

NAACL HLT 2007

**Human Language Technologies 2007:
The Conference of the
North American Chapter
of the
Association for Computational Linguistics**

Companion Volume: Short Papers

Candace Sidner, General Chair
Tanja Schultz, Matthew Stone, and ChengXiang Zhai
Program Committee Chairs

22–27 April 2007
Rochester, New York, USA

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

©2007 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: +1-507-476-0860
acl@aclweb.org

Table of Contents

<i>Comparing User Simulation Models For Dialog Strategy Learning</i> Hua Ai, Joel Tetreault and Diane Litman	1
<i>Automatic Acquisition of Grammatical Types for Nouns</i> Nuria Bel, Sergio Espeja and Montserrat Marimon	5
<i>Conquest—An Open-Source Dialog System for Conferences</i> Dan Bohus, Sergio Grau Puerto, David Huggins-Daines, Venkatesh Keri, Gopala Krishna, Rohit Kumar, Antoine Raux and Stefanie Tomko	9
<i>Joint Versus Independent Phonological Feature Models within CRF Phone Recognition</i> Ilana Bromberg, Jeremy Morris and Eric Fosler-Lussier	13
<i>K-Best Suffix Arrays</i> Kenneth Church, Bo Thiesson and Robert Ragno	17
<i>Translation Model Pruning via Usage Statistics for Statistical Machine Translation</i> Matthias Eck, Stephan Vogel and Alex Waibel	21
<i>Combination of Statistical Word Alignments Based on Multiple Preprocessing Schemes</i> Jakob Elming and Nizar Habash	25
<i>A Fast Method for Parallel Document Identification</i> Jessica Enright and Grzegorz Kondrak	29
<i>Generalized Graphical Abstractions for Statistical Machine Translation</i> Karim Filali and Jeff Bilmes	33
<i>Situated Models of Meaning for Sports Video Retrieval</i> Michael Fleischman and Deb Roy	37
<i>Exploring Affect-Context Dependencies for Adaptive System Development</i> Kate Forbes-Riley, Mihai Rotaru, Diane Litman and Joel Tetreault	41
<i>A Filter-Based Approach to Detect End-of-Utterances from Prosody in Dialog Systems</i> Olac Fuentes, David Vera and Thamar Solorio	45
<i>Document Similarity Measures to Distinguish Native vs. Non-Native Essay Writers</i> Olga Gurevich and Paul Deane	49
<i>Arabic Diacritization through Full Morphological Tagging</i> Nizar Habash and Owen Rambow	53
<i>Are Very Large N-Best Lists Useful for SMT?</i> Saša Hasan, Richard Zens and Hermann Ney	57

<i>Relationship between Non-Projective Edges, Their Level Types, and Well-Nestedness</i>	
Jiří Havelka	61
<i>iROVER: Improving System Combination with Classification</i>	
Dustin Hillard, Bjoern Hoffmeister, Mari Ostendorf, Ralf Schlueter and Hermann Ney	65
<i>Clustered Sub-Matrix Singular Value Decomposition</i>	
Fang Huang and Yorick Wilks	69
<i>Implicitly Supervised Language Model Adaptation for Meeting Transcription</i>	
David Huggins-Daines and Alexander I. Rudnicky	73
<i>ILR-Based MT Comprehension Test with Multi-Level Questions</i>	
Douglas Jones, Martha Herzog, Hussny Ibrahim, Arvind Jairam, Wade Shen, Edward Gibson and Michael Emonts	77
<i>Semi-Supervised Learning for Semantic Parsing using Support Vector Machines</i>	
Rohit Kate and Raymond Mooney	81
<i>Discriminative Alignment Training without Annotated Data for Machine Translation</i>	
Patrik Lambert, Rafael E. Banchs and Josep M. Crego	85
<i>A Geometric Interpretation of Non-Target-Normalized Maximum Cross-Channel Correlation for Vocal Activity Detection in Meetings</i>	
Kornel Laskowski and Tanja Schultz	89
<i>Detection of Non-Native Sentences Using Machine-Translated Training Data</i>	
John Lee, Ming Zhou and Xiaohua Liu	93
<i>Exploiting Rich Syntactic Information for Relationship Extraction from Biomedical Articles</i>	
Yudong Liu, Zhongmin Shi and Anoop Sarkar	97
<i>Look Who is Talking: Soundbite Speaker Name Recognition in Broadcast News Speech</i>	
Feifan Liu and Yang Liu	101
<i>Tagging Icelandic Text using a Linguistic and a Statistical Tagger</i>	
Hrafn Loftsson	105
<i>Efficient Computation of Entropy Gradient for Semi-Supervised Conditional Random Fields</i>	
Gideon Mann and Andrew McCallum	109
<i>Hybrid Document Indexing with Spectral Embedding</i>	
Irina Matveeva and Gina-Anne Levow	113
<i>On using Articulatory Features for Discriminative Speaker Adaptation</i>	
Florian Metze	117
<i>RH: A Retro-Hybrid Parser</i>	
Paula Newman	121

