

HLT-NAACL 2006

**Human Language Technology
Conference of the
North American Chapter of the
Association of Computational Linguistics**

Short Papers

Robert C. Moore, General Chair
Jeff Bilmes, Jennifer Chu-Carroll and Mark Sanderson
Program Committee Chairs

June 4-9, 2006
New York, New York, USA

Published by the Association for Computational Linguistics
<http://www.aclweb.org>

Production and Manufacturing by
Omnipress Inc.
2600 Anderson Street
Madison, WI 53704

©2006 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Table of Contents

<i>Factored Neural Language Models</i>	
Andrei Alexandrescu and Katrin Kirchhoff	1
<i>The MILE Corpus for Less Commonly Taught Languages</i>	
Alison Alvarez, Lori Levin, Robert Frederking, Simon Fung, Donna Gates and Jeff Good	5
<i>Museli: A Multi-Source Evidence Integration Approach to Topic Segmentation of Spontaneous Dialogue</i>	
Jaime Arguello and Carolyn Rose	9
<i>Measuring Semantic Relatedness Using People and WordNet</i>	
Beata Beigman Klebanov	13
<i>Thai Grapheme-Based Speech Recognition</i>	
Paisarn Charoenpornasawat, Sanjika Hewavitharana and Tanja Schultz	17
<i>Class Model Adaptation for Speech Summarisation</i>	
Pierre Chatain, Edward Whittaker, Joanna Mrozinski and Sadaoki Furui	21
<i>Semi-supervised Relation Extraction with Label Propagation</i>	
Jinxu Chen, Donghong Ji, Chew Lim Tan and Zhengyu Niu	25
<i>Temporal Classification of Text and Automatic Document Dating</i>	
Angelo Dalli	29
<i>Answering the question you wish they had asked: The impact of paraphrasing for Question Answering</i>	
Pablo Duboue and Jennifer Chu-Carroll	33
<i>Gesture Improves Coreference Resolution</i>	
Jacob Eisenstein and Randall Davis	37
<i>Spectral Clustering for Example Based Machine Translation</i>	
Rashmi Gangadharaiah, Ralf Brown and Jaime Carbonell	41
<i>A Finite-State Model of Georgian Verbal Morphology</i>	
Olga Gurevich	45
<i>Arabic Preprocessing Schemes for Statistical Machine Translation</i>	
Nizar Habash and Fatiha Sadat	49
<i>Agreement/Disagreement Classification: Exploiting Unlabeled Data using Contrast Classifiers</i>	
Sangyun Hahn, Richard Ladner and Mari Ostendorf	53
<i>OntoNotes: The 90% Solution</i>	
Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw and Ralph Weischedel	57
<i>Investigating Cross-Language Speech Retrieval for a Spontaneous Conversational Speech Collection</i>	
Diana Inkpen, Muath Alzghool, Gareth Jones and Douglas Oard	61

<i>Evaluating Centering for Sentence Ordering in Two New Domains</i> Nikiforos Karamanis	65
<i>MMR-based Active Machine Learning for Bio Named Entity Recognition</i> Seokhwan Kim, Yu Song, Kyungduk Kim, Jeong-Won Cha and Gary Geunbae Lee	69
<i>Early Deletion of Fillers In Processing Conversational Speech</i> Matthew Lease and Mark Johnson	73
<i>Evaluation of Utility of LSA for Word Sense Discrimination</i> Esther Levin, Mehrbod Sharifi and Jerry Ball	77
<i>Initial Study on Automatic Identification of Speaker Role in Broadcast News Speech</i> Yang Liu	81
<i>Automatic Recognition of Personality in Conversation</i> Franois Mairesse and Marilyn Walker	85
<i>Summarizing Speech Without Text Using Hidden Markov Models</i> Sameer Maskey and Julia Hirschberg	89
<i>NER Systems that Suit User’s Preferences: Adjusting the Recall-Precision Trade-off for Entity Extraction</i> Einat Minkov, Richard Wang, Anthony Tomasic and William Cohen	93
<i>Syntactic Kernels for Natural Language Learning: the Semantic Role Labeling Case</i> Alessandro Moschitti	97
<i>Accurate Parsing of the Proposition Bank</i> Gabriele Musillo and Paola Merlo	101
<i>Using Semantic Authoring for Blissymbols Communication Boards</i> Yael Netzer and Michael Elhadad	105
<i>Extracting Salient Keywords from Instructional Videos Using Joint Text, Audio and Visual Cues</i> Youngja Park and Ying Li	109
<i>Exploiting Variant Corpora for Machine Translation</i> Michael Paul and Eiichiro Sumita	113
<i>Quantitative Methods for Classifying Writing Systems</i> Gerald Penn and Travis Choma	117
<i>Computational Modelling of Structural Priming in Dialogue</i> David Reitter, Frank Keller and Johanna D. Moore	121
<i>Story Segmentation of Broadcast News in English, Mandarin and Arabic</i> Andrew Rosenberg and Julia Hirschberg	125
<i>Parser Combination by Reparsing</i> Kenji Sagae and Alon Lavie	129

<i>Using Phrasal Patterns to Identify Discourse Relations</i> Manami Saito, Kazuhide Yamamoto and Satoshi Sekine	133
<i>Weblog Classification for Fast Splog Filtering: A URL Language Model Segmentation Approach</i> Franco Salvetti and Nicolas Nicolov	137
<i>Word Domain Disambiguation via Word Sense Disambiguation</i> Antonio Sanfilippo, Stephen Tratz and Michelle Gregory	141
<i>Selecting relevant text subsets from web-data for building topic specific language models</i> Abhinav Sethy, Panayiotis Georgiou and Shrikanth Narayanan	145
<i>A Comparison of Tagging Strategies for Statistical Information Extraction</i> Christian Siefkes	149
<i>Unsupervised Induction of Modern Standard Arabic Verb Classes</i> Neal Snider and Mona Diab	153
<i>Sentence Planning for Realtime Navigational Instruction</i> Laura Stoia, Donna Byron, Darla Shockley and Eric Fosler-Lussier	157
<i>Using the Web to Disambiguate Acronyms</i> Eiichiro Sumita and Fumiaki Sugaya	161
<i>Word Pronunciation Disambiguation using the Web</i> Eiichiro Sumita and Fumiaki Sugaya	165
<i>Illuminating Trouble Tickets with Sublanguage Theory</i> Svetlana Symonenko, Steven Rowe and Elizabeth D. Liddy	169
<i>Evolving optimal inspectable strategies for spoken dialogue systems</i> Dave Toney, Johanna Moore and Oliver Lemon	173
<i>Lycos Retriever: An Information Fusion Engine</i> Brian Ulicny	177
<i>Improved Affinity Graph Based Multi-Document Summarization</i> Xiaojun Wan and Jianwu Yang	181
<i>A Maximum Entropy Framework that Integrates Word Dependencies and Grammatical Relations for Reading Comprehension</i> Kui Xu, Helen Meng and Fuliang Weng	185
<i>BioEx: A Novel User-Interface that Accesses Images from Abstract Sentences</i> Hong Yu and Minsuk Lee	189
<i>Subword-based Tagging by Conditional Random Fields for Chinese Word Segmentation</i> Ruiqiang Zhang, Genichiro Kikui and Eiichiro Sumita	193

Comparing the roles of textual, acoustic and spoken-language features on spontaneous-conversation summarization

Xiaodan Zhu and Gerald Penn 197

Bridging the Inflection Morphology Gap for Arabic Statistical Machine Translation

Andreas Zollmann, Ashish Venugopal and Stephan Vogel 201