

Table of Contents

PREFACE: GENERAL CHAIR	iii
PREFACE: PROGRAM CO-CHAIRS	v
CONFERENCE ORGANIZERS	vii
LIST OF AREA CHAIRS AND REVIEWERS	viii
CONFERENCE PROGRAM	xi
TABLE OF CONTEXTS	xvii
<i>Offline Strategies for Online Question Answering: Answering Questions Before They Are Asked</i>	
Michael Fleischman, Eduard Hovy and Abdessamad Echihabi	1
<i>Using Predicate-Argument Structures for Information Extraction</i>	
Mihai Surdeanu, Sanda Harabagiu, John Williams and Paul Aarseth	8
<i>A Noisy-Channel Approach to Question Answering</i>	
Abdessamad Echihabi and Daniel Marcu	16
<i>Fast Methods for Kernel-Based Text Analysis</i>	
Taku Kudo and Yuji Matsumoto	24
<i>Hierarchical Directed Acyclic Graph Kernel: Methods for Structured Natural Language Data</i>	
Jun Suzuki, Tsutomu Hirao, Yutaka Sasaki and Eisaku Maeda	32
<i>Generalized Algorithms for Constructing Statistical Language Models</i>	
Cyril Allauzen, Mehryar Mohri and Brian Roark	40
<i>Improving Subcategorization Acquisition Using Word Sense Disambiguation</i>	
Anna Korhonen and Judita Preiss	48
<i>Syntactic Features and Word Similarity for Supervised Metonymy Resolution</i>	
Malvina Nissim and Katja Markert	56
<i>Clustering Polysemic Subcategorization Frame Distributions Semantically</i>	
Anna Korhonen, Yuval Krymolowski and Zvika Marx	64
<i>Reliable Measures for Aligning Japanese-English News Articles and Sentences</i>	
Masao Utiyama and Hitoshi Isahara	72
<i>Loosely Tree-Based Alignment for Machine Translation</i>	
Daniel Gildea	80
<i>A Probability Model to Improve Word Alignment</i>	
Colin Cherry and Dekang Lin	88
<i>Probabilistic Parsing for German Using Sister-Head Dependencies</i>	
Amit Dubey and Frank Keller	96

<i>Integrated Shallow and Deep Parsing: TopP Meets HPSG</i> Anette Frank, Markus Becker, Berthold Crysmann, Bernd Kiefer and Ulrich Schäfer .	104
<i>Combining Deep and Shallow Approaches in Parsing German</i> Michael Schiehlen .	112
<i>Synonymous Collocation Extraction Using Translation Information</i> Hua Wu and Ming Zhou .	120
<i>Constructing Semantic Space Models from Parsed Corpora</i> Sebastian Padó and Mirella Lapata .	128
<i>Orthogonal Negation in Vector Spaces for Modelling Word-Meanings and Document Retrieval</i> Dominic Widdows .	136
<i>A Comparative Study on Reordering Constraints in Statistical Machine Translation</i> Richard Zens and Hermann Ney .	144
<i>tRuEcasIng</i> Lucian Vlad Lita, Abe Ittycheriah, Salim Roukos and Nanda Kambhatla .	152
<i>Minimum Error Rate Training in Statistical Machine Translation</i> Franz Josef Och .	160
<i>A Machine Learning Approach to Pronoun Resolution in Spoken Dialogue</i> Michael Strube and Christoph Müller .	168
<i>Coreference Resolution Using Competition Learning Approach</i> Xiaofeng Yang, Guodong Zhou, Jian Su and Chew Lim Tan .	176
<i>Generating Parallel Multilingual LFG-TAG Grammars from a MetaGrammar</i> Lionel Clément and Alexandra Kinyon .	184
<i>Compounding and Derivational Morphology in a Finite-State Setting</i> Jonas Kuhn .	192
<i>A Tabulation-Based Parsing Method that Reduces Copying</i> Gerald Penn and Cosmin Munteanu .	200
<i>Recognizing Expressions of Commonsense Psychology in English Text</i> Andrew Gordon, Abe Kazemzadeh, Anish Nair and Milena Petrova .	208
<i>Closing the Gap: Learning-Based Information Extraction Rivaling Knowledge-Engineering Methods</i> Hai Leong Chieu, Hwee Tou Ng and Yoong Keok Lee .	216
<i>An Improved Extraction Pattern Representation Model for Automatic IE Pattern Acquisition</i> Kiyoshi Sudo, Satoshi Sekine and Ralph Grishman .	224
<i>Optimizing Story Link Detection is not Equivalent to Optimizing New Event Detection</i> Ayman Farahat, Francine Chen and Thorsten Brants .	232

