48th Annual Meeting of the
Association for Computational Linguistics

Proceedings of the Conference Short Papers

11-16 July 2010
Uppsala University
Uppsala, Sweden
# Table of Contents

**Paraphrase Lattice for Statistical Machine Translation**  
Takashi Onishi, Masao Utiyama and Eiichiro Sumita .......................... 1

**A Joint Rule Selection Model for Hierarchical Phrase-Based Translation**  
Lei Cui, Dongdong Zhang, Mu Li, Ming Zhou and Tiejun Zhao .................. 6

**Learning Lexicalized Reordering Models from Reordering Graphs**  
Jinsong Su, Yang Liu, Yajuan Lv, Haitao Mi and Qun Liu ...................... 12

**Filtering Syntactic Constraints for Statistical Machine Translation**  
Hailong Cao and Eiichiro Sumita .............................................. 17

**Diversify and Combine: Improving Word Alignment for Machine Translation on Low-Resource Languages**  
Bing Xiang, Yonggang Deng and Bowen Zhou .................................. 22

**Efficient Path Counting Transducers for Minimum Bayes-Risk Decoding of Statistical Machine Translation Lattices**  
Graeme Blackwood, Adrià de Gispert and William Byrne ..................... 27

**The Same-Head Heuristic for Coreference**  
Micha Elsner and Eugene Charniak ......................................... 33

**Authorship Attribution Using Probabilistic Context-Free Grammars**  
Sindhu Raghavan, Adriana Kovashka and Raymond Mooney .................. 38

**The Impact of Interpretation Problems on Tutorial Dialogue**  
Myroslava O. Dzikovska, Johanna D. Moore, Natalie Steinhauser and Gwendolyn Campbell .. 43

**The Prevalence of Descriptive Referring Expressions in News and Narrative**  
Raquel Hervas and Mark Finlayson .......................................... 49

**Preferences versus Adaptation during Referring Expression Generation**  
Martijn Goudbeek and Emiel Krahmer ...................................... 55

**Cognitively Plausible Models of Human Language Processing**  
Frank Keller ................................................................. 60

**The Manually Annotated Sub-Corpus: A Community Resource for and by the People**  
Nancy Ide, Collin Baker, Christiane Fellbaum and Rebecca Passonneau .......... 68

**Correcting Errors in a Treebank Based on Synchronous Tree Substitution Grammar**  
Yoshihide Kato and Shigeki Matsubara .................................... 74

**Evaluating Machine Translations Using mNCD**  
Marcus Dobrinskait, Tero Tapiovaara, Jaakko Väyrynen and Kimmo Kettunen ........ 80

**Tackling Sparse Data Issue in Machine Translation Evaluation**  
Ondřej Bojar, Kamil Kos and David Mareček ................................ 86

**Exemplar-Based Models for Word Meaning in Context**  
Katrin Erk and Sebastian Pado ............................................. 92
A Structured Model for Joint Learning of Argument Roles and Predicate Senses
Yotaro Watanabe, Masayuki Asahara and Yuji Matsumoto .................................................. 98

Semantics-Driven Shallow Parsing for Chinese Semantic Role Labeling
Weiwei Sun ......................................................................................................................... 103

Collocation Extraction beyond the Independence Assumption
Gerlof Bouma .................................................................................................................. 109

Automatic Collocation Suggestion in Academic Writing
Jian-Cheng Wu, Yu-Chia Chang, Teruko Mitamura and Jason S. Chang ............................ 115

Event-Based Hyperspace Analogue to Language for Query Expansion
Tingxu Yan, Tamsin Maxwell, Dawei Song, Yuexian Hou and Peng Zhang ....................... 120

Automatically Generating Term Frequency Induced Taxonomies
Karin Murthy, Tanveer A Faruquie, L Venkata Subramaniam, Hima Prasad K and Mukesh Mohania 126

Complexity Assumptions in Ontology Verbalisation
Richard Power ................................................................. 132

Word Alignment with Synonym Regularization
Hiroyuki Shindo, Akinori Fujino and Masaaki Nagata ................................................... 137

Better Filtration and Augmentation for Hierarchical Phrase-Based Translation Rules
Zhiyang Wang, Yajuan Lv, Qun Liu and Young-Sook Hwang ............................ 142

Fixed Length Word Suffix for Factored Statistical Machine Translation
Narges Sharif Razavian and Stephan Vogel .......................................................... 147

Unsupervised Discourse Segmentation of Documents with Inherently Parallel Structure
Minwoo Jeong and Ivan Titov ......................................................................................... 151

Coreference Resolution with Reconcile
Veselin Stoyanov, Claire Cardie, Nathan Gilbert, Ellen Riloff, David Butler and David Hysom 156

Predicate Argument Structure Analysis Using Transformation Based Learning
Hirotoshi Taira, Sanae Fujita and Masaaki Nagata .................................................. 162

Improving Chinese Semantic Role Labeling with Rich Syntactic Features
Weiwei Sun ......................................................................................................................... 168

Balancing User Effort and Translation Error in Interactive Machine Translation via Confidence Measures
Jesús González Rubio, Daniel Ortiz Martínez and Francisco Casacuberta ....................... 173

Improving Arabic-to-English Statistical Machine Translation by Reordering Post-Verbal Subjects for Alignment
Marine Carpuat, Yuval Marton and Nizar Habash ................................................ 178

Learning Common Grammar from Multilingual Corpus
Tomoharu Iwata, Daichi Mochihashi and Hiroshi Sawada ........................................... 184

