

**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

ISBN: 978-1-61738-258-1

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© (2009) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2010)

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

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
Web: www.proceedings.com

Table of Contents

<i>Heterogeneous Transfer Learning for Image Clustering via the SocialWeb</i> Qiang Yang, Yuqiang Chen, Gui-Rong Xue, Wenyuan Dai and Yong Yu	1
<i>Investigations on Word Senses and Word Usages</i> Katrin Erk, Diana McCarthy and Nicholas Gaylord	10
<i>A Comparative Study on Generalization of Semantic Roles in FrameNet</i> Yuichiroh Matsubayashi, Naoaki Okazaki and Jun'ichi Tsujii	19
<i>Unsupervised Argument Identification for Semantic Role Labeling</i> Omri Abend, Roi Reichart and Ari Rappoport	28
<i>Brutus: A Semantic Role Labeling System Incorporating CCG, CFG, and Dependency Features</i> Stephen Boxwell, Dennis Mehay and Chris Brew	37
<i>Exploiting Heterogeneous Treebanks for Parsing</i> Zheng-Yu Niu, Haifeng Wang and Hua Wu	46
<i>Cross Language Dependency Parsing using a Bilingual Lexicon</i> Hai Zhao, Yan Song, Chunyu Kit and Guodong Zhou	55
<i>Topological Field Parsing of German</i> Jackie Chi Kit Cheung and Gerald Penn	64
<i>Unsupervised Multilingual Grammar Induction</i> Benjamin Snyder, Tahira Naseem and Regina Barzilay	73
<i>Reinforcement Learning for Mapping Instructions to Actions</i> S.R.K. Branavan, Harr Chen, Luke Zettlemoyer and Regina Barzilay	82
<i>Learning Semantic Correspondences with Less Supervision</i> Percy Liang, Michael Jordan and Dan Klein	91
<i>Bayesian Unsupervised Word Segmentation with Nested Pitman-Yor Language Modeling</i> Daichi Mochihashi, Takeshi Yamada and Naonori Ueda	100
<i>Knowing the Unseen: Estimating Vocabulary Size over Unseen Samples</i> Suma Bhat and Richard Sproat	109
<i>A Ranking Approach to Stress Prediction for Letter-to-Phoneme Conversion</i> Qing Dou, Shane Bergsma, Sittichai Jiampojarn and Grzegorz Kondrak	118
<i>Reducing the Annotation Effort for Letter-to-Phoneme Conversion</i> Kenneth Dwyer and Grzegorz Kondrak	127
<i>Transliteration Alignment</i> Vladimir Pervouchine, Haizhou Li and Bo Lin	136
<i>Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike</i> Bart Jongejan and Hercules Dalianis	145

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

<i>A Syntax-Driven Bracketing Model for Phrase-Based Translation</i> Deyi Xiong, Min Zhang, Aiti Aw and Haizhou Li	315
<i>Topological Ordering of Function Words in Hierarchical Phrase-based Translation</i> Hendra Setiawan, Min Yen Kan, Haizhou Li and Philip Resnik	324
<i>Phrase-Based Statistical Machine Translation as a Traveling Salesman Problem</i> Mikhail Zaslavskiy, Marc Dymetman and Nicola Cancedda	333
<i>Concise Integer Linear Programming Formulations for Dependency Parsing</i> Andre Martins, Noah Smith and Eric Xing	342
<i>Non-Projective Dependency Parsing in Expected Linear Time</i> Joakim Nivre	351
<i>Semi-supervised Learning of Dependency Parsers using Generalized Expectation Criteria</i> Gregory Druck, Gideon Mann and Andrew McCallum	360
<i>Dependency Grammar Induction via Bitext Projection Constraints</i> Kuzman Ganchev, Jennifer Gillenwater and Ben Taskar	369
<i>Cross-Domain Dependency Parsing Using a Deep Linguistic Grammar</i> Yi Zhang and Rui Wang	378
<i>A Chinese-English Organization Name Translation System Using Heuristic Web Mining and Asymmetric Alignment</i> Fan Yang, Jun Zhao and Kang Liu	387
<i>Reducing Semantic Drift with Bagging and Distributional Similarity</i> Tara McIntosh and James R. Curran	396
<i>Jointly Identifying Temporal Relations with Markov Logic</i> Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara and Yuji Matsumoto	405
<i>Profile Based Cross-Document Coreference Using Kernelized Fuzzy Relational Clustering</i> Jian Huang, Sarah M. Taylor, Jonathan L. Smith, Konstantinos A. Fotiadis and C. Lee Giles ..	414
<i>Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task</i> 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
<i>Bilingual Co-Training for Monolingual Hyponymy-Relation Acquisition</i> Jong-Hoon Oh, Kiyotaka Uchimoto and Kentaro Torisawa	432
<i>Automatic Set Instance Extraction using the Web</i> Richard C. Wang and William W. Cohen	441
<i>Extracting Lexical Reference Rules from Wikipedia</i> Eyal Shnarch, Libby Barak and Ido Dagan	450
<i>Employing Topic Models for Pattern-based Semantic Class Discovery</i> Huibin Zhang, Mingjie Zhu, Shuming Shi and Ji-Rong Wen	459
<i>Paraphrase Identification as Probabilistic Quasi-Synchronous Recognition</i> Dipanjan Das and Noah A. Smith	468