<i>Corpus-Based Discourse Understanding in Spoken Dialogue Systems</i> Ryuichiro Higashinaka, Mikio Nakano and Kiyooki Aikawa	240
<i>Extracting Key Semantic Terms from Chinese Speech Query for Web Searches</i> Gang Wang, Tat-Seng Chua and Yong-Cheng Wang	248
<i>Flexible Guidance Generation Using User Model in Spoken Dialogue Systems</i> Kazunori Komatani, Shinichi Ueno, Tatsuya Kawahara and Hiroshi G. Okuno	256
<i>Integrating Discourse Markers into a Pipelined Natural Language Generation Architecture</i> Charles B. Callaway	264
<i>Improved Source-Channel Models for Chinese Word Segmentation</i> Jianfeng Gao, Mu Li and Chang-Ning Huang	272
<i>Unsupervised Segmentation of Words Using Prior Distributions of Morph Length and Frequency</i> Mathias Creutz	280
<i>Parametric Models of Linguistic Count Data</i> Martin Jansche	288
<i>Self-Organizing Markov Models and Their Application to Part-of-Speech Tagging</i> Jin-Dong Kim, Hae-Chang Rim and Jun'ichi Tsujii	296
<i>Chunk-Based Statistical Translation</i> Taro Watanabe, Eiichiro Sumita and Hiroshi G. Okuno	303
<i>Feature-Rich Statistical Translation of Noun Phrases</i> Philipp Koehn and Kevin Knight	311
<i>Effective Phrase Translation Extraction from Alignment Models</i> Ashish Venugopal, Stephan Vogel and Alex Waibel	319
<i>Uncertainty Reduction in Collaborative Bootstrapping: Measure and Algorithm</i> Yunbo Cao, Hang Li and Li Lian	327
<i>A Bootstrapping Approach to Named Entity Classification Using Successive Learners</i> Cheng Niu, Wei Li, Jihong Ding and Rohini Srihari	335
<i>Counter-Training in Discovery of Semantic Patterns</i> Roman Yangarber	343
<i>k-Valued Non-Associative Lambek Categorical Grammars are not Learnable from Strings</i> Denis B�chet and Annie Foret	351
<i>Parsing with Generative Models of Predicate-Argument Structure</i> Julia Hockenmaier	359
<i>Bridging the Gap Between Underspecification Formalisms: Minimal Recursion Semantics as Dominance Constraints</i> Joachim Niehren and Stefan Thater	367

<i>Evaluation Challenges in Large-Scale Document Summarization</i> Dragomir R. Radev, Simone Teufel, Horacio Saggion, Wai Lam, John Blitzer, Hong Qi, Arda Çelebi, Danyu Liu and Elliott Drabek	375
<i>Analysis of Source Identified Text Corpora: Exploring the Statistics of the Reused Text and Authorship</i> Akiko Aizawa.....	383
<i>Unsupervised Learning of Arabic Stemming Using a Parallel Corpus</i> Monica Rogati, Scott McCarley and Yiming Yang.....	391
<i>Language Model Based Arabic Word Segmentation</i> Young-Suk Lee, Kishore Papineni, Salim Roukos, Ossama Emam and Hany Hassan..	399
<i>Acquiring Vocabulary for Predictive Text Entry through Dynamic Reuse of a Small User Corpus</i> Kumiko Tanaka-Ishii, Daichi Hayakawa and Masato Takeichi.....	407
<i>A Word-Order Database for Testing Computational Models of Language Acquisition</i> William Gregory Sakas.....	415
<i>Accurate Unlexicalized Parsing</i> Dan Klein and Christopher D. Manning.....	423
<i>Deep Syntactic Processing by Combining Shallow Methods</i> Péter Dienes and Amit Dubey	431
<i>Is it Harder to Parse Chinese, or the Chinese Treebank?</i> Roger Levy and Christopher Manning.....	439
<i>Feedback Cleaning of Machine Translation Rules Using Automatic Evaluation</i> Kenji Imamura, Eiichiro Sumita and Yuji Matsumoto	447
<i>Exploiting Parallel Texts for Word Sense Disambiguation: An Empirical Study</i> Hwee Tou Ng, Bin Wang and Yee Seng Chan	455
<i>Learning the Countability of English Nouns from Corpus Data</i> Timothy Baldwin and Francis Bond.....	463
<i>A Syllable Based Word Recognition Model for Korean Noun Extraction</i> Do-Gil Lee, Hae-Chang Rim and Heui-Seok Lim	471
<i>Morphological Analysis of a Large Spontaneous Speech Corpus in Japanese</i> Kiyotaka Uchimoto, Chikashi Nobata, Atsushi Yamada, Satoshi Sekine and Hitoshi Isahara	479
<i>Learning to Predict Pitch Accents and Prosodic Boundaries in Dutch</i> Erwin Marsi, Martin Reynaert, Antal van den Bosch, Walter Daelemans and Véronique Hoste	489
<i>Text Chunking by Combining Hand-Crafted Rules and Memory-Based Learning</i> Seong-Bae Park and Byoung-Tak Zhang.....	497

<i>A SNoW Based Supertagger with Application to NP Chunking</i> Libin Shen and Aravind K. Joshi	505
<i>An Expert Lexicon Approach to Identifying English Phrasal Verbs</i> Wei Li, Xiuhong Zhang, Cheng Niu, Yuankai Jiang and Rohini K. Srihari	513
<i>Unsupervised Learning of Dependency Structure for Language Modeling</i> Jianfeng Gao and Hisami Suzuki	521
<i>Using Model-Theoretic Semantic Interpretation to Guide Statistical Parsing and Word Recognition in a Spoken Language Interface</i> William Schuler	529
<i>Towards a Resource for Lexical Semantics: A Large German Corpus with Extensive Semantic Annotation</i> Katrin Erk, Andrea Kowalski, Sebastian Padó and Manfred Pinkal	537
<i>Probabilistic Text Structuring: Experiments with Sentence Ordering</i> Mirella Lapata	545
<i>Towards a Model of Face-to-Face Grounding</i> Yukiko Nakano, Gabe Reinstein, Tom Stocky and Justine Cassell	553
<i>Discourse Segmentation of Multi-Party Conversation</i> Michel Galley, Kathleen R. McKeown, Eric Fosler-Lussier and Hongyan Jing	562
AUTHOR INDEX	570