

# **24th International Conference on Computational Linguistics**

## **(COLING 2012)**

**Posters**

**Mumbai, India  
8-15 December 2012**

**Volume 1 of 2**

**Editors:**

**Martin Kay  
Christian Boitet**

**ISBN: 978-1-62748-390-2**

**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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Table of Contents

<i>K-Best Spanning Tree Dependency Parsing With Verb Valency Lexicon Reranking</i> Zeljko Agic .....	1
<i>A Best-First Anagram Hashing Filter for Approximate String Matching with Generalized Edit Distance</i> Malin Ahlberg and Gerlof Bouma .....	13
<i>Automatic Bilingual Phrase Extraction from Comparable Corpora</i> Ahmet Aker, Yang Feng and Robert Gaizauskas .....	23
<i>A Formalized Reference Grammar for UNL-Based Machine Translation between English and Arabic</i> Sameh Alansary .....	33
<i>Mapping Arabic Wikipedia into the Named Entities Taxonomy</i> Fahd Alotaibi and Mark Lee .....	43
<i>Probabilistic Refinement Algorithms for the Generation of Referring Expressions</i> Romina Altamirano, Carlos Areces and Luciana Benotti .....	53
<i>Measuring the Adequacy of Cross-Lingual Paraphrases in a Machine Translation Setting</i> Marianna Apidianaki .....	63
<i>Cross-Lingual Sentiment Analysis for Indian Languages using Linked WordNets</i> Balamurali A.R., Aditya Joshi and Pushpak Bhattacharyya .....	73
<i>The Creation of Large-Scale Annotated Corpora of Minority Languages using UniParser and the EANC platform</i> Timofey Arkhangeskiy, Oleg Belyaev and Arseniy Vydrin .....	83
<i>Collocation Extraction using Parallel Corpus</i> Kavosh Asadi Atui, Hesham Faili and Kaveh Assadi Atui .....	93
<i>Improved Spelling Error Detection and Correction for Arabic</i> Mohammed Attia, Pavel Pecina, Younes Samih, Khaled Shaalan and Josef van Genabith .	103
<i>Heloise — A Reengineering of Ariane-G5 SLLPs for Application to <math>\pi</math>-languages</i> Vincent Berment and Christian Boitet .....	113
<i>Machine Translation for Language Preservation</i> Steven Bird and David Chiang .....	125
<i>Comparing Non-projective Strategies for Labeled Graph-Based Dependency Parsing</i> Anders Björkelund and Jonas Kuhn .....	135
<i>Phrase Structures and Dependencies for End-to-End Coreference Resolution</i> Anders Björkelund and Jonas Kuhn .....	145
<i>The Language of Power and its Cultural Influence</i> David Bracewell and Marc Tomlinson .....	155

