Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing 2009

(ACL-IJCNLP 2009)

Singapore
2-7 August 2009

Volume 1 of 2
Table of Contents

Heterogeneous Transfer Learning for Image Clustering via the SocialWeb
Qiang Yang, Yuqiang Chen, Gui-Rong Xue, Wenyuan Dai and Yong Yu .................................................. 1

Investigations on Word Senses and Word Usages
Katrin Erk, Diana McCarthy and Nicholas Gaylord .......................................................... 10

A Comparative Study on Generalization of Semantic Roles in FrameNet
Yuichiroh Matsubayashi, Naoaki Okazaki and Jun’ichi Tsujii .................................................. 19

Unsupervised Argument Identification for Semantic Role Labeling
Omri Abend, Roi Reichart and Ari Rappoport .......................................................... 28

Brutus: A Semantic Role Labeling System Incorporating CCG, CFG, and Dependency Features
Stephen Boxwell, Dennis Mehay and Chris Brew .......................................................... 37

Exploiting Heterogeneous Treebanks for Parsing
Zheng-Yu Niu, Haifeng Wang and Hua Wu .......................................................... 46

Cross Language Dependency Parsing using a Bilingual Lexicon
Hai Zhao, Yan Song, Chunyu Kit and Guodong Zhou .......................................................... 55

Topological Field Parsing of German
Jackie Chi Kit Cheung and Gerald Penn .......................................................... 64

Unsupervised Multilingual Grammar Induction
Benjamin Snyder, Tahira Naseem and Regina Barzilay .......................................................... 73

Reinforcement Learning for Mapping Instructions to Actions

Learning Semantic Correspondences with Less Supervision
Percy Liang, Michael Jordan and Dan Klein .......................................................... 91

Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling
Daichi Mochihashi, Takeshi Yamada and Naonori Ueda .......................................................... 100

Knowing the Unseen: Estimating Vocabulary Size over Unseen Samples
Suma Bhat and Richard Sproat .......................................................... 109

A Ranking Approach to Stress Prediction for Letter-to-Phoneme Conversion
Qing Dou, Shane Bergsma, Sittichai Jiampojamarn and Grzegorz Kondrak .................................................. 118

Reducing the Annotation Effort for Letter-to-Phoneme Conversion
Kenneth Dwyer and Grzegorz Kondrak .......................................................... 127

Transliteration Alignment
Vladimir Pervouchine, Haizhou Li and Bo Lin .......................................................... 136

Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike
Bart Jongejan and Hercules Dalianis .......................................................... 145
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revisiting Pivot Language Approach for Machine Translation</td>
<td>Hua Wu and Haifeng Wang</td>
<td>154</td>
</tr>
<tr>
<td>Efficient Minimum Error Rate Training and Minimum Bayes-Risk Decoding for Translation Hypergraphs and Lattices</td>
<td>Shankar Kumar, Wolfgang Macherey, Chris Dyer and Franz Och</td>
<td>163</td>
</tr>
<tr>
<td>Forest-based Tree Sequence to String Translation Model</td>
<td>Hui Zhang, Min Zhang, Haizhou Li, Aiti Aw and Chew Lim Tan</td>
<td>172</td>
</tr>
<tr>
<td>Active Learning for Multilingual Statistical Machine Translation</td>
<td>Gholamreza Haffari and Anoop Sarkar</td>
<td>181</td>
</tr>
<tr>
<td>DEPEVAL(summ): Dependency-based Evaluation for Automatic Summaries</td>
<td>Karolina Owczarzak</td>
<td>190</td>
</tr>
<tr>
<td>Summarizing Definition from Wikipedia</td>
<td>Shiren Ye, Tat-Seng Chua and Jie LU</td>
<td>199</td>
</tr>
<tr>
<td>Automatically Generating Wikipedia Articles: A Structure-Aware Approach</td>
<td>Christina Sauper and Regina Barzilay</td>
<td>208</td>
</tr>
<tr>
<td>Learning to Tell Tales: A Data-driven Approach to Story Generation</td>
<td>Neil McIntyre and Mirella Lapata</td>
<td>217</td>
</tr>
<tr>
<td>Recognizing Stances in Online Debates</td>
<td>Swapna Somasundaran and Janyce Wiebe</td>
<td>226</td>
</tr>
<tr>
<td>Co-Training for Cross-Lingual Sentiment Classification</td>
<td>Xiaojun Wan</td>
<td>235</td>
</tr>
<tr>
<td>A Non-negative Matrix Tri-factorization Approach to Sentiment Classification with Lexical Prior Knowledge</td>
<td>Tao Li, Yi Zhang and Vikas Sindhwani</td>
<td>244</td>
</tr>
<tr>
<td>Discovering the Discriminative Views: Measuring Term Weights for Sentiment Analysis</td>
<td>Jungi Kim, Jin-Ji Li and Jong-Hyeok Lee</td>
<td>253</td>
</tr>
<tr>
<td>Compiling a Massive, Multilingual Dictionary via Probabilistic Inference</td>
<td>Mausam, Stephen Soderland, Oren Etzioni, Daniel Weld, Michael Skinner and Jeff Bilmes</td>
<td>262</td>
</tr>
<tr>
<td>A Metric-based Framework for Automatic Taxonomy Induction</td>
<td>Hui Yang and Jamie Callan</td>
<td>271</td>
</tr>
<tr>
<td>Learning with Annotation Noise</td>
<td>Eyal Beigman and Beata Beigman Klebanov</td>
<td>280</td>
</tr>
<tr>
<td>Abstraction and Generalisation in Semantic Role Labels: PropBank, VerbNet or both?</td>
<td>Paola Merlo and Lonneke van der Plas</td>
<td>288</td>
</tr>
<tr>
<td>Robust Machine Translation Evaluation with Entailment Features</td>
<td>Sebastian Pado, Michel Galley, Dan Jurafsky and Christopher D. Manning</td>
<td>297</td>
</tr>
<tr>
<td>The Contribution of Linguistic Features to Automatic Machine Translation Evaluation</td>
<td>Enrique Amigó, Jesús Giménez, Julio Gonzalo and Felisa Verdejo</td>
<td>306</td>
</tr>
</tbody>
</table>
A Syntax-Driven Bracketing Model for Phrase-Based Translation
Deyi Xiong, Min Zhang, Aiti Aw and Haizhou Li ........................................... 315