<i>Stochastic Gradient Descent Training for L1-regularized Log-linear Models with Cumulative Penalty</i> Yoshimasa Tsuruoka, Jun'ichi Tsujii and Sophia Ananiadou	477
<i>A global model for joint lemmatization and part-of-speech prediction</i> Kristina Toutanova and Colin Cherry	486
<i>Distributional Representations for Handling Sparsity in Supervised Sequence-Labeling</i> Fei Huang and Alexander Yates	495
<i>Minimized Models for Unsupervised Part-of-Speech Tagging</i> Sujith Ravi and Kevin Knight	504
<i>An Error-Driven Word-Character Hybrid Model for Joint Chinese Word Segmentation and POS Tagging</i> Canasai Kruengkrai, Kiyotaka Uchimoto, Jun'ichi Kazama, Yiou Wang, Kentaro Torisawa and Hitoshi Isahara	513
<i>Automatic Adaptation of Annotation Standards: Chinese Word Segmentation and POS Tagging – A Case Study</i> Wenbin Jiang, Liang Huang and Qun Liu	522
<i>Linefeed Insertion into Japanese Spoken Monologue for Captioning</i> Tomohiro Ohno, Masaki Murata and Shigeki Matsubara	531
<i>Semi-supervised Learning for Automatic Prosodic Event Detection Using Co-training Algorithm</i> Je Hun Jeon and Yang Liu	540
<i>Summarizing multiple spoken documents: finding evidence from untranscribed audio</i> Xiaodan Zhu, Gerald Penn and Frank Rudzicz	549
<i>Improving Tree-to-Tree Translation with Packed Forests</i> Yang Liu, Yajuan Lü and Qun Liu	558
<i>Fast Consensus Decoding over Translation Forests</i> John DeNero, David Chiang and Kevin Knight	567
<i>Joint Decoding with Multiple Translation Models</i> Yang Liu, Haitao Mi, Yang Feng and Qun Liu	576
<i>Collaborative Decoding: Partial Hypothesis Re-ranking Using Translation Consensus between De- coders</i> Mu Li, Nan Duan, Dongdong Zhang, Chi-Ho Li and Ming Zhou	585
<i>Variational Decoding for Statistical Machine Translation</i> Zhifei Li, Jason Eisner and Sanjeev Khudanpur	593
<i>Unsupervised Learning of Narrative Schemas and their Participants</i> Nathanael Chambers and Dan Jurafsky	602
<i>Learning a Compositional Semantic Parser using an Existing Syntactic Parser</i> Ruifang Ge and Raymond Mooney	611
<i>Latent Variable Models of Concept-Attribute Attachment</i> Joseph Reisinger and Marius Pasca	620
<i>The Chinese Aspect Generation Based on Aspect Selection Functions</i> Guowen Yang and John Bateman	629