<i>Learning Opinionated Patterns for Contextual Opinion Detection</i> Caroline Brun .....	165
<i>Does Similarity Matter? The Case of Answer Extraction from Technical Discussion Forums</i> Rose Catherine, Amit Singh, Rashmi Gangadharaiah, Dinesh Raghu and Karthik Visweswariah	175
<i>Chinese Noun Phrase Coreference Resolution: Insights into the State of the Art</i> Chen Chen and Vincent Ng .....	185
<i>Linguistic and Statistical Traits Characterising Plagiarism</i> Miranda Chong and Lucia Specia .....	195
<i>Impact of Less Skewed Distributions on Efficiency and Effectiveness of Biomedical Relation Extraction</i> Md. Faisal Mahbub Chowdhury and Alberto Lavelli .....	205
<i>Lattice Rescoring for Speech Recognition using Large Scale Distributed Language Models</i> Euisok Chung, Hyung-Bae Jeon, Jeon-Gue Park and Yun-Keun Lee .....	217
<i>Morphological Analyzer for Affix Stacking Languages: A Case Study of Marathi</i> Raj Dabre, Archana Amberkar and Pushpak Bhattacharyya .....	225
<i>Modelling the Organization and Processing of Bangla Polymorphemic Words in the Mental Lexicon: A Computational Approach</i> Tirthankar Dasgupta, Manjira Sinha and Anupam Basu .....	235
<i>Coreference Clustering using Column Generation</i> Jan De Belder and Marie-Francine Moens .....	245
<i>Metric Learning for Graph-Based Domain Adaptation</i> Paramveer Dhillon, Partha Talukdar and Koby Crammer .....	255
<i>Automatic Hashtag Recommendation for Microblogs using Topic-Specific Translation Model</i> Zhuoye Ding, Qi Zhang and Xuanjing Huang .....	265
<i>Unsupervised Feature-Rich Clustering</i> Vladimir Eidelman .....	275
<i>Token Level Identification of Linguistic Code Switching</i> Heba Elfardy and Mona Diab .....	287
<i>Parenthetical Classification for Information Extraction</i> Ismail El Maarouf and Jeanne Villaneau .....	297
<i>A Dictionary-Based Approach to Identifying Aspects Implied by Adjectives for Opinion Mining</i> Geli Fei, Bing Liu, Meichun Hsu, Malu Castellanos and Riddhiman Ghosh .....	309
<i>Dealing with Input Noise in Statistical Machine Translation</i> Lluís Formiga and José A. R. Fonollosa .....	319
<i>A Comparison of Knowledge-based Algorithms for Graded Word Sense Assignment</i> Annemarie Friedrich, Nikos Engonopoulos, Stefan Thater and Manfred Pinkal .....	329

<i>Leveraging Statistical Transliteration for Dictionary-Based English-Bengali CLIR of OCR'd Text</i> Utpal Garain, Arjun Das, David Doermann and Douglas Oard .....	339
<i>RU-EVAL-2012: Evaluating Dependency Parsers for Russian</i> Anastasia Gareyshina, Maxim Ionov, Olga Lyashevskaya, Dmitry Privoznov, Elena Sokolova and Svetlana Toldova .....	349
<i>Assessing Sentiment Strength in Words Prior Polarities</i> Lorenzo Gatti and Marco Guerini .....	361
<i>Improving Dependency Parsing with Interlinear Glossed Text and Syntactic Projection</i> Ryan Georgi, Fei Xia and William Lewis .....	371
<i>Diachronic Variation in Grammatical Relations</i> Aaron Gerow and Khurshid Ahmad .....	381
<i>Relation Classification using Entity Sequence Kernels</i> Debanjan Ghosh and Smaranda Muresan .....	391
<i>Translating Questions to SQL Queries with Generative Parsers Discriminatively Reranked</i> Alessandra Giordani and Alessandro Moschitti .....	401
<i>Classifier-Based Tense Model for SMT</i> Zhengxian Gong, Min Zhang, Chew-lim Tan and Guodong Zhou .....	411
<i>Extracting and Normalizing Entity-Actions from Users' Comments</i> Swapna Gottipati and Jing Jiang .....	421
<i>Expected Divergence Based Feature Selection for Learning to Rank</i> Parth Gupta and Paolo Rosso .....	431
<i>LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors</i> Aaron L. F. Han, Derek F. Wong and Lidia S. Chao .....	441
<i>Predicting Stance in Ideological Debate with Rich Linguistic Knowledge</i> Kazi Saidul Hasan and Vincent Ng .....	451
<i>FeatureForge: A Novel Tool for Visually Supported Feature Engineering and Corpus Revision</i> Florian Heimerl, Charles Jochim, Steffen Koch and Thomas Ertl .....	461
<i>Verb Temporality Analysis using Reichenbach's Tense System</i> André Horie, Kumiko Tanaka-Ishii and Mitsuru Ishizuka .....	471
<i>A Metric for Evaluating Discourse Coherence based on Coreference Resolution</i> Ryu Iida and Takenobu Tokunaga .....	483
<i>Comparing Word Relatedness Measures Based on Google n-grams</i> Aminul Islam, Evangelos Milios and Vlado Keselj .....	495
<i>Two-Stage Bootstrapping for Anaphora Resolution</i> Balaji Jagan, T V Geetha and Ranjani Parthasarathi .....	507
<i>Explorations in the Speakers' Interaction Experience and Self-Assessments</i> Kristiina Jokinen .....	517

