

Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics

(NAACL HLT 2010)

**Los Angeles, California, USA
2-4 June 2010**

ISBN: 978-1-61738-669-5

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© (2010) 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>Invited Talk: Recognition and Understanding of Meetings</i>	
Steve Renals	1
<i>Chart Mining-based Lexical Acquisition with Precision Grammars</i>	
Yi Zhang, Timothy Baldwin, Valia Kordoni, David Martinez and Jeremy Nicholson	10
<i>Products of Random Latent Variable Grammars</i>	
Slav Petrov	19
<i>Automatic Domain Adaptation for Parsing</i>	
David McClosky, Eugene Charniak and Mark Johnson	28
<i>Appropriately Handled Prosodic Breaks Help PCFG Parsing</i>	
Zhongqiang Huang and Mary Harper	37
<i>Using Confusion Networks for Speech Summarization</i>	
Shasha Xie and Yang Liu	46
<i>Qme! : A Speech-based Question-Answering system on Mobile Devices</i>	
Taniya Mishra and Srinivas Bangalore	55
<i>Dialogue-Oriented Review Summary Generation for Spoken Dialogue Recommendation Systems</i>	
Jingjing Liu, Stephanie Seneff and Victor Zue	64
<i>Minimally-Supervised Extraction of Entities from Text Advertisements</i>	
Sameer Singh, Dustin Hillard and Chris Leggetter	73
<i>Taxonomy Learning Using Word Sense Induction</i>	
Ioannis P. Klapaftis and Suresh Manandhar	82
<i>Visual Information in Semantic Representation</i>	
Yansong Feng and Mirella Lapata	91
<i>Automatic Evaluation of Topic Coherence</i>	
David Newman, Jey Han Lau, Karl Grieser and Timothy Baldwin	100
<i>Multi-Prototype Vector-Space Models of Word Meaning</i>	
Joseph Reisinger and Raymond J. Mooney	109
<i>Unsupervised Syntactic Alignment with Inversion Transduction Grammars</i>	
Adam Pauls, Dan Klein, David Chiang and Kevin Knight	118
<i>Joint Parsing and Alignment with Weakly Synchronized Grammars</i>	
David Burkett, John Blitzer and Dan Klein	127
<i>Learning Translation Boundaries for Phrase-Based Decoding</i>	
Deyi Xiong, Min Zhang and Haizhou Li	136

<i>Hitting the Right Paraphrases in Good Time</i> Stanley Kok and Chris Brockett	145
<i>Training Paradigms for Correcting Errors in Grammar and Usage</i> Alla Rozovskaya and Dan Roth	154
<i>Using Mostly Native Data to Correct Errors in Learners' Writing</i> Michael Gamon	163
<i>Unsupervised Modeling of Twitter Conversations</i> Alan Ritter, Colin Cherry and Bill Dolan	172
<i>Streaming First Story Detection with application to Twitter</i> Saša Petrović, Miles Osborne and Victor Lavrenko	181
<i>Unsupervised Model Adaptation using Information-Theoretic Criterion</i> Ariya Rastrow, Frederick Jelinek, Abhinav Sethy and Bhuvana Ramabhadran	190
<i>Formatting Time-Aligned ASR Transcripts for Readability</i> Maria Shugrina	198
<i>Cheap, Fast and Good Enough: Automatic Speech Recognition with Non-Expert Transcription</i> Scott Novotney and Chris Callison-Burch	207
<i>Contextual Information Improves OOV Detection in Speech</i> Carolina Parada, Mark Dredze, Denis Filimonov and Frederick Jelinek	216
<i>Improved Extraction Assessment through Better Language Models</i> Arun Ahuja and Doug Downey	225
<i>Language Identification: The Long and the Short of the Matter</i> Timothy Baldwin and Marco Lui	229
<i>Inducing Synchronous Grammars with Slice Sampling</i> Phil Blunsom and Trevor Cohn	238
<i>Task-based Evaluation of Multiword Expressions: a Pilot Study in Statistical Machine Translation</i> Marine Carpuat and Mona Diab	242
<i>Improving Semantic Role Labeling with Word Sense</i> Wanxiang Che, Ting Liu and Yongqiang Li	246
<i>Extending the METEOR Machine Translation Evaluation Metric to the Phrase Level</i> Michael Denkowski and Alon Lavie	250
<i>Testing a Grammar Customization System with Sahaptin</i> Scott Drellishak	254
<i>Two monolingual parses are better than one (synchronous parse)</i> Chris Dyer	263

