

**Joint Conference on Empirical
Methods in Natural Language
Processing and Computational
Natural Language Learning 2012**

(EMNLP 2012)

**Jeju Island, Korea
12 – 14 July 2012**

Volume 1 of 2

ISBN: 978-1-62276-503-4

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© (2012) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2013)

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>Syntactic Transfer Using a Bilingual Lexicon</i> Greg Durrett, Adam Pauls and Dan Klein	1
<i>Regularized Interlingual Projections: Evaluation on Multilingual Transliteration</i> Jagadeesh Jagarlamudi and Hal Daume III	12
<i>Bilingual Lexicon Extraction from Comparable Corpora Using Label Propagation</i> Akihiro Tamura, Taro Watanabe and Eiichiro Sumita	24
<i>Lexical Differences in Autobiographical Narratives from Schizophrenic Patients and Healthy Controls</i> Kai Hong, Christian G. Kohler, Mary E. March, Amber A. Parker and Ani Nenkova	37
<i>Streaming Analysis of Discourse Participants</i> Benjamin Van Durme	48
<i>Detecting Subgroups in Online Discussions by Modeling Positive and Negative Relations among Participants</i> Ahmed Hassan, Amjad Abu-Jbara and Dragomir Radev	59
<i>Generative Goal-Driven User Simulation for Dialog Management</i> Aciel Eshky, Ben Allison and Mark Steedman	71
<i>Optimising Incremental Dialogue Decisions Using Information Density for Interactive Systems</i> Nina Dethlefs, Helen Hastie, Verena Rieser and Oliver Lemon	82
<i>Mixed Membership Markov Models for Unsupervised Conversation Modeling</i> Michael J. Paul	94
<i>An Entity-Topic Model for Entity Linking</i> Xianpei Han and Le Sun	105
<i>Linking Named Entities to Any Database</i> Avirup Sil, Ernest Cronin, Penghai Nie, Yinfei Yang, Ana-Maria Popescu and Alexander Yates	116
<i>Towards Efficient Named-Entity Rule Induction for Customizability</i> Ajay Nagesh, Ganesh Ramakrishnan, Laura Chiticariu, Rajasekar Krishnamurthy, Ankush Dharkar and Pushpak Bhattacharyya	128
<i>Active Learning for Imbalanced Sentiment Classification</i> Shoushan Li, Shengfeng Ju, Guodong Zhou and Xiaojun Li	139
<i>A Weakly Supervised Model for Sentence-Level Semantic Orientation Analysis with Multiple Experts</i> Lizhen Qu, Rainer Gemulla and Gerhard Weikum	149

<i>Collocation Polarity Disambiguation Using Web-based Pseudo Contexts</i> Yanyan Zhao, Bing Qin and Ting Liu	160
<i>Aligning Predicates across Monolingual Comparable Texts using Graph-based Clustering</i> Michael Roth and Anette Frank	171
<i>Local and Global Context for Supervised and Unsupervised Metonymy Resolution</i> Vivi Nastase, Alex Judea, Katja Markert and Michael Strube	183
<i>Learning Verb Inference Rules from Linguistically-Motivated Evidence</i> Hila Weisman, Jonathan Berant, Idan Szpektor and Ido Dagan	194
<i>Spectral Dependency Parsing with Latent Variables</i> Paramveer Dhillon, Jordan Rodu, Michael Collins, Dean Foster and Lyle Ungar	205
<i>A Phrase-Discovering Topic Model Using Hierarchical Pitman-Yor Processes</i> Robert Lindsey, William Headden and Michael Stipicevic	214
<i>A Bayesian Model for Learning SCFGs with Discontiguous Rules</i> Abby Levenberg, Chris Dyer and Phil Blunsom	223
<i>Multiple Aspect Summarization Using Integer Linear Programming</i> Kristian Woodsend and Mirella Lapata	233
<i>Minimal Dependency Length in Realization Ranking</i> Michael White and Rajakrishnan Rajkumar	244
<i>Framework of Automatic Text Summarization Using Reinforcement Learning</i> Seonggi Ryang and Takeshi Abekawa	256
<i>Large Scale Decipherment for Out-of-Domain Machine Translation</i> Qing Dou and Kevin Knight	266
<i>N-gram-based Tense Models for Statistical Machine Translation</i> Zhengxian Gong, Min Zhang, Chew Lim Tan and Guodong Zhou	276
<i>Source Language Adaptation for Resource-Poor Machine Translation</i> Pidong Wang, Preslav Nakov and Hwee Tou Ng	286
<i>Exploiting Reducibility in Unsupervised Dependency Parsing</i> David Mareček and Zdeněk Žabokrtský	297
<i>Improving Transition-Based Dependency Parsing with Buffer Transitions</i> Daniel Fernández-González and Carlos Gómez-Rodríguez	308
<i>Generalized Higher-Order Dependency Parsing with Cube Pruning</i> Hao Zhang and Ryan McDonald	320
<i>Universal Grapheme-to-Phoneme Prediction Over Latin Alphabets</i> Young-Bum Kim and Benjamin Snyder	332