<i>Multimodal Signals and Holistic Interaction Structuring</i> Kristiina Jokinen and Graham Wilcock.....	527
<i>New Insights from Coarse Word Sense Disambiguation in the Crowd</i> Adam Kapelner, Krishna Kaliannan, H. Andrew Schwartz, Lyle Ungar and Dean Foster	539
<i>A Unified Sentence Space for Categorical Distributional-Compositional Semantics: Theory and Experiments</i> Dimitri Kartsaklis, Mehrnoosh Sadrzadeh and Stephen Pulman.....	549
<i>A Knowledge-Based Approach to Syntactic Disambiguation of Biomedical Noun Compounds</i> Ramakanth Kavuluru and Daniel Harris.....	559
<i>Classification of Inconsistent Sentiment Words using Syntactic Constructions</i> Wiltrud Kessler and Hinrich Schütze.....	569
<i>Learning Semantics with Deep Belief Network for Cross-Language Information Retrieval</i> Jungi Kim, Jinseok Nam and Iryna Gurevych.....	579
<i>Detection of Acoustic-Phonetic Landmarks in Mismatched Conditions using a Biomimetic Model of Human Auditory Processing</i> Sarah King and Mark Hasegawa-Johnson.....	589
<i>Learning Verbs on the Fly</i> Zornitsa Kozareva.....	599
<i>Decoder-based Discriminative Training of Phrase Segmentation for Statistical Machine Translation</i> Hyoung-Gyu Lee and Hae-Chang Rim.....	611
<i>Glimpses of Ancient China from Classical Chinese Poems</i> John Lee and Tak-sum Wong.....	621
<i>Conversion between Scripts of Punjabi: Beyond Simple Transliteration</i> Gurpreet Singh Lehal and Tejinder Singh Saini.....	633
<i>Development of a Complete Urdu-Hindi Transliteration System</i> Gurpreet Singh Lehal and Tejinder Singh Saini.....	643
<i>Random Walks on Context-Aware Relation Graphs for Ranking Social Tags</i> Han Li, Zhiyuan Liu and Maosong Sun.....	653
<i>Phrase-Based Evaluation for Machine Translation</i> Liangyou Li, Zhengxian Gong and Guodong Zhou.....	663
<i>A Beam Search Algorithm for ITG Word Alignment</i> Peng Li, Yang Liu and Maosong Sun.....	673
<i>Active Learning for Chinese Word Segmentation</i> Shoushan Li, Guodong Zhou and Chu-Ren Huang.....	683
<i>Fine-Grained Classification of Named Entities by Fusing Multi-Features</i> Wenjie Li, Jiwei Li, Ye Tian and Zhifang Sui.....	693
<i>Expert Finding for Microblog Misinformation Identification</i> Chen Liang, Zhiyuan Liu and Maosong Sun.....	703