<i>Quantitative modeling of the neural representation of adjective-noun phrases to account for fMRI activation</i>	
Kai-min K. Chang, Vladimir L. Cherkassky, Tom M. Mitchell and Marcel Adam Just	638
<i>Capturing Saliency with a Trainable Cache Model for Zero-anaphora Resolution</i>	
Ryu Iida, Kentaro Inui and Yuji Matsumoto	647
<i>Conundrums in Noun Phrase Coreference Resolution: Making Sense of the State-of-the-Art</i>	
Veselin Stoyanov, Nathan Gilbert, Claire Cardie and Ellen Riloff	656
<i>A Novel Discourse Parser Based on Support Vector Machine Classification</i>	
David duVerle and Helmut Prendinger	665
<i>Genre distinctions for discourse in the Penn TreeBank</i>	
Bonnie Webber	674
<i>Automatic sense prediction for implicit discourse relations in text</i>	
Emily Pitler, Annie Louis and Ani Nenkova	683
<i>A Framework of Feature Selection Methods for Text Categorization</i>	
Shoushan Li, Rui Xia, Chengqing Zong and Chu-Ren Huang	692
<i>Mine the Easy, Classify the Hard: A Semi-Supervised Approach to Automatic Sentiment Classification</i>	
Sajib Dasgupta and Vincent Ng	701
<i>Modeling Latent Biographic Attributes in Conversational Genres</i>	
Nikesh Garera and David Yarowsky	710
<i>A Graph-based Semi-Supervised Learning for Question-Answering</i>	
Asli Celikyilmaz, Marcus Thint and Zhiheng Huang	719
<i>Combining Lexical Semantic Resources with Question & Answer Archives for Translation-Based Answer Finding</i>	
Delphine Bernhard and Iryna Gurevych	728
<i>Answering Opinion Questions with Random Walks on Graphs</i>	
Fangtao Li, Yang Tang, Minlie Huang and Xiaoyan Zhu	737
<i>What lies beneath: Semantic and syntactic analysis of manually reconstructed spontaneous speech</i>	
Erin Fitzgerald, Frederick Jelinek and Robert Frank	746
<i>Discriminative Lexicon Adaptation for Improved Character Accuracy - A New Direction in Chinese Language Modeling</i>	
Yi-cheng Pan, Lin-shan Lee and Sadaoki Furui	755
<i>Improving Automatic Speech Recognition for Lectures through Transformation-based Rules Learned from Minimal Data</i>	
Cosmin Munteanu, Gerald Penn and Xiaodan Zhu	764
<i>Quadratic-Time Dependency Parsing for Machine Translation</i>	
Michel Galley and Christopher D. Manning	773
<i>A Gibbs Sampler for Phrasal Synchronous Grammar Induction</i>	
Phil Blunsom, Trevor Cohn, Chris Dyer and Miles Osborne	782

<i>Source-Language Entailment Modeling for Translating Unknown Terms</i>	
Shachar Mirkin, Lucia Specia, Nicola Cancedda, Ido Dagan, Marc Dymetman and Idan Szpektor	
	791
<i>Case markers and Morphology: Addressing the crux of the fluency problem in English-Hindi SMT</i>	
Ananthakrishnan Ramanathan, Hansraj Choudhary, Avishek Ghosh and Pushpak Bhattacharyya	800
<i>Dependency Based Chinese Sentence Realization</i>	
Wei He, Haifeng Wang, Yuqing Guo and Ting Liu	809
<i>Incorporating Information Status into Generation Ranking</i>	
Aoife Cahill and Arndt Riester	817
<i>A Syntax-Free Approach to Japanese Sentence Compression</i>	
Tsutomu Hirao, Jun Suzuki and Hideki Isozaki	826
<i>Application-driven Statistical Paraphrase Generation</i>	
Shiqi Zhao, Xiang Lan, Ting Liu and Sheng Li	834
<i>Semi-Supervised Cause Identification from Aviation Safety Reports</i>	
Isaac Persing and Vincent Ng	843
<i>SMS based Interface for FAQ Retrieval</i>	
Govind Kothari, Sumit Negi, Tanveer A. Faruque, Venkatesan T. Chakaravarthy and L. Venkata Subramaniam	852
<i>Semantic Tagging of Web Search Queries</i>	
Mehdi Manshadi and Xiao Li	861
<i>Mining Bilingual Data from the Web with Adaptively Learnt Patterns</i>	
Long Jiang, Shiquan Yang, Ming Zhou, Xiaohua Liu and Qingsheng Zhu	870
<i>Comparing Objective and Subjective Measures of Usability in a Human-Robot Dialogue System</i>	
Mary Ellen Foster, Manuel Giuliani and Alois Knoll	879
<i>Setting Up User Action Probabilities in User Simulations for Dialog System Development</i>	
Hua Ai and Diane Litman	888
<i>Dialogue Segmentation with Large Numbers of Volunteer Internet Annotators</i>	
T. Daniel Midgley	897
<i>Robust Approach to Abbreviating Terms: A Discriminative Latent Variable Model with Global Information</i>	
Xu Sun, Naoaki Okazaki and Jun'ichi Tsujii	905
<i>A non-contiguous Tree Sequence Alignment-based Model for Statistical Machine Translation</i>	
Jun Sun, Min Zhang and Chew Lim Tan	914
<i>Better Word Alignments with Supervised ITG Models</i>	
Aria Haghighi, John Blitzer, John DeNero and Dan Klein	923
<i>Confidence Measure for Word Alignment</i>	
Fei Huang	932