<i>Fast Query for Large Treebanks</i> Sumukh Ghodke and Steven Bird	267
<i>Efficient Parsing of Well-Nested Linear Context-Free Rewriting Systems</i> Carlos Gómez-Rodríguez, Marco Kuhlmann and Giorgio Satta	276
<i>Utility Evaluation of Cross-document Information Extraction</i> Heng Ji, Zheng Chen, Jonathan Feldman, Antonio Gonzalez, Ralph Grishman and Vivek Upadhyay	285
<i>Evaluation Metrics for the Lexical Substitution Task</i> Sanaz Jabbari, Mark Hepple and Louise Guthrie	289
<i>Movie Reviews and Revenues: An Experiment in Text Regression</i> Mahesh Joshi, Dipanjan Das, Kevin Gimpel and Noah A. Smith	293
<i>Using Gaussian Mixture Models to Detect Figurative Language in Context</i> Linlin Li and Caroline Sporleder	297
<i>Improving Phrase-Based Translation with Prototypes of Short Phrases</i> Frank Liberato, Behrang Mohit and Rebecca Hwa	301
<i>Putting the User in the Loop: Interactive Maximal Marginal Relevance for Query-Focused Summarization</i> Jimmy Lin, Nitin Madnani and Bonnie Dorr	305
<i>Improving Blog Polarity Classification via Topic Analysis and Adaptive Methods</i> Feifan Liu, Dong Wang, Bin Li and Yang Liu	309
<i>Creating Local Coherence: An Empirical Assessment</i> Annie Louis and Ani Nenkova	313
<i>Time-Efficient Creation of an Accurate Sentence Fusion Corpus</i> Kathleen McKeown, Sara Rosenthal, Kapil Thadani and Coleman Moore	317
<i>Towards Cross-Lingual Textual Entailment</i> Yashar Mehdad, Matteo Negri and Marcello Federico	321
<i>A Comparative Study of Word Co-occurrence for Term Clustering in Language Model-based Sentence Retrieval</i> Saeedeh Momtazi, Sanjeev Khudanpur and Dietrich Klakow	325
<i>Information Content Measures of Semantic Similarity Perform Better Without Sense-Tagged Text</i> Ted Pedersen	329
<i>Generating Expository Dialogue from Monologue: Motivation, Corpus and Preliminary Rules</i> Paul Piwek and Svetlana Stoyanchev	333
<i>The Simple Truth about Dependency and Phrase Structure Representations: An Opinion Piece</i> Owen Rambow	337

<i>Word Alignment with Stochastic Bracketing Linear Inversion Transduction Grammar</i> Markus Saers, Joakim Nivre and Dekai Wu	341
<i>Crowdsourcing the evaluation of a domain-adapted named entity recognition system</i> Asad B. Sayeed, Timothy J. Meyer, Hieu C. Nguyen, Olivia Buzek and Amy Weinberg.....	345
<i>Generalizing Hierarchical Phrase-based Translation using Rules with Adjacent Nonterminals</i> Hendra Setiawan and Philip Resnik	349
<i>The Effect of Ambiguity on the Automated Acquisition of WSD Examples</i> Mark Stevenson and Yikun Guo	353
<i>Word Sense Subjectivity for Cross-lingual Lexical Substitution</i> Fangzhong Su and Katja Markert	357
<i>Query Ambiguity Revisited: Clickthrough Measures for Distinguishing Informational and Ambiguous Queries</i> Yu Wang and Eugene Agichtein.....	361
<i>For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia</i> Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil and Lillian Lee.....	365
<i>Predicting Human-Targeted Translation Edit Rate via Untrained Human Annotators</i> Omar F. Zaidan and Chris Callison-Burch.....	369
<i>Improving Semantic Role Classification with Selectional Preferences</i> Beñat Zepirain, Eneko Agirre, Lluís Màrquez and Mihai Surdeanu	373
<i>Generalizing Syntactic Structures for Product Attribute Candidate Extraction</i> Yanyan Zhao, Bing Qin, Shen Hu and Ting Liu.....	377
<i>“cba to check the spelling”: Investigating Parser Performance on Discussion Forum Posts</i> Jennifer Foster	381
<i>Coreference Resolution in a Modular, Entity-Centered Model</i> Aria Haghighi and Dan Klein	385
<i>Stream-based Translation Models for Statistical Machine Translation</i> Abby Levenberg, Chris Callison-Burch and Miles Osborne	394
<i>Extracting Parallel Sentences from Comparable Corpora using Document Level Alignment</i> Jason R. Smith, Chris Quirk and Kristina Toutanova	403
<i>Statistical Machine Translation of Texts with Misspelled Words</i> Nicola Bertoldi, Mauro Cettolo and Marcello Federico	412
<i>Everybody loves a rich cousin: An empirical study of transliteration through bridge languages</i> Mitesh M. Khapra, A Kumaran and Pushpak Bhattacharyya	420
<i>Discriminative Learning over Constrained Latent Representations</i> Ming-Wei Chang, Dan Goldwasser, Dan Roth and Vivek Srikumar	429