<i>Improving Relative-Entropy Pruning using Statistical Significance</i> Wang Ling, Nadi Tomeh, Guang Xiang, Isabel Trancoso and Alan Black .....	713
<i>Expected Error Minimization with Ultraconservative Update for SMT</i> Lemao Liu, Tiejun Zhao, Taro Watanabe, Hailong Cao and Conghui Zhu .....	723
<i>Generalized Sentiment-Bearing Expression Features for Sentiment Analysis</i> Shizhu Liu, Gady Agam and David Grossman .....	733
<i>Unsupervised Domain Adaptation for Joint Segmentation and POS-Tagging</i> Yang Liu and Yue Zhang .....	745
<i>Tag Dispatch Model with Social Network Regularization for Microblog User Tag Suggestion</i> Zhiyuan Liu, Cunchao Tu and Maosong Sun .....	755
<i>Summarization of Business-Related Tweets: A Concept-Based Approach</i> Annie Louis and Todd Newman .....	765
<i>Towards the Automatic Detection of the Source Language of a Literary Translation.</i> Gerard Lynch and Carl Vogel .....	775
<i>Fourth-Order Dependency Parsing</i> Xuezhe Ma and Hai Zhao .....	785
<i>A Subjective Logic Framework for Multi-Document Summarization</i> Sukanya Manna, Byron J. Gao and Reed Coke .....	797
<i>Manual Corpus Annotation: Giving Meaning to the Evaluation Metrics</i> Yann Mathet, Antoine Widlöcher, Karën Fort, Claire François, Olivier Galibert, Cyril Grouin, Juliette Kahn, Sophie Rosset and Pierre Zweigenbaum .....	809
<i>Discriminative Boosting from Dictionary and Raw Text – A Novel Approach to Build A Chinese Word Segmenter</i> Fandong Meng, Wenbin Jiang, Hao Xiong and Qun Liu .....	819
<i>Lost in Translations? Building Sentiment Lexicons using Context Based Machine Translation</i> Xinfan Meng, Furu Wei, Ge Xu, Longkai Zhang, Xiaohua Liu, Ming Zhou and Houfeng Wang .....	829
<i>How Does the Granularity of an Annotation Scheme Influence Dependency Parsing Performance?</i> Simon Mille, Alicia Burga, Gabriela Ferraro and Leo Wanner .....	839
<i>Does Tectogramatics Help the Annotation of Discourse?</i> Jiří Mírovský, Pavlína Jínová and Lucie Poláková .....	853
<i>The Effect of Learner Corpus Size in Grammatical Error Correction of ESL Writings</i> Tomoya Mizumoto, Yuta Hayashibe, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto	863
<i>GRAFIX: Automated Rule-Based Post Editing System to Improve English-Persian SMT Output</i> Mahsa Mohaghegh, Abdolhossein Sarrafzadeh and Mehdi Mohammadi .....	873
<i>Relational Structures and Models for Coreference Resolution</i> Truc-Vien T. Nguyen and Massimo Poesio .....	883

<i>Text Summarization Model based on Redundancy-Constrained Knapsack Problem</i> Hitoshi Nishikawa, Tsutomu Hirao, Toshiro Makino and Yoshihiro Matsuo .....	893
<i>Lexical Categories for Improved Parsing of Web Data</i> Lilja Øvrelid and Arne Skjærholt .....	903
<i>Text-To-Speech for Languages without an Orthography</i> Sukhada Palkar, Alan Black and Alok Parlikar .....	913
<i>Part of Speech (POS) Tagger for Kokborok</i> Braja Gopal Patra, Khumbar Debbarma, Dipankar Das and Sivaji Bandyopadhyay ...	923
<i>Forced Derivations for Hierarchical Machine Translation</i> Stephan Peitz, Arne Mauser, Joern Wuebker and Hermann Ney .....	933
<i>On Panini and the Generative Capacity of Contextualized Replacement Systems</i> Gerald Penn and Paul Kiparsky .....	943
<i>Joint Segmentation and Tagging with Coupled Sequences Labeling</i> Xipeng Qiu, Feng Ji, Jiayi Zhao and Xuanjing Huang .....	951
<i>Defining Syntax for Learner Language Annotation</i> Marwa Ragheb and Markus Dickinson .....	965
<i>How Good are Typological Distances for Determining Genealogical Relationships among Languages?</i> Taraka Rama and Prasanth Kolachina .....	975
<i>Sentence Boundary Detection: A Long Solved Problem?</i> Jonathon Read, Rebecca Dridan, Stephan Oepen and Lars Jørgen Solberg .....	985
<i>Document and Corpus Level Inference For Unsupervised and Transductive Learning of Information Structure of Scientific Documents</i> Roi Reichart and Anna Korhonen .....	995
<i>Light Textual Inference for Semantic Parsing</i> Kyle Richardson and Jonas Kuhn .....	1007
<i>Korektor -- A System for Contextual Spell-Checking and Diacritics Completion</i> Michal Richter, Pavel Straňák and Alexandr Rosen .....	1019
<i>Using Qualia Information to Identify Lexical Semantic Classes in an Unsupervised Clustering Task</i> Lauren Romeo, Sara Mendes and Núria Bel .....	1029
<i>A Strategy of Mapping Polish WordNet onto Princeton WordNet</i> Ewa Rudnicka, Marek Maziarz, Maciej Piasecki and Stan Szpakowicz .....	1039
<i>A Hierarchical Domain Model-Based Multi-Domain Selection Framework for Multi-Domain Dialog Systems</i> Seonghan Ryu, Donghyeon Lee, Injae Lee, Sangdo Han, Gary Geunbae Lee, Myungjae Kim and Kyungduk Kim .....	1049
<i>A Fully Coreference-annotated Corpus of Scholarly Papers from the ACL Anthology</i> Ulrich Schäfer, Christian Spurk and Jörg Steffen .....	1059



