24th International Conference on Computational Linguistics

(COLING 2012)

Posters

Mumbai, India
8-15 December 2012

Volume 1 of 2

Editors:

Martin Kay
Christian Boitet

# Table of Contents

**K-Best Spanning Tree Dependency Parsing With Verb Valency Lexicon Reranking**  
Zeljko Agic ................................................................. 1

**A Best-First Anagram Hashing Filter for Approximate String Matching with Generalized Edit Distance**  
Malin Ahlberg and Gerlof Bouma ........................................ 13

**Automatic Bilingual Phrase Extraction from Comparable Corpora**  
Ahmet Aker, Yang Feng and Robert Gaizauskas .......................... 23

**A Formalized Reference Grammar for UNL-Based Machine Translation between English and Arabic**  
Sameh Alansary .......................................................... 33

**Mapping Arabic Wikipedia into the Named Entities Taxonomy**  
Fahd Alotaibi and Mark Lee .............................................. 43

**Probabilistic Refinement Algorithms for the Generation of Referring Expressions**  
Romina Altamirano, Carlos Areces and Luciana Benotti ................. 53

**Measuring the Adequacy of Cross-Lingual Paraphrases in a Machine Translation Setting**  
Marianna Apidianaki ....................................................... 63

**Cross-Lingual Sentiment Analysis for Indian Languages using Linked WordNets**  
Balamurali A.R., Aditya Joshi and Pushpak Bhattacharyya ............. 73

**The Creation of Large-Scale Annotated Corpora of Minority Languages using UniParser and the EANC platform**  
Timofey Arkhangeskiy, Oleg Belyaev and Arseniy Vydrin .............. 83

**Collocation Extraction using Parallel Corpus**  
Kavosh Asadi Atui, Heshaam Faili and Kaveh Assadi Atuie ............. 93

**Improved Spelling Error Detection and Correction for Arabic**  
Mohammed Attia, Pavel Pecina, Younes Samih, Khaled Shaalan and Josef van Genabith 103

**Heloise —- A Reengineering of Ariane-G5 SLLPs for Application to \( \pi \)-languages**  
Vincent Berment and Christian Boitet ...................................... 113

**Machine Translation for Language Preservation**  
Steven Bird and David Chiang ............................................ 125

**Comparing Non-projective Strategies for Labeled Graph-Based Dependency Parsing**  
Anders Björkelund and Jonas Kuhn ....................................... 135

**Phrase Structures and Dependencies for End-to-End Coreference Resolution**  
Anders Björkelund and Jonas Kuhn ....................................... 145

**The Language of Power and its Cultural Influence**  
David Bracewell and Marc Tomlinson .................................... 155
Learning Opinionated Patterns for Contextual Opinion Detection
Caroline Brun ................................................................. 165

Does Similarity Matter? The Case of Answer Extraction from Technical Discussion Forums
Rose Catherine, Amit Singh, Rashmi Gangadharaiyah, Dinesh Raghu and Karthik Visweswariah
175

Chinese Noun Phrase Coreference Resolution: Insights into the State of the Art
Chen Chen and Vincent Ng .................................................. 185

Linguistic and Statistical Traits Characterising Plagiarism
Miranda Chong and Lucia Specia ........................................... 195

Impact of Less Skewed Distributions on Efficiency and Effectiveness of Biomedical Relation Extraction
Md. Faisal Mahbub Chowdhury and Alberto Lavelli .................. 205

Lattice Rescoring for Speech Recognition using Large Scale Distributed Language Models
Euisok Chung, Hyung-Bae Jeon, Jeong-Gue Park and Yun-Keun Lee 217

Morphological Analyzer for Affix Stacking Languages: A Case Study of Marathi
Raj Dabre, Archana Amberkar and Pushpak Bhattacharyya ............ 225

Modelling the Organization and Processing of Bangla Polymorphemic Words in the Mental Lexicon: A Computational Approach
Tirthankar Dasgupta, Manjira Sinha and Anupam Basu ................. 235

Coreference Clustering using Column Generation
Jan De Belder and Marie-Francine Moens ................................ 245

Metric Learning for Graph-Based Domain Adaptation
Paramveer Dhillon, Partha Talukdar and Koby Crammer ............... 255

Automatic Hashtag Recommendation for Microblogs using Topic-Specific Translation Model
Zhuoye Ding, Qi Zhang and Xuanjing Huang .............................. 265

Unsupervised Feature-Rich Clustering
Vladimir Eidelman .................................................................. 275

Token Level Identification of Linguistic Code Switching
Heba Elfardy and Mona Diab ................................................. 287