Tree-Based Deterministic Dependency Parsing — An Application to Nivre’s Method —
Kotaro Kitagawa and Kumiko Tanaka-Ishii .................................................................... 189
Sparsity in Dependency Grammar Induction
Jennifer Gillenwater, Kuzman Ganchev, João Graça, Fernando Pereira and Ben Taskar ........................... 194

Top-Down K-Best A* Parsing
Adam Pauls, Dan Klein and Chris Quirk ...................................................... 200

Simple Semi-Supervised Training of Part-Of-Speech Taggers
Anders Søgaard ......................................................................................... 205

Efficient Optimization of an MDL-Inspired Objective Function for Unsupervised Part-Of-Speech Tagging
Ashish Vaswani, Adam Pauls and David Chiang ........................................... 209

SVD and Clustering for Unsupervised POS Tagging
Michael Lamar, Yariv Maron, Mark Johnson and Elie Bienenstock .............. 215

Intelligent Selection of Language Model Training Data
Robert C. Moore and William Lewis ......................................................... 220

Blocked Inference in Bayesian Tree Substitution Grammars
Trevor Cohn and Phil Blunsom ................................................................. 225

Online Generation of Locality Sensitive Hash Signatures
Benjamin Van Durme and Ashwin Lall ..................................................... 231

Optimizing Question Answering Accuracy by Maximizing Log-Likelihood
Matthias H. Heie, Edward W. D. Whittaker and Sadaoki Furui ...................... 236

Generating Entailment Rules from FrameNet
Roni Ben Aharon, Idan Szpektor and Ido Dagan ......................................... 241

Don’t ‘Have a Clue’? Unsupervised Co-Learning of Downward-Entailing Operators.
Cristian Danescu-Niculescu-Mizil and Lillian Lee ........................................ 247

Vocabulary Choice as an Indicator of Perspective
Beata Beigman Klebanov, Eyal Beigman and Daniel Diermeier ....................... 253

Cross Lingual Adaptation: An Experiment on Sentiment Classifications
Bin Wei and Christopher Pal .................................................................. 258

Using Anaphora Resolution to Improve Opinion Target Identification in Movie Reviews
Niklas Jakob and Iryna Gurevych ............................................................. 263

Hierarchical Sequential Learning for Extracting Opinions and Their Attributes
Yejin Choi and Claire Cardie .................................................................... 269

Jointly Optimizing a Two-Step Conditional Random Field Model for Machine Transliteration and Its Fast Decoding Algorithm
Dong Yang, Paul Dixon and Sadaoki Furui .................................................. 275

Arabic Named Entity Recognition: Using Features Extracted from Noisy Data
Yassine Benajiba, Imed Zitouni, Mona Diab and Paolo Rosso ....................... 281

Extracting Sequences from the Web
Anthony Fader, Stephen Soderland and Oren Etzioni .................................. 286
An Entity-Level Approach to Information Extraction
Aria Haghighi and Dan Klein ................................................................. 291

A Semi-Supervised Key Phrase Extraction Approach: Learning from Title Phrases through a Document Semantic Network
Decong Li, Suijuan Li, Wenjie Li, Wei Wang and Weiguang Qu ....................... 296

Domain Adaptation of Maximum Entropy Language Models
Tanel Alumäe and Mikko Kurimo ........................................................... 301

Decision Detection Using Hierarchical Graphical Models
Trung H. Bui and Stanley Peters .............................................................. 307

Using Speech to Reply to SMS Messages While Driving: An In-Car Simulator User Study
Yun-Cheng Ju and Tim Paek ................................................................. 313

Classification of Feedback Expressions in Multimodal Data
Costanza Navarretta and Patrizia Paggio ............................................... 318

Optimizing Informativeness and Readability for Sentiment Summarization
Hitoshi Nishikawa, Takaaki Hasegawa, Yoshihiro Matsuo and Genichiro Kikui 325

Last but Definitely Not Least: On the Role of the Last Sentence in Automatic Polarity-Classification
Israel Becker and Vered Aharonson ......................................................... 331

Automatically Generating Annotator Rationales to Improve Sentiment Classification
Ainur Yessenalina, Yejin Choi and Claire Cardie ........................................ 336

Simultaneous Tokenization and Part-Of-Speech Tagging for Arabic without a Morphological Analyzer
Seth Kulick .................................................................................................. 342

Hierarchical A* Parsing with Bridge Outside Scores
Adam Pauls and Dan Klein ........................................................................ 348

Using Parse Features for Preposition Selection and Error Detection
Joel Tetreault, Jennifer Foster and Martin Chodorow .................................. 353

Distributional Similarity vs. PU Learning for Entity Set Expansion
Xiaoli Li, Lei Zhang, Bing Liu and See-Kiong Ng ........................................... 359

Active Learning-Based Elicitation for Semi-Supervised Word Alignment
Vamshi Ambati, Stephan Vogel and Jaime Carbonell .................................... 365

An Active Learning Approach to Finding Related Terms
David Vickrey, Oscar Kipersztok and Daphne Koller .................................... 371

Learning Better Data Representation Using Inference-Driven Metric Learning
Paramveer S. Dhillon, Partha Pratim Talukdar and Koby Crammer .................. 377

Wrapping up a Summary: From Representation to Generation
Josef Steinberger, Marco Turchi, Mijail Kabadjov, Ralf Steinberger and Nello Cristianini ........ 382