<i>Some Empirical Evidence for Annotation Noise in a Benchmarked Dataset</i>	
Beata Beigman Klebanov and Eyal Beigman	438
<i>Bayesian Inference for Finite-State Transducers</i>	
David Chiang, Jonathan Graehl, Kevin Knight, Adam Pauls and Sujith Ravi	447
<i>Distributed Training Strategies for the Structured Perceptron</i>	
Ryan McDonald, Keith Hall and Gideon Mann	456
<i>Term Weighting Schemes for Latent Dirichlet Allocation</i>	
Andrew T. Wilson and Peter A. Chew	465
<i>Learning Dense Models of Query Similarity from User Click Logs</i>	
Fabio De Bona, Stefan Riezler, Keith Hall, Massimiliano Ciaramita, Amaç Herdağdelen and Maria Holmqvist	474
<i>Learning to Link Entities with Knowledge Base</i>	
Zhicheng Zheng, Fangtao Li, Minlie Huang and Xiaoyan Zhu	483
<i>Improving the Multilingual User Experience of Wikipedia Using Cross-Language Name Search</i>	
Raghavendra Udupa and Mitesh M. Khapra	492
<i>Learning Words and Their Meanings from Unsegmented Child-directed Speech</i>	
Bevan K. Jones, Mark Johnson and Michael C. Frank	501
<i>Subword Variation in Text Message Classification</i>	
Robert Munro and Christopher D. Manning	510
<i>Automatic Diacritization for Low-Resource Languages Using a Hybrid Word and Consonant CMM</i>	
Robbie Haertel, Peter McClanahan and Eric K. Ringger	519
<i>Urdu Word Segmentation</i>	
Nadir Durrani and Sarmad Hussain	528
<i>Enabling Monolingual Translators: Post-Editing vs. Options</i>	
Philipp Koehn	537
<i>Online Learning for Interactive Statistical Machine Translation</i>	
Daniel Ortiz-Martínez, Ismael García-Varea and Francisco Casacuberta	546
<i>The Best Lexical Metric for Phrase-Based Statistical MT System Optimization</i>	
Daniel Cer, Christopher D. Manning and Daniel Jurafsky	555
<i>Variational Inference for Adaptor Grammars</i>	
Shay B. Cohen, David M. Blei and Noah A. Smith	564
<i>Type-Based MCMC</i>	
Percy Liang, Michael I. Jordan and Dan Klein	573
<i>Painless Unsupervised Learning with Features</i>	
Taylor Berg-Kirkpatrick, Alexandre Bouchard-Côté, John DeNero and Dan Klein	582