Parenthetical Classification for Information Extraction
Ismail El Maarouf and Jeanne Villaneau ................................... 297

A Dictionary-Based Approach to Identifying Aspects Implied by Adjectives for Opinion Mining
Geli Fei, Bing Liu, Meichun Hsu, Malu Castellanos and Riddhiman Ghosh 309

Dealing with Input Noise in Statistical Machine Translation
Lluís Formiga and José A. R. Fonollosa ..................................... 319

A Comparison of Knowledge-based Algorithms for Graded Word Sense Assignment
Annamarie Friedrich, Nikos Engonopoulos, Stefan Thater and Manfred Pinkal .... 329
Leveraging Statistical Transliteration for Dictionary-Based English-Bengali CLIR of OCR’d Text
Utpal Garain, Arjun Das, David Doermann and Douglas Oard ............................. 339

RU-EVAL-2012: Evaluating Dependency Parsers for Russian
Anastasia Gareyshina, Maxim Ionov, Olga Lyshevskaya, Dmitry Privoznov, Elena Sokolova and Svetlana Toldova................................................................. 349

Assessing Sentiment Strength in Words Prior Polarities
Lorenzo Gatti and Marco Guerini ................................................................. 361

Improving Dependency Parsing with Interlinear Glossed Text and Syntactic Projection
Ryan Georgi, Fei Xia and William Lewis .................................................... 371

Diachronic Variation in Grammatical Relations
Aaron Gerow and Khurshid Ahmad .................................................. 381

Relation Classification using Entity Sequence Kernels
Debanjan Ghosh and Smaranda Muresan .............................................. 391

Translating Questions to SQL Queries with Generative Parsers Discriminatively Reranked
Alessandra Giordani and Alessandro Moschitti ........................................ 401

Classifier-Based Tense Model for SMT
Zhengxian Gong, Min Zhang, Chew-lim Tan and Guodong Zhou..................... 411

Extracting and Normalizing Entity-Actions from Users’ Comments
Swapna Gottipati and Jing Jiang ..................................................... 421

Expected Divergence Based Feature Selection for Learning to Rank
Parth Gupta and Paolo Rosso .................................................................. 431

LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors
Aaron L. F. Han, Derek F. Wong and Lidia S. Chao .................................... 441

Predicting Stance in Ideological Debate with Rich Linguistic Knowledge
Kazi Saidul Hasan and Vincent Ng ....................................................... 451

FeatureForge: A Novel Tool for Visually Supported Feature Engineering and Corpus Revision
Florian Heimerl, Charles Jochim, Steffen Koch and Thomas Ertl ....................... 461

Verb Temporality Analysis using Reichenbach’s Tense System
André Horie, Kumiko Tanaka-Ishii and Mitsuru Ishizuka ..................................... 471

A Metric for Evaluating Discourse Coherence based on Coreference Resolution
Ryu Iida and Takenobu Tokunaga ........................................................... 483

Comparing Word Relatedness Measures Based on Google n-grams
Aminul Islam, Evangelos Milios and Vlado Keselj ........................................... 495

Two-Stage Bootstrapping for Anaphora Resolution
Balaji Jagan, T V Geetha and Ranjani Parthasarathi ........................................ 507

Explorations in the Speakers’ Interaction Experience and Self-Assessments
Kristiina Jokinen .................................................................. 517
Multimodal Signals and Holistic Interaction Structuring
Kristiina Jokinen and Graham Wilcock ............................................. 527

New Insights from Coarse Word Sense Disambiguation in the Crowd
Adam Kapelner, Krishna Kaliannan, H. Andrew Schwartz, Lyle Ungar and Dean Foster 539

A Unified Sentence Space for Categorical Distributional-Compositional Semantics: Theory and Experiments
Dimitri Kartsaklis, Mehrnoosh Sadrzadeh and Stephen Pulman .............................. 549

A Knowledge-Based Approach to Syntactic Disambiguation of Biomedical Noun Compounds
Ramakanth Kavuluru and Daniel Harris ............................................ 559

Classification of Inconsistent Sentiment Words using Syntactic Constructions
Wiltrud Kessler and Hinrich Schütze .................................................. 569

Learning Semantics with Deep Belief Network for Cross-Language Information Retrieval
Jungi Kim, Jinseok Nam and Iryna Gurevych ........................................... 579

Detection of Acoustic-Phonetic Landmarks in Mismatched Conditions using a Biomimetic Model of Human Auditory Processing
Sarah King and Mark Hasegawa-Johnson ........................................... 589