<i>Name Phylogeny: A Generative Model of String Variation</i> Nicholas Andrews, Jason Eisner and Mark Dredze	344
<i>Syntactic Surprisal Affects Spoken Word Duration in Conversational Contexts</i> Vera Demberg, Asad Sayeed, Philip Gorinski and Nikolaos Engonopoulos	356
<i>Why Question Answering using Sentiment Analysis and Word Classes</i> Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Takuya Kawada, Stijn De Saeger, Jun'ichi Kazama and You Wang	368
<i>Natural Language Questions for the Web of Data</i> Mohamed Yahya, Klaus Berberich, Shady Elbassuoni, Maya Ramanath, Volker Tresp and Gerhard Weikum	379
<i>Answering Opinion Questions on Products by Exploiting Hierarchical Organization of Consumer Reviews</i> Jianxing Yu, Zheng-Jun Zha and Tat-Seng Chua	391
<i>Locally Training the Log-Linear Model for SMT</i> Lemao Liu, Hailong Cao, Taro Watanabe, Tiejun Zhao, Mo Yu and Conghui Zhu	402
<i>Iterative Annotation Transformation with Predict-Self Reestimation for Chinese Word Segmentation</i> Wenbin Jiang, Fandong Meng, Qun Liu and Yajuan Lü	412
<i>Automatically Constructing a Normalisation Dictionary for Microblogs</i> Bo Han, Paul Cook and Timothy Baldwin	421
<i>Unsupervised PCFG Induction for Grounded Language Learning with Highly Ambiguous Supervision</i> Joohyun Kim and Raymond Mooney	433
<i>Forced Derivation Tree based Model Training to Statistical Machine Translation</i> Nan Duan, Mu Li and Ming Zhou	445
<i>Multi-instance Multi-label Learning for Relation Extraction</i> Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati and Christopher D. Manning	455
<i>An “AI readability” Formula for French as a Foreign Language</i> Thomas François and Cédric Fairon	466
<i>Dynamic Programming for Higher Order Parsing of Gap-Minding Trees</i> Emily Pitler, Sampath Kannan and Mitchell Marcus	478
<i>Joint Entity and Event Coreference Resolution across Documents</i> Heeyoung Lee, Marta Recasens, Angel Chang, Mihai Surdeanu and Dan Jurafsky	489
<i>Joint Chinese Word Segmentation, POS Tagging and Parsing</i> Xian Qian and Yang Liu	501

<i>Translation Model Based Cross-Lingual Language Model Adaptation: from Word Models to Phrase Models</i>	
Shixiang Lu, Wei Wei, Xiaoyin Fu and Bo Xu	512
<i>Open Language Learning for Information Extraction</i>	
Mausam, Michael Schmitz, Stephen Soderland, Robert Bart and Oren Etzioni	523
<i>Modelling Sequential Text with an Adaptive Topic Model</i>	
Lan Du, Wray Buntine and Huidong Jin	535
<i>A Comparison of Vector-based Representations for Semantic Composition</i>	
William Blacoe and Mirella Lapata	546
<i>Exploiting Chunk-level Features to Improve Phrase Chunking</i>	
Junsheng Zhou, Weiguang Qu and Fen Zhang	557
<i>A Beam-Search Decoder for Grammatical Error Correction</i>	
Daniel Dahlmeier and Hwee Tou Ng	568
<i>A Statistical Relational Learning Approach to Identifying Evidence Based Medicine Categories</i>	
Mathias Verbeke, Vincent Van Asch, Roser Morante, Paolo Frasconi, Walter Daelemans and Luc De Raedt	579
<i>Lyrics, Music, and Emotions</i>	
Rada Mihalcea and Carlo Strapparava	590
<i>Assessment of ESL Learners' Syntactic Competence Based on Similarity Measures</i>	
Su-Youn Yoon and Suma Bhat	600
<i>A Unified Approach to Transliteration-based Text Input with Online Spelling Correction</i>	
Hisami Suzuki and Jianfeng Gao	609
<i>Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web</i>	
Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jong-Hoon Oh and Jun'ichi Kazama	619
<i>Enlarging Paraphrase Collections through Generalization and Instantiation</i>	
Atsushi Fujita, Pierre Isabelle and Roland Kuhn	631
<i>Concurrent Acquisition of Word Meaning and Lexical Categories</i>	
Afra Alishahi and Grzegorz Chrupala	643
<i>Do Neighbours Help? An Exploration of Graph-based Algorithms for Cross-domain Sentiment Classification</i>	
Natalia Ponomareva and Mike Thelwall	655
<i>Learning Lexicon Models from Search Logs for Query Expansion</i>	
Jianfeng Gao, Shasha Xie, Xiaodong He and Alnur Ali	666