Topological Ordering of Function Words in Hierarchical Phrase-based Translation
Hendra Setiawan, Min Yen Kan, Haizhou Li and Philip Resnik ................................ 324

Phrase-Based Statistical Machine Translation as a Traveling Salesman Problem
Mikhail Zaslavskiy, Marc Dymetman and Nicola Cancedda ................................... 333

Concise Integer Linear Programming Formulations for Dependency Parsing
Andre Martins, Noah Smith and Eric Xing ......................................................... 342

Non-Projective Dependency Parsing in Expected Linear Time
Joakim Nivre ........................................................................................................... 351

Semi-supervised Learning of Dependency Parsers using Generalized Expectation Criteria
Gregory Druck, Gideon Mann and Andrew McCallum ............................................. 360

Dependency Grammar Induction via Bitext Projection Constraints
Kuzman Ganchev, Jennifer Gillenwater and Ben Taskar ......................................... 369

Cross-Domain Dependency Parsing Using a Deep Linguistic Grammar
Yi Zhang and Rui Wang ......................................................................................... 378

A Chinese-English Organization Name Translation System Using Heuristic Web Mining and Asymmetric Alignment
Fan Yang, Jun Zhao and Kang Liu ........................................................................... 387

Reducing Semantic Drift with Bagging and Distributional Similarity
Tara McIntosh and James R. Curran ......................................................................... 396

Jointly Identifying Temporal Relations with Markov Logic
Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara and Yuji Matsumoto ........ 405

Profile Based Cross-Document Coreference Using Kernelized Fuzzy Relational Clustering
Jian Huang, Sarah M. Taylor, Jonathan L. Smith, Konstantinos A. Fotiadis and C. Lee Giles . . 414

Kristen Parton, Kathleen R. McKeown, Bob Coyne, Mona T. Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Wei Yun Ma, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu and Sibel Yaman .................................................. 423

Bilingual Co-training for Monolingual Hyponymy-Relation Acquisition
Jong-Hoon Oh, Kiyotaka Uchimoto and Kentaro Torisawa ......................................... 432

Automatic Set Instance Extraction using the Web
Richard C. Wang and William W. Cohen ............................................................... 441

Extracting Lexical Reference Rules from Wikipedia
Eyal Shnarch, Libby Barak and Ido Dagan ............................................................... 450

Employing Topic Models for Pattern-based Semantic Class Discovery
Huibin Zhang, Mingjie Zhu, Shuming Shi and Ji-Rong Wen ..................................... 459

Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition
Dipanjan Das and Noah A. Smith ............................................................................ 468
Quantitative modeling of the neural representation of adjective-noun phrases to account for fMRI activation
  Kai-min K. Chang, Vladimir L. Cherkassky, Tom M. Mitchell and Marcel Adam Just ........ 638

Capturing Salience with a Trainable Cache Model for Zero-anaphora Resolution
  Ryu Iida, Kentaro Inui and Yuji Matsumoto .................................................. 647

Conundrums in Noun Phrase Coreference Resolution: Making Sense of the State-of-the-Art
  Veselin Stoyanov, Nathan Gilbert, Claire Cardie and Ellen Riloff ........................ 656

