

# **International Conference on Recent Advances in Natural Language Processing 2011**

**(RANLP 2011)**

**Hissar, Bulgaria  
12-14 September 2011**

**Editors:**

**Galia Angelova  
Ruslan Mitkov**

**Kalina Bontcheva  
Nikolai Nikolov**

**ISBN: 978-1-62276-511-9**

**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© (2011) 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>Extracting STRIPS Representations of Actions and Events</i> Avirup Sil and Alexander Yates .....	1
<i>Acquiring Topic Features to improve Event Extraction: in Pre-selected and Balanced Collections</i> Shasha Liao and Ralph Grishman .....	9
<i>Minimally Supervised Rule Learning for the Extraction of Biographic Information from Various Social Domains</i> Hong Li, Feiyu Xu and Hans Uszkoreit .....	17
<i>Extracting Relations Within and Across Sentences</i> Kumutha Swampillai and Mark Stevenson .....	25
<i>Knowledge-Poor Approach to Shallow Parsing: Contribution of Unsupervised Part-of-Speech Induction</i> Marie Guégan and Claude de Loupy .....	33
<i>Fast Domain Adaptation for Part of Speech Tagging for Dialogues</i> Sandra Kübler and Eric Baucom .....	41
<i>Using a Morphological Database to Increase the Accuracy in POS Tagging</i> Hrafn Loftsson, Sigrún Helgadóttir and Eiríkur Rögnvaldsson .....	49
<i>Actions Speak Louder than Words: Evaluating Parsers in the Context of Natural Language Understanding Systems for Human-Robot Interaction</i> Sandra Kübler, Rachael Cantrell and Matthias Scheutz .....	56
<i>Constructing Linguistically Motivated Structures from Statistical Grammars</i> Ali Basirat and Heshaam Faily .....	63
<i>An Open Source Punjabi Resource Grammar</i> Shafqat Mumtaz Virk, Muhammad Humayoun and Aarne Ranta .....	70
<i>Multi-Document Summarization by Capturing the Information Users are Interested in</i> Elena Lloret, Laura Plaza and Ahmet Aker .....	77
<i>Efficient algorithm for Context Sensitive Aggregation in Natural Language generation</i> Hemanth Sagar Bayyarapu .....	84
<i>Enriching a statistical machine translation system trained on small parallel corpora with rule-based bilingual phrases</i> V́ctor M. Sánchez-Cartagena, Felipe Sánchez-Martínez and Juan Antonio Pérez-Ortiz .....	90
<i>Assessing the Post-Editing Effort for Automatic and Semi-Automatic Translations of DVD Subtitles</i> Sheila C. M. de Sousa, Wilker Aziz and Lucia Specia .....	97
<i>JRC-NAMES: A Freely Available, Highly Multilingual Named Entity Resource</i> Ralf Steinberger, Bruno Pouliquen, Mijail Kabadjov, Jenya Belyaeva and Erik van der Goot .....	104
<i>MDL-based Models for Alignment of Etymological Data</i> Hannes Wettig, Suvi Hiltunen and Roman Yangarber .....	111
<i>Building a Multilingual Named Entity-Annotated Corpus Using Annotation Projection</i> Maud Ehrmann, Marco Turchi and Ralf Steinberger .....	118