<i>Joint Inference for Event Timeline Construction</i>	
Quang Do, Wei Lu and Dan Roth	677
<i>Three Dependency-and-Boundary Models for Grammar Induction</i>	
Valentin I. Spitzkovsky, Hiyan Alshawi and Daniel Jurafsky	688
<i>Exploring Adaptor Grammars for Native Language Identification</i>	
Sze-Meng Jojo Wong, Mark Dras and Mark Johnson	699
<i>Discovering Diverse and Salient Threads in Document Collections</i>	
Jennifer Gillenwater, Alex Kulesza and Ben Taskar	710
<i>Generalizing Sub-sentential Paraphrase Acquisition across Original Signal Type of Text Pairs</i>	
Aurélien Max, Houda Bouamor and Anne Vilnat	721
<i>Parse, Price and Cut—Delayed Column and Row Generation for Graph Based Parsers</i>	
Sebastian Riedel, David Smith and Andrew McCallum	732
<i>Domain Adaptation for Coreference Resolution: An Adaptive Ensemble Approach</i>	
Jian Bo Yang, Qi Mao, Qiao Liang Xiang, Ivor Wai-Hung Tsang, Kian Ming Adam Chai and Hai Leong Chieu	744
<i>Weakly Supervised Training of Semantic Parsers</i>	
Jayant Krishnamurthy and Tom Mitchell	754
<i>Cross-Lingual Language Modeling with Syntactic Reordering for Low-Resource Speech Recognition</i>	
Ping Xu and Pascale Fung	766
<i>Resolving Complex Cases of Definite Pronouns: The Winograd Schema Challenge</i>	
Altaf Rahman and Vincent Ng	777
<i>A Sequence Labelling Approach to Quote Attribution</i>	
Timothy O’Keefe, Silvia Pareti, James R. Curran, Irena Koprinska and Matthew Honnibal ...	790
<i>SSHLDA: A Semi-Supervised Hierarchical Topic Model</i>	
Xian-Ling Mao, Zhao-Yan Ming, Tat-Seng Chua, Si Li, Hongfei Yan and Xiaoming Li	800
<i>Improving NLP through Marginalization of Hidden Syntactic Structure</i>	
Jason Naradowsky, Sebastian Riedel and David Smith	810
<i>Type-Supervised Hidden Markov Models for Part-of-Speech Tagging with Incomplete Tag Dictionaries</i>	
Dan Garrette and Jason Baldridge	821
<i>Explore Person Specific Evidence in Web Person Name Disambiguation</i>	
Liwei Chen, Yansong Feng, Lei Zou and Dongyan Zhao	832
<i>Inducing a Discriminative Parser to Optimize Machine Translation Reordering</i>	
Graham Neubig, Taro Watanabe and Shinsuke Mori	843