A Novel Discourse Parser Based on Support Vector Machine Classification
  David duVerle and Helmut Prendinger ............................................................... 665

Genre distinctions for discourse in the Penn TreeBank
  Bonnie Webber ................................................................................................. 674

Automatic sense prediction for implicit discourse relations in text
  Emily Pitler, Annie Louis and Ani Nenkova .................................................... 683

A Framework of Feature Selection Methods for Text Categorization
  Shoushan Li, Rui Xia, Chengqing Zong and Chu-Ren Huang .............................. 692

Mine the Easy, Classify the Hard: A Semi-Supervised Approach to Automatic Sentiment Classification
  Sajib Dasgupta and Vincent Ng ................................................................. 701

Modeling Latent Biographic Attributes in Conversational Genres
  Nikesh Garera and David Yarowsky ................................................................. 710

A Graph-based Semi-Supervised Learning for Question-Answering
  Asli Celikyilmaz, Marcus Thint and Zhiheng Huang ....................................... 719

Combining Lexical Semantic Resources with Question & Answer Archives for Translation-Based Answer Finding
  Delphine Bernhard and Iryna Gurevych ....................................................... 728

Answering Opinion Questions with Random Walks on Graphs
  Fangtao Li, Yang Tang, Minlie Huang and Xiaoyan Zhu .................................. 737

What lies beneath: Semantic and syntactic analysis of manually reconstructed spontaneous speech
  Erin Fitzgerald, Frederick Jelinek and Robert Frank ......................................... 746

Discriminative Lexicon Adaptation for Improved Character Accuracy - A New Direction in Chinese Language Modeling
  Yi-cheng Pan, Lin-shan Lee and Sadaoki Furui ............................................... 755

Improving Automatic Speech Recognition for Lectures through Transformation-based Rules Learned from Minimal Data
  Cosmin Munteanu, Gerald Penn and Xiaodan Zhu ........................................... 764

Quadratic-Time Dependency Parsing for Machine Translation
  Michel Galley and Christopher D. Manning ..................................................... 773

A Gibbs Sampler for Phrasal Synchronous Grammar Induction
  Phil Blunsom, Trevor Cohn, Chris Dyer and Miles Osborne ............................. 782
Source-Language Entailment Modeling for Translating Unknown Terms
Shachar Mirkin, Lucia Specia, Nicola Cancedda, Ido Dagan, Marc Dymetman and Idan Szpektor

Case markers and Morphology: Addressing the crux of the fluency problem in English-Hindi SMT
Ananthakrishnan Ramanathan, Hansraj Choudhary, Avishek Ghosh and Pushpak Bhattacharyya

Dependency Based Chinese Sentence Realization
Wei He, Haifeng Wang, Yuqing Guo and Ting Liu

Incorporating Information Status into Generation Ranking
Aoife Cahill and Arndt Riester

A Syntax-Free Approach to Japanese Sentence Compression
Tsutomu Hirao, Jun Suzuki and Hideki Isozaki

Application-driven Statistical Paraphrase Generation
Shiqi Zhao, Xiang Lan, Ting Liu and Sheng Li

Semi-Supervised Cause Identification from Aviation Safety Reports
Isaac Persing and Vincent Ng

SMS based Interface for FAQ Retrieval
Govind Kothari, Sumit Negi, Tanveer A. Faruquie, Venkatesan T. Chakaravarthy and L. Venkata Subramaniam

Semantic Tagging of Web Search Queries
Mehdi Manshadi and Xiao Li

Mining Bilingual Data from the Web with Adaptively Learnt Patterns
Long Jiang, Shiquan Yang, Ming Zhou, Xiaohua Liu and Qingsheng Zhu

Comparing Objective and Subjective Measures of Usability in a Human-Robot Dialogue System
Mary Ellen Foster, Manuel Giuliani and Alois Knoll

Setting Up User Action Probabilities in User Simulations for Dialog System Development
Hua Ai and Diane Litman

Dialogue Segmentation with Large Numbers of Volunteer Internet Annotators
T. Daniel Midgley

Robust Approach to Abbreviating Terms: A Discriminative Latent Variable Model with Global Information
Xu Sun, Naoaki Okazaki and Jun’ichi Tsujii

A non-contiguous Tree Sequence Alignment-based Model for Statistical Machine Translation
Jun Sun, Min Zhang and Chew Lim Tan

Better Word Alignments with Supervised ITG Models
Aria Haghighi, John Blitzer, John DeNero and Dan Klein

