

# **5th International Joint Conference on Natural Language Processing 2011**

**(IJCNLP 2011)**

**Chiang Mai, Thailand  
8-13 November 2011**

**Volume 1 of 2**

ISBN: 978-1-62276-505-8

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2011) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2013)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Table of Contents

## Part A: Full Papers

<i>Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers</i> Sonal Gupta and Christopher Manning .....	1
<i>Dependency-directed Tree Kernel-based Protein-Protein Interaction Extraction from Biomedical Literature</i> Longhua Qian and Guodong Zhou .....	10
<i>Learning Logical Structures of Paragraphs in Legal Articles</i> Ngo Xuan Bach, Nguyen Le Minh, Tran Thi Oanh and Akira Shimazu .....	20
<i>Extracting Pre-ordering Rules from Predicate-Argument Structures</i> Xianchao Wu, Katsuhito Sudoh, Kevin Duh, Hajime Tsukada and Masaaki Nagata .....	29
<i>Context-Sensitive Syntactic Source-Reordering by Statistical Transduction</i> Maxim Khalilov and Khalil Sima'an .....	38
<i>Discriminative Phrase-based Lexicalized Reordering Models using Weighted Reordering Graphs</i> Wang Ling, João Graça, David Martins de Matos, Isabel Trancoso and Alan W Black .....	47
<i>Active Learning Strategies for Support Vector Machines, Application to Temporal Relation Classification</i> Seyed Abolghasem Mirroshandel, Gholamreza Ghassem-Sani and Alexis Nasr .....	56
<i>A Fast Accurate Two-stage Training Algorithm for L1-regularized CRFs with Heuristic Line Search Strategy</i> Jinlong Zhou, Xipeng Qiu and Xuanjing Huang .....	65
<i>Automatic Topic Model Adaptation for Sentiment Analysis in Structured Domains</i> Geoffrey Levine and Gerald DeJong .....	75
<i>Multi-modal Reference Resolution in Situated Dialogue by Integrating Linguistic and Extra-Linguistic Clues</i> Ryu Iida, Masaaki Yasuhara and Takenobu Tokunaga .....	84
<i>Single and multi-objective optimization for feature selection in anaphora resolution</i> Sriparna Saha, Asif Ekbal, Olga Uryupina and Massimo Poesio .....	93
<i>A Unified Event Coreference Resolution by Integrating Multiple Resolvers</i> Bin Chen, Jian Su, Sinno Jialin Pan and Chew Lim Tan .....	102
<i>Handling verb phrase morphology in highly inflected Indian languages for Machine Translation</i> Ankur Gandhe, Rashmi Gangadharaiyah, Karthik Visweswariah and Ananthakrishnan Ramanathan	111
<i>Japanese Pronunciation Prediction as Phrasal Statistical Machine Translation</i> Jun Hatori and Hisami Suzuki .....	120
<i>Comparing Two Techniques for Learning Transliteration Models Using a Parallel Corpus</i> Hassan Sajjad, Nadir Durrani, Helmut Schmid and Alexander Fraser .....	129