<i>Continuous Space Translation Models for Phrase-Based Statistical Machine Translation</i> Holger Schwenk .....	1071
<i>Data-driven Dependency Parsing With Empty Heads</i> Wolfgang Seeker, Richárd Farkas, Bernd Bohnet, Helmut Schmid and Jonas Kuhn .	1081
<i>Extension of TSVM to Multi-Class and Hierarchical Text Classification Problems With General Losses</i> Sathiya Keerthi Selvaraj, Sundararajan Sellamanickam and Shirish Shevade .....	1091
<i>Calculation of Phrase Probabilities for Statistical Machine Translation by using Belief Functions</i> Christophe Servan and Simon Petitrenaud .....	1101
<i>Sense and Reference Disambiguation in Wikipedia</i> Hui Shen, Razvan Bunescu and Rada Mihalcea .....	1111
<i>Unsupervised Metaphor Paraphrasing using a Vector Space Model</i> Ekaterina Shutova, Tim van de Cruys and Anna Korhonen .....	1121
<i>Memory-Efficient Katakana Compound Segmentation using Conditional Random Fields</i> Krauchanka Siarhei and Artsimemya Artsiom .....	1131
<i>New Readability Measures for Bangla and Hindi Texts</i> Manjira Sinha, Sakshi Sharma, Tirthankar Dasgupta and Anupam Basu .....	1141
<i>Automatic Question Generation in Multimedia-Based Learning</i> Yvonne Skalban, Le An Ha, Lucia Specia and Ruslan Mitkov .....	1151
<i>A More Cohesive Summarizer</i> Christian Smith, Henrik Danielsson and Arne Jönsson .....	1161
<i>Robust Learning in Random Subspaces: Equipping NLP for OOV Effects</i> Anders Søgaard and Anders Johannsen .....	1171
<i>An Empirical Etudy of Non-Lexical Extensions to Delexicalized Transfer</i> Anders Søgaard and Julie Wulff .....	1181
<i>Entropy-based Training Data Selection for Domain Adaptation</i> Yan Song, Prescott Klassen, Fei Xia and Chunyu Kit .....	1191
<i>Corpus-based Explorations of Affective Load Differences in Arabic-Hebrew-English</i> Carlo Strapparava, Oliviero Stock and Ilai Alon .....	1201
<i>Acquiring and Generalizing Causal Inference Rules from Deverbal Noun Constructions</i> Shohei Tanaka, Naoaki Okazaki and Mitsuru Ishizuka .....	1209
<i>Advertising Legality Recognition</i> Yi-jie Tang, Cong-kai Lin and Hsin-Hsi Chen .....	1219
<i>A Joint Phrasal and Dependency Model for Paraphrase Alignment</i> Kapil Thadani, Scott Martin and Michael White .....	1229
<i>Sourcing the Crowd for a Few Good Ones: Event Type Detection</i> Caselli Tommaso and Huang Chu-Ren .....	1239

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