<i>Re-training Monolingual Parser Bilingually for Syntactic SMT</i>	
Shujie Liu, Chi-Ho Li, Mu Li and Ming Zhou	854
<i>Transforming Trees to Improve Syntactic Convergence</i>	
David Burkett and Dan Klein	863
<i>Learning Constraints for Consistent Timeline Extraction</i>	
David McClosky and Christopher D. Manning	873
<i>Identifying Constant and Unique Relations by using Time-Series Text</i>	
Yohei Takaku, Nobuhiro Kaji, Naoki Yoshinaga and Masashi Toyoda	883
<i>No Noun Phrase Left Behind: Detecting and Typing Unlinkable Entities</i>	
Thomas Lin, Mausam and Oren Etzioni	893
<i>A Novel Discriminative Framework for Sentence-Level Discourse Analysis</i>	
Shafiq Joty, Giuseppe Carenini and Raymond Ng	904
<i>Using Discourse Information for Paraphrase Extraction</i>	
Michaela Regneri and Rui Wang	916
<i>Generating Non-Projective Word Order in Statistical Linearization</i>	
Bernd Bohnet, Anders Björkelund, Jonas Kuhn, Wolfgang Seeker and Sina Zarriess	928
<i>Learning Syntactic Categories Using Paradigmatic Representations of Word Context</i>	
Mehmet Ali Yatbaz, Enis Sert and Deniz Yuret	940
<i>Exploring Topic Coherence over Many Models and Many Topics</i>	
Keith Stevens, Philip Kegelmeyer, David Andrzejewski and David Buttler	952
<i>Entropy-based Pruning for Phrase-based Machine Translation</i>	
Wang Ling, João Graça, Isabel Trancoso and Alan Black	962
<i>A Systematic Comparison of Phrase Table Pruning Techniques</i>	
Richard Zens, Daisy Stanton and Peng Xu	972
<i>Probabilistic Finite State Machines for Regression-based MT Evaluation</i>	
Mengqiu Wang and Christopher D. Manning	984
<i>An Empirical Investigation of Statistical Significance in NLP</i>	
Taylor Berg-Kirkpatrick, David Burkett and Dan Klein	995
<i>Employing Compositional Semantics and Discourse Consistency in Chinese Event Extraction</i>	
Peifeng Li, Guodong Zhou, Qiaoming Zhu and Libin Hou	1006
<i>Reading The Web with Learned Syntactic-Semantic Inference Rules</i>	
Ni Lao, Amarnag Subramanya, Fernando Pereira and William W. Cohen	1017
<i>Ensemble Semantics for Large-scale Unsupervised Relation Extraction</i>	
Bonan Min, Shuming Shi, Ralph Grishman and Chin-Yew Lin	1027

<i>Forest Reranking through Subtree Ranking</i>	
Richard Farkas and Helmut Schmid	1038
<i>Parser Showdown at the Wall Street Corral: An Empirical Investigation of Error Types in Parser Output</i>	
Jonathan K. Kummerfeld, David Hall, James R. Curran and Dan Klein	1048
<i>Extending Machine Translation Evaluation Metrics with Lexical Cohesion to Document Level</i>	
Billy T. M. Wong and Chunyu Kit	1060
<i>Fast Large-Scale Approximate Graph Construction for NLP</i>	
Amit Goyal, Hal Daume III and Raul Guerra	1069
<i>Building a Lightweight Semantic Model for Unsupervised Information Extraction on Short Listings</i>	
Doo Soon Kim, Kunal Verma and Peter Yeh	1081
<i>Sketch Algorithms for Estimating Point Queries in NLP</i>	
Amit Goyal, Hal Daume III and Graham Cormode	1093
<i>Monte Carlo MCMC: Efficient Inference by Approximate Sampling</i>	
Sameer Singh, Michael Wick and Andrew McCallum	1104
<i>On Amortizing Inference Cost for Structured Prediction</i>	
Vivek Srikumar, Gourab Kundu and Dan Roth	1114
<i>Exact Sampling and Decoding in High-Order Hidden Markov Models</i>	
Simon Carter, Marc Dymetman and Guillaume Bouchard	1125
<i>PATTY: A Taxonomy of Relational Patterns with Semantic Types</i>	
Ndapandula Nakashole, Gerhard Weikum and Fabian Suchanek	1135
<i>Training Factored PCFGs with Expectation Propagation</i>	
David Hall and Dan Klein	1146
<i>A Coherence Model Based on Syntactic Patterns</i>	
Annie Louis and Ani Nenkova	1157
<i>Language Model Rest Costs and Space-Efficient Storage</i>	
Kenneth Heafield, Philipp Koehn and Alon Lavie	1169
<i>Document-Wide Decoding for Phrase-Based Statistical Machine Translation</i>	
Christian Hardmeier, Joakim Nivre and Jörg Tiedemann	1179
<i>Left-to-Right Tree-to-String Decoding with Prediction</i>	
Yang Feng, Yang Liu, Qun Liu and Trevor Cohn	1191
<i>Semantic Compositionality through Recursive Matrix-Vector Spaces</i>	
Richard Socher, Brody Huval, Christopher D. Manning and Andrew Y. Ng	1201
<i>Polarity Inducing Latent Semantic Analysis</i>	
Wen-tau Yih, Geoffrey Zweig and John Platt	1212