<i>A Semantic-Specific Model for Chinese Named Entity Translation</i> Yufeng Chen and Chengqing Zong .....	138
<i>Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners</i> Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto .....	147
<i>Modality Specific Meta Features for Authorship Attribution in Web Forum Posts</i> Thamar Solorio, Sangita Pillay, Sindhu Raghavan and Manuel Montes-Gomez .....	156
<i>Keyphrase Extraction from Online News Using Binary Integer Programming</i> Zhuoye Ding, Qi Zhang and Xuanjing Huang .....	165
<i>Improving Related Entity Finding via Incorporating Homepages and Recognizing Fine-grained Entities</i> Youzheng Wu, Chiori Hori, Hisashi Kawai and Hideki Kashioka .....	174
<i>Enhancing Active Learning for Semantic Role Labeling via Compressed Dependency Trees</i> Chenhua Chen, Alexis Palmer and Caroline Sporleder .....	183
<i>Semantic Role Labeling Without Treebanks?</i> Stephen Boxwell, Chris Brew, Jason Baldridge, Dennis Mehay and Sujith Ravi .....	192
<i>Japanese Predicate Argument Structure Analysis Exploiting Argument Position and Type</i> Yuta Hayashibe, Mamoru Komachi and Yuji Matsumoto .....	201
<i>An Empirical Study on Compositionality in Compound Nouns</i> Siva Reddy, Diana McCarthy and Suresh Manandhar .....	210
<i>Feature-Rich Log-Linear Lexical Model for Latent Variable PCFG Grammars</i> Zhongqiang Huang and Mary Harper .....	219
<i>Improving Dependency Parsing with Fined-Grained Features</i> Guangyou Zhou, Li Cai, Kang Liu and Jun Zhao .....	228
<i>Natural Language Programming Using Class Sequential Rules</i> Cohan Sujay Carlos .....	237
<i>Treeblazing: Using External Treebanks to Filter Parse Forests for Parse Selection and Treebanking</i> Andrew MacKinlay, Rebecca Dridan, Dan Flickinger, Stephan Oepen and Timothy Baldwin ..	246
<i>Cross-Language Entity Linking</i> Paul McNamee, James Mayfield, Dawn Lawrie, Douglas Oard and David Doermann .....	255
<i>Generating Chinese Named Entity Data from a Parallel Corpus</i> Ruiji Fu, Bing Qin and Ting Liu .....	264
<i>Learning the Latent Topics for Question Retrieval in Community QA</i> Li Cai, Guangyou Zhou, Kang Liu and Jun Zhao .....	273
<i>Identifying Event Descriptions using Co-training with Online News Summaries</i> William Yang Wang, Kapil Thadani and Kathleen McKeown .....	282
<i>Automatic Labeling of Voiced Consonants for Morphological Analysis of Modern Japanese Literature</i> Teruaki Oka, Mamoru Komachi, Toshinobu Ogiso and Yuji Matsumoto .....	292

<i>S<sup>3</sup> - Statistical Sandhi Splitting</i>	
Abhiram Natarajan and Eugene Charniak .....	301
<i>Improving Chinese Word Segmentation and POS Tagging with Semi-supervised Methods Using Large Auto-Analyzed Data</i>	
Yiou Wang, Jun'ichi Kazama, Yoshimasa Tsuruoka, Wenliang Chen, Yujie Zhang and Kentaro Torisawa .....	309
<i>CODACT: Towards Identifying Orthographic Variants in Dialectal Arabic</i>	
Pradeep Dasigi and Mona Diab .....	318
<i>Enhancing the HL-SOT Approach to Sentiment Analysis via a Localized Feature Selection Framework</i>	
Wei and Jon Atle Gulla.....	327
<i>Fine-Grained Sentiment Analysis with Structural Features</i>	
Cäcilia Zirn, Mathias Niepert, Heiner Stuckenschmidt and Michael Strube .....	336
<i>Predicting Opinion Dependency Relations for Opinion Analysis</i>	
Lun-Wei Ku, Ting-Hao Huang and Hsin-Hsi Chen .....	345
<i>Detecting and Blocking False Sentiment Propagation</i>	
Hye-Jin Min and Jong C. Park .....	354
<i>Efficient induction of probabilistic word classes with LDA</i>	
Grzegorz Chrupala.....	363
<i>Quality-biased Ranking of Short Texts in Microblogging Services</i>	
Minlie Huang, Yi Yang and Xiaoyan Zhu .....	373
<i>Labeling Unlabeled Data using Cross-Language Guided Clustering</i>	
Sachindra Joshi, Danish Contractor and Sumit Negi .....	383
<i>Extracting Relation Descriptors with Conditional Random Fields</i>	
Yaliang Li, Jing Jiang, Hai Leong Chieu and Kian Ming A. Chai .....	392
<i>Attribute Extraction from Synthetic Web Search Queries</i>	
Marius Pasca.....	401
<i>Japanese Abbreviation Expansion with Query and Clickthrough Logs</i>	
Kei Uchiumi, Mamoru Komachi, Keigo Machinaga, Toshiyuki Maezawa, Toshinori Satou and Yoshinori Kobayashi.....	410
<i>Mining Parallel Documents Using Low Bandwidth and High Precision CLIR from the Heterogeneous Web</i>	
Simon Shi, Pascale Fung, Emmanuel Prochasson, Chi-kiu Lo and Dekai Wu.....	420
<i>Crawling Back and Forth: Using Back and Out Links to Locate Bilingual Sites</i>	
Luciano Barbosa, Srinivas Bangalore and Vivek Kumar Rangarajan Sridhar .....	429
<i>Grammar Induction from Text Using Small Syntactic Prototypes</i>	
Prachya Boonkwan and Mark Steedman .....	438
<i>Transferring Syntactic Relations from English to Hindi Using Alignments on Local Word Groups</i>	
Aswarth Dara, Prashanth Mannem, Hemanth Sagar Bayyarapu and Avinesh PVS .....	447