Learning Verbs on the Fly
Zornitsa Kozareva ................................................................. 599

Decoder-based Discriminative Training of Phrase Segmentation for Statistical Machine Translation
Hyoung-Gyu Lee and Hae-Chang Rim ............................................. 611

Glimpses of Ancient China from Classical Chinese Poems
John Lee and Tak-sum Wong .......................................................... 621

Conversion between Scripts of Punjabi: Beyond Simple Transliteration
Gurpreet Singh Lehal and Tejinder Singh Saini ........................................... 633

Development of a Complete Urdu-Hindi Transliteration System
Gurpreet Singh Lehal and Tejinder Singh Saini ........................................... 643

Random Walks on Context-Aware Relation Graphs for Ranking Social Tags
Han Li, Zhiyuan Liu and Maosong Sun ............................................. 653

Phrase-Based Evaluation for Machine Translation
Liangyou Li, Zhengxian Gong and Guodong Zhou ........................................... 663

A Beam Search Algorithm for ITG Word Alignment
Peng Li, Yang Liu and Maosong Sun .................................................. 673

Active Learning for Chinese Word Segmentation
Shoushan Li, Guodong Zhou and Chu-Ren Huang ........................................... 683

Fine-Grained Classification of Named Entities by Fusing Multi-Features
Wenjie Li, Jiwei Li, Ye Tian and Zhifang Sui ........................................... 693

Expert Finding for Microblog Misinformation Identification
Chen Liang, Zhiyuan Liu and Maosong Sun ........................................... 703
Improving Relative-Entropy Pruning using Statistical Significance
Wang Ling, Nadi Tomeh, Guang Xiang, Isabel Trancoso and Alan Black.............. 713

Expected Error Minimization with Ultraconservative Update for SMT
Lemao Liu, Tiejun Zhao, Taro Watanabe, Hailong Cao and Conghui Zhu............ 723

Generalized Sentiment-Bearing Expression Features for Sentiment Analysis
Shizhu Liu, Gady Agam and David Grossman........................................... 733

Unsupervised Domain Adaptation for Joint Segmentation and POS-Tagging
Yang Liu and Yue Zhang........................................................................... 745

Tag Dispatch Model with Social Network Regularization for Microblog User Tag Suggestion
Zhiyuan Liu, Cunchao Tu and Maosong Sun........................................... 755

Summarization of Business-Related Tweets: A Concept-Based Approach
Annie Louis and Todd Newman.............................................................. 765

Towards the Automatic Detection of the Source Language of a Literary Translation.
Gerard Lynch and Carl Vogel.................................................................... 775

Fourth-Order Dependency Parsing
Xuezhe Ma and Hai Zhao........................................................................... 785

A Subjective Logic Framework for Multi-Document Summarization
Sukanya Manna, Byron J. Gao and Reed Coke.......................................... 797

Manual Corpus Annotation: Giving Meaning to the Evaluation Metrics
Yann Mathet, Antoine Widlöcher, Karên Fort, Claire François, Olivier Galibert, Cyril Grouin, Juliette Kahn, Sophie Rosset and Pierre Zweigenbaum................................. 809

Discriminative Boosting from Dictionary and Raw Text – A Novel Approach to Build A Chinese Word Segmenter
Fandong Meng, Wenbin Jiang, Hao Xiong and Qun Liu............................. 819

Lost in Translations? Building Sentiment Lexicons using Context Based Machine Translation
Xinfan Meng, Furu Wei, Ge Xu, Longkai Zhang, Xiaohua Liu, Ming Zhou and Houfeng Wang................................................................. 829

How Does the Granularity of an Annotation Scheme Influence Dependency Parsing Performance?
Simon Mille, Alicia Burga, Gabriela Ferraro and Leo Wanner...................... 839

Does Tectogrammatics Help the Annotation of Discourse?
Jiří Mírovský, Pavlína Jínová and Lucie Poláková.................................... 853

The Effect of Learner Corpus Size in Grammatical Error Correction of ESL Writings
Tomoya Mizumoto, Yuta Hayashibe, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto 863

GRAFIX: Automated Rule-Based Post Editing System to Improve English-Persian SMT Output
Mahsa Mohaghegh, Abdolhossein Sarrafzadeh and Mehdi Mohammadi........... 873

Relational Structures and Models for Coreference Resolution
Truc-Vien T. Nguyen and Massimo Poesio............................................. 883
Text Summarization Model based on Redundancy-Constrained Knapsack Problem
Hitoshi Nishikawa, Tsutomu Hirao, Toshiro Makino and Yoshihiro Matsuo .......... 893