Confidence Measure for Word Alignment
Fei Huang
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Comparative Study of Hypothesis Alignment and its Improvement for Machine Translation System Combination</td>
<td>941</td>
</tr>
<tr>
<td>Boxing Chen, Min Zhang, Haizhou Li and Aiti Aw</td>
<td></td>
</tr>
<tr>
<td>Incremental HMM Alignment for MT System Combination</td>
<td>949</td>
</tr>
<tr>
<td>Chi-Ho Li, Xiaodong He, Yupeng Liu and Ning Xi</td>
<td></td>
</tr>
<tr>
<td>K-Best A* Parsing</td>
<td>958</td>
</tr>
<tr>
<td>Adam Pauls and Dan Klein</td>
<td></td>
</tr>
<tr>
<td>Coordinate Structure Analysis with Global Structural Constraints and Alignment-Based Local Features</td>
<td>967</td>
</tr>
<tr>
<td>Kazuo Hara, Masashi Shimbo, Hideharu Okuma and Yuji Matsumoto</td>
<td></td>
</tr>
<tr>
<td>Learning Context-Dependent Mappings from Sentences to Logical Form</td>
<td>976</td>
</tr>
<tr>
<td>Luke Zettlemoyer and Michael Collins</td>
<td></td>
</tr>
<tr>
<td>An Optimal-Time Binarization Algorithm for Linear Context-Free Rewriting Systems with Fan-Out Two</td>
<td>985</td>
</tr>
<tr>
<td>Carlos Gómez-Rodríguez and Giorgio Satta</td>
<td></td>
</tr>
<tr>
<td>A Polynomial-Time Parsing Algorithm for TT-MCTAG</td>
<td>994</td>
</tr>
<tr>
<td>Laura Kallmeyer and Giorgio Satta</td>
<td></td>
</tr>
<tr>
<td>Distant supervision for relation extraction without labeled data</td>
<td>1003</td>
</tr>
<tr>
<td>Mike Mintz, Steven Bills, Rion Snow and Daniel Jurafsky</td>
<td></td>
</tr>
<tr>
<td>Multi-Task Transfer Learning for Weakly-Supervised Relation Extraction</td>
<td>1012</td>
</tr>
<tr>
<td>Jing Jiang</td>
<td></td>
</tr>
<tr>
<td>Unsupervised Relation Extraction by Mining Wikipedia Texts Using Information from the Web</td>
<td>1021</td>
</tr>
<tr>
<td>Yulan Yan, Naoaki Okazaki, Yutaka Matsuo, Zhenglu Yang and Mitsuru Ishizuka</td>
<td></td>
</tr>
<tr>
<td>Phrase Clustering for Discriminative Learning</td>
<td>1030</td>
</tr>
<tr>
<td>Dekang Lin and Xiaoyun Wu</td>
<td></td>
</tr>
<tr>
<td>Semi-Supervised Active Learning for Sequence Labeling</td>
<td>1039</td>
</tr>
<tr>
<td>Katrin Tomanek and Udo Huhn</td>
<td></td>
</tr>
<tr>
<td>Word or Phrase? Learning Which Unit to Stress for Information Retrieval</td>
<td>1048</td>
</tr>
<tr>
<td>Young-In Song, Jung-Tae Lee and Hae-Chang Rim</td>
<td></td>
</tr>
<tr>
<td>A Generative Blog Post Retrieval Model that Uses Query Expansion based on External Collections</td>
<td>1057</td>
</tr>
<tr>
<td>Wouter Weerkamp, Krisztian Balog and Maarten de Rijke</td>
<td></td>
</tr>
<tr>
<td>Language Identification of Search Engine Queries</td>
<td>1066</td>
</tr>
<tr>
<td>Hakan Ceylan and Yookyung Kim</td>
<td></td>
</tr>
<tr>
<td>Exploiting Bilingual Information to Improve Web Search</td>
<td>1075</td>
</tr>
<tr>
<td>Wei Gao, John Blitzer, Ming Zhou and Kam-Fai Wong</td>
<td></td>
</tr>
</tbody>
</table>
Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing 2009

(ACL-IJCNLP 2009)

Singapore
2-7 August 2009

Companion

ISBN: 978-1-61738-258-1
Proceedings of the Conference Short Papers
# Table of Contents

**Variational Inference for Grammar Induction with Prior Knowledge**  
Shay Cohen and Noah A Smith .......................................................... 1

**Bypassed alignment graph for learning coordination in Japanese sentences**  
Hideharu Okuma, Kazuo Hara, Masashi Shimbo and Yuji Matsumoto .................. 5

**An Earley Parsing Algorithm for Range Concatenation Grammars**  
Laura Kallmeyer, Wolfgang Maier and Yannick Parmentier .......................... 9

**Using Syntax to Disambiguate Explicit Discourse Connectives in Text**  
Emily Pitler and Ani Nenkova ....................................................... 13

**Hybrid Approach to User Intention Modeling for Dialog Simulation**  
Sangkeun Jung, Cheongjae Lee, Kyungduk Kim and Gary Geunbae Lee ............. 17

**Homophones and Tonal Patterns in English-Chinese Transliteration**  
Oi Yee Kwong .................................................................................. 21

