

12th Conference of the European Chapter of the Association for Computational Linguistics 2009

(EACL 2009)

**Athens, Greece
30 March - 3 April 2009**

ISBN: 978-1-63266-013-8

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© (2009) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2014)

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: Slacker Semantics: Why Superficiality, Dependency and Avoidance of Commitment can be the Right Way to Go</i> Ann Copestake	1
<i>Invited Talk: NLP and the Humanities: The Revival of an Old Liaison</i> Francisca de Jong	10
<i>On the Use of Comparable Corpora to Improve SMT performance</i> Sadaf Abdul-Rauf and Holger Schwenk	16
<i>Contextual Phrase-Level Polarity Analysis Using Lexical Affect Scoring and Syntactic N-Grams</i> Apoorv Agarwal, Fadi Biadisy and Kathleen Mckeown	24
<i>Personalizing PageRank for Word Sense Disambiguation</i> Eneko Agirre and Aitor Soroa	33
<i>Supervised Domain Adaption for WSD</i> Eneko Agirre and Oier Lopez de Lacalle	42
<i>Clique-Based Clustering for Improving Named Entity Recognition Systems</i> Julien Ah-Pine and Guillaume Jacquet	51
<i>Correcting Automatic Translations through Collaborations between MT and Monolingual Target-Language Users</i> Joshua Albrecht, Rebecca Hwa and G. Elisabeta Marai	60
<i>Incremental Parsing with Parallel Multiple Context-Free Grammars</i> Krasimir Angelov	69
<i>Data-Driven Semantic Analysis for Multilingual WSD and Lexical Selection in Translation</i> Marianna Apidianaki	77
<i>Syntactic Phrase Reordering for English-to-Arabic Statistical Machine Translation</i> Ibrahim Badr, Rabih Zbib and James Glass	86
<i>Incremental Parsing Models for Dialog Task Structure</i> Srinivas Bangalore and Amanda Stent	94
<i>Bayesian Word Sense Induction</i> Samuel Brody and Mirella Lapata	103
<i>Human Evaluation of a German Surface Realisation Ranker</i> Aoife Cahill and Martin Forst	112
<i>Large-Coverage Root Lexicon Extraction for Hindi</i> Cohan Sujay Carlos, Monojit Choudhury and Sandipan Dandapat	121
<i>Lexical Morphology in Machine Translation: A Feasibility Study</i> Bruno Cartoni	130
<i>Predicting the Fluency of Text with Shallow Structural Features: Case Studies of Machine Translation and Human-Written Text</i> Jieun Chae and Ani Nenkova	139

<i>EM Works for Pronoun Anaphora Resolution</i> Eugene Charniak and Micha Elsner	148
<i>Web Augmentation of Language Models for Continuous Speech Recognition of SMS Text Messages</i> Mathias Creutz, Sami Virpioja and Anna Kovaleva	157
<i>An Alignment Algorithm Using Belief Propagation and a Structure-Based Distortion Model</i> Fabien Cromierès and Sadao Kurohashi	166
<i>Translation and Extension of Concepts Across Languages</i> Dmitry Davidov and Ari Rappoport	175
<i>Learning to Interpret Utterances Using Dialogue History</i> David DeVault and Matthew Stone	184
<i>Correcting Dependency Annotation Errors</i> Markus Dickinson	193
<i>Re-Ranking Models for Spoken Language Understanding</i> Marco Dinarelli, Alessandro Moschitti and Giuseppe Riccardi	202
<i>Inference Rules and their Application to Recognizing Textual Entailment</i> Georgiana Dinu and Rui Wang	211
<i>Semi-Supervised Semantic Role Labeling</i> Hagen Fürstenau and Mirella Lapata	220
<i>Cognitively Motivated Features for Readability Assessment</i> Lijun Feng, Noémie Elhadad and Matt Huenerfauth	229
<i>Effects of Word Confusion Networks on Voice Search</i> Junlan Feng and Srinivas Bangalore	238
<i>Company-Oriented Extractive Summarization of Financial News</i> Katja Filippova, Mihai Surdeanu, Massimiliano Ciaramita and Hugo Zaragoza	246
<i>Reconstructing False Start Errors in Spontaneous Speech Text</i> Erin Fitzgerald, Keith Hall and Frederick Jelinek	255
<i>TBL-Improved Non-Deterministic Segmentation and POS Tagging for a Chinese Parser</i> Martin Forst and Ji Fang	264
<i>Who is “You”? Combining Linguistic and Gaze Features to Resolve Second-Person References in Dialogue</i> Matthew Frampton, Raquel Fernández, Patrick Ehlen, Mario Christoudias, Trevor Darrell and Stanley Peters	273
<i>Rich Bitext Projection Features for Parse Reranking</i> Alexander Fraser, Renjing Wang and Hinrich Schütze	282
<i>Parsing Mildly Non-Projective Dependency Structures</i> Carlos Gómez-Rodríguez, David Weir and John Carroll	291
<i>Structural, Transitive and Latent Models for Biographic Fact Extraction</i> Nikesh Garera and David Yarowsky	300