Lexical Categories for Improved Parsing of Web Data
Lilja Øvrelid and Arne Skjærholt ................................................... 903

Text-To-Speech for Languages without an Orthography
Sukhada Palkar, Alan Black and Alok Parlikar ...................................... 913

Part of Speech (POS) Tagger for Kokborok
Braja Gopal Patra, Khumbar Debbarma, Dipankar Das and Sivaji Bandyopadhyay . 923

Forced Derivations for Hierarchical Machine Translation
Stephan Peitz, Arne Mauser, Joern Wuebker and Hermann Ney ..................... 933

On Panini and the Generative Capacity of Contextualized Replacement Systems
Gerald Penn and Paul Kiparsky .................................................... 943

Joint Segmentation and Tagging with Coupled Sequences Labeling
Xipeng Qiu, Feng Ji, Jiayi Zhao and Xuanjing Huang ............................... 951

Defining Syntax for Learner Language Annotation
Marwa Ragheb and Markus Dickinson ............................................. 965

How Good are Typological Distances for Determining Genealogical Relationships among Languages?
Taraka Rama and Prasanth Kolachina ............................................. 975

Sentence Boundary Detection: A Long Solved Problem?
Jonathon Read, Rebecca Dridan, Stephan Oepen and Lars Jørgen Solberg ........ 985

Document and Corpus Level Inference For Unsupervised and Transductive Learning of Information Structure of Scientific Documents
Roi Reichart and Anna Korhonen .................................................. 995

Light Textual Inference for Semantic Parsing
Kyle Richardson and Jonas Kuhn ................................................. 1007

Korektor — A System for Contextual Spell-Checking and Diacritics Completion
Michal Richter, Pavel Straňák and Alexandr Rosen ................................. 1019

Using Qualia Information to Identify Lexical Semantic Classes in an Unsupervised Clustering Task
Lauren Romeo, Sara Mendes and Núria Bel ....................................... 1029

A Strategy of Mapping Polish WordNet onto Princeton WordNet
Ewa Rudnicka, Marek Maziarz, Maciej Piasecki and Stan Szpakowicz .......... 1039

A Hierarchical Domain Model-Based Multi-Domain Selection Framework for Multi-Domain Dialog Systems
Seonghan Ryu, Donghyeon Lee, Injae Lee, Sangdo Han, Gary Geunbae Lee, Myungjae Kim and Kyungduk Kim ............................................ 1049

A Fully Coreference-annotated Corpus of Scholarly Papers from the ACL Anthology
Ulrich Schäfer, Christian Spurk and Jörg Steffen .................................. 1059
Continuous Space Translation Models for Phrase-Based Statistical Machine Translation
Holger Schwenk ................................................................. 1071

Data-driven Dependency Parsing With Empty Heads
Wolfgang Seeker, Richárd Farkas, Bernd Bohnet, Helmut Schmid and Jonas Kuhn .... 1081

Extension of TSVM to Multi-Class and Hierarchical Text Classification Problems With General Losses
Sathiya Keerthi Selvaraj, Sundararajan Sellamanickam and Shirish Shevade...... 1091

Calculation of Phrase Probabilities for Statistical Machine Translation by using Belief Functions
Christophe Servan and Simon Petitrenaud ........................................ 1101

Sense and Reference Disambiguation in Wikipedia
Hui Shen, Razvan Bunescu and Rada Mihalcea .................................. 1111

Unsupervised Metaphor Paraphrasing using a Vector Space Model
Ekaterina Shutova, Tim van de Cruys and Anna Korhonen ...................... 1121

Memory-Efficient Katakana Compound Segmentation using Conditional Random Fields
Krauchanka Siarhei and Artsimenya Artsiom ..................................... 1131

New Readability Measures for Bangla and Hindi Texts
Manjira Sinha, Sakshi Sharma, Tirthankar Dasgupta and Anupam Basu ........... 1141

Automatic Question Generation in Multimedia-Based Learning
Yvonne Skalban, Le An Ha, Lucia Specia and Ruslan Mitkov ...................... 1151

A More Cohesive Summarizer
Christian Smith, Henrik Danielsson and Arne Jönsson .......................... 1161

Robust Learning in Random Subspaces: Equipping NLP for OOV Effects
Anders Søgaard and Anders Johannsen ......................................... 1171

An Empirical Etudy of Non-Lexical Extensions to Delexicalized Transfer
Anders Søgaard and Julie Wulff .................................................. 1181