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

**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

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© (2009) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2010)

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

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
Web: www.proceedings.com

Proceedings of the Conference Short Papers

Table of Contents

<i>Variational Inference for Grammar Induction with Prior Knowledge</i> Shay Cohen and Noah A Smith	1
<i>Bypassed alignment graph for learning coordination in Japanese sentences</i> Hideharu Okuma, Kazuo Hara, Masashi Shimbo and Yuji Matsumoto	5
<i>An Earley Parsing Algorithm for Range Concatenation Grammars</i> Laura Kallmeyer, Wolfgang Maier and Yannick Parmentier	9
<i>Using Syntax to Disambiguate Explicit Discourse Connectives in Text</i> Emily Pitler and Ani Nenkova	13
<i>Hybrid Approach to User Intention Modeling for Dialog Simulation</i> Sangkeun Jung, Cheongjae Lee, Kyungduk Kim and Gary Geunbae Lee	17
<i>Homophones and Tonal Patterns in English-Chinese Transliteration</i> Oi Yee Kwong	21
<i>Capturing Errors in Written Chinese Words</i> Chao-Lin Liu, Kan-Wen Tien, Min-Hua Lai, Yi-Hsuan Chuang and Shih-Hung Wu	25
<i>A Novel Word Segmentation Approach for Written Languages with Word Boundary Markers</i> Han-Cheol Cho, Do-Gil Lee, Jung-Tae Lee, Pontus Stenetorp, Jun'ichi Tsujii and Hae-Chang Rim	29
<i>Part of Speech Tagger for Assamese Text</i> Navanath Saharia, Dhrubajyoti Das, Utpal Sharma and Jugal Kalita	33
<i>Improving data-driven dependency parsing using large-scale LFG grammars</i> Lilja Øvrelid, Jonas Kuhn and Kathrin Spreyer	37
<i>Incremental Parsing with Monotonic Adjoining Operation</i> Yoshihide Kato and Shigeki Matsubara	41
<i>Bayesian Learning of a Tree Substitution Grammar</i> Matt Post and Daniel Gildea	45
<i>A Unified Single Scan Algorithm for Japanese Base Phrase Chunking and Dependency Parsing</i> Manabu Sassano and Sadao Kurohashi	49
<i>Comparing the Accuracy of CCG and Penn Treebank Parsers</i> Stephen Clark and James R. Curran	53
<i>A Framework for Entailed Relation Recognition</i> Dan Roth, Mark Sammons and V.G.Vinod Vydiswaran	57
<i>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</i> Makoto Imamura, Yasuhiro Takayama, Nobuhiro Kaji, Masashi Toyoda and Masaru Kitsuregawa	61
<i>Detecting Compositionality in Multi-Word Expressions.</i> Ioannis Korkontzelos and Suresh Manandhar	65