<i>Generative Modeling of Coordination by Factoring Parallelism and Selectional Preferences</i> Daisuke Kawahara and Sadao Kurohashi .....	456
<i>Syntactic Parsing for Ranking-Based Coreference Resolution</i> Altaf Rahman and Vincent Ng .....	465
<i>TriS: A Statistical Sentence Simplifier with Log-linear Models and Margin-based Discriminative Training</i> Nguyen Bach, Qin Gao, Stephan Vogel and Alex Waibel .....	474
<i>Social Summarization via Automatically Discovered Social Context</i> Po Hu, Cheng Sun, Longfei Wu, Donghong Ji and Chong Teng .....	483
<i>Simultaneous Clustering and Noise Detection for Theme-based Summarization</i> Xiaoyan Cai, Renxian Zhang, Dehong Gao and Wenjie Li .....	491
<i>Extractive Summarization Method for Contact Center Dialogues based on Call Logs</i> Akihiro Tamura, Kai Ishikawa, Masahiro Saikou and Masaaki Tsuchida .....	500
<i>Indexing Spoken Documents with Hierarchical Semantic Structures: Semantic Tree-to-string Alignment Models</i> Xiaodan Zhu, Colin Cherry and Gerald Penn .....	509
<i>Structured and Extended Named Entity Evaluation in Automatic Speech Transcriptions</i> Olivier Galibert, Sophie Rosset, Cyril Grouin, Pierre Zweigenbaum and Ludovic Quintard .....	518
<i>Normalising Audio Transcriptions for Unwritten Languages</i> Adel Foda and Steven Bird .....	527
<i>Similarity Based Language Model Construction for Voice Activated Open-Domain Question Answering</i> Istvan Varga, Kiyonori Ohtake, Kentaro Torisawa, Stijn De Saeger, Teruhisa Misu, Shigeki Matsuda and Jun'ichi Kazama .....	536
<i>The application of chordal graphs to inferring phylogenetic trees of languages</i> Jessica Enright and Grzegorz Kondrak .....	545
<i>Cross-domain Feature Selection for Language Identification</i> Marco Lui and Timothy Baldwin .....	553
<i>A Wikipedia-LDA Model for Entity Linking with Batch Size Changing Instance Selection</i> Wei Zhang, Jian Su and Chew-Lim Tan .....	562
<i>Discovering Latent Concepts and Exploiting Ontological Features for Semantic Text Search</i> Vuong M. Ngo and Tru H. Cao .....	571
<i>CLGVSM: Adapting Generalized Vector Space Model to Cross-lingual Document Clustering</i> Guoyu Tang, Yunqing Xia, Min Zhang, Haizhou Li and Fang Zheng .....	580
<i>Thread Cleaning and Merging for Microblog Topic Detection</i> Jianfeng Zhang, Yunqing Xia, Bin Ma, Jianmin Yao and Yu Hong .....	589
<i>Training a BN-based user model for dialogue simulation with missing data</i> Stéphane Rossignol, Olivier Pietquin and Michel Iannotto .....	598
<i>Automatic identification of general and specific sentences by leveraging discourse annotations</i> Annie Louis and Ani Nenkova .....	605