<i>Bilingual lexicon extraction from comparable corpora for closely related languages</i> Darja Fišer and Nikola Ljubešić .....	125
<i>Sentiments and Opinions in Health-related Web messages</i> Marina Sokolova and Victoria Bobicev .....	132
<i>An Exploration into the Use of Contextual Document Clustering for Cluster Sentiment Analysis</i> Niall Rooney, Hui Wang, Fiona Browne, Fergal Monaghan, Jann Müller, Alan Sergeant, Zhiwei Lin, Philip Taylor and Vladimir Dobrynin .....	140
<i>Pause and Stop Labeling for Chinese Sentence Boundary Detection</i> Hen-Hsen Huang and Hsin-Hsi Chen .....	146
<i>Multilabel Tagging of Discourse Relations in Ambiguous Temporal Connectives</i> Yannick Versley .....	154
<i>Noun Compound and Named Entity Recognition and their Usability in Keyphrase Extraction</i> István Nagy T., Gábor Berend and Veronika Vincze .....	162
<i>A Named Entity Recognition Method using Rules Acquired from Unlabeled Data</i> Tomoya Iwakura .....	170
<i>An Incremental Entity-Mention Model for Coreference Resolution with Restrictive Antecedent Accessibility</i> Manfred Klenner and Don Tugener .....	178
<i>Cross-Domain Dutch Coreference Resolution</i> Orphée De Clercq, Véronique Hoste and Iris Hendrickx .....	186
<i>Finding the Best Approach for Multi-lingual Text Summarisation: A Comparative Analysis</i> Elena Lloret and Manuel Palomar .....	194
<i>Automatically Creating General-Purpose Opinion Summaries from Text</i> Veselin Stoyanov and Claire Cardie .....	202
<i>Exploring the Usefulness of Cross-lingual Information Fusion for Refining Real-time News Event Extraction: A Preliminary Study</i> Jakub Piskorski, Jenya Belayeva and Martin Atkinson .....	210
<i>Temporal Relation Extraction Using Expectation Maximization</i> Seyed Abolghasem Mirroshandel and Gholamreza Ghassem-Sani .....	218
<i>Improving Chunk-based Semantic Role Labeling with Lexical Features</i> Wilker Aziz, Miguel Rios and Lucia Specia .....	226
<i>Improving WSD using ISR-WN with Relevant Semantic Trees and SemCor Senses Frequency</i> Yoan Gutiérrez, Sonia Vázquez and Andrés Montoyo .....	233
<i>Investigating Advanced Techniques for Document Content Similarity Applied to External Plagiarism Analysis</i> Daniel Micol, Rafael Muñoz and Óscar Ferrández .....	240
<i>Using Cognates in a French-Romanian Lexical Alignment System: A Comparative Study</i> Mirabela Navlea and Amalia Todiraşcu .....	247

<i>Highly Multilingual Coreference Resolution Exploiting a Mature Entity Repository</i>	
Josef Steinberger, Jenya Belyaeva, Jonathan Crawley, Leonida Della-Rocca, Mohamed Ebrahim, Maud Ehrmann, Mijail Kabadjov, Ralf Steinberger and Erik Van-der-Goot . . . . .	254
<i>Singletons and Coreference Resolution Evaluation</i>	
Sandra Kübler and Desislava Zhekova . . . . .	261
<i>Modelling Entity Instantiations</i>	
Andrew McKinlay and Katja Markert . . . . .	268
<i>A New Scheme for Annotating Semantic Relations between Named Entities in Corpora</i>	
Mani Ezzat and Thierry Poibeau . . . . .	275
<i>Prototypical Opinion Holders: What We can Learn from Experts and Analysts</i>	
Michael Wiegand and Dietrich Klakow . . . . .	282
<i>Multiword Expressions and Named Entities in the Wiki50 Corpus</i>	
Veronika Vincze, István Nagy T. and Gábor Berend . . . . .	289
<i>Towards the Automatic Merging of Lexical Resources: Automatic Mapping</i>	
Muntsa Padró, Núria Bel and Silvia Neculescu . . . . .	296
<i>Unsupervised Learning for Persian WordNet Construction</i>	
Mortaza Montazery and Hesham Faili . . . . .	302
<i>Domain Independent Authorship Attribution without Domain Adaptation</i>	
Rohith Menon and Yejin Choi . . . . .	309
<i>Cultural Configuration of Wikipedia: measuring Autoreferentiality in Different Languages</i>	
Marc Miquel Ribé and Horacio Rodríguez . . . . .	316
<i>Combining Relational and Attributional Similarity for Semantic Relation Classification</i>	
Preslav Nakov and Zornitsa Kozareva . . . . .	323
<i>In Search of Missing Arguments: A Linguistic Approach</i>	
Josef Ruppenhofer, Philip Gorinski and Caroline Sporleder . . . . .	331
<i>Enlarging Monolingual Dictionaries for Machine Translation with Active Learning and Non-Expert Users</i>	
Miquel Esplà-Gomis, Víctor M. Sánchez-Cartagena and Juan Antonio Pérez-Ortiz . . . . .	339
<i>Morphological Analysis of Biomedical Terminology with Analogy-Based Alignment</i>	
Vincent Claveau and Ewa Kijak . . . . .	347
<i>Adaptability of Lexical Acquisition for Large-scale Grammars</i>	
Kostadin Cholakov, Gertjan van Noord, Valia Kordoni and Yi Zhang . . . . .	355
<i>Integration of Data from a Syntactic Lexicon into Generative and Discriminative Probabilistic Parsers</i>	
Anthony Sigogne, Matthieu Constant and Éric Laporte . . . . .	363
<i>Pattern Learning for Event Extraction using Monolingual Statistical Machine Translation</i>	
Marco Turchi, Vanni Zavarella and Hristo Tanev . . . . .	371
<i>META-DARE: Monitoring the Minimally Supervised ML of Relation Extraction Rules</i>	
Hong Li, Feiyu Xu and Hans Uszkoreit . . . . .	378

