

13th Conference of the European Chapter of the Association for Computational Linguistics 2012

(EACL 2012)

**Avignon, France
23-27 April 2012**

ISBN: 978-1-62276-042-8

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

Printed by Curran Associates, Inc. (2012)

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>Speech Communication in the Wild</i>	
Martin Cooke	1
<i>Power-Law Distributions for Paraphrases Extracted from Bilingual Corpora</i>	
Spyros Martzoukos and Christof Monz	2
<i>A Bayesian Approach to Unsupervised Semantic Role Induction</i>	
Ivan Titov and Alexandre Klementiev	12
<i>Entailment above the word level in distributional semantics</i>	
Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do and Chung-chieh Shan	23
<i>Evaluating Distributional Models of Semantics for Syntactically Invariant Inference</i>	
Jackie Chi Kit Cheung and Gerald Penn	33
<i>Cross-Framework Evaluation for Statistical Parsing</i>	
Reut Tsarfaty, Joakim Nivre and Evelina Andersson	44
<i>Dependency Parsing of Hungarian: Baseline Results and Challenges</i>	
Richárd Farkas, Veronika Vincze and Helmut Schmid	55
<i>Dependency Parsing with Undirected Graphs</i>	
Carlos Gómez-Rodríguez and Daniel Fernández-González	66
<i>The Best of Both Worlds – A Graph-based Completion Model for Transition-based Parsers</i>	
Bernd Bohnet and Jonas Kuhn	77
<i>Answer Sentence Retrieval by Matching Dependency Paths acquired from Question/Answer Sentence Pairs</i>	
Michael Kaiser	88
<i>Can Click Patterns across User’s Query Logs Predict Answers to Definition Questions?</i>	
Alejandro Figueroa	99
<i>Adaptation of Statistical Machine Translation Model for Cross-Lingual Information Retrieval in a Service Context</i>	
Vassilina Nikoulina, Bogomil Kovachev, Nikolaos Lagos and Christof Monz	109
<i>Computing Lattice BLEU Oracle Scores for Machine Translation</i>	
Artem Sokolov, Guillaume Wisniewski and Francois Yvon	120
<i>Toward Statistical Machine Translation without Parallel Corpora</i>	
Alexandre Klementiev, Ann Irvine, Chris Callison-Burch and David Yarowsky	130
<i>Character-Based Pivot Translation for Under-Resourced Languages and Domains</i>	
Jörg Tiedemann	141
<i>Does more data always yield better translations?</i>	
Guillem Gascó, Martha-Alicia Rocha, Germán Sanchis-Trilles, Jesús Andrés-Ferrer and Francisco Casacuberta	152
<i>Recall-Oriented Learning of Named Entities in Arabic Wikipedia</i>	
Behrang Mohit, Nathan Schneider, Rishav Bhowmick, Kemal Oflazer and Noah A. Smith	162

<i>Tree Representations in Probabilistic Models for Extended Named Entities Detection</i>	
Marco Dinarelli and Sophie Rosset	174
<i>When Did that Happen? — Linking Events and Relations to Timestamps</i>	
Dirk Hovy, James Fan, Alfio Gliozzo, Siddharth Patwardhan and Christopher Welty	185
<i>Compensating for Annotation Errors in Training a Relation Extractor</i>	
Bonan Min and Ralph Grishman	194
<i>Incorporating Lexical Priors into Topic Models</i>	
Jagadeesh Jagarlamudi, Hal Daume III and Raghavendra Udupa	204
<i>DualSum: a Topic-Model based approach for update summarization</i>	
Jean-Yves Delort and Enrique Alfonseca	214
<i>Large-Margin Learning of Submodular Summarization Models</i>	
Ruben Sipos, Pannaga Shivaswamy and Thorsten Joachims	224
<i>A Probabilistic Model of Syntactic and Semantic Acquisition from Child-Directed Utterances and their Meanings</i>	
Tom Kwiatkowski, Sharon Goldwater, Luke Zettlemoyer and Mark Steedman	234
<i>Active learning for interactive machine translation</i>	
Jesús González-Rubio, Daniel Ortiz-Martínez and Francisco Casacuberta	245
<i>Adapting Translation Models to Translationese Improves SMT</i>	
Gennadi Lembersky, Noam Ordan and Shuly Wintner	255
<i>Aspectual Type and Temporal Relation Classification</i>	
Francisco Costa and António Branco	266
<i>Automatic generation of short informative sentiment summaries</i>	
Andrea Glaser and Hinrich Schütze	276
<i>Bootstrapped Training of Event Extraction Classifiers</i>	
Ruihong Huang and Ellen Riloff	286
<i>Bootstrapping Events and Relations from Text</i>	
Ting Liu and Tomasz Strzalkowski	296
<i>CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language</i>	
Svitlana Volkova, William B. Dolan and Theresa Wilson	306
<i>Extending the Entity-based Coherence Model with Multiple Ranks</i>	
Vanessa Wei Feng and Graeme Hirst	315
<i>Generalization Methods for In-Domain and Cross-Domain Opinion Holder Extraction</i>	
Michael Wiegand and Dietrich Klakow	325
<i>Skip N-grams and Ranking Functions for Predicting Script Events</i>	
Bram Jans, Steven Bethard, Ivan Vulić and Marie-Francine Moens	336
<i>The Problem with Kappa</i>	
David Martin Ward Powers	345