<i>Linguistic Steganography Using Automatically Generated Paraphrases</i> Ching-Yun Chang and Stephen Clark	591
<i>Prenominal Modifier Ordering via Multiple Sequence Alignment</i> Aaron Dunlop, Margaret Mitchell and Brian Roark	600
<i>Good Question! Statistical Ranking for Question Generation</i> Michael Heilman and Noah A. Smith	609
<i>Not All Seeds Are Equal: Measuring the Quality of Text Mining Seeds</i> Zornitsa Kozareva and Eduard Hovy	618
<i>Extracting Glosses to Disambiguate Word Senses</i> Weisi Duan and Alexander Yates	627
<i>Can Recognising Multiword Expressions Improve Shallow Parsing?</i> Ioannis Korkontzelos and Suresh Manandhar	636
<i>A Simple Approach for HPSG Supertagging Using Dependency Information</i> Yao-zhong Zhang, Takuya Matsuzaki and Jun'ichi Tsujii	645
<i>Ensemble Models for Dependency Parsing: Cheap and Good?</i> Mihai Surdeanu and Christopher D. Manning	649
<i>Enlarged Search Space for SITG Parsing</i> Guillem Gascó, Joan-Andreu Sánchez and José-Miguel Benedí	653
<i>Improving Data Driven Dependency Parsing using Clausal Information</i> Phani Gadde, Karan Jindal, Samar Husain, Dipti Misra Sharma and Rajeev Sangal	657
<i>A Treebank Query System Based on an Extracted Tree Grammar</i> Seth Kulick and Ann Bies	661
<i>Reranking the Berkeley and Brown Parsers</i> Mark Johnson and Ahmet Engin Ural	665
<i>An Exploration of Off Topic Conversation</i> Whitney L. Cade, Blair A. Lehman and Andrew Olney	669
<i>Making Conversational Structure Explicit: Identification of Initiation-response Pairs within Online Discussions</i> Yi-Chia Wang and Carolyn P. Rosé	673
<i>Engaging learning groups using Social Interaction Strategies</i> Rohit Kumar and Carolyn P. Rosé	677
<i>Using Entity-Based Features to Model Coherence in Student Essays</i> Jill Burstein, Joel Tetreault and Slava Andreyev	681
<i>Summarizing Microblogs Automatically</i> Beaux Sharifi, Mark-Anthony Hutton and Jugal Kalita	685

<i>Automatic Generation of Personalized Annotation Tags for Twitter Users</i>	
Wei Wu, Bin Zhang and Mari Ostendorf	689
<i>Language identification of names with SVMs</i>	
Aditya Bhargava and Grzegorz Kondrak	693
<i>Integrating Joint n-gram Features into a Discriminative Training Framework</i>	
Sittichai Jiampojamarn, Colin Cherry and Grzegorz Kondrak	697
<i>A Hybrid Morphologically Decomposed Factored Language Models for Arabic LVCSR</i>	
Amr El-Desoky, Ralf Schlüter and Hermann Ney	701
<i>Is Arabic Part of Speech Tagging Feasible Without Word Segmentation?</i>	
Emad Mohamed and Sandra Kübler	705
<i>Arabic Mention Detection: Toward Better Unit of Analysis</i>	
Yassine Benajiba and Imed Zitouni	709
<i>An MDL-based approach to extracting subword units for grapheme-to-phoneme conversion</i>	
Sravana Reddy and John Goldsmith	713
<i>Extracting Phrase Patterns with Minimum Redundancy for Unsupervised Speaker Role Classification</i>	
Bin Zhang, Brian Hutchinson, Wei Wu and Mari Ostendorf	717
<i>Classification of Prosodic Events using Quantized Contour Modeling</i>	
Andrew Rosenberg	721
<i>Investigations into the Crandem Approach to Word Recognition</i>	
Rohit Prabhavalkar, Preethi Jyothi, William Hartmann, Jeremy Morris and Eric Fosler-Lussier	725
<i>Constraint-Driven Rank-Based Learning for Information Extraction</i>	
Sameer Singh, Limin Yao, Sebastian Riedel and Andrew McCallum	729
<i>Softmax-Margin CRFs: Training Log-Linear Models with Cost Functions</i>	
Kevin Gimpel and Noah A. Smith	733
<i>Bitext-Based Resolution of German Subject-Object Ambiguities</i>	
Florian Schwarck, Alexander Fraser and Hinrich Schütze	737
<i>Invited Talk: Music, Language, and Computational Modeling: Lessons from the Key-Finding Problem</i>	
David Temperley	741
<i>An Efficient Algorithm for Easy-First Non-Directional Dependency Parsing</i>	
Yoav Goldberg and Michael Elhadad	742
<i>From Baby Steps to Leapfrog: How “Less is More” in Unsupervised Dependency Parsing</i>	
Valentin I. Spitzkovsky, Hiyan Alshawi and Daniel Jurafsky	751

