

## Table of Contents

Foreword .....	iii
Organizers .....	iv

### Invited and Plenary Talks

#### Symposium on Language Resources

<i>PodCastle: A Spoken Document Retrieval Service Improved by Anonymous User Contributions</i> Masataka Goto and Jun Ogata .....	3
<i>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?)</i> Julia Hockenmaier .....	13
<i>The Acquisition of Word Order in a Topic-prominent Language: Corpus Findings and Experimental Investigation</i> Thomas Hun-Tak Lee .....	15

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

<i>How and Why Two Strangers Can Co-create a Story: An Application of the 'Ba'-Theory Based Approach to the Discourse</i> Sachiko Ide .....	19
<i>Situated Language: Case of Japanese</i> Hideyuki Nakashima .....	25
<i>Listening In</i> Stanley Peters .....	31

#### Plenary Talks

<i>Basic Units of Lexicons and Ontologies: Words, Senses and Concepts</i> Hee-Rahk Chae .....	35
<i>A Morphosyntactic Analysis of the Pronominal System of Philippine Languages</i> Shirley N. Dita .....	45
<i>Statistical Translation Model Based On Source Syntax Structure</i> Qun Liu, Yang Liu and Haitai Mi .....	61

## Regular Papers

<i>Lexical Semantics-Syntactic Model for Defining and Subcategorizing Attribute Noun Class</i> Xiaopeng Bai and Hui Wang .....	65
<i>Benglish Verbs: A Case of Code-mixing in Bengali</i> Shishir Bhattacharja .....	75
<i>Enhanced Genre Classification through Linguistically Fine-Grained POS Tags</i> Alex Chengyu Fang and Jing Cao .....	85
<i>Identifying Emotional Expressions, Intensities and Sentence Level Emotion Tags Using a Supervised Framework</i> Dipankar Das and Sivaji Bandyopadhyay .....	95
<i>Decision Theory and Discourse Particles: A Case Study from a Large Japanese Sentiment Corpus</i> Christopher Davis .....	105
<i>Finding Appropriate Subset of Votes Per Classifier Using Multiobjective Optimization: Application to Named Entity Recognition</i> Asif Ekbal, Sriparna Saha and Md. Hasanuzzaman .....	115
<i>Experimental Approach to Reference in Discourse: Working Memory Capacity and Language Comprehension in Russian</i> Olga Fedorova, Ekaterina Delikishkina and Anna Uspenskaya .....	125
<i>A Computational Model of Language Generation Applied to Japanese Wh-questions</i> Jason Ginsburg .....	133
<i>Chains in Syntax and Morphology</i> Thomas Gross .....	143
<i>Feature Subset Selection Using Genetic Algorithm for Named Entity Recognition</i> Md. Hasanuzzaman, Sriparna Saha and Asif Ekbal .....	153
<i>Developing Punjabi Morphology, Corpus and Lexicon</i> Muhammad Humayoun and Aarne Ranta .....	163
<i>Effective Use of Linguistic Features for Sentiment Analysis of Korean</i> Hayeon Jang and Hyopil Shin .....	173
<i>Word Order and NP Structure in Korean: A Constraint Based Approach</i> Jong-Bok Kim, Nam-Guen Lee and Yae-shiek Lee .....	183
<i>Evidentials and Epistemic Modal in Korean: Evidence from Their Interactions</i> Chungmin Lee .....	193