<i>User Edits Classification Using Document Revision Histories</i> Amit Bronner and Christof Monz	356
<i>User Participation Prediction in Online Forums</i> Zhonghua Qu and Yang Liu	367
<i>Inferring Selectional Preferences from Part-Of-Speech N-grams</i> Hyeju Jang and Jack Mostow	377
<i>WebCAGe – A Web-Harvested Corpus Annotated with GermaNet Senses</i> Verena Henrich, Erhard Hinrichs and Tatiana Vodolazova	387
<i>Learning to Behave by Reading</i> Regina Barzilay	397
<i>Lexical surprisal as a general predictor of reading time</i> Irene Fernandez Monsalve, Stefan L. Frank and Gabriella Vigliocco	398
<i>Spectral Learning for Non-Deterministic Dependency Parsing</i> Franco M. Luque, Ariadna Quattoni, Borja Balle and Xavier Carreras	409
<i>Combining Tree Structures, Flat Features and Patterns for Biomedical Relation Extraction</i> Md. Faisal Mahbub Chowdhury and Alberto Lavelli	420
<i>Coordination Structure Analysis using Dual Decomposition</i> Atsushi Hanamoto, Takuya Matsuzaki and Jun'ichi Tsujii	430
<i>Cutting the Long Tail: Hybrid Language Models for Translation Style Adaptation</i> Arianna Bisazza and Marcello Federico	439
<i>Detecting Highly Confident Word Translations from Comparable Corpora without Any Prior Knowledge</i> Ivan Vulić and Marie-Francine Moens	449
<i>Efficient parsing with Linear Context-Free Rewriting Systems</i> Andreas van Cranenburgh	460
<i>Evaluating language understanding accuracy with respect to objective outcomes in a dialogue system</i> Myroslava O. Dzikovska, Peter Bell, Amy Isard and Johanna D. Moore	471
<i>Experimenting with Distant Supervision for Emotion Classification</i> Matthew Purver and Stuart Battersby	482
<i>Feature-Rich Part-of-speech Tagging for Morphologically Complex Languages: Application to Bulgarian</i> Georgi Georgiev, Valentin Zhikov, Kiril Simov, Petya Osenova and Preslav Nakov	492
<i>Instance-Driven Attachment of Semantic Annotations over Conceptual Hierarchies</i> Janara Christensen and Marius Pasca	503
<i>Joint Satisfaction of Syntactic and Pragmatic Constraints Improves Incremental Spoken Language Understanding</i> Andreas Peldszus, Okko Buß, Timo Baumann and David Schlangen	514
<i>Learning How to Conjugate the Romanian Verb. Rules for Regular and Partially Irregular Verbs</i> Liviu P. Dinu, Vlad Niculae and Octavia-Maria Sulea	524

