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
# Table of Contents

*Speech Communication in the Wild*  
Martin Cooke ................................................................. 1

*Power-Law Distributions for Paraphrases Extracted from Bilingual Corpora*  
Spyros Martzoukos and Christof Monz ..................................... 2

*A Bayesian Approach to Unsupervised Semantic Role Induction*  
Ivan Titov and Alexandre Klementiev ...................................... 12

*Entailment above the word level in distributional semantics*  
Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do and Chung-chieh Shan .......... 23

*Evaluating Distributional Models of Semantics for Syntactically Invariant Inference*  
Jackie Chi Kit Cheung and Gerald Penn .................................. 33

*Cross-Framework Evaluation for Statistical Parsing*  
Reut Tsarfaty, Joakim Nivre and Evelina Andersson ......................... 44

*Dependency Parsing of Hungarian: Baseline Results and Challenges*  
Richárd Farkas, Veronika Vincze and Helmut Schmid ...................... 55

*Dependency Parsing with Undirected Graphs*  
Carlos Gómez-Rodríguez and Daniel Fernández-González .................... 66

*The Best of Both Worlds – A Graph-based Completion Model for Transition-based Parsers*  
Bernd Bohnet and Jonas Kuhn ............................................... 77

*Answer Sentence Retrieval by Matching Dependency Paths acquired from Question/Answer Sentence Pairs*  
Michael Kaisser ........................................................................ 88

*Can Click Patterns across User's Query Logs Predict Answers to Definition Questions?*  
Alejandro Figueroa .................................................................. 99

*Adaptation of Statistical Machine Translation Model for Cross-Lingual Information Retrieval in a Service Context*  
Vassilina Nikoulina, Bogomil Kovachev, Nikolaos Lagos and Christof Monz .......... 109

*Computing Lattice BLEU Oracle Scores for Machine Translation*  
Artem Sokolov, Guillaume Wisniewski and Francois Yvon .................... 120

*Toward Statistical Machine Translation without Parallel Corpora*  
Alexandre Klementiev, Ann Irvine, Chris Callison-Burch and David Yarowsky .......... 130

*Character-Based Pivot Translation for Under-Resourced Languages and Domains*  
Jörg Tiedemann ................................................................. 141

*Does more data always yield better translations?*  
Guillem Gascó, Martha-Alicia Rocha, Germán Sanchis-Trilles, Jesús Andrés-Ferrer and Francisco Casacuberta ................................................................. 152

*Recall-Oriented Learning of Named Entities in Arabic Wikipedia*  
Behrang Mohit, Nathan Schneider, Rishav Bhowmick, Kemal Oflazer and Noah A. Smith .... 162
Tree Representations in Probabilistic Models for Extended Named Entities Detection
Marco Dinarelli and Sophie Rosset

When Did that Happen? — Linking Events and Relations to Timestamps
Dirk Hovy, James Fan, Alfio Gliozzo, Siddharth Patwardhan and Christopher Welty

Compensating for Annotation Errors in Training a Relation Extractor
Bonan Min and Ralph Grishman

Incorporating Lexical Priors into Topic Models
Jagadeesh Jagarlamudi, Hal Daume III and Raghavendra Udupa

DualSum: a Topic-Model based approach for update summarization
Jean-Yves Delort and Enrique Alfonseca

Large-Margin Learning of Submodular Summarization Models
Ruben Sipos, Pannaga Shivashwamy and Thorsten Joachims

A Probabilistic Model of Syntactic and Semantic Acquisition from Child-Directed Utterances and their Meanings
Tom Kwiatkowski, Sharon Goldwater, Luke Zettlemoyer and Mark Steedman

Active learning for interactive machine translation
Jesús González-Rubio, Daniel Ortiz-Martinez and Francisco Casacuberta

Adapting Translation Models to Translationese Improves SMT
Gennadi Lembersky, Noam Ordan and Shuly Wintner

Aspectual Type and Temporal Relation Classification
Francisco Costa and António Branco

Automatic generation of short informative sentiment summaries
Andrea Glaser and Hinrich Schütze

Bootstrapped Training of Event Extraction Classifiers
Ruihong Huang and Ellen Riloff

Bootstrapping Events and Relations from Text
Ting Liu and Tomek Strzalkowski

CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language
Svitlana Volkova, William B. Dolan and Theresa Wilson

Extending the Entity-based Coherence Model with Multiple Ranks
Vanessa Wei Feng and Graeme Hirst

Generalization Methods for In-Domain and Cross-Domain Opinion Holder Extraction
Michael Wiegand and Dietrich Klakow

Skip N-grams and Ranking Functions for Predicting Script Events
Bram Jans, Steven Bethard, Ivan Vulić and Marie-Francine Moens