<i>Subtree Mining for Relation Extraction from Wikipedia</i>	
Dat P.T. Nguyen, Yutaka Matsuo and Mitsuru Ishizuka	125
<i>Advances in the CMU/Interact Arabic GALE Transcription System</i>	
Mohamed Noamany, Thomas Schaaf and Tanja Schultz	129
<i>An Integrated Architecture for Speech-Input Multi-Target Machine Translation</i>	
Alicia Pérez, M. Teresa González, M. Inés Torres and Francisco Casacuberta	133
<i>Analysis and System Combination of Phrase- and N-Gram-Based Statistical Machine Translation Systems</i>	
Marta R. Costa-jussà, Josep M. Crego, David Vilar, José A. R. Fonollosa, José B. Mariño and Hermann Ney	137
<i>Stating with Certainty or Stating with Doubt: Intercoder Reliability Results for Manual Annotation of Epistemically Modalized Statements</i>	
Victoria L. Rubin	141
<i>Joint Morphological-Lexical Language Modeling for Machine Translation</i>	
Ruhi Sarikaya and Yonggang Deng	145
<i>Agenda-Based User Simulation for Bootstrapping a POMDP Dialogue System</i>	
Jost Schatzmann, Blaise Thomson, Karl Weilhammer, Hui Ye and Steve Young	149
<i>Reversible Sound-to-Letter/Letter-to-Sound Modeling Based on Syllable Structure</i>	
Stephanie Seneff	153
<i>Are Some Speech Recognition Errors Easier to Detect than Others?</i>	
Yongmei Shi and Lina Zhou	157
<i>Simultaneous Identification of Biomedical Named-Entity and Functional Relation Using Statistical Parsing Techniques</i>	
Zhongmin Shi, Anoop Sarkar and Fred Popowich	161
<i>Virtual Evidence for Training Speech Recognizers Using Partially Labeled Data</i>	
Amarnag Subramanya and Jeff Bilmes	165
<i>A High Accuracy Method for Semi-Supervised Information Extraction</i>	
Stephen Tratz and Antonio Sanfilippo	169
<i>The Effects of Word Prediction on Communication Rate for AAC</i>	
Keith Trnka, Debra Yarrington, John McCaw, Kathleen F. McCoy and Christopher Pennington	173
<i>Language Modeling for Determiner Selection</i>	
Jenine Turner and Eugene Charniak	177
<i>Entity Extraction is a Boring Solved Problem—Or is it?</i>	
Marc Vilain, Jennifer Su and Suzi Lubar	181

<i>Kernel Regression Based Machine Translation</i>	
Zhuoran Wang, John Shawe-Taylor and Sandor Szedmak	185
<i>Modifying SO-PMI for Japanese Weblog Opinion Mining by Using a Balancing Factor and Detecting Neutral Expressions</i>	
Guangwei Wang and Kenji Araki	189
<i>Combined Use of Speaker- and Tone-Normalized Pitch Reset with Pause Duration for Automatic Story Segmentation in Mandarin Broadcast News</i>	
Lei Xie, Chuan Liu and Helen Meng	193
<i>Chinese Named Entity Recognition with Cascaded Hybrid Model</i>	
Xiaofeng Yu	197
<i>A Three-Step Deterministic Parser for Chinese Dependency Parsing</i>	
Kun Yu, Sadao Kurohashi and Hao Liu	201
<i>Comparing Wikipedia and German Wordnet by Evaluating Semantic Relatedness on Multiple Datasets</i>	
Torsten Zesch, Iryna Gurevych and Max Mühlhäuser	205
<i>Selective Phrase Pair Extraction for Improved Statistical Machine Translation</i>	
Luke Zettlemoyer and Robert Moore	209
<i>Speech Summarization Without Lexical Features for Mandarin Broadcast News</i>	
Jian Zhang and Pascale Fung	213
<i>A Semi-Automatic Evaluation Scheme: Automated Nuggetization for Manual Annotation</i>	
Liang Zhou, Namhee Kwon and Eduard Hovy	217