<i>Relaxed Marginal Inference and its Application to Dependency Parsing</i> Sebastian Riedel and David A. Smith	760
<i>Optimal Parsing Strategies for Linear Context-Free Rewriting Systems</i> Daniel Gildea	769
<i>The viability of web-derived polarity lexicons</i> Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan and Ryan McDonald	777
<i>Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables</i> Tetsuji Nakagawa, Kentaro Inui and Sadao Kurohashi	786
<i>Convolution Kernels for Opinion Holder Extraction</i> Michael Wiegand and Dietrich Klakow	795
<i>An Unsupervised Aspect-Sentiment Model for Online Reviews</i> Samuel Brody and Noemie Elhadad	804
<i>Joint Inference for Knowledge Extraction from Biomedical Literature</i> Hoifung Poon and Lucy Vanderwende	813
<i>Clinical Information Retrieval using Document and PICO Structure</i> Florian Boudin, Jian-Yun Nie and Martin Dawes	822
<i>Topic Models for Image Annotation and Text Illustration</i> Yansong Feng and Mirella Lapata	831
<i>Learning about Voice Search for Spoken Dialogue Systems</i> Rebecca Passonneau, Susan L. Epstein, Tiziana Ligorio, Joshua B. Gordon and Pravin Bhutada	840
<i>A Direct Syntax-Driven Reordering Model for Phrase-Based Machine Translation</i> Niyu Ge	849
<i>Context-free reordering, finite-state translation</i> Chris Dyer and Philip Resnik	858
<i>Improved Models of Distortion Cost for Statistical Machine Translation</i> Spence Green, Michel Galley and Christopher D. Manning	867
<i>Why Synchronous Tree Substitution Grammars?</i> Andreas Maletti	876
<i>An extractive supervised two-stage method for sentence compression</i> Dimitrios Galanis and Ion Androutsopoulos	885
<i>Interpretation and Transformation for Abstracting Conversations</i> Gabriel Murray, Giuseppe Carenini and Raymond Ng	894
<i>Quantifying the Limits and Success of Extractive Summarization Systems Across Domains</i> Hakan Ceylan, Rada Mihalcea, Umut Özertem, Elena Lloret and Manuel Palomar	903

<i>Multi-document Summarization via Budgeted Maximization of Submodular Functions</i>	
Hui Lin and Jeff Bilmes	912
<i>Cross-lingual Induction of Selectional Preferences with Bilingual Vector Spaces</i>	
Yves Peirsman and Sebastian Padó	921
<i>Latent SVMs for Semantic Role Labeling using LTAG Derivation Trees</i>	
Yudong Liu, Gholamreza Haffari and Anoop Sarkar	930
<i>Unsupervised Induction of Semantic Roles</i>	
Joel Lang and Mirella Lapata	939
<i>Probabilistic Frame-Semantic Parsing</i>	
Dipanjan Das, Nathan Schneider, Desai Chen and Noah A. Smith	948
<i>Expected Sequence Similarity Maximization</i>	
Cyril Allauzen, Shankar Kumar, Wolfgang Macherey, Mehryar Mohri and Michael Riley . . .	957
<i>Accurate Non-Hierarchical Phrase-Based Translation</i>	
Michel Galley and Christopher D. Manning	966
<i>Model Combination for Machine Translation</i>	
John DeNero, Shankar Kumar, Ciprian Chelba and Franz Och	975
<i>Detecting Emails Containing Requests for Action</i>	
Andrew Lampert, Robert Dale and Cecile Paris	984
<i>Evaluating Hierarchical Discourse Segmentation</i>	
Lucien Carroll	993
<i>Reformulating Discourse Connectives for Non-Expert Readers</i>	
Advaith Siddharthan and Napoleon Katsos	1002
<i>Tree Edit Models for Recognizing Textual Entailments, Paraphrases, and Answers to Questions</i>	
Michael Heilman and Noah A. Smith	1011
<i>Syntactic/Semantic Structures for Textual Entailment Recognition</i>	
Yashar Mehdad, Alessandro Moschitti and Fabio Massimo Zanzotto	1020
<i>Automatic Metaphor Interpretation as a Paraphrasing Task</i>	
Ekaterina Shutova	1029