The Problem with Kappa
David Martin Ward Powers
User Edits Classification Using Document Revision Histories
Amit Bronner and Christof Monz ................................................................. 356

User Participation Prediction in Online Forums
Zhonghua Qu and Yang Liu ................................................................. 367

Inferring Selectional Preferences from Part-Of-Speech N-grams
Hyeju Jang and Jack Mostow ............................................................... 377

WebCAGE – A Web-Harvested Corpus Annotated with GermaNet Senses
Verena Henrich, Erhard Hinrichs and Tatiana Vodolazova .......................... 387

Learning to Behave by Reading
Regina Barzilay ...................................................................................... 397

Lexical surprisal as a general predictor of reading time
Irene Fernandez Monsalve, Stefan L. Frank and Gabriella Vigliocco .............. 398

Spectral Learning for Non-Deterministic Dependency Parsing
Franco M. Luque, Ariadna Quattoni, Borja Balle and Xavier Carreras ............ 409

Combining Tree Structures, Flat Features and Patterns for Biomedical Relation Extraction
Md. Faisal Mahbub Chowdhury and Alberto Lavelli .................................. 420

Coordination Structure Analysis using Dual Decomposition
Atsushi Hanamoto, Takuya Matsuzaki and Jun’ichi Tsujii ............................. 430

Cutting the Long Tail: Hybrid Language Models for Translation Style Adaptation
Arianna Bisazza and Marcello Federico ..................................................... 439

Detecting Highly Confident Word Translations from Comparable Corpora without Any Prior Knowledge
Ivan Vulić and Marie-Francine Moens ....................................................... 449

Efficient parsing with Linear Context-Free Rewriting Systems
Andreas van Cranenburgh ........................................................................ 460

Evaluating language understanding accuracy with respect to objective outcomes in a dialogue system
Myroslava O. Dzikovska, Peter Bell, Amy Isard and Johanna D. Moore ........... 471

Experimenting with Distant Supervision for Emotion Classification
Matthew Purver and Stuart Battersby ........................................................ 482

Feature-Rich Part-of-speech Tagging for Morphologically Complex Languages: Application to Bulgarian
Georgi Georgiev, Valentin Zhikov, Kiril Simov, Petya Osenova and Preslav Nakov .... 492

Instance-Driven Attachment of Semantic Annotations over Conceptual Hierarchies
Janara Christensen and Marius Pasca ....................................................... 503

Joint Satisfaction of Syntactic and Pragmatic Constraints Improves Incremental Spoken Language Understanding
Andreas Peldszus, Okko Buß, Timo Baumann and David Schlangen ................. 514