<i>Directional Distributional Similarity for Lexical Expansion</i>	
Lili Kotlerman, Ido Dagan, Idan Szpektor and Maayan Zhitomirsky-Geffet	69
<i>Generalizing over Lexical Features: Selectional Preferences for Semantic Role Classification</i>	
Beñat Zapiain, Eneko Agirre and Lluís Màrquez	73
<i>A Syntactic and Lexical-Based Discourse Segmenter</i>	
Milan Tofiloski, Julian Brooke and Maite Taboada	77
<i>Realistic Grammar Error Simulation using Markov Logic</i>	
Sungjin Lee and Gary Geunbae Lee	81
<i>Discriminative Approach to Predicate-Argument Structure Analysis with Zero-Anaphora Resolution</i>	
Kenji Imamura, Kuniko Saito and Tomoko Izumi	85
<i>Predicting Barge-in Utterance Errors by using Implicitly-Supervised ASR Accuracy and Barge-in Rate per User</i>	
Kazunori Komatani and Alexander I. Rudnicky	89
<i>Automatic Generation of Information-seeking Questions Using Concept Clusters</i>	
Shuguang Li and Suresh Manandhar	93
<i>Correlating Human and Automatic Evaluation of a German Surface Realiser</i>	
Aoife Cahill	97
<i>Leveraging Structural Relations for Fluent Compressions at Multiple Compression Rates</i>	
Sourish Chaudhuri, Naman K. Gupta, Noah A. Smith and Carolyn P. Rose	101
<i>Query-Focused Summaries or Query-Biased Summaries?</i>	
Rahul Katragadda and Vasudeva Varma	105
<i>Using Generation for Grammar Analysis and Error Detection</i>	
Michael Goodman and Francis Bond	109
<i>An Integrated Multi-document Summarization Approach based on Word Hierarchical Representation</i>	
You Ouyang, Wenjie Li and Qin Lu	113
<i>Co-Feedback Ranking for Query-Focused Summarization</i>	
Furu Wei, Wenjie Li and Yanxiang He	117
<i>Reducing SMT Rule Table with Monolingual Key Phrase</i>	
Zhongjun He, Yao Meng, Yajuan Lü, Hao Yu and Qun Liu	121
<i>A Statistical Machine Translation Model Based on a Synthetic Synchronous Grammar</i>	
Hongfei Jiang, Muyun Yang, Tiejun Zhao, Sheng Li and Bo Wang	125
<i>English-Chinese Bi-Directional OOV Translation based on Web Mining and Supervised Learning</i>	
Yuejie Zhang, Yang Wang and Xiangyang Xue	129
<i>The Backtranslation Score: Automatic MT Evaluation at the Sentence Level without Reference Translations</i>	
Reinhard Rapp	133
<i>Sub-Sentence Division for Tree-Based Machine Translation</i>	
Hao Xiong, Wenwen Xu, Haitao Mi, Yang Liu and Qun Liu	137

<i>Asynchronous Binarization for Synchronous Grammars</i>	
John DeNero, Adam Pauls and Dan Klein	141
<i>Hidden Markov Tree Model in Dependency-based Machine Translation</i>	
Zdenek Zabokrtsky and Martin Popel	145
<i>Word to Sentence Level Emotion Tagging for Bengali Blogs</i>	
Dipankar Das and Sivaji Bandyopadhyay	149
<i>Extracting Comparative Sentences from Korean Text Documents Using Comparative Lexical Patterns and Machine Learning Techniques</i>	
Seon Yang and Youngjoong Ko	153
<i>Opinion and Generic Question Answering Systems: a Performance Analysis</i>	
Alexandra Balahur, Ester Boldrini, Andrés Montoyo and Patricio Martínez-Barco	157
<i>Automatic Satire Detection: Are You Having a Laugh?</i>	
Clint Burfoot and Timothy Baldwin	161
<i>Hierarchical Multi-Label Text Categorization with Global Margin Maximization</i>	
Xipeng Qiu, Wenjun Gao and Xuanjing Huang	165
<i>Toward finer-grained sentiment identification in product reviews through linguistic and ontological analyses</i>	
Hye-Jin Min and Jong C. Park	169
<i>Finding Hedges by Chasing Weasels: Hedge Detection Using Wikipedia Tags and Shallow Linguistic Features</i>	
Viola Ganter and Michael Strube	173
<i>Mining User Reviews: from Specification to Summarization</i>	
Xinfan Meng and Houfeng Wang	177
<i>An Ontology-Based Approach for Key Phrase Extraction</i>	
Chau Q. Nguyen and Tuoi T. Phan	181
<i>Query Segmentation Based on Eigenspace Similarity</i>	
Chao Zhang, Nan Sun, Xia Hu, Tingzhu Huang and Tat-Seng Chua	185
<i>Learning Semantic Categories from Clickthrough Logs</i>	
Mamoru Komachi, Shimpei Makimoto, Kei Uchiumi and Manabu Sassano	189
<i>A Rose is a Roos is a Ruusu: Querying Translations for Web Image Search</i>	
Janara Christensen, Mausam and Oren Etzioni	193
<i>Extracting Paraphrases of Technical Terms from Noisy Parallel Software Corpora</i>	
Xiaoyin Wang, David Lo, Jing Jiang, Lu Zhang and Hong Mei	197
<i>Mining Association Language Patterns for Negative Life Event Classification</i>	
Liang-Chih Yu, Chien-Lung Chan, Chung-Hsien Wu and Chao-Cheng Lin	201
<i>Automatic Compilation of Travel Information from Automatically Identified Travel Blogs</i>	
Hidetsugu Nanba, Haruka Taguma, Takahiro Ozaki, Daisuke Kobayashi, Aya Ishino and Toshiyuki Takezawa	205