<i>A POS-based Ensemble Model for Cross-domain Sentiment Classification</i> Rui Xia and Chengqing Zong .....	614
<i>Ensemble-style Self-training on Citation Classification</i> Cailing Dong and Ulrich Schäfer .....	623
<i>Back to the Roots of Genres: Text Classification by Language Function</i> Henning Wachsmuth and Kathrin Bujna .....	632
<i>Transductive Minimum Error Rate Training for Statistical Machine Translation</i> Yingcong Zhao, Shujie Liu, Yangsheng Ji, Jiajun Chen and Guodong Zhou .....	641
<i>Distributed Minimum Error Rate Training of SMT using Particle Swarm Optimization</i> Jun Suzuki, Kevin Duh and Masaaki Nagata .....	649
<i>Going Beyond Word Cooccurrences in Global Lexical Selection for Statistical Machine Translation using a Multilayer Perceptron</i> Alexandre Patry and Philippe Langlais .....	658
<i>System Combination Using Discriminative Cross-Adaptation</i> Jacob Devlin, Antti-Veikko Rosti, Sankaranarayanan Ananthakrishnan and Spyros Matsoukas .....	667
<i>Word Sense Disambiguation by Combining Labeled Data Expansion and Semi-Supervised Learning Method</i> Sanae Fujita and Akinori Fujino .....	676
<i>Combining ConceptNet and WordNet for Word Sense Disambiguation</i> Junpeng Chen and Juan Liu .....	686
<i>It Takes Two to Tango: A Bilingual Unsupervised Approach for Estimating Sense Distributions using Expectation Maximization</i> Mitesh M Khapra, Salil Joshi and Pushpak Bhattacharyya .....	695
<i>Dynamic and Static Prototype Vectors for Semantic Composition</i> Siva Reddy, Ioannis Klapaftis, Diana McCarthy and Suresh Manandhar .....	705
<i>Using Prediction from Sentential Scope to Build a Pseudo Co-Testing Learner for Event Extraction</i> Shasha Liao and Ralph Grishman .....	714
<i>Text Segmentation and Graph-based Method for Template Filling in Information Extraction</i> Ludovic Jean-Louis, Romaric Besançon and Olivier Ferret .....	723
<i>Joint Distant and Direct Supervision for Relation Extraction</i> Truc-Vien T. Nguyen and Alessandro Moschitti .....	732
<i>A Cross-lingual Annotation Projection-based Self-supervision Approach for Open Information Extraction</i> Seokhwan Kim, Minwoo Jeong, Jonghoon Lee and Gary Geunbae Lee .....	741
<i>Exploring Difficulties in Parsing Imperatives and Questions</i> Tadayoshi Hara, Takuya Matsuzaki, Yusuke Miyao and Jun'ichi Tsujii .....	749
<i>A Discriminative Approach to Japanese Zero Anaphora Resolution with Large-scale Lexicalized Case Frames</i> Ryohei Sasano and Sadao Kurohashi .....	758

<i>An Empirical Comparison of Unknown Word Prediction Methods</i> Kostadin Cholakov, Gertjan van Noord, Valia Kordoni and Yi Zhang .....	767
<i>Training Dependency Parsers from Partially Annotated Corpora</i> Daniel Flannery, Yusuke Miayo, Graham Neubig and Shinsuke Mori .....	776
<i>A Breadth-First Representation for Tree Matching in Large Scale Forest-Based Translation</i> Sumukh Ghodke, Steven Bird and Rui Zhang .....	785
<i>Bayesian Subtree Alignment Model based on Dependency Trees</i> Toshiaki Nakazawa and Sadao Kurohashi .....	794
<i>Enriching SMT Training Data via Paraphrasing</i> Wei He, Shiqi Zhao, Haifeng Wang and Ting Liu .....	803
<i>Translation Quality Indicators for Pivot-based Statistical MT</i> Michael Paul and Eiichiro Sumita .....	811
<i>Source Error-Projection for Sample Selection in Phrase-Based SMT for Resource-Poor Languages</i> Sankaranarayanan Ananthkrishnan, Shiv Vitaladevuni, Rohit Prasad and Prem Natarajan .....	819
<i>A Named Entity Recognition Method based on Decomposition and Concatenation of Word Chunks</i> Tomoya Iwakura, Hiroya Takamura and Manabu Okumura .....	828
<i>Extract Chinese Unknown Words from a Large-scale Corpus Using Morphological and Distributional Evidences</i> Kaixu Zhang, Ruining Wang, Ping Xue and Maosong Sun .....	837
<i>Entity Disambiguation Using a Markov-Logic Network</i> Hong-Jie Dai, Richard Tzong-Han Tsai and Wen-Lian Hsu .....	846
<i>Named Entity Recognition in Chinese News Comments on the Web</i> Xiaojun Wan, Liang Zong, Xiaojiang Huang, Tengfei Ma, Houping Jia, Yuqian Wu and Jianguo Xiao .....	856
<i>Clustering Semantically Equivalent Words into Cognate Sets in Multilingual Lists</i> Bradley Hauer and Grzegorz Kondrak .....	865
<i>Extending WordNet with Hypernyms and Siblings Acquired from Wikipedia</i> Ichiro Yamada, Jong-Hoon Oh, Chikara Hashimoto, Kentaro Torisawa, Jun'ichi Kazama, Stijn De Saeger and Takuya Kawada .....	874
<i>What Psycholinguists Know About Chemistry: Aligning Wiktionary and WordNet for Increased Domain Coverage</i> Christian M. Meyer and Iryna Gurevych .....	883
<i>From News to Comment: Resources and Benchmarks for Parsing the Language of Web 2.0</i> Jennifer Foster, Ozlem Cetinoglu, Joachim Wagner, Joseph Le Roux, Joakim Nivre, Deirdre Hogan and Josef van Genabith .....	893
<i>Toward Finding Semantic Relations not Written in a Single Sentence: An Inference Method using Auto-Discovered Rules</i> Masaaki Tsuchida, Kentaro Torisawa, Stijn De Saeger, Jong Hoon Oh, Jun'ichi Kazama, Chikara Hashimoto and Hayato Ohwada .....	902



