

# **48th Annual Meeting of the Association for Computational Linguistics 2010 (ACL 2010)**

## **Short Papers**

**Uppsala, Sweden  
11-16 July 2010**

ISBN: 978-1-61738-825-5

**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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Paraphrase Lattice for Statistical Machine Translation</i> Takashi Onishi, Masao Utiyama and Eiichiro Sumita .....	1
<i>A Joint Rule Selection Model for Hierarchical Phrase-Based Translation</i> Lei Cui, Dongdong Zhang, Mu Li, Ming Zhou and Tiejun Zhao .....	6
<i>Learning Lexicalized Reordering Models from Reordering Graphs</i> Jinsong Su, Yang Liu, Yajuan Lv, Haitao Mi and Qun Liu .....	12
<i>Filtering Syntactic Constraints for Statistical Machine Translation</i> Hailong Cao and Eiichiro Sumita .....	17
<i>Diversify and Combine: Improving Word Alignment for Machine Translation on Low-Resource Languages</i> Bing Xiang, Yonggang Deng and Bowen Zhou .....	22
<i>Efficient Path Counting Transducers for Minimum Bayes-Risk Decoding of Statistical Machine Translation Lattices</i> Graeme Blackwood, Adrià de Gispert and William Byrne .....	27
<i>The Same-Head Heuristic for Coreference</i> Micha Elsner and Eugene Charniak .....	33
<i>Authorship Attribution Using Probabilistic Context-Free Grammars</i> Sindhu Raghavan, Adriana Kovashka and Raymond Mooney .....	38
<i>The Impact of Interpretation Problems on Tutorial Dialogue</i> Myroslava O. Dzikovska, Johanna D. Moore, Natalie Steinhauser and Gwendolyn Campbell ...	43
<i>The Prevalence of Descriptive Referring Expressions in News and Narrative</i> Raquel Hervas and Mark Finlayson .....	49
<i>Preferences versus Adaptation during Referring Expression Generation</i> Martijn Goudbeek and Emiel Krahmer .....	55
<i>Cognitively Plausible Models of Human Language Processing</i> Frank Keller .....	60
<i>The Manually Annotated Sub-Corpus: A Community Resource for and by the People</i> Nancy Ide, Collin Baker, Christiane Fellbaum and Rebecca Passonneau .....	68
<i>Correcting Errors in a Treebank Based on Synchronous Tree Substitution Grammar</i> Yoshihide Kato and Shigeki Matsubara .....	74
<i>Evaluating Machine Translations Using mNCD</i> Marcus Dobrinkat, Tero Tapiovaara, Jaakko Väyrynen and Kimmo Kettunen .....	80
<i>Tackling Sparse Data Issue in Machine Translation Evaluation</i> Ondřej Bojar, Kamil Kos and David Mareček .....	86
<i>Exemplar-Based Models for Word Meaning in Context</i> Katrin Erk and Sebastian Pado .....	92

<i>A Structured Model for Joint Learning of Argument Roles and Predicate Senses</i> Yotaro Watanabe, Masayuki Asahara and Yuji Matsumoto .....	98
<i>Semantics-Driven Shallow Parsing for Chinese Semantic Role Labeling</i> Weiwei Sun .....	103
<i>Collocation Extraction beyond the Independence Assumption</i> Gerlof Bouma .....	109
<i>Automatic Collocation Suggestion in Academic Writing</i> Jian-Cheng Wu, Yu-Chia Chang, Teruko Mitamura and Jason S. Chang .....	115
<i>Event-Based Hyperspace Analogue to Language for Query Expansion</i> Tingxu Yan, Tamsin Maxwell, Dawei Song, Yuexian Hou and Peng Zhang .....	120
<i>Automatically Generating Term Frequency Induced Taxonomies</i> Karin Murthy, Tanveer A Faruque, L Venkata Subramaniam, Hima Prasad K and Mukesh Mohania 126	
<i>Complexity Assumptions in Ontology Verbalisation</i> Richard Power .....	132
<i>Word Alignment with Synonym Regularization</i> Hiroyuki Shindo, Akinori Fujino and Masaaki Nagata .....	137
<i>Better Filtration and Augmentation for Hierarchical Phrase-Based Translation Rules</i> Zhiyang Wang, Yajuan Lv, Qun Liu and Young-Sook Hwang .....	142
<i>Fixed Length Word Suffix for Factored Statistical Machine Translation</i> Narges Sharif Razavian and Stephan Vogel .....	147
<i>Unsupervised Discourse Segmentation of Documents with Inherently Parallel Structure</i> Minwoo Jeong and Ivan Titov .....	151
<i>Coreference Resolution with Reconcile</i> Veselin Stoyanov, Claire Cardie, Nathan Gilbert, Ellen Riloff, David Buttler and David Hysom	156
<i>Predicate Argument Structure Analysis Using Transformation Based Learning</i> Hirotoishi Taira, Sanae Fujita and Masaaki Nagata .....	162
<i>Improving Chinese Semantic Role Labeling with Rich Syntactic Features</i> Weiwei Sun .....	168
<i>Balancing User Effort and Translation Error in Interactive Machine Translation via Confidence Measures</i> Jesús González Rubio, Daniel Ortiz Martínez and Francisco Casacuberta .....	173
<i>Improving Arabic-to-English Statistical Machine Translation by Reordering Post-Verbal Subjects for Alignment</i> Marine Carpuat, Yuval Marton and Nizar Habash .....	178
<i>Learning Common Grammar from Multilingual Corpus</i> Tomoharu Iwata, Daichi Mochihashi and Hiroshi Sawada .....	184
<i>Tree-Based Deterministic Dependency Parsing — An Application to Nivre’s Method —</i> Kotaro Kitagawa and Kumiko Tanaka-Ishii .....	189