**Capturing Errors in Written Chinese Words**  
Chao-Lin Liu, Kan-Wen Tien, Min-Hua Lai, Yi-Hsuan Chuang and Shih-Hung Wu ........ 25

**A Novel Word Segmentation Approach for Written Languages with Word Boundary Markers**  
Han-Cheol Cho, Do-Gil Lee, Jung-Tae Lee, Pontus Stenetorp, Jun’ichi Tsujii and Hae-Chang Rim 29

**Part of Speech Tagger for Assamese Text**  
Navanath Saharia, Dhrubajyoti Das, Utpal Sharma and Jugal Kalita .................. 33

**Improving data-driven dependency parsing using large-scale LFG grammars**  
Lilja Øvrelid, Jonas Kuhn and Kathrin Spreyer ...................................... 37

**Incremental Parsing with Monotonic Adjoining Operation**  
Yoshihide Kato and Shigeki Matsubara ............................................. 41

**Bayesian Learning of a Tree Substitution Grammar**  
Matt Post and Daniel Gildea ............................................................... 45

**A Unified Single Scan Algorithm for Japanese Base Phrase Chunking and Dependency Parsing**  
Manabu Sassano and Sadao Kurohashi ............................................... 49

**Comparing the Accuracy of CCG and Penn Treebank Parsers**  
Stephen Clark and James R. Curran .................................................. 53

**A Framework for Entailed Relation Recognition**  
Dan Roth, Mark Sammons and V.G. Vinod Vydiswaran ............................... 57

**A Combination of Active Learning and Semi-supervised Learning Starting with Positive and Unlabeled Examples for Word Sense Disambiguation: An Empirical Study on Japanese Web Search Query**  
Makoto Imamura, Yasuhiro Takayama, Nobuhiro Kaji, Masashi Toyoda and Masaru Kitsuregawa 61

**Detecting Compositionality in Multi-Word Expressions.**  
Ioannis Korkontzelos and Suresh Manandhar ........................................ 65
Directional Distributional Similarity for Lexical Expansion
   Lili Kotlerman, Ido Dagan, Idan Szpektor and Maayan Zhitomirsky-Geffet ..................... 69

Generalizing over Lexical Features: Selectional Preferences for Semantic Role Classification
   Beñat Zapirain, Eneko Agirre and Lluís Marquez ..................................................... 73

A Syntactic and Lexical-Based Discourse Segmenter
   Milan Tofiloski, Julian Brooke and Maite Taboada .................................................. 77

Realistic Grammar Error Simulation using Markov Logic
   Sungjin Lee and Gary Geunbae Lee ................................................................. 81

Discriminative Approach to Predicate-Argument Structure Analysis with Zero-Anaphora Resolution
   Kenji Imamura, Kuniko Saito and Tomoko Izumi ...................................................... 85

Predicting Barge-in Utterance Errors by using Implicitly-Supervised ASR Accuracy and Barge-in Rate per User
   Kazunori Komatani and Alexander I. Rudnicky ....................................................... 89

Automatic Generation of Information-seeking Questions Using Concept Clusters
   Shuguang Li and Suresh Manandhar ................................................................. 93

Correlating Human and Automatic Evaluation of a German Surface Realiser
   Aoife Cahill .......................................................... ................................. 97

Leveraging Structural Relations for Fluent Compressions at Multiple Compression Rates
   Sourish Chaudhuri, Naman K. Gupta, Noah A. Smith and Carolyn P. Rose .......................... 101

Query-Focused Summaries or Query-Biased Summaries?
   Rahul Katragadda and Vasudeva Varma ............................................................... 105

Using Generation for Grammar Analysis and Error Detection
   Michael Goodman and Francis Bond ............................................................... 109

An Integrated Multi-document Summarization Approach based on Word Hierarchical Representation
   You Ouyang, Wenjie Li and Qin Lu ............................................................... 113

Co-Feedback Ranking for Query-Focused Summarization
   Furu Wei, Wenjie Li and Yanxiang He ............................................................... 117

Reducing SMT Rule Table with Monolingual Key Phrase
   Zhongjun He, Yao Meng, Yajuan Lü, Hao Yu and Qun Liu ........................................... 121

A Statistical Machine Translation Model Based on a Synthetic Synchronous Grammar
   Hongfei Jiang, Muyun Yang, Tiejun Zhao, Sheng Li and Bo Wang .............................. 125

English-Chinese Bi-Directional OOV Translation based on Web Mining and Supervised Learning
   Yuejie Zhang, Yang Wang and Xiangyang Xue .................................................... 129

The Backtranslation Score: Automatic MT Evaluation at the Sentence Level without Reference Translations
   Reinhard Rapp .......................................................... ................................. 133