<i>Mining Transliterations from Wikipedia using Dynamic Bayesian Networks</i>	
Peter Nabende .....	385
<i>Detecting Opinions Using Deep Syntactic Analysis</i>	
Caroline Brun .....	392
<i>Using Visual Information to Predict Lexical Preference</i>	
Shane Bergsma and Randy Goebel .....	399
<i>Systematic Knowledge Acquisition for Question Analysis</i>	
Dat Quoc Nguyen, Dai Quoc Nguyen and Son Bao Pham .....	406
<i>A Semi-Automatic, Iterative Method for Creating a Domain-Specific Treebank</i>	
Corina Dima and Erhard Hinrichs .....	413
<i>Determining Immediate Constituents of Compounds in GermaNet</i>	
Verena Henrich and Erhard Hinrichs .....	420
<i>Segmentation and Clustering of Textual Sequences: a Typological Approach</i>	
Christelle Cocco, Raphaël Pittier, François Bavaud and Aris Xanthos .....	427
<i>A Contextual Classification Strategy for Polarity Analysis of Direct Quotations from Financial News</i>	
Brett Drury, Gaël Dias and Luís Torgo .....	434
<i>On the Contribution of MWE-based Features to a Readability Formula for French as a Foreign Language</i>	
Thomas François and Patrick Watrin .....	441
<i>Exploiting Hidden Morphophonemic Constraints for Finding the Underlying Forms of 'weak' Arabic Verbs</i>	
Allan Ramsay and Hanady Mansour .....	448
<i>A Confidence Model for Syntactically-Motivated Entailment Proofs</i>	
Asher Stern and Ido Dagan .....	455
<i>Learning Script Participants from Unlabeled Data</i>	
Michaela Regneri, Alexander Koller, Josef Ruppenhofer and Manfred Pinkal .....	463
<i>Towards Minimal Recursion Semantics over Bulgarian Dependency Parsing</i>	
Kiril Simov and Petya Osenova .....	471
<i>Discourse Structures to Reduce Discourse Incoherence in Blog Summarization</i>	
Shamima Mithun and Leila Kosseim .....	479
<i>Parallel Suffix Arrays for Linguistic Pattern Search</i>	
Johannes Goller .....	487
<i>A Mechanism to Restrict the Scope of Clause-Bounded Quantifiers in 'Continuation' Semantics</i>	
Anca Dinu .....	495
<i>A Support Tool for Deriving Domain Taxonomies from Wikipedia</i>	
Lili Kotlerman, Zemer Avital, Ido Dagan, Amnon Lotan and Ofer Weintraub .....	503
<i>Barrier Features for Classification of Semantic Relations</i>	
Anita Alicante and Anna Corazza .....	509