<i>Fleshing it out: A Supervised Approach to MWE-token and MWE-type Classification</i> Richard Fothergill and Timothy Baldwin .....	911
<i>Identification of relations between answers with global constraints for Community-based Question Answering services</i> Hikaru Yokono, Takaaki Hasegawa, Genichiro Kikui and Manabu Okumura .....	920
<i>Automatically Generating Questions from Queries for Community-based Question Answering</i> Shiqi Zhao, Haifeng Wang, Chao Li, Ting Liu and Yi Guan .....	929
<i>Question classification based on an extended class sequential rule model</i> Zijing Hui, Juan Liu and Lumei Ouyang .....	938
<i>K2Q: Generating Natural Language Questions from Keywords with User Refinements</i> Zhicheng Zheng, Xiance Si, Edward Chang and Xiaoyan Zhu .....	947
<i>Answering Complex Questions via Exploiting Social Q&amp;A Collection</i> Youzheng Wu, Chiori Hori, Hisashi Kawai and Hideki Kashioka .....	956
<i>Safety Information Mining — What can NLP do in a disaster—</i> Graham Neubig, Yuichiroh Matsubayashi, Masato Hagiwara and Koji Murakami .....	965
<i>A Character-Level Machine Translation Approach for Normalization of SMS Abbreviations</i> Deana Pennell and Yang Liu .....	974
<i>Using Text Reviews for Product Entity Completion</i> Mrinmaya Sachan, Tanveer Faruque, L. V. Subramaniam and Mukesh Mohania .....	983
<i>Mining bilingual topic hierarchies from unaligned text</i> Sumit Negi .....	992
<i>Efficient Near-Duplicate Detection for Q&amp;A Forum</i> Yan Wu, Qi Zhang and Xuanjing Huang .....	1001
<i>A Graph-based Method for Entity Linking</i> Yuhang Guo, Wanxiang Che, Ting Liu and Sheng Li .....	1010
<i>Harvesting Related Entities with a Search Engine</i> Shuqi Sun, Shiqi Zhao, Muyun Yang, Haifeng Wang and Sheng Li .....	1019
<i>Acquiring Strongly-related Events using Predicate-argument Co-occurring Statistics and Case Frames</i> Tomohide Shibata and Sadao Kurohashi .....	1028
<i>Relevance Feedback using Latent Information</i> Jun Harashima and Sadao Kurohashi .....	1037
<i>Passage Retrieval for Information Extraction using Distant Supervision</i> Wei Xu, Ralph Grishman and Le Zhao .....	1046
<i>Using Context Inference to Improve Sentence Ordering for Multi-document Summarization</i> Peifeng Li, Guangxi Deng and Qiaoming Zhu .....	1055
<i>Enhancing extraction based summarization with outside word space</i> Christian Smith and Arne Jönsson .....	1062