Sub-Sentence Division for Tree-Based Machine Translation
   Hao Xiong, Wenwen Xu, Haitao Mi, Yang Liu and Qun Liu ........................................... 137
Asynchronous Binarization for Synchronous Grammars
   John DeNero, Adam Pauls and Dan Klein .......................... 141

Hidden Markov Tree Model in Dependency-based Machine Translation
   Zdenek Zabokrtsky and Martin Popel ............................. 145

Word to Sentence Level Emotion Tagging for Bengali Blogs
   Dipankar Das and Sivaji Bandyopadhyay .......................... 149

Extracting Comparative Sentences from Korean Text Documents Using Comparative Lexical Patterns and Machine Learning Techniques
   Seon Yang and Youngjoong Ko .................................... 153

Opinion and Generic Question Answering Systems: a Performance Analysis
   Alexandra Balahur, Ester Boldrini, Andrés Montoyo and Patricio Martínez-Barco .............. 157

Automatic Satire Detection: Are You Having a Laugh?
   Clint Burfoot and Timothy Baldwin .............................. 161

Hierarchical Multi-Label Text Categorization with Global Margin Maximization
   Xipeng Qiu, Wenjun Gao and Xuanjing Huang .......................... 165

Toward finer-grained sentiment identification in product reviews through linguistic and ontological analyses
   Hye-Jin Min and Jong C. Park ........................................ 169

Finding Hedges by Chasing Weasels: Hedge Detection Using Wikipedia Tags and Shallow Linguistic Features
   Viola Ganter and Michael Strube .................................. 173

Mining User Reviews: from Specification to Summarization
   Xinfan Meng and Houfeng Wang .................................... 177

An Ontology-Based Approach for Key Phrase Extraction
   Chau Q. Nguyen and Tuoi T. Phan .................................. 181

Query Segmentation Based on Eigenspace Similarity
   Chao Zhang, Nan Sun, Xia Hu, Tingzhu Huang and Tat-Seng Chua ............................. 185

Learning Semantic Categories from Clickthrough Logs
   Mamoru Komachi, Shimpei Makimoto, Kei Uchiumi and Manabu Sassano .......................... 189

A Rose is a Roos is a Ruusu: Querying Translations for Web Image Search
   Janara Christensen, Mausam and Oren Etzioni .......................... 193

Extracting Paraphrases of Technical Terms from Noisy Parallel Software Corpora
   Xiaoyin Wang, David Lo, Jing Jiang, Lu Zhang and Hong Mei ................................ 197

Mining Association Language Patterns for Negative Life Event Classification
   Liang-Chih Yu, Chien-Lung Chan, Chung-Hsien Wu and Chao-Cheng Lin ..................... 201

Automatic Compilation of Travel Information from Automatically Identified Travel Blogs
   Hidetsugu Nanba, Haruka Taguma, Takahiro Ozaki, Daisuke Kobayashi, Aya Ishino and Toshiyuki Takezawa .......................... 205
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Play the Language: Play Coreference</td>
<td>Barbora Hladká, Jiří Mírovský and Pavel Schlesinger</td>
<td>209</td>
</tr>
<tr>
<td>Chinese Term Extraction Using Different Types of Relevance</td>
<td>Yuhang Yang, Tiejun Zhao, Qin Lu, Dequan Zheng and Hao Yu</td>
<td>213</td>
</tr>
<tr>
<td>iChi: a bilingual dictionary generating tool</td>
<td>István Varga and Shoichi Yokoyama</td>
<td>217</td>
</tr>
<tr>
<td>CATiB: The Columbia Arabic Treebank</td>
<td>Nizar Habash and Ryan Roth</td>
<td>221</td>
</tr>
<tr>
<td>A Beam-Search Extraction Algorithm for Comparable Data</td>
<td>Christoph Tillmann</td>
<td>225</td>
</tr>
<tr>
<td>Optimizing Word Alignment Combination For Phrase Table Training</td>
<td>Yonggang Deng and Bowen Zhou</td>
<td>229</td>
</tr>
<tr>
<td>Bridging Morpho-Syntactic Gap between Source and Target Sentences for English-Korean Statistical Machine Translation</td>
<td>Gumwon Hong, Seung-Wook Lee and Hae-Chang Rim</td>
<td>233</td>
</tr>
<tr>
<td>Toward Smaller, Faster, and Better Hierarchical Phrase-based SMT</td>
<td>Mei Yang and Jing Zheng</td>
<td>237</td>
</tr>
<tr>
<td>Handling phrase reorderings for machine translation</td>
<td>Yizhao Ni, Craig Saunders, Sandor Szedmak and Mahesan Niranjan</td>
<td>241</td>
</tr>
<tr>
<td>Syntax is from Mars while Semantics from Venus! Insights from Spectral Analysis of Distributional Similarity Networks</td>
<td>Chris Biemann, Monojit Choudhury and Animesh Mukherjee</td>
<td>245</td>
</tr>
<tr>
<td>Introduction of a new paraphrase generation tool based on Monte-Carlo sampling</td>
<td>Jonathan Chevelu, Thomas Lavergne, Yves Lepage and Thierry Moudenc</td>
<td>249</td>
</tr>
<tr>
<td>Prediction of Thematic Rank for Structured Semantic Role Labeling</td>
<td>Weiwei Sun, Zhifang Sui and Meng Wang</td>
<td>253</td>
</tr>
<tr>
<td>Transfer Learning, Feature Selection and Word Sense Disambiguation</td>
<td>Paramveer S. Dhillon and Lyle H. Ungar</td>
<td>257</td>
</tr>
<tr>
<td>From Extractive to Abstractive Meeting Summaries: Can It Be Done by Sentence Compression?</td>
<td>Fei Liu and Yang Liu</td>
<td>261</td>
</tr>
<tr>
<td>Automatic Story Segmentation using a Bayesian Decision Framework for Statistical Models of Lexical Chain Features</td>
<td>Wai-Kit Lo, Wenying Xiong and Helen Meng</td>
<td>265</td>
</tr>
<tr>
<td>Investigating Pitch Accent Recognition in Non-native Speech</td>
<td>Gina-Anne Levow</td>
<td>269</td>
</tr>
<tr>
<td>A Stochastic Finite-State Morphological Parser for Turkish</td>
<td>Hasan Sak, Tunga Güngör and Murat Saraclar</td>
<td>273</td>
</tr>
<tr>
<td>Parsing Speech Repair without Specialized Grammar Symbols</td>
<td>Tim Miller, Luan Nguyen and William Schuler</td>
<td>277</td>
</tr>
</tbody>
</table>
Efficient Inference of CRFs for Large-Scale Natural Language Data
Minwoo Jeong, Chin-Yew Lin and Gary Geunbae Lee .................................................. 281