<i>Measuring Contextual Fitness Using Error Contexts Extracted from the Wikipedia Revision History</i> Torsten Zesch	529
<i>Perplexity Minimization for Translation Model Domain Adaptation in Statistical Machine Translation</i> Rico Sennrich	539
<i>Subcat-LMF: Fleshing out a standardized format for subcategorization frame interoperability</i> Judith Eckle-Kohler and Iryna Gurevych	550
<i>The effect of domain and text type on text prediction quality</i> Suzan Verberne, Antal van den Bosch, Helmer Strik and Lou Boves	561
<i>The Impact of Spelling Errors on Patent Search</i> Benno Stein, Dennis Hoppe and Tim Gollub	570
<i>UBY - A Large-Scale Unified Lexical-Semantic Resource Based on LMF</i> Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer and Christian Wirth	580
<i>Word Sense Induction for Novel Sense Detection</i> Jey Han Lau, Paul Cook, Diana McCarthy, David Newman and Timothy Baldwin	591
<i>Learning Language from Perceptual Context</i> Raymond Mooney	602
<i>Learning for Microblogs with Distant Supervision: Political Forecasting with Twitter</i> Micol Marchetti-Bowick and Nathanael Chambers	603
<i>Learning from evolving data streams: online triage of bug reports</i> Grzegorz Chrupala	613
<i>Towards a model of formal and informal address in English</i> Manaal Faruqui and Sebastian Pado	623
<i>Character-based kernels for novelistic plot structure</i> Micha Elsner	634
<i>Smart Paradigms and the Predictability and Complexity of Inflectional Morphology</i> Grégoire Détrez and Aarne Ranta	645
<i>Probabilistic Hierarchical Clustering of Morphological Paradigms</i> Burcu Can and Suresh Manandhar	654
<i>Modeling Inflection and Word-Formation in SMT</i> Alexander Fraser, Marion Weller, Aoife Cahill and Fabienne Cap	664
<i>Identifying Broken Plurals, Irregular Gender, and Rationality in Arabic Text</i> Sarah Alkuhlani and Nizar Habash	675
<i>Framework of Semantic Role Assignment based on Extended Lexical Conceptual Structure: Comparison with VerbNet and FrameNet</i> Yuichiroh Matsubayashi, Yusuke Miyao and Akiko Aizawa	686
<i>Unsupervised Detection of Downward-Entailing Operators By Maximizing Classification Certainty</i> Jackie Chi Kit Cheung and Gerald Penn	696

<i>Elliphant: Improved Automatic Detection of Zero Subjects and Impersonal Constructions in Spanish</i> Luz Rello, Ricardo Baeza-Yates and Ruslan Mitkov	706
<i>Validation of sub-sentential paraphrases acquired from parallel monolingual corpora</i> Houda Bouamor, Aurélien Max and Anne Vilnat	716
<i>Determining the placement of German verbs in English-to-German SMT</i> Anita Gojun and Alexander Fraser	726
<i>Syntax-Based Word Ordering Incorporating a Large-Scale Language Model</i> Yue Zhang, Graeme Blackwood and Stephen Clark	736
<i>Midge: Generating Image Descriptions From Computer Vision Detections</i> Margaret Mitchell, Jesse Dodge, Amit Goyal, Kota Yamaguchi, Karl Stratos, Xufeng Han, Alyssa Mensch, Alex Berg, Tamara Berg and Hal Daume III	747
<i>Generation of landmark-based navigation instructions from open-source data</i> Markus Dräger and Alexander Koller	757
<i>To what extent does sentence-internal realisation reflect discourse context? A study on word order</i> Sina Zarrieß, Aoife Cahill and Jonas Kuhn	767
<i>Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages</i> Oliver Ferschke, Iryna Gurevych and Yevgen Chebotar	777
<i>An Unsupervised Dynamic Bayesian Network Approach to Measuring Speech Style Accommodation</i> Mahaveer Jain, John McDonough, Gahgene Gweon, Bhiksha Raj and Carolyn Penstein Rosé .	787
<i>Learning the Fine-Grained Information Status of Discourse Entities</i> Altaf Rahman and Vincent Ng	798
<i>Composing extended top-down tree transducers</i> Aurelie Lagoutte, Fabienne Braune, Daniel Quernheim and Andreas Maletti	808
<i>Structural and Topical Dimensions in Multi-Task Patent Translation</i> Katharina Waeschle and Stefan Riezler	818
<i>Not as Awful as it Seems: Explaining German Case through Computational Experiments in Fluid Construction Grammar</i> Remi van Trijp	829
<i>Managing Uncertainty in Semantic Tagging</i> Silvie Cinková, Martin Holub and Vincent Kríž	840
<i>Parallel and Nested Decomposition for Factoid Questions</i> Aditya Kalyanpur, Siddharth Patwardhan, Branimir Boguraev, Jennifer Chu-Carroll and Adam Lally	851