<i>Semitic Morphological Analysis and Generation Using Finite State Transducers with Feature Structures</i> Michael Gasser	309
<i>Cube Summing, Approximate Inference with Non-Local Features, and Dynamic Programming without Semirings</i> Kevin Gimpel and Noah A. Smith	318
<i>Enhancing Unlexicalized Parsing Performance Using a Wide Coverage Lexicon, Fuzzy Tag-Set Mapping, and EM-HMM-Based Lexical Probabilities</i> Yoav Goldberg, Reut Tsarfaty, Meni Adler and Michael Elhadad	327
<i>Person Identification from Text and Speech Genre Samples</i> Jade Goldstein-Stewart, Ransom Winder and Roberta Sabin	336
<i>End-to-End Evaluation in Simultaneous Translation</i> Olivier Hamon, Christian Fügen, Djamel Mostefa, Victoria Arranz, Muntsin Kolss, Alex Waibel and Khalid Choukri	345
<i>Learning-Based Named Entity Recognition for Morphologically-Rich, Resource-Scarce Languages</i> Kazi Saidul Hasan, Md. Altaf ur Rahman and Vincent Ng	354
<i>Weakly Supervised Part-of-Speech Tagging for Morphologically-Rich, Resource-Scarce Languages</i> Kazi Saidul Hasan and Vincent Ng	363
<i>Improving Mid-Range Re-Ordering Using Templates of Factors</i> Hieu Hoang and Philipp Koehn	372
<i>Rule Filtering by Pattern for Efficient Hierarchical Translation</i> Gonzalo Iglesias, Adrià de Gispert, Eduardo R. Banga and William Byrne	380
<i>An Empirical Study on Class-Based Word Sense Disambiguation</i> Rubén Izquierdo, Armando Suárez and German Rigau	389
<i>Generating a Non-English Subjectivity Lexicon: Relations That Matter</i> Valentin Jijkoun and Katja Hofmann	398
<i>Parsing Coordinations</i> Sandra Kübler, Erhard Hinrichs, Wolfgang Maier and Eva Klett	406
<i>Automatic Single-Document Key Fact Extraction from Newswire Articles</i> Itamar Kastner and Christof Monz	415
<i>N-Gram-Based Statistical Machine Translation versus Syntax Augmented Machine Translation: Comparison and System Combination</i> Maxim Khalilov and José A. R. Fonollosa	424
<i>Lightly Supervised Transliteration for Machine Translation</i> Amit Kirschenbaum and Shuly Wintner	433
<i>Optimization in Coreference Resolution is not Needed: A Nearly-Optimal Algorithm with Intensional Constraints</i> Manfred Klenner and Étienne Ailloud	442
<i>A Logic of Semantic Representations for Shallow Parsing</i> Alexander Koller and Alex Lascarides	451

<i>Dependency Trees and the Strong Generative Capacity of CCG</i> Alexander Koller and Marco Kuhlmann	460
<i>Lattice Parsing to Integrate Speech Recognition and Rule-Based Machine Translation</i> Selçuk Köprü and Adnan Yazıcı	469
<i>Trebank Grammar Techniques for Non-Projective Dependency Parsing</i> Marco Kuhlmann and Giorgio Satta	478
<i>Improvements in Analogical Learning: Application to Translating Multi-Terms of the Medical Domain</i> Philippe Langlais, François Yvon and Pierre Zweigenbaum	487
<i>Language-Independent Bilingual Terminology Extraction from a Multilingual Parallel Corpus</i> Els Lefever, Lieve Macken and Veronique Hoste	496
<i>User Simulations for Context-Sensitive Speech Recognition in Spoken Dialogue Systems</i> Oliver Lemon and Ioannis Konstas	505
<i>Sentiment Summarization: Evaluating and Learning User Preferences</i> Kevin Lerman, Sasha Blair-Goldensohn and Ryan McDonald	514
<i>Correcting a POS-Tagged Corpus Using Three Complementary Methods</i> Hrafn Loftsson	523
<i>Translation as Weighted Deduction</i> Adam Lopez	532
<i>Performance Confidence Estimation for Automatic Summarization</i> Annie Louis and Ani Nenkova	541
<i>Bilingually Motivated Domain-Adapted Word Segmentation for Statistical Machine Translation</i> Yanjun Ma and Andy Way	549
<i>Evaluating the Inferential Utility of Lexical-Semantic Resources</i> Shachar Mirkin, Ido Dagan and Eyal Shnarch	558
<i>Text-to-Text Semantic Similarity for Automatic Short Answer Grading</i> Michael Mohler and Rada Mihalcea	567
<i>Syntactic and Semantic Kernels for Short Text Pair Categorization</i> Alessandro Moschitti	576
<i>Discovering Global Patterns in Linguistic Networks through Spectral Analysis: A Case Study of the Consonant Inventories</i> Animesh Mukherjee, Monojit Choudhury and Ravi Kannan	585
<i>Using Cycles and Quasi-Cycles to Disambiguate Dictionary Glosses</i> Roberto Navigli	594
<i>Deterministic Shift-Reduce Parsing for Unification-Based Grammars by Using Default Unification</i> Takashi Ninomiya, Takuya Matsuzaki, Nobuyuki Shimizu and Hiroshi Nakagawa	603
<i>Analysing Wikipedia and Gold-Standard Corpora for NER Training</i> Joel Nothman, Tara Murphy and James R. Curran	612