Entropy-based Training Data Selection for Domain Adaptation
Yan Song, Prescott Klassen, Fei Xia and Chunyu Kit ............................. 1191

Corpus-based Explorations of Affective Load Differences in Arabic-Hebrew-English
Carlo Strapparava, Oliviero Stock and Ilai Alon ................................ 1201

Acquiring and Generalizing Causal Inference Rules from Deverbal Noun Constructions
Shohei Tanaka, Naoki Okazaki and Mitsuru Ishizuka ............................. 1209

Advertising Legality Recognition
Yi-jie Tang, Cong-kai Lin and Hsin-Hsi Chen .................................... 1219

A Joint Phrasal and Dependency Model for Paraphrase Alignment
Kapil Thadani, Scott Martin and Michael White .................................. 1229

Sourcing the Crowd for a Few Good Ones: Event Type Detection
Caselli Tommaso and Huang Chu-Ren ............................................. 1239
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combining Multiple Alignments to Improve Machine Translation</td>
<td>Zhaopeng Tu, Yang Liu, Yifan He, Josef van Genabith, Qun Liu and Shouxun Lin</td>
<td>1249</td>
</tr>
<tr>
<td>A New Search Approach for Interactive-Predictive Computer-Assisted Translation</td>
<td>Zeinab Vakil and Shahram Khadivi</td>
<td>1261</td>
</tr>
<tr>
<td>Automatic Extraction of Polar Adjectives for the Creation of Polarity Lexicons</td>
<td>Silvia Vázquez, Muntsa Padró, Núria Bel and Julio Gonzalo</td>
<td>1271</td>
</tr>
<tr>
<td>Optimal Scheduling of Information Extraction Algorithms</td>
<td>Henning Wachsmuth and Benno Stein</td>
<td>1281</td>
</tr>
<tr>
<td>Update Summarization Based on Co-Ranking with Constraints</td>
<td>Xiaojun Wan</td>
<td>1291</td>
</tr>
<tr>
<td>Sentence Realization with Unlexicalized Tree Linearization Grammars</td>
<td>Rui Wang and Yi Zhang</td>
<td>1301</td>
</tr>
<tr>
<td>Exploiting Discourse Relations for Sentiment Analysis</td>
<td>Fei Wang, Yunfang Wu and Likun Qiu</td>
<td>1311</td>
</tr>
<tr>
<td>Expansion Methods for Job-Candidate Matching Amidst Unreliable and Sparse Data</td>
<td>Jerome White, Krishna Kummamuru and Nitendra Rajput</td>
<td>1321</td>
</tr>
<tr>
<td>A Unified Framework for Discourse Argument Identification via Shallow Semantic Parsing</td>
<td>Fan Xu, Qiaoming Zhu and Guodong Zhou</td>
<td>1331</td>
</tr>
<tr>
<td>Using Deep Linguistic Features for Finding Deceptive Opinion Spam</td>
<td>Qiongkai Xu and Hai Zhao</td>
<td>1341</td>
</tr>
<tr>
<td>Latent Community Discovery with Network Regularization for Core Actors Clustering</td>
<td>Guangxu Xun, Yujiu Yang, Liangwei Wang and Wenhuang Liu</td>
<td>1351</td>
</tr>
<tr>
<td>HYENA: Hierarchical Type Classification for Entity Names</td>
<td>Mohamed Amir Yosef, Sandro Bauer, Johannes Hoffart, Marc Spaniol and Gerhard Weikum</td>
<td>1361</td>
</tr>
<tr>
<td>Identifying Temporal Relations by Sentence and Document Optimizations</td>
<td>Katsumasa Yoshikawa, Masayuki Asahara and Ryu Iida</td>
<td>1371</td>
</tr>
<tr>
<td>Affect Detection from Semantic Interpretation of Drama Improvisation</td>
<td>Li Zhang and Ming Jiang</td>
<td>1381</td>
</tr>
<tr>
<td>Analyzing the Effect of Global Learning and Beam-Search on Transition-Based Dependency Parsing</td>
<td>Yue Zhang and Joakim Nivre</td>
<td>1391</td>
</tr>
<tr>
<td>Chinese Word Sense Disambiguation based on Context Expansion</td>
<td>Yang Zhizhuo and Huang Heyan</td>
<td>1401</td>
</tr>
<tr>
<td>Cross-Lingual Identification of Ambiguous Discourse Connectives for Resource-Poor Language</td>
<td>Lanjun Zhou, Wei Gao, Binyang Li, Zhongyu Wei and Kam-Fai Wong</td>
<td>1409</td>
</tr>
</tbody>
</table>