and Thai
Kingkarn Thepkanjana and Satoshi Uehara 291

Expanding Chinese Sentiment Dictionaries from Large Scale Unlabeled Corpus
Hongzhi Xu, Kai Zhao, Likun Qiu and Changjian Hu 301

Representing Events in Japanese Complex Predicates
Yoshie Yamamori 311

Fault-Tolerant Learning for Term Extraction
Yuhang Yang, Hao Yu, Yao Meng, Yingliang Lu and Yingju Xia 321

Syntactically Complex Demonstratives and Sortal Inherency
Richard Zuber 331

Poster Papers

On Automated Hyperonym Hierarchy Construction Using an Internet Search Engine
Sergey Afonin 341

An Ontological Analysis of Japanese and Chinese Kinship Terms
Songiy Baik and Hee-Rahk Chae 349

GRASP: Grammar- and Syntax-based Pattern-Finder for Collocation and Phrase Learning
Mei-hua Chen, Chung-chi Huang, Shih-ting Huang and Jason S. Chang 357

Mitigating Problems in Analogy-based EBMT with SMT and vice versa: A Case Study with Named Entity Transliteration
Sandipan Dandapat, Sara Morrissey, Sudip Kumar Naskar and Harold Somers 365

Why the English Easiest Type Became the Hardest in Russian, or Russian Adults' Comprehension of before *and* after *sentences*
Olga Fedorova 373

Simpler Is Better: Re-evaluation of Default Word Alignment Models in Statistical MT
Mark Fishel 381

The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme
Serge Heiden 389

Using Corpus-based Linguistic Approaches in Sense Prediction Study
Jia-Fei Hong, Sue-Jin Ker, Chu-Ren Huang and Kathleen Ahrens 399

Synthetic Compounds in Mandarin
Shuang Hong 409

The Specialized Vocabulary of Modern Patent Language: Semantic Associations in Patent Lexis
Darren Hsin-hung Lin and Shelley Ching-yu Hsieh 417

The KOLON System: Tools for Ontological Natural Language Processing in Korean
Juliano Paiva Junho, Yumi Jo and Hyopil Shin 425

An Approach toward Register Classification of Book Samples in the Balanced Corpus of Contemporary Written Japanese
Wakako Kashino and Manabu Okumura 433

A Computational Treatment of Korean Serial Verb Constructions
Jong-Bok Kim, Jaehyung Yang and Sanghoun Song 439

A Supervised Machine Learning Approach for Event-Event Relation Identification
Anup Kumar Kolya, Asif Ekbal and Sivaji Bandyopadhyay 447

Arguments for Parallel Distributed Parsing: Toward the Integration of Lexical and Sublexical (Semantic) Parsings
Kow Kuroda 455

Scrambling of Wh*-phrases in Japanese*
Paul Law 463

A Comparative Study on the Aspectual Classification of Korean and Japanese Verbs in Relation to '-ess-' '-essess-' *in Korean and* 'sudeni' *in Japanese*
Younghee Lee and Byong-Rae Ryu 471

A Modular Architecture for the Wide-Coverage Translation of Natural Language Texts into Predicate Logic Formulas
Yusuke Miyao, Alastair Butler, Kei Yoshimoto and Jun'ichi Tsujii 481

Toward a Unified Computational Model of Quantificational Scope Readings
Prakash Mondal 489

CCG of Japanese Sentence-final Particles
Sumiyo Nishiguchi 497

A Multi-Dimensional Analysis of Japanese Benefactives: The Case of the Yaru*-Construction*
Akira Otani and Mark Steedman 503

Typed Dependency Relations for Syntactic Analysis of Thai Sentences
Siripong Potisuk 511

Exploiting a Multilingual Web-based Encyclopedia for Bilingual Terminology Extraction
Fatiha Sadat 519

Language Homogeneity in the Japanese Wikipedia
Karl-André Skevik 527

Incorporate Credibility into Context for the Best Social Media Answers
Qi Su, Helen Kai-yun Chen and Chu-Ren Huang 535

Student Award Papers

A Query Focused Multi Document Automatic Summarization
Pinaki Bhaskar and Sivaji Bandyopadhyay 545

Through Low-Cost Annotation to Reliable Parsing Evaluation
Marek Grác, Miloš Jakubíček and Vojtěch Kovář 555

A Look into the Acquisition of English Motion Event Conflation by Native Speakers of Chinese and Japanese
Ryan Spring 563

Stage/Individual-level Predicates, Topics and Indefinite Subjects
Shuiying Yao 573

How Well Conditional Random Fields Can be Used in Novel Term Recognition
Xing Zhang, Yan Song and Alex Chengyu Fang 583

Student Papers

Clitics in Arabic Language: A Statistical Study
Fahad Alotaiby, Salah Foda and Ibrahim Alkharashi 595

AUTOLEX: An Automatic Lexicon Builder for Minority Languages Using an Open Corpus
Evan Liz C. Buhay, Marie Joy P. Evardone, Hansel B. Nocon, Davis Muhajereen D. Dimalen and Rachel Edita O. Roxas 603