<i>Using Lexical and Relational Similarity to Classify Semantic Relations</i>	
Diarmuid Ó Séaghdha and Ann Copestake	621
<i>Empirical Evaluations of Animacy Annotation</i>	
Lilja Øvrelid	630
<i>Outclassing Wikipedia in Open-Domain Information Extraction: Weakly-Supervised Acquisition of Attributes over Conceptual Hierarchies</i>	
Marius Paşca	639
<i>Predicting Strong Associations on the Basis of Corpus Data</i>	
Yves Peirsman and Dirk Geeraerts	648
<i>Measuring Frame Relatedness</i>	
Marco Pennacchiotti and Michael Wirth	657
<i>Flexible Answer Typing with Discriminative Preference Ranking</i>	
Christopher Pinchak, Dekang Lin and Davood Rafiei	666
<i>Semi-Supervised Polarity Lexicon Induction</i>	
Delip Rao and Deepak Ravichandran	675
<i>Natural Language Generation as Planning Under Uncertainty for Spoken Dialogue Systems</i>	
Verena Rieser and Oliver Lemon	683
<i>Tagging Urdu Text with Parts of Speech: A Tagger Comparison</i>	
Hassan Sajjad and Helmut Schmid	692
<i>Unsupervised Methods for Head Assignments</i>	
Federico Sangati and Willem Zuidema	701
<i>A General, Abstract Model of Incremental Dialogue Processing</i>	
David Schlangen and Gabriel Skantze	710
<i>Word Lattices for Multi-Source Translation</i>	
Josh Schroeder, Trevor Cohn and Philipp Koehn	719
<i>Frequency Matters: Pitch Accents and Information Status</i>	
Katrin Schweitzer, Michael Walsh, Bernd Möbius, Arndt Riester, Antje Schweitzer and Hinrich Schütze	728
<i>Using Non-Lexical Features to Identify Effective Indexing Terms for Biomedical Illustrations</i>	
Matthew Simpson, Dina Demner-Fushman, Charles Sneiderman, Sameer K. Antani and George R. Thoma	737
<i>Incremental Dialogue Processing in a Micro-Domain</i>	
Gabriel Skantze and David Schlangen	745
<i>Unsupervised Recognition of Literal and Non-Literal Use of Idiomatic Expressions</i>	
Caroline Sporleder and Linlin Li	754
<i>Semi-Supervised Training for the Averaged Perceptron POS Tagger</i>	
Drahomíra “johanka” Spoustová, Jan Hajič, Jan Raab and Miroslav Spousta	763
<i>Sequential Labeling with Latent Variables: An Exact Inference Algorithm and its Efficient Approximation</i>	
Xu Sun and Jun’ichi Tsujii	772

<i>Text Summarization Model Based on Maximum Coverage Problem and its Variant</i> Hiroya Takamura and Manabu Okumura	781
<i>Fast Full Parsing by Linear-Chain Conditional Random Fields</i> Yoshimasa Tsuruoka, Jun'ichi Tsujii and Sophia Ananiadou	790
<i>MINT: A Method for Effective and Scalable Mining of Named Entity Transliterations from Large Comparable Corpora</i> Raghavendra Udupa, K Saravanan, A Kumaran and Jagadeesh Jagarlamudi.....	799
<i>Deriving Generalized Knowledge from Corpora Using WordNet Abstraction</i> Benjamin Van Durme, Phillip Michalak and Lenhart Schubert.....	808
<i>Learning Efficient Parsing</i> Gertjan van Noord.....	817
<i>A Robust and Extensible Exemplar-Based Model of Thematic Fit</i> Bram Vandekerckhove, Dominiek Sandra and Walter Daelemans	826
<i>Growing Finely-Discriminating Taxonomies from Seeds of Varying Quality and Size</i> Tony Veale, Guofu Li and Yanfen Hao	835
<i>Feature-Based Method for Document Alignment in Comparable News Corpora</i> Thuy Vu, Ai Ti Aw and Min Zhang.....	843
<i>Improving Grammaticality in Statistical Sentence Generation: Introducing a Dependency Spanning Tree Algorithm with an Argument Satisfaction Model</i> Stephen Wan, Mark Dras, Robert Dale and Cécile Paris	852
<i>Co-Dispersion: A Windowless Approach to Lexical Association</i> Justin Washtell	861
<i>Language ID in the Context of Harvesting Language Data off the Web</i> Fei Xia, William Lewis and Hoifung Poon	870
<i>Character-Level Dependencies in Chinese: Usefulness and Learning</i> Hai Zhao	879