Learning How to Conjugate the Romanian Verb. Rules for Regular and Partially Irregular Verbs
Liviu P. Dinu, Vlad Niculae and Octavia-Maria Sulea ................................... 524
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measuring Contextual Fitness Using Error Contexts Extracted from the Wikipedia Revision History</td>
<td>529</td>
</tr>
<tr>
<td>Torsten Zesch</td>
<td></td>
</tr>
<tr>
<td>Perplexity Minimization for Translation Model Domain Adaptation in Statistical Machine Translation</td>
<td>539</td>
</tr>
<tr>
<td>Rico Sennrich</td>
<td></td>
</tr>
<tr>
<td>Subcat-LMF: Fleshing out a standardized format for subcategorization frame interoperability</td>
<td>550</td>
</tr>
<tr>
<td>Judith Eckle-Kohler and Iryna Gurevych</td>
<td></td>
</tr>
<tr>
<td>The effect of domain and text type on text prediction quality</td>
<td>561</td>
</tr>
<tr>
<td>Suzan Verberne, Antal van den Bosch, Helmer Strik and Lou Boves</td>
<td></td>
</tr>
<tr>
<td>The Impact of Spelling Errors on Patent Search</td>
<td>570</td>
</tr>
<tr>
<td>Benno Stein, Dennis Hoppe and Tim Gollub</td>
<td></td>
</tr>
<tr>
<td>UBY - A Large-Scale Unified Lexical-Semantic Resource Based on LMF</td>
<td>580</td>
</tr>
<tr>
<td>Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer and Christian Wirth</td>
<td></td>
</tr>
<tr>
<td>Word Sense Induction for Novel Sense Detection</td>
<td>591</td>
</tr>
<tr>
<td>Jey Han Lau, Paul Cook, Diana McCarthy, David Newman and Timothy Baldwin</td>
<td></td>
</tr>
<tr>
<td>Learning Language from Perceptual Context</td>
<td>602</td>
</tr>
<tr>
<td>Raymond Mooney</td>
<td></td>
</tr>
<tr>
<td>Learning for Microblogs with Distant Supervision: Political Forecasting with Twitter</td>
<td>603</td>
</tr>
<tr>
<td>Micol Marchetti-Bowick and Nathanael Chambers</td>
<td></td>
</tr>
<tr>
<td>Learning from evolving data streams: online triage of bug reports</td>
<td>613</td>
</tr>
<tr>
<td>Grzegorz Chrupala</td>
<td></td>
</tr>
<tr>
<td>Towards a model of formal and informal address in English</td>
<td>623</td>
</tr>
<tr>
<td>Manaal Faruqui and Sebastian Pado</td>
<td></td>
</tr>
<tr>
<td>Character-based kernels for novelistic plot structure</td>
<td>634</td>
</tr>
<tr>
<td>Micha Elsner</td>
<td></td>
</tr>
<tr>
<td>Smart Paradigms and the Predictability and Complexity of Inflectional Morphology</td>
<td>645</td>
</tr>
<tr>
<td>Grégoire Détrez and Aarne Ranta</td>
<td></td>
</tr>
<tr>
<td>Probabilistic Hierarchical Clustering of Morphological Paradigms</td>
<td>654</td>
</tr>
<tr>
<td>Burcu Can and Suresh Manandhar</td>
<td></td>
</tr>
<tr>
<td>Modeling Inflection and Word-Formation in SMT</td>
<td>664</td>
</tr>
<tr>
<td>Alexander Fraser, Marion Weller, Aoife Cahill and Fabienne Cap</td>
<td></td>
</tr>
<tr>
<td>Identifying Broken Plurals, Irregular Gender, and Rationality in Arabic Text</td>
<td>675</td>
</tr>
<tr>
<td>Sarah Alkuhlani and Nizar Habash</td>
<td></td>
</tr>
<tr>
<td>Framework of Semantic Role Assignment based on Extended Lexical Conceptual Structure: Comparison with VerbNet and FrameNet</td>
<td>686</td>
</tr>
<tr>
<td>Yuichihiro Matsubayashi, Yusuke Miyao and Akiko Aizawa</td>
<td></td>
</tr>
<tr>
<td>Unsupervised Detection of Downward-Entailing Operators By Maximizing Classification Certainty</td>
<td>696</td>
</tr>
<tr>
<td>Jackie Chi Kit Cheung and Gerald Penn</td>
<td></td>
</tr>
</tbody>
</table>
Elliphant: Improved Automatic Detection of Zero Subjects and Impersonal Constructions in Spanish
Luz Rello, Ricardo Baeza-Yates and Ruslan Mitkov ..................................................... 706

Validation of sub-sentential paraphrases acquired from parallel monolingual corpora
Houda Bouamor, Aurélien Max and Anne Vilnat ........................................................... 716

Determining the placement of German verbs in English–to–German SMT
Anita Gojun and Alexander Fraser ..................................................................................... 726

Syntax-Based Word Ordering Incorporating a Large-Scale Language Model
Yue Zhang, Graeme Blackwood and Stephen Clark ......................................................... 736

Midge: Generating Image Descriptions From Computer Vision Detections
Margaret Mitchell, Jesse Dodge, Amit Goyal, Kota Yamaguchi, Karl Stratos, Xufeng Han, Alyssa
Mensch, Alex Berg, Tamara Berg and Hal Daume III ....................................................... 747

Generation of landmark-based navigation instructions from open-source data
Markus Dräger and Alexander Koller ................................................................................ 757

To what extent does sentence-internal realisation reflect discourse context? A study on word order
Sina Zarrieß, Aoife Cahill and Jonas Kuhn ........................................................................ 767

Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages
Oliver Ferschke, Iryna Gurevych and Yevgen Chebotar ................................................ 777

An Unsupervised Dynamic Bayesian Network Approach to Measuring Speech Style Accommodation
Mahaveer Jain, John McDonough, Gahgene Gweon, Bhiksha Raj and Carolyn Penstein Rosé . 787

Learning the Fine-Grained Information Status of Discourse Entities
Altaf Rahman and Vincent Ng ......................................................................................... 798

Composing extended top-down tree transducers
Aurélie Lagoutte, Fabienne Braune, Daniel Quernheim and Andreas Maletti ..................... 808

Structural and Topical Dimensions in Multi-Task Patent Translation
Katharina Waeschle and Stefan Riezler ............................................................................. 818

Not as Awful as it Seems: Explaining German Case through Computational Experiments in Fluid Con-
struction Grammar
Remi van Trijp ................................................................................................................... 829

Managing Uncertainty in Semantic Tagging
Silvie Cinková, Martin Holub and Vincent Kríž .............................................................. 840

Parallel and Nested Decomposition for Factoid Questions
Aditya Kalyanpur, Siddharth Patwardhan, Branimir Boguraev, Jennifer Chu-Carroll and Adam
Lally .................................................................................................................................... 851