Iterative Scaling and Coordinate Descent Methods for Maximum Entropy
Fang-Lan Huang, Cho-Jui Hsieh, Kai-Wei Chang and Chih-Jen Lin ................................ 285

Automatic Cost Estimation for Tree Edit Distance Using Particle Swarm Optimization
Yashar Mehdad .................................................................................................................. 289

Markov Random Topic Fields
Hal Daume III .................................................................................................................. 293

Multi-Document Summarization using Sentence-based Topic Models
Dingding Wang, Shenghuo Zhu, Tao Li and Yihong Gong ................................................. 297

Validating the web-based evaluation of NLG systems
Alexander Koller, Kristina Striegnitz, Donna Byron, Justine Cassell, Robert Dale, Sara Dalzel-Job,
Johanna Moore and Jon Oberlander .............................................................................. 301

Extending a Surface Realizer to Generate Coherent Discourse
Eva Banik ......................................................................................................................... 305

The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language
Rada Mihalcea and Carlo Strapparava ............................................................................ 309

Generalizing Dependency Features for Opinion Mining
Mahesh Joshi and Carolyn Penstein-Rosé ....................................................................... 313

Graph Ranking for Sentiment Transfer
Qiong Wu, Songbo Tan and Xueqi Cheng ........................................................................ 317

The Contribution of Stylistic Information to Content-based Mobile Spam Filtering
Dae-Neung Sohn, Jung-Tae Lee and Hae-Chang Rim ...................................................... 321

Learning foci for Question Answering over Topic Maps
Alexander Mikhailian, Tiphaine Dalmas and Rani Pinchuk ............................................. 325

Do Automatic Annotation Techniques Have Any Impact on Supervised Complex Question Answering?
Yllias Chali, Sadid Hasan and Shafiq Joty ...................................................................... 329

Where’s the Verb? Correcting Machine Translation During Question Answering
Wei-Yun Ma and Kathy McKeown .................................................................................. 333

A Note on the Implementation of Hierarchical Dirichlet Processes
Phil Blunsom, Trevor Cohn, Sharon Goldwater and Mark Johnson ................................. 337

A Succinct N-gram Language Model
Taro Watanabe, Hajime Tsukada and Hideki Isozaki ...................................................... 341

Modeling Morphologically Rich Languages Using Split Words and Unstructured Dependencies
Deniz Yuret and Ergun Bicici ......................................................................................... 345

Improved Smoothing for N-gram Language Models Based on Ordinary Counts
Robert C. Moore and Chris Quirk .................................................................................. 349
Updating a Name Tagger Using Contemporary Unlabeled Data
Cristina Mota and Ralph Grishman ................................................................. 353

Arabic Cross-Document Coreference Resolution
Asad Sayeed, Tamer Elsayed, Nikesh Garera, David Alexander, Tan Xu, Doug Oard, David Yarowsky and Christine Piatko ................................................................. 357