<i>Shallow Discourse Parsing with Conditional Random Fields</i>	
Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi and Sara Tonelli .....	1071
<i>Relational Lasso —An Improved Method Using the Relations Among Features—</i>	
Kotaro Kitagawa and Kumiko Tanaka-Ishii .....	1080
<i>Enhance Top-down method with Meta-Classification for Very Large-scale Hierarchical Classification</i>	
Xiao-lin Wang, Hai Zhao and Bao-Liang Lu .....	1089
<i>Using Syntactic and Shallow Semantic Kernels to Improve Multi-Modality Manifold-Ranking for Topic-Focused Multi-Document Summarization</i>	
Yllias Chali, Sadid A. Hasan and Kaisar Imam .....	1098
<i>Automatic Determination of a Domain Adaptation Method for Word Sense Disambiguation Using Decision Tree Learning</i>	
Kanako Komiya and Manabu Okumura .....	1107
<i>Learning from Chinese-English Parallel Data for Chinese Tense Prediction</i>	
Feifan Liu, Fei Liu and Yang Liu .....	1116
<i>Jointly Extracting Japanese Predicate-Argument Relation with Markov Logic</i>	
Katsumasa Yoshikawa, Masayuki Asahara and Yuji Matsumoto .....	1125
<i>Word Meaning in Context: A Simple and Effective Vector Model</i>	
Stefan Thater, Hagen Fürstenaу and Manfred Pinkal .....	1134
<i>Automatic Analysis of Semantic Coherence in Academic Abstracts Written in Portuguese</i>	
Vinicius Mourão Alves de Souza and Valéria Delisandra Feltrim .....	1144
<i>Sentence Subjectivity Detection with Weakly-Supervised Learning</i>	
Chenghua Lin, Yulan He and Richard Everson .....	1153
<i>Opinion Expression Mining by Exploiting Keyphrase Extraction</i>	
Gábor Berend .....	1162
<i>Extracting Resource Terms for Sentiment Analysis</i>	
Lei Zhang and Bing Liu .....	1171
<i>Towards Context-Based Subjectivity Analysis</i>	
Farah Benamara, Baptiste Chardon, Yannick Mathieu and Vladimir Popescu .....	1180
<i>Compression Methods by Code Mapping and Code Dividing for Chinese Dictionary Stored in a Double-Array Trie</i>	
Huidan Liu, Minghua Nuo, Longlong Ma, Jian Wu and Yeping He .....	1189
<i>Functional Elements and POS Categories</i>	
Qiuye Zhao and Mitch Marcus .....	1198
<i>Joint Alignment and Artificial Data Generation: An Empirical Study of Pivot-based Machine Transliteration</i>	
Min Zhang, Xiangyu Duan, Ming Liu, Yunqing Xia and Haizhou Li .....	1207
<i>Incremental Joint POS Tagging and Dependency Parsing in Chinese</i>	
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao and Jun'ichi Tsujii .....	1216

<i>Extending the adverbial coverage of a NLP oriented resource for French</i> Elsa Tolone and Stavroula Voyatzi .....	1225
<i>Linguistic Phenomena, Analyses, and Representations: Understanding Conversion between Treebanks</i> Rajesh Bhatt, Owen Rambow and Fei Xia .....	1234
<i>Automatic Transformation of the Thai Categorical Grammar Treebank to Dependency Trees</i> Christian Rishøj, Taneth Ruangrajitpakorn, Prachya Boonkwan and Thepchai Supnithi .....	1243
<i>Parse Reranking Based on Higher-Order Lexical Dependencies</i> Zhiguo Wang and Chengqing Zong .....	1251
<i>Improving Part-of-speech Tagging for Context-free Parsing</i> Xiao Chen and Chunyu Kit .....	1260
<i>Models Cascade for Tree-Structured Named Entity Detection</i> Marco Dinarelli and Sophie Rosset .....	1269
<i>Clausal parsing helps data-driven dependency parsing: Experiments with Hindi</i> Samar Husain, Phani Gadde, Joakim Nivre and Rajeev Sangal .....	1279
<i>Word-reordering for Statistical Machine Translation Using Trigram Language Model</i> Jing He and Hongyu Liang .....	1288
<i>Extracting Hierarchical Rules from a Weighted Alignment Matrix</i> Zhaopeng Tu, Yang Liu, Qun Liu and Shouxun Lin .....	1294
<i>Integration of Reduplicated Multiword Expressions and Named Entities in a Phrase Based Statistical Machine Translation System</i> Thoudam Doren Singh and Sivaji Bandyopadhyay .....	1304
<i>Regularizing Mono- and Bi-Word Models for Word Alignment</i> Thomas Schoenemann .....	1313
<i>Parametric Weighting of Parallel Data for Statistical Machine Translation</i> Kashif Shah, Loïc Barrault and Holger Schwenk .....	1323
<i>An Effective and Robust Framework for Transliteration Exploration</i> EA-EE JAN, Niyu Ge, Shih-Hsiang Lin and Berlin Chen .....	1332