<i>Implementation of Korean Syllable Structures in the Typed Feature Structure Formalism</i> Gyu-hyung Lee, Ye-seul Park and Yong-hun Lee .....	203
<i>Focus Types and Subject-Object Asymmetry in Korean Case Ellipsis: A New Look at Focus Effects</i> Hanjung Lee and Haejeong Choi .....	213
<i>eSpatialML: An Event-Driven Spatial Annotation Framework</i> Kiyong Lee, Jonathan Webster and Alex Chengyu Fang .....	223
<i>Domain-Independent Novel Event Discovery and Semi-Automatic Event Annotation</i> Hao Li, Xiang Li, Heng Ji and Yuval Marton .....	233
<i>Developing an Online Indonesian Corpora Repository</i> Ruli Manurung, Bayu Distiawan and Desmond Darma Putra .....	243
<i>Topicalization and Truth Conditions: A Categorical Account</i> Hiroaki Nakamura .....	251
<i>Unsupervised Classification of Biomedical Abstracts Using Lexical Association</i> Jonathon Read, Jonathan Webster and Alex Chengyu Fang .....	261
<i>Licensing Nominals in the Multiple Nominative Constructions in Korean: A Mereological Perspective</i> Byong-Rae Ryu .....	271
<i>Combination of 3 Types of Speech Recognizers for Anaphora Resolution</i> Kazutaka Shimada, Noriko Tanamachi and Tsutomu Endo .....	281
<i>Syntactic and Semantic Discrepancies among the Verbs for 'kill' in English, Chinese and Thai</i> Kingkarn Thepkanjana and Satoshi Uehara .....	291
<i>Expanding Chinese Sentiment Dictionaries from Large Scale Unlabeled Corpus</i> Hongzhi Xu, Kai Zhao, Likun Qiu and Changjian Hu .....	301
<i>Representing Events in Japanese Complex Predicates</i> Yoshie Yamamori .....	311
<i>Fault-Tolerant Learning for Term Extraction</i> Yuhang Yang, Hao Yu, Yao Meng, Yingliang Lu and Yingju Xia .....	321
<i>Syntactically Complex Demonstratives and Sortal Inherency</i> Richard Zuber .....	331

## **Poster Papers**

<i>On Automated Hyperonym Hierarchy Construction Using an Internet Search Engine</i> Sergey Afonin .....	341
---	-----

<i>An Ontological Analysis of Japanese and Chinese Kinship Terms</i> Songji Baik and Hee-Rahk Chae .....	349
<i>GRASP: Grammar- and Syntax-based Pattern-Finder for Collocation and Phrase Learning</i> Mei-hua Chen, Chung-chi Huang, Shih-ting Huang and Jason S. Chang .....	357
<i>Mitigating Problems in Analogy-based EBMT with SMT and vice versa: A Case Study with Named Entity Transliteration</i> Sandipan Dandapat, Sara Morrissey, Sudip Kumar Naskar and Harold Somers .....	365
<i>Why the English Easiest Type Became the Hardest in Russian, or Russian Adults' Comprehension of before and after sentences</i> Olga Fedorova .....	373
<i>Simpler Is Better: Re-evaluation of Default Word Alignment Models in Statistical MT</i> Mark Fishel .....	381
<i>The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme</i> Serge Heiden .....	389
<i>Using Corpus-based Linguistic Approaches in Sense Prediction Study</i> Jia-Fei Hong, Sue-Jin Ker, Chu-Ren Huang and Kathleen Ahrens .....	399
<i>Synthetic Compounds in Mandarin</i> Shuang Hong .....	409
<i>The Specialized Vocabulary of Modern Patent Language: Semantic Associations in Patent Lexis</i> Darren Hsin-hung Lin and Shelley Ching-yu Hsieh .....	417
<i>The KOLON System: Tools for Ontological Natural Language Processing in Korean</i> Juliano Paiva Junho, Yumi Jo and Hyopil Shin .....	425
<i>An Approach toward Register Classification of Book Samples in the Balanced Corpus of Contemporary Written Japanese</i> Wakako Kashino and Manabu Okumura .....	433
<i>A Computational Treatment of Korean Serial Verb Constructions</i> Jong-Bok Kim, Jaehyung Yang and Sanghoun Song .....	439
<i>A Supervised Machine Learning Approach for Event-Event Relation Identification</i> Anup Kumar Kolya, Asif Ekbal and Sivaji Bandyopadhyay .....	447
<i>Arguments for Parallel Distributed Parsing: Toward the Integration of Lexical and Sublexical (Semantic) Parsings</i> Kow Kuroda .....	455