The Impact of Query Refinement in the Web People Search Task
Javier Artiles, Julio Gonzalo and Enrique Amigó ........................................... 361

Composite Kernels For Relation Extraction
Frank Reichartz, Hannes Korte and Gerhard Paass ................................. 365

Predicting Unknown Time Arguments based on Cross-Event Propagation
Prashant Gupta and Heng Ji ............................................................................ 369
Proceedings of the Student Research Workshop
# Table of Contents

*Sense-based Interpretation of Logical Metonymy Using a Statistical Method*
Ekaterina Shutova .......................................................... 1

*Insights into Non-projectivity in Hindi*
Prashanth Mannem, Himani Chaudhry and Akshar Bharati .......... 10

*Annotating and Recognising Named Entities in Clinical Notes*
Yefeng Wang .................................................................. 18

*Paraphrase Recognition Using Machine Learning to Combine Similarity Measures*
Prodromos Malakasiotis ...................................................... 27

*A System for Semantic Analysis of Chemical Compound Names*
Henriette Engelken .......................................................... 36

*Sentence diagram generation using dependency parsing*
Elijah Mayfield ............................................................... 45

*Accurate Learning for Chinese Function Tags from Minimal Features*
Caixia Yuan, Fuji Ren and Xiaojie Wang ............................ 54

*Optimizing Language Model Information Retrieval System with Expectation Maximization Algorithm*
Justin Liang-Te Chiu and Jyun-Wei Huang ......................... 63

*Data Cleaning for Word Alignment*
Tsuyoshi Okita ............................................................... 72

*The Modulation of Cooperation and Emotion in Dialogue: The REC Corpus*
Federica Cavicchio .......................................................... 81

*Clustering Technique in Multi-Document Personal Name Disambiguation*
Chen Chen, Hu Junfeng and Wang Houfeng ......................... 88

*Creating a Gold Standard for Sentence Clustering in Multi-Document Summarization*
Johanna Geiss ............................................................... 96
Proceedings of Software Demonstrations
# Table of Contents

**WISDOM: A Web Information Credibility Analysis Systematic**
Susumu Akamine, Daisuke Kawahara, Yoshikiyo Kato, Tetsuji Nakagawa, Kentaro Inui, Sadao Kurohashi and Yutaka Kidawara ................................................................. 1

**LX-Center: a center of online linguistic services**
António Branco, Francisco Costa, Eduardo Ferreira, Pedro Martins, Filipe Nunes, João Silva and Sara Silveira ................................................................. 5

**A Tool for Deep Semantic Encoding of Narrative Texts**
David K. Elson and Kathleen R. McKeown ................................................................. 9

**ProLiV - a Tool for Teaching by Viewing Computational Linguistics**
Monica Gavrila and Cristina Vertan ................................................................. 13

**A Web-Based Interactive Computer Aided Translation Tool**
Philipp Koehn .......................... 17

**MARS: Multilingual Access and Retrieval System with Enhanced Query Translation and Document Retrieval**
Lianhau Lee, Aiti Aw, Thuy Vu, Sharifah Aljunied Mahani, Min Zhang and Haizhou Li ........ 21

**Demonstration of Joshua: An Open Source Toolkit for Parsing-based Machine Translation**
Zhifei Li, Chris Callison-Burch, Chris Dyery, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz, Wren N. G. Thornton, Jonathan Weese and Omar F. Zaidan ................................. 25

**WikiBABEL: A Wiki-style Platform for Creation of Parallel Data**
A. Kumaran, K. Saravanan, Naren Datha, B. Ashok and Vikram Dendi ........................ 29

**System for Querying Syntactically Annotated Corpora**
Petr Pajas and Jan Štěpánek ................................................................. 33

**A NLG-based Application for Walking Directions**
Michael Roth and Anette Frank ................................................................. 37

**Combining POMDPs trained with User Simulations and Rule-based Dialogue Management in a Spoken Dialogue System**
Sebastian Varges, Silvia Quarteroni, Giuseppe Riccardi, Alexei V. Ivanov and Pierluigi Roberti . 41
Tutorial Abstracts
## Table of Contents

*Fundamentals of Chinese Language Processing*
  Chu-Ren Huang and Qin Lu ................................................................. 1

*Topics in Statistical Machine Translation*
  Kevin Knight and Philipp Koehn ......................................................... 2

*Semantic Role Labeling: Past, Present and Future*
  Lluís Márquez ..................................................................................... 3

*Computational Modeling of Human Language Acquisition*
  Afra Alishahi ...................................................................................... 4

*Learning to Rank*
  Hang Li ............................................................................................... 5

*State-of-the-art NLP Approaches to Coreference Resolution: Theory and Practical Recipes*
  Simone Paolo Ponzetto and Massimo Poesio ........................................ 6