<i>First Order vs. Higher Order Modification in Distributional Semantics</i>	
Gemma Boleda, Eva Maria Vecchi, Miquel Cornudella and Louise McNally	1223
<i>Learning-based Multi-Sieve Co-reference Resolution with Knowledge</i>	
Lev Ratinov and Dan Roth	1234
<i>Joint Learning for Coreference Resolution with Markov Logic</i>	
Yang Song, Jing Jiang, Wayne Xin Zhao, Sujian Li and Houfeng Wang	1245
<i>Resolving “This-issue” Anaphora</i>	
Varada Kolhatkar and Graeme Hirst	1255
<i>Entity based Q&A Retrieval</i>	
Amit Singh	1266
<i>Constructing Task-Specific Taxonomies for Document Collection Browsing</i>	
Hui Yang	1278
<i>Besting the Quiz Master: Crowdsourcing Incremental Classification Games</i>	
Jordan Boyd-Graber, Brianna Satinoff, He He and Hal Daume III	1290
<i>Multi-Domain Learning: When Do Domains Matter?</i>	
Mahesh Joshi, Mark Dredze, William W. Cohen and Carolyn Rose	1302
<i>Biased Representation Learning for Domain Adaptation</i>	
Fei Huang and Alexander Yates	1313
<i>Unambiguity Regularization for Unsupervised Learning of Probabilistic Grammars</i>	
Kewei Tu and Vasant Honavar	1324
<i>Extracting Opinion Expressions with semi-Markov Conditional Random Fields</i>	
Bishan Yang and Claire Cardie	1335
<i>Opinion Target Extraction Using Word-Based Translation Model</i>	
Kang Liu, Liheng Xu and Jun Zhao	1346
<i>Word Salad: Relating Food Prices and Descriptions</i>	
Victor Chahuneau, Kevin Gimpel, Bryan R. Routledge, Lily Scherlis and Noah A. Smith ...	1357
<i>Learning to Map into a Universal POS Tagset</i>	
Yuan Zhang, Roi Reichart, Regina Barzilay and Amir Globerson	1368
<i>Part-of-Speech Tagging for Chinese-English Mixed Texts with Dynamic Features</i>	
Jiayi Zhao, Xipeng Qiu, Shu Zhang, Feng Ji and Xuanjing Huang	1379
<i>Wiki-ly Supervised Part-of-Speech Tagging</i>	
Shen Li, João Graça and Ben Taskar	1389
<i>Joining Forces Pays Off: Multilingual Joint Word Sense Disambiguation</i>	
Roberto Navigli and Simone Paolo Ponzetto	1399

<i>A New Minimally-Supervised Framework for Domain Word Sense Disambiguation</i>	
Stefano Faralli and Roberto Navigli	1411
<i>Grounded Models of Semantic Representation</i>	
Carina Silberer and Mirella Lapata	1423
<i>Improved Parsing and POS Tagging Using Inter-Sentence Consistency Constraints</i>	
Alexander Rush, Roi Reichart, Michael Collins and Amir Globerson	1434
<i>Unified Dependency Parsing of Chinese Morphological and Syntactic Structures</i>	
Zhongguo Li and Guodong Zhou	1445
<i>A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing</i>	
Bernd Bohnet and Joakim Nivre	1455
<i>Identifying Event-related Bursts via Social Media Activities</i>	
Xin Zhao, Baihan Shu, Jing Jiang, Yang Song, Hongfei Yan and Xiaoming Li	1466
<i>User Demographics and Language in an Implicit Social Network</i>	
Katja Filippova	1478
<i>Revisiting the Predictability of Language: Response Completion in Social Media</i>	
Bo Pang and Sujith Ravi	1489
<i>Supervised Text-based Geolocation Using Language Models on an Adaptive Grid</i>	
Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing and Jason Baldrige ...	1500
<i>A Discriminative Model for Query Spelling Correction with Latent Structural SVM</i>	
Huizhong Duan, Yanen Li, ChengXiang Zhai and Dan Roth	1511
<i>Characterizing Stylistic Elements in Syntactic Structure</i>	
Song Feng, Ritwik Banerjee and Yejin Choi	1522