<i>Scrambling of Wh-phrases in Japanese</i> Paul Law .....	463
<i>A Comparative Study on the Aspectual Classification of Korean and Japanese Verbs in Relation to ‘-ess-’ ‘-essess-’ in Korean and ‘sudeni’ in Japanese</i> Younghee Lee and Byong-Rae Ryu .....	471
<i>A Modular Architecture for the Wide-Coverage Translation of Natural Language Texts into Predicate Logic Formulas</i> Yusuke Miyao, Alastair Butler, Kei Yoshimoto and Jun’ichi Tsujii .....	481
<i>Toward a Unified Computational Model of Quantificational Scope Readings</i> Prakash Mondal .....	489
<i>CCG of Japanese Sentence-final Particles</i> Sumiyo Nishiguchi .....	497
<i>A Multi-Dimensional Analysis of Japanese Benefactives: The Case of the Yaru-Construction</i> Akira Otani and Mark Steedman .....	503
<i>Typed Dependency Relations for Syntactic Analysis of Thai Sentences</i> Siripong Potisuk .....	511
<i>Exploiting a Multilingual Web-based Encyclopedia for Bilingual Terminology Extraction</i> Fatiha Sadat .....	519
<i>Language Homogeneity in the Japanese Wikipedia</i> Karl-André Skevik .....	527
<i>Incorporate Credibility into Context for the Best Social Media Answers</i> Qi Su, Helen Kai-yun Chen and Chu-Ren Huang .....	535

## **Student Award Papers**

<i>A Query Focused Multi Document Automatic Summarization</i> Pinaki Bhaskar and Sivaji Bandyopadhyay .....	545
<i>Through Low-Cost Annotation to Reliable Parsing Evaluation</i> Marek Grác, Miloš Jakubíček and Vojtěch Kovář .....	555
<i>A Look into the Acquisition of English Motion Event Conflation by Native Speakers of Chinese and Japanese</i> Ryan Spring .....	563
<i>Stage/Individual-level Predicates, Topics and Indefinite Subjects</i> Shuiying Yao .....	573
<i>How Well Conditional Random Fields Can be Used in Novel Term Recognition</i> Xing Zhang, Yan Song and Alex Chengyu Fang .....	583

## Student Papers

<i>Clitics in Arabic Language: A Statistical Study</i> Fahad Alotaiby, Salah Foda and Ibrahim Alkharashi . . . . .	595
<i>AUTOLEX: An Automatic Lexicon Builder for Minority Languages Using an Open Corpus</i> Evan Liz C. Buhay, Marie Joy P. Evardone, Hansel B. Nocon, Davis Muhajereen D. Dimalen and Rachel Edita O. Roxas . . . . .	603
<i>Improving Information Extraction Using Knowledge Model</i> Yue Chen . . . . .	613
<i>Finding Emotion Holder from Bengali Blog Texts: An Unsupervised Syntactic Approach</i> Dipankar Das and Sivaji Bandyopadhyay . . . . .	621
<i>Terminological Ontology and Cognitive Processes in Translation</i> Fumiko Kano Glückstad . . . . .	629
<i>Detection of Users Suspected of Pretending to Be Other Users in a Community Site by Using Messages Submitted to Non-Target Categories</i> Naoki Ishikawa, Ryo Nishimura, Yasuhiko Watanabe, Masaki Murata and Yoshihiro Okada . . . . .	637
<i>Detecting Nasty Comments from BBS Posts</i> Tatsuya Ishisaka and Kazuhide Yamamoto . . . . .	645
<i>Using Various Features in Machine Learning to Obtain High Levels of Performance for Recognition of Japanese Notational Variants</i> Masahiro Kojima, Masaki Murata, Jun'ichi Kazama, Kow Kuroda, Atsushi Fujita, Eiji Aramaki, Masaaki Tsuchida, Yasuhiko Watanabe and Kentaro Torisawa . . . . .	653
<i>Korean Kes-Relative Clauses as Stages</i> Sang-Hee Park . . . . .	661
<i>Finite State Morphology and Sindhi Noun Inflections</i> Mutee U Rahman Muslim and Mohammad Iqbal Bhatti . . . . .	669
<i>The Acquisition of Imperfective Aspect Marking in Korean as a Second language by Japanese Learners</i> Ju-Yeon Ryu, Kaoru Horie and Yasuhiro Shirai . . . . .	677
<i>Multi-aspects Review Summarization Based on Identification of Important Opinions and their Similarity</i> Ryosuke Tadano, Kazutaka Shimada and Tsutomu Endo . . . . .	685
<i>Generation of Summaries that Appropriately and Adequately Express the Contents of Original Documents Using Word-Association Knowledge</i> Kazuki Takigawa, Masaki Murata, Masaaki Tsuchida, Stijn De Saeger, Kazuhide Yamamoto and Kentaro Torisawa . . . . .	693