<i>Sparsity in Dependency Grammar Induction</i>	
Jennifer Gillenwater, Kuzman Ganchev, João Graça, Fernando Pereira and Ben Taskar . . . . .	194
<i>Top-Down K-Best A* Parsing</i>	
Adam Pauls, Dan Klein and Chris Quirk . . . . .	200
<i>Simple Semi-Supervised Training of Part-Of-Speech Taggers</i>	
Anders Søgaard . . . . .	205
<i>Efficient Optimization of an MDL-Inspired Objective Function for Unsupervised Part-Of-Speech Tagging</i>	
Ashish Vaswani, Adam Pauls and David Chiang . . . . .	209
<i>SVD and Clustering for Unsupervised POS Tagging</i>	
Michael Lamar, Yariv Maron, Mark Johnson and Elie Bienenstock . . . . .	215
<i>Intelligent Selection of Language Model Training Data</i>	
Robert C. Moore and William Lewis . . . . .	220
<i>Blocked Inference in Bayesian Tree Substitution Grammars</i>	
Trevor Cohn and Phil Blunsom . . . . .	225
<i>Online Generation of Locality Sensitive Hash Signatures</i>	
Benjamin Van Durme and Ashwin Lall . . . . .	231
<i>Optimizing Question Answering Accuracy by Maximizing Log-Likelihood</i>	
Matthias H. Heie, Edward W. D. Whittaker and Sadaoki Furui . . . . .	236
<i>Generating Entailment Rules from FrameNet</i>	
Roni Ben Aharon, Idan Szepktor and Ido Dagan . . . . .	241
<i>Don't 'Have a Clue'? Unsupervised Co-Learning of Downward-Entailing Operators.</i>	
Cristian Danescu-Niculescu-Mizil and Lillian Lee . . . . .	247
<i>Vocabulary Choice as an Indicator of Perspective</i>	
Beata Beigman Klebanov, Eyal Beigman and Daniel Diermeier . . . . .	253
<i>Cross Lingual Adaptation: An Experiment on Sentiment Classifications</i>	
Bin Wei and Christopher Pal . . . . .	258
<i>Using Anaphora Resolution to Improve Opinion Target Identification in Movie Reviews</i>	
Niklas Jakob and Iryna Gurevych . . . . .	263
<i>Hierarchical Sequential Learning for Extracting Opinions and Their Attributes</i>	
Yejin Choi and Claire Cardie . . . . .	269
<i>Jointly Optimizing a Two-Step Conditional Random Field Model for Machine Transliteration and Its Fast Decoding Algorithm</i>	
Dong Yang, Paul Dixon and Sadaoki Furui . . . . .	275
<i>Arabic Named Entity Recognition: Using Features Extracted from Noisy Data</i>	
Yassine Benajiba, Imed Zitouni, Mona Diab and Paolo Rosso . . . . .	281
<i>Extracting Sequences from the Web</i>	
Anthony Fader, Stephen Soderland and Oren Etzioni . . . . .	286

<i>An Entity-Level Approach to Information Extraction</i> Aria Haghighi and Dan Klein .....	291
<i>A Semi-Supervised Key Phrase Extraction Approach: Learning from Title Phrases through a Document Semantic Network</i> Decong Li, Sujian Li, Wenjie Li, Wei Wang and Weiguang Qu .....	296
<i>Domain Adaptation of Maximum Entropy Language Models</i> Tanel Alumäe and Mikko Kurimo .....	301
<i>Decision Detection Using Hierarchical Graphical Models</i> Trung H. Bui and Stanley Peters .....	307
<i>Using Speech to Reply to SMS Messages While Driving: An In-Car Simulator User Study</i> Yun-Cheng Ju and Tim Paek .....	313
<i>Classification of Feedback Expressions in Multimodal Data</i> Costanza Navarretta and Patrizia Paggio .....	318
<i>Optimizing Informativeness and Readability for Sentiment Summarization</i> Hitoshi Nishikawa, Takaaki Hasegawa, Yoshihiro Matsuo and Genichiro Kikui .....	325
<i>Last but Definitely Not Least: On the Role of the Last Sentence in Automatic Polarity-Classification</i> Israela Becker and Vered Aharonson .....	331
<i>Automatically Generating Annotator Rationales to Improve Sentiment Classification</i> Ainur Yessenalina, Yejin Choi and Claire Cardie .....	336
<i>Simultaneous Tokenization and Part-Of-Speech Tagging for Arabic without a Morphological Analyzer</i> Seth Kulick .....	342
<i>Hierarchical A* Parsing with Bridge Outside Scores</i> Adam Pauls and Dan Klein .....	348
<i>Using Parse Features for Preposition Selection and Error Detection</i> Joel Tetreault, Jennifer Foster and Martin Chodorow .....	353
<i>Distributional Similarity vs. PU Learning for Entity Set Expansion</i> Xiao-Li Li, Lei Zhang, Bing Liu and See-Kiong Ng .....	359
<i>Active Learning-Based Elicitation for Semi-Supervised Word Alignment</i> Vamshi Ambati, Stephan Vogel and Jaime Carbonell .....	365
<i>An Active Learning Approach to Finding Related Terms</i> David Vickrey, Oscar Kipersztok and Daphne Koller .....	371
<i>Learning Better Data Representation Using Inference-Driven Metric Learning</i> Paramveer S. Dhillon, Partha Pratim Talukdar and Koby Crammer .....	377
<i>Wrapping up a Summary: From Representation to Generation</i> Josef Steinberger, Marco Turchi, Mijail Kabadjov, Ralf Steinberger and Nello Cristianini .....	382