## Part B: Short Papers

<i>An Evaluation of Alternative Strategies for Implementing Dialogue Policies Using Statistical Classification and Hand-Authored Rules</i>	
David DeVault, Anton Leuski and Kenji Sagae .....	1341
<i>Reducing Asymmetry between language-pairs to Improve Alignment and Translation Quality</i>	
Rashmi Gangadharaiah .....	1346
<i>Clause-Based Reordering Constraints to Improve Statistical Machine Translation</i>	
Ananthakrishnan Ramanathan, Pushpak Bhattacharyya, Karthik Visweswariah, Kushal Ladha and Ankur Gandhe .....	1351
<i>Generalized Minimum Bayes Risk System Combination</i>	
Kevin Duh, Katsuhito Sudoh, Xianchao Wu, Hajime Tsukada and Masaaki Nagata .....	1356
<i>Enhancing scarce-resource language translation through pivot combinations</i>	
Marta R. Costa-jussà, Carlos Henríquez and Rafael E. Banchs .....	1361
<i>A Baseline System for Chinese Near-Synonym Choice</i>	
Liang-Chih Yu, Wei-Nan Chien and Shih-Ting Chen .....	1366
<i>Cluster Labelling based on Concepts in a Machine-Readable Dictionary</i>	
Fumiyo Fukumoto and Yoshimi Suzuki .....	1371
<i>Text Patterns and Compression Models for Semantic Class Learning</i>	
Chung-Yao Chuang, Yi-Hsun Lee and Wen-Lian Hsu .....	1376
<i>Potts Model on the Case Fillers for Word Sense Disambiguation</i>	
Hiroya Takamura and Manabu Okumura .....	1382
<i>Improving Word Sense Induction by Exploiting Semantic Relevance</i>	
Zhenzhong Zhang and Le Sun .....	1387
<i>Predicting Word Clipping with Latent Semantic Analysis</i>	
Julian Brooke, Tong Wang and Graeme Hirst .....	1392
<i>A Semantic Relatedness Measure Based on Combined Encyclopedic, Ontological and Collocational Knowledge</i>	
Yannis Haralambous and Vitaly Klyuev .....	1397
<i>Going Beyond Text: A Hybrid Image-Text Approach for Measuring Word Relatedness</i>	
Chee Wee Leong and Rada Mihalcea .....	1403
<i>Domain Independent Model for Product Attribute Extraction from User Reviews using Wikipedia</i>	
Sudheer Kovelamudi, Sethu Ramalingam, Arpit Sood and Vasudeva Varma .....	1408
<i>Finding Problem Solving Threads in Online Forum</i>	
Zhonghua Qu and Yang Liu .....	1413

<i>Compiling Learner Corpus Data of Linguistic Output and Language Processing in Speaking, Listening, Writing, and Reading</i>	
Katsunori Kotani, Takehiko Yoshimi, Hiroaki Nanjo and Hitoshi Isahara .....	1418
<i>Mining the Sentiment Expectation of Nouns Using Bootstrapping Method</i>	
Miaomiao Wen and Yunfang Wu .....	1423
<i>An Analysis of Questions in a Q&amp;A Site Resubmitted Based on Indications of Unclear Points of Original Questions</i>	
Masahiro Kojima, Yasuhiko Watanabe and Yoshihiro Okada .....	1428
<i>Diversifying Information Needs in Results of Question Retrieval</i>	
Yaoyun Zhang, Xiaolong Wang, Xuan Wang, Ruifeng Xu, Jun Xu and ShiXi Fan .....	1432
<i>Beyond Normalization: Pragmatics of Word Form in Text Messages</i>	
Tyler Baldwin and Joyce Chai .....	1437
<i>Chinese Discourse Relation Recognition</i>	
Hen-Hsen Huang and Hsin-Hsi Chen .....	1442
<i>Improving Chinese POS Tagging with Dependency Parsing</i>	
Zhenghua Li, Wanxiang Che and Ting Liu .....	1447
<i>Exploring self training for Hindi dependency parsing</i>	
Rahul Goutam and Bharat Ram Ambati .....	1452
<i>Reduction of Search Space to Annotate Monolingual Corpora</i>	
Prajol Shrestha, Christine Jacquin and Beatrice Daille .....	1457
<i>Toward a Parallel Corpus of Spoken Cantonese and Written Chinese</i>	
John Lee .....	1462
<i>Query Expansion for IR using Knowledge-Based Relatedness</i>	
Arantxa Otegi, Xabier Arregi and Eneko Agirre .....	1467
<i>Word Sense Disambiguation Corpora Acquisition via Confirmation Code</i>	
Wanxiang Che and Ting Liu .....	1472