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

<i>Efficient Inference of CRFs for Large-Scale Natural Language Data</i>	
Minwoo Jeong, Chin-Yew Lin and Gary Geunbae Lee	281
<i>Iterative Scaling and Coordinate Descent Methods for Maximum Entropy</i>	
Fang-Lan Huang, Cho-Jui Hsieh, Kai-Wei Chang and Chih-Jen Lin	285
<i>Automatic Cost Estimation for Tree Edit Distance Using Particle Swarm Optimization</i>	
Yashar Mehdad	289
<i>Markov Random Topic Fields</i>	
Hal Daume III	293
<i>Multi-Document Summarization using Sentence-based Topic Models</i>	
Dingding Wang, Shenghuo Zhu, Tao Li and Yihong Gong	297
<i>Validating the web-based evaluation of NLG systems</i>	
Alexander Koller, Kristina Striegnitz, Donna Byron, Justine Cassell, Robert Dale, Sara Dalzel-Job, Johanna Moore and Jon Oberlander	301
<i>Extending a Surface Realizer to Generate Coherent Discourse</i>	
Eva Banik	305
<i>The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language</i>	
Rada Mihalcea and Carlo Strapparava	309
<i>Generalizing Dependency Features for Opinion Mining</i>	
Mahesh Joshi and Carolyn Penstein-Rosé	313
<i>Graph Ranking for Sentiment Transfer</i>	
Qiong Wu, Songbo Tan and Xueqi Cheng	317
<i>The Contribution of Stylistic Information to Content-based Mobile Spam Filtering</i>	
Dae-Neung Sohn, Jung-Tae Lee and Hae-Chang Rim	321
<i>Learning foci for Question Answering over Topic Maps</i>	
Alexander Mikhailian, Tiphaine Dalmás and Rani Pinchuk	325
<i>Do Automatic Annotation Techniques Have Any Impact on Supervised Complex Question Answering?</i>	
Yllias Chali, Sadid Hasan and Shafiq Joty	329
<i>Where's the Verb? Correcting Machine Translation During Question Answering</i>	
Wei-Yun Ma and Kathy McKeown	333
<i>A Note on the Implementation of Hierarchical Dirichlet Processes</i>	
Phil Blunsom, Trevor Cohn, Sharon Goldwater and Mark Johnson	337
<i>A Succinct N-gram Language Model</i>	
Taro Watanabe, Hajime Tsukada and Hideki Isozaki	341
<i>Modeling Morphologically Rich Languages Using Split Words and Unstructured Dependencies</i>	
Deniz Yuret and Ergun Bici	345
<i>Improved Smoothing for N-gram Language Models Based on Ordinary Counts</i>	
Robert C. Moore and Chris Quirk	349

<i>Updating a Name Tagger Using Contemporary Unlabeled Data</i>	
Cristina Mota and Ralph Grishman	353
<i>Arabic Cross-Document Coreference Resolution</i>	
Asad Sayeed, Tamer Elsayed, Nikesh Garera, David Alexander, Tan Xu, Doug Oard, David Yarowsky and Christine Piatko	357
<i>The Impact of Query Refinement in the Web People Search Task</i>	
Javier Artiles, Julio Gonzalo and Enrique Amigó	361
<i>Composite Kernels For Relation Extraction</i>	
Frank Reichartz, Hannes Korte and Gerhard Paass	365
<i>Predicting Unknown Time Arguments based on Cross-Event Propagation</i>	
Prashant Gupta and Heng Ji	369