<i>Adjectival Modification to Nouns in Mandarin Chinese: Case Studies on “cháng+noun” and “adjective+tú shū gu n”</i> Shan Wang and Chu-Ren Huang .....	701
<i>Compositional Operations of Mandarin Chinese Perception Verb “kàn”: A Generative Lexicon Approach</i> Shan Wang and Chu-Ren Huang .....	707

### **Workshop on Advanced Corpus Solutions**

<i>Workshop on Advanced Corpus Solutions</i> Janne Bondi Johannessen .....	717
<i>Degrees of Orality in Speech-like Corpora: Comparative Annotation of Chat and E-mail Corpora</i> Eckhard Bick .....	721
<i>Parallel Suffix Arrays for Corpus Exploration</i> Johannes Goller .....	731
<i>Fast Syntactic Searching in Very Large Corpora for Many Languages</i> Miloš Jakubíček, Adam Kilgarriff, Diana McCarthy and Pavel Rychlý .....	741
<i>A Multilingual Speech Resource: The Nordic Dialect Corpus</i> Janne Bondi Johannessen, Joel Priestley and Anders Nøklestad .....	749
<i>Dialect Corpora Taken Further: The DynaSAND Corpus and Its Application in Newer Tools</i> Jan Pieter Kunst and Franca Wesseling .....	759
<i>Advanced Corpus Solutions for Humanities Researchers</i> James Wilson, Anthony Hartley, Serge Sharoff and Paul Stephenson .....	769
<i>Entity Coherence in Comparable Learner Corpora: Seeking Pedagogical Insights</i> Mitsuko Yamura-Takei, Miho Fujiwara and Etsuko Yoshida .....	779

### **Workshop on Model and Measurement of Meaning (M3)**

<i>The Use of a Cultural Protocol for Quantifying Cultural Variations in Verb Semantic between Chinese and French</i> Hintat Cheung, Yann Desalle, Karine Duvignau, Bruno Gaume, Chunhan Chang and Pierre Magistry .....	791
<i>Towards the Global SentiWordNet</i> Amitava Das and Sivaji Bandyopadhyay .....	799
<i>Towards an Automatic Measurement of Verbal Lexicon Acquisition: The Case for a Young Children-versus-Adults Classification in French and Mandarin</i> Yann Desalle, Shu-Kai Hsieh, Bruno Gaume and Hintat Cheung .....	809

<i>Graph Representation of Synonymy and Translation Resources for Crosslinguistic Modelisation of Meaning</i> Benoît Gaillard, Yannick Chudy, Pierre Magistry, Shu-Kai Hsieh and Emmanuel Navarro .....	819
<i>Exploring Chinese Verbal Lexicon Developmental Trend with Semantic Space</i> Ching-Fen Pan .....	831
<i>Computational Modeling of Verb Acquisition, from a Monolingual to a Bilingual Study</i> Laurent Prévot, Chun-Han Chang and Yann Desalle .....	841
<i>Cross-sortal Predication and Polysemy</i> Petr Šimon and Chu-Ren Huang .....	853
<i>Verb Use in Chinese Children: Extensibility of Instrument</i> Te-Hsin Wu .....	863

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

<i>Natural Language Production in Database Semantics</i> Roland Hausser .....	875
<i>Change of Location and Change of State: How Telicity is Attained</i> Chungmin Lee .....	885
<i>Various Evidentials in Korean</i> Kyung An Song .....	895
<i>A Note on Pseudo-comparatives like “John is rich like X!” and “Like X, John is rich!”</i> Benjamin T’sou .....	907