<i>A Reflective View on Text Similarity</i>	
Daniel Bär, Torsten Zesch and Iryna Gurevych .....	515
<i>Evaluating the Robustness of EmotiBlog for Sentiment Analysis and Opinion Mining</i>	
Ester Boldrini, Javi Fernández, José Manuel Gómez and Patricio Martínez-Barco .....	521
<i>Hybrid System For Plagiarism Detection</i>	
Javier R. Bru, Patricio Martínez-Barco and Rafael Muñoz .....	527
<i>Data-Driven Approach Using Semantics for Recognizing and Classifying TimeML Events in Italian</i>	
Tommaso Caselli, Hector Llorens, Borja Navarro-Colorado and Estela Saquete .....	533
<i>Can Alternations Be Learned? A Machine Learning Approach To Romanian Verb Conjugation</i>	
Liviu P. Dinu, Emil Ionescu, Vlad Niculae and Octavia-Maria Şulea .....	539
<i>A New Representation Model for the Automatic Recognition and Translation of Arabic Named Entities with NooJ</i>	
Héla Fehri, Kais Haddar and Abdelmajid Ben Hamadou .....	545
<i>Training Data in Statistical Machine Translation - the More, the Better?</i>	
Monica Gavrila and Cristina Vertan .....	551
<i>Towards a Corpus-based Approach to Modelling Language Production of Foreign Language Learners in Communicative Contexts</i>	
Voula Gotsoulia and Bessie Dendrinou .....	557
<i>Parsing a Polysynthetic Language</i>	
Petr Homola .....	562
<i>An algorithm of Identifying Semantic Arguments of a Verb From Structured Data</i>	
Minhua Huang and Robert M. Haralick .....	568
<i>Construction of an HPSG Grammar for the Arabic Relative Sentences</i>	
Ines Zalila and Kais Haddar .....	574
<i>Automatically Selected Skip Edges in Conditional Random Fields for Named Entity Recognition</i>	
Roman Klinger .....	580
<i>Negation Naive Bayes for Categorization of Product Pages on the Web</i>	
Kanako Komiya, Naoto Sato, Koji Fujimoto and Yoshiyuki Kotani .....	586
<i>A Hybrid Approach for Event Extraction and Event Actor Identification</i>	
Anup Kumar Kolya, Asif Ekbal and Sivaji Bandyopadhyay .....	592
<i>Evaluating Human Correction Quality for Machine Translation from Crowdsourcing</i>	
Shasha Liao, Cheng Wu and Juan Huerta .....	598
<i>Multi-class SVM for Relation Extraction from Clinical Reports</i>	
Anne-Lyse Minard, Anne-Laure Ligozat and Brigitte Grau .....	604
<i>Discovering coreference using image-grounded verb models</i>	
Amitabha Mukerjee, Kruti Neema and Sushobhan Nayak .....	610
<i>Word and Phrase Learning based on Prior Semantics</i>	
Amitabha Mukerjee and Nikhil Joshi .....	616