Improving Information Extraction Using Knowledge Model
Yue Chen 613

Finding Emotion Holder from Bengali Blog Texts: An Unsupervised Syntactic Approach
Dipankar Das and Sivaji Bandyopadhyay 621

Terminological Ontology and Cognitive Processes in Translation
Fumiko Kano Glückstad 629

Detection of Users Suspected of Pretending to Be Other Users in a Community Site by Using Messages Submitted to Non-Target Categories
Naoki Ishikawa, Ryo Nishimura, Yasuhiko Watanabe, Masaki Murata and Yoshihiro Okada 637

Detecting Nasty Comments from BBS Posts
Tatsuya Ishisaka and Kazuhide Yamamoto 645

Using Various Features in Machine Learning to Obtain High Levels of Performance for Recognition of Japanese Notational Variants
Masahiro Kojima, Masaki Murata, Jun'ichi Kazama, Kow Kuroda, Atsushi Fujita, Eiji Aramaki, Masaaki Tsuchida, Yasuhiko Watanabe and Kentaro Torisawa 653

Korean Kes-*Relative Clauses as Stages*
Sang-Hee Park 661

Finite State Morphology and Sindhi Noun Inflections
Mutee U Rahman Muslim and Mohammad Iqbal Bhatti 669

The Acquisition of Imperfective Aspect Marking in Korean as a Second language by Japanese Learners
Ju-Yeon Ryu, Kaoru Horie and Yasuhiro Shirai 677

Multi-aspects Review Summarization Based on Identification of Important Opinions and their Similarity
Ryosuke Tadano, Kazutaka Shimada and Tsutomu Endo 685

Generation of Summaries that Appropriately and Adequately Express the Contents of Original Documents Using Word-Association Knowledge
Kazuki Takigawa, Masaki Murata, Masaaki Tsuchida, Stijn De Saeger, Kazuhide Yamamoto and Kentaro Torisawa 693

Adjectival Modification to Nouns in Mandarin Chinese:
Case Studies on "cháng+*noun*" *and* "*adjective*+tú shū gu n"
Shan Wang and Chu-Ren Huang 701

Compositional Operations of Mandarin Chinese Perception Verb "kàn":
A Generative Lexicon Approach
Shan Wang and Chu-Ren Huang 707

Workshop on Advanced Corpus Solutions

Workshop on Advanced Corpus Solutions
Janne Bondi Johannessen 717

Degrees of Orality in Speech-like Corpora:
Comparative Annotation of Chat and E-mail Corpora
Eckhard Bick 721

Parallel Suffix Arrays for Corpus Exploration
Johannes Goller 731

Fast Syntactic Searching in Very Large Corpora for Many Languages
Miloš Jakubíček, Adam Kilgarriff, Diana McCarthy and Pavel Rychlý 741

A Multilingual Speech Resource: The Nordic Dialect Corpus
Janne Bondi Johannessen, Joel Priestley and Anders Nøklestad 749

Dialect Corpora Taken Further:
The DynaSAND Corpus and Its Application in Newer Tools
Jan Pieter Kunst and Franca Wesseling 759

Advanced Corpus Solutions for Humanities Researchers
James Wilson, Anthony Hartley, Serge Sharoff and Paul Stephenson 769

Entity Coherence in Comparable Learner Corpora: Seeking Pedagogical Insights
Mitsuko Yamura-Takei, Miho Fujiwara and Etsuko Yoshida 779

Workshop on Model and Measurement of Meaning (M3)

The Use of a Cultural Protocol for Quantifying Cultural Variations
in Verb Semantic between Chinese and French
Hintat Cheung, Yann Desalle, Karine Duvignau, Bruno Gaume,
Chunhan Chang and Pierre Magistry 791

Towards the Global SentiWordNet
Amitava Das and Sivaji Bandyopadhyay 799

Towards an Automatic Measurement of Verbal Lexicon Acquisition:
The Case for a Young Children-versus-Adults Classification in French and Mandarin
Yann Desalle, Shu-Kai Hsieh, Bruno Gaume and Hintat Cheung 809

Graph Representation of Synonymy and Translation Resources for Crosslinguistic Modelisation of Meaning
Benoît Gaillard, Yannick Chudy, Pierre Magistry, Shu-Kai Hsieh and Emmanuel Navarro 819

Exploring Chinese Verbal Lexicon Developmental Trend with Semantic Space
Ching-Fen Pan 831

Computational Modeling of Verb Acquisition, from a Monolingual to a Bilingual Study
Laurent Prévot, Chun-Han Chang and Yann Desalle 841

Cross-sortal Predication and Polysemy
Petr Šimon and Chu-Ren Huang 853

Verb Use in Chinese Children: Extensibility of Instrument
Te-Hsin Wu 863

Workshop on Semantics and Logical Flavor: In Honor of Professor Akira Ikeya

Natural Language Production in Database Semantics
Roland Hausser 875

Change of Location and Change of State: How Telicity is Attained
Chungmin Lee 885

Various Evidentials in Korean
Kyung An Song 895

A Note on Pseudo-comparatives like "John is rich like X!" and "Like X, John is rich!"
Benjamin T'sou 907