Proceedings of the Student Research Workshop

Table of Contents

<i>Sense-based Interpretation of Logical Metonymy Using a Statistical Method</i> Ekaterina Shutova	1
<i>Insights into Non-projectivity in Hindi</i> Prashanth Mannem, Himani Chaudhry and Akshar Bharati	10
<i>Annotating and Recognising Named Entities in Clinical Notes</i> Yefeng Wang	18
<i>Paraphrase Recognition Using Machine Learning to Combine Similarity Measures</i> Prodromos Malakasiotis	27
<i>A System for Semantic Analysis of Chemical Compound Names</i> Henriette Engelken	36
<i>Sentence diagram generation using dependency parsing</i> Elijah Mayfield	45
<i>Accurate Learning for Chinese Function Tags from Minimal Features</i> Caixia Yuan, Fuji Ren and Xiaojie Wang	54
<i>Optimizing Language Model Information Retrieval System with Expectation Maximization Algorithm</i> Justin Liang-Te Chiu and Jyun-Wei Huang	63
<i>Data Cleaning for Word Alignment</i> Tsuyoshi Okita	72
<i>The Modulation of Cooperation and Emotion in Dialogue: The REC Corpus</i> Federica Cavicchio	81
<i>Clustering Technique in Multi-Document Personal Name Disambiguation</i> Chen Chen, Hu Junfeng and Wang Houfeng	88
<i>Creating a Gold Standard for Sentence Clustering in Multi-Document Summarization</i> Johanna Geiss	96

Proceedings of Software Demonstrations

Table of Contents

<i>WISDOM: A Web Information Credibility Analysis Systematic</i>	
Susumu Akamine, Daisuke Kawahara, Yoshikiyo Kato, Tetsuji Nakagawa, Kentaro Inui, Sadao Kurohashi and Yutaka Kidawara	1
<i>LX-Center: a center of online linguistic services</i>	
António Branco, Francisco Costa, Eduardo Ferreira, Pedro Martins, Filipe Nunes, João Silva and Sara Silveira	5
<i>A Tool for Deep Semantic Encoding of Narrative Texts</i>	
David K. Elson and Kathleen R. McKeown	9
<i>ProLiV - a Tool for Teaching by Viewing Computational Linguistics</i>	
Monica Gavrilă and Cristina Vertan	13
<i>A Web-Based Interactive Computer Aided Translation Tool</i>	
Philipp Koehn	17
<i>MARS: Multilingual Access and Retrieval System with Enhanced Query Translation and Document Retrieval</i>	
Lianhau Lee, Aiti Aw, Thuy Vu, Sharifah Aljunied Mahani, Min Zhang and Haizhou Li	21
<i>Demonstration of Joshua: An Open Source Toolkit for Parsing-based Machine Translation</i>	
Zhifei Li, Chris Callison-Burch, Chris Dyery, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz, Wren N. G. Thornton, Jonathan Weese and Omar F. Zaidan	25
<i>WikiBABEL: A Wiki-style Platform for Creation of Parallel Data</i>	
A. Kumaran, K. Saravanan, Naren Datha, B. Ashok and Vikram Dendi	29
<i>System for Querying Syntactically Annotated Corpora</i>	
Petr Pajas and Jan Štěpánek	33
<i>A NLG-based Application for Walking Directions</i>	
Michael Roth and Anette Frank	37
<i>Combining POMDPs trained with User Simulations and Rule-based Dialogue Management in a Spoken Dialogue System</i>	
Sebastian Varges, Silvia Quarteroni, Giuseppe Riccardi, Alexei V. Ivanov and Pierluigi Roberti	41

Tutorial Abstracts

Table of Contents

<i>Fundamentals of Chinese Language Processing</i>	
Chu-Ren Huang and Qin Lu	1
<i>Topics in Statistical Machine Translation</i>	
Kevin Knight and Philipp Koehn	2
<i>Semantic Role Labeling: Past, Present and Future</i>	
Lluís Màrquez	3
<i>Computational Modeling of Human Language Acquisition</i>	
Afra Alishahi	4
<i>Learning to Rank</i>	
Hang Li	5
<i>State-of-the-art NLP Approaches to Coreference Resolution: Theory and Practical Recipes</i>	
Simone Paolo Ponzetto and Massimo Poesio	6