<i>Domain-Dependent Identification of Multiword Expressions</i>	
István Nagy T., Veronika Vincze and Gábor Berend . . . . .	622
<i>Robust Semantic Analysis for Unseen Data in FrameNet</i>	
Alexis Palmer, Afra Alishahi and Caroline Sporleder . . . . .	628
<i>Studying Translationese at the Character Level</i>	
Marius Popescu . . . . .	634
<i>Linear Transduction Grammars and Zipper Finite-State Transducers</i>	
Markus Saers and Dekai Wu . . . . .	640
<i>Finding Negative Key Phrases for Internet Advertising Campaigns using Wikipedia</i>	
Martin Scaiano and Diana Inkpen . . . . .	648
<i>Establishing Implementation Priorities in Aiding Writers of Controlled Crisis Management Texts</i>	
Irina Temnikova . . . . .	654
<i>TechWatchTool: Innovation and Trend Monitoring</i>	
Hong Li, Feiyu Xu and Hans Uszkoreit . . . . .	660
<i>”Yes we can?”: Subjectivity Annotation and Tagging for the Health Domain</i>	
Muhammad Abdul-Mageed, Mohammed Korayem and Ahmed YoussefAgha . . . . .	666
<i>Wordnets: State of the Art and Perspectives. Case Study: the Romanian Wordnet</i>	
Verginica Barbu Mititelu . . . . .	672
<i>Creation and Development of the Romanian Lexical Resources</i>	
Elena Boian, Constantin Ciubotaru, Svetlana Cojocaru, Alexandru Colesnicov, Ludmila Malahov and Mircea Petic . . . . .	678
<i>Analyses Tools for Non-head Structures</i>	
Sirine Boukedi and Kais Haddar . . . . .	686
<i>Visualization for Coreference Annotation</i>	
Andre Burkovski and Gunther Heidemann . . . . .	692
<i>The RST Spanish Treebank On-line Interface</i>	
Iria da Cunha, Juan-Manuel Torres-Moreno, Gerardo Sierra, Luis Adrián Cabrera-Diego, Brenda Gabriela Castro Rolón and Juan Miguel Rolland Bartilotti . . . . .	698
<i>Lexical Generalisation for Word-level Matching in Plagiarism Detection</i>	
Miranda Chong and Lucia Specia . . . . .	704
<i>Multiple Evidence for Term Extraction in Broad Domains</i>	
Boris Dobrov and Natalia Loukachevitch . . . . .	710
<i>Language Modeling for Document Selection in Question Answering</i>	
Nicolas Foucault, Gilles Adda and Sophie Rosset . . . . .	716
<i>Evaluating Various Linguistic Features on Semantic Relation Extraction</i>	
Marcos Garcia and Pablo Gamallo . . . . .	721
<i>Automatic titling of Articles Using Position and Statistical Information</i>	
Cédric Lopez, Violaine Prince and Mathieu Roche . . . . .	727



<i>Unsupervised Domain Adaptation based on Text Relatedness</i>	
Georgios Petasis .....	733
<i>Bilingual Experiments with an Arabic-English Corpus for Opinion Mining</i>	
Mohammed Rushdi-Saleh, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López and José M. Perea-Ortega .....	740
<i>Experiments on Term Extraction using Noun Phrase Subclassifications</i>	
Merley da Silva Conrado, Walter Koza, Josuka Díaz-Labrador, Joseba Abaitua, Solange Oliveira Rezende, Thiago Pardo and Zulema Solana .....	746
<i>Adaptive Feedback Message Generation for Second Language Learners of Arabic</i>	
Khaled Shaalan and Marwa Magdy .....	752
<i>Building a Patient-based Ontology for User-written Web Messages</i>	
Marina Sokolova and David Schramm .....	758
<i>Recognition and Classification of Numerical Entities in Basque</i>	
Ander Soraluze, Iñaki Alegria, Olatz Ansa, Olatz Arregi and Xabier Arregi .....	764
<i>Multilingual Entity-Centered Sentiment Analysis Evaluated by Parallel Corpora</i>	
Josef Steinberger, Polina Lenkova, Mijail Kabadjoy, Ralf Steinberger and Erik van der Goot ..	770
<i>Term Validation for Vocabulary Construction and Key Term Extraction</i>	
Alexander Ulanov and Andrey Simanovsky .....	776
<i>Agreement: How to Reach it? Defining Language Features Leading to Agreement in Discourse</i>	
Tatiana Zidraşco, Victoria Bobicev, Shun Shiramatsu, Tadachika Ozono and Toramatsu Shintani	781