

International Conference on Recent Advances in Natural Language Processing

(RANLP 2013)

**Hissar, Bulgaria
9 – 11 September 2013**

Editors:

**G. Angelova
K. Bontcheva
R. Mitkov**

ISBN: 978-1-62993-555-3

Table of Contents

<i>ASMA: A System for Automatic Segmentation and Morpho-Syntactic Disambiguation of Modern Standard Arabic</i> Muhammad Abdul-Mageed, Mona Diab and Sandra Kübler	1
<i>Optimising Tree Edit Distance with Subtrees for Textual Entailment</i> Maytham Alabbas and Allan Ramsay	9
<i>Opinion Learning from Medical Forums</i> Tanveer Ali, Marina Sokolova, David Schramm and Diana Inkpen	18
<i>Annotating events, Time and Place Expressions in Arabic Texts</i> Hassina Aliane, Wassila Guendouzi and Amina Mokrani	25
<i>A Semi-supervised Learning Approach to Arabic Named Entity Recognition</i> Maha Althobaiti, Udo Kruschwitz and Massimo Poesio	32
<i>An NLP-based Reading Tool for Aiding Non-native English Readers</i> Mahmoud Azab, Ahmed Salama, Kemal Oflazer, Hideki Shima, Jun Araki and Teruko Mitamura	41
<i>Improving Sentiment Analysis in Twitter Using Multilingual Machine Translated Data</i> Alexandra Balahur and Marco Turchi	49
<i>Domain Adaptation for Parsing</i> Eric Baucom, Levi King and Sandra Kübler	56
<i>Towards a Structured Representation of Generic Concepts and Relations in Large Text Corpora</i> Archana Bhattarai and Vasile Rus	65
<i>Authorship Attribution in Health Forums</i> Victoria Bobicev, Marina Sokolova, Khaled El Emam and Stan Matwin	74
<i>TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text</i> Kalina Bontcheva, Leon Derczynski, Adam Funk, Mark Greenwood, Diana Maynard and Niraj Aswani	83
<i>A unified lexical processing framework based on the Margin Infused Relaxed Algorithm. A case study on the Romanian Language</i> Tiberiu Boros	91
<i>Automatic extraction of contextual valence shifters.</i> Noémi Boubel, Thomas François and Hubert Naets	98
<i>Grammar-Based Lexicon Extension for Aligning German Radiology Text and Images</i> Claudia Bretschneider, Sonja Zillner and Matthias Hammon	105
<i>Recognising and Interpreting Named Temporal Expressions</i> Matteo Brucato, Leon Derczynski, Hector Llorens, Kalina Bontcheva and Christian S. Jensen .	113
<i>Unsupervised Improving of Sentiment Analysis Using Global Target Context</i> Tomáš Brychcín and Ivan Habernal	122

<i>An Agglomerative Hierarchical Clustering Algorithm for Labelling Morphs</i> Burcu Can and Suresh Manandhar	129
<i>Temporal Text Classification for Romanian Novels set in the Past</i> Alina Maria Ciobanu, Liviu P. Dinu, Octavia-Maria Şulea, anca dinu and Vlad Niculae	136
<i>A Dictionary-Based Approach for Evaluating Orthographic Methods in Cognates Identification</i> Alina Maria Ciobanu and Liviu Petrisor Dinu	141
<i>A Pilot Study on the Semantic Classification of Two German Prepositions: Combining Monolingual and Multilingual Evidence</i> Simon Clematide and Manfred Klenner	148
<i>Semantic Relations between Events and their Time, Locations and Participants for Event Coreference Resolution</i> Agata Cybulska and Piek Vossen	156
<i>Sense Clustering Using Wikipedia</i> Bharath Dandala, Chris Hokamp, Rada Mihalcea and Razvan Bunescu	164
<i>Effective Spell Checking Methods Using Clustering Algorithms</i> Renato Cordeiro de Amorim and Marcos Zampieri	172
<i>Normalization of Dutch User-Generated Content</i> Orphee De Clercq, Sarah Schulz, Bart Desmet, Els Lefever and Veronique Hoste	179
<i>Linguistic Profiling of Texts Across Textual Genres and Readability Levels. An Exploratory Study on Italian Fictional Prose</i> Felice Dell'Orletta, Simonetta Montemagni and Giulia Venturi	189
<i>Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data</i> Leon Derczynski, Alan Ritter, Sam Clark and Kalina Bontcheva	198
<i>Weighted maximum likelihood loss as a convenient shortcut to optimizing the F-measure of maximum entropy classifiers</i> Georgi Dimitroff, Laura Tolosi, Borislav Popov and Georgi Georgiev	207
<i>Sequence Tagging for Verb Conjugation in Romanian</i> Liviu Dinu, Octavia-Maria Şulea and Vlad Niculae	215
<i>A Tagging Approach to Identify Complex Constituents for Text Simplification</i> Justin Dornescu, Richard Evans and Constantin Orasan	221
<i>Automatic Evaluation Metric for Machine Translation that is Independent of Sentence Length</i> Hiroshi Echizen'ya, Kenji Araki and Eduard Hovy	230
<i>Acronym recognition and processing in 22 languages</i> Maud Ehrmann, Leo della Rocca, Ralf Steinberger and Hristo Tanev	237
<i>An Evaluation Summary Method Based on a Combination of Content and Linguistic Metrics</i> Samira Ellouze, Maher Jaoua and Lamia Hadrich Belguith	245
<i>Hierarchy Identification for Automatically Generating Table-of-Contents</i> Nicolai Erbs, Iryna Gurevych and Torsten Zesch	252

<i>Temporal Relation Classification in Persian and English contexts</i> Mahbaneh Eshaghzadeh Torbati, Gholamreza Ghassem-sani, Seyed Abolghasem Mirroshandel, Yadollah Yaghoobzadeh and Negin Karimi Hosseini	261
<i>The Extended Lexicon: Language Processing as Lexical Description</i> Roger Evans	270
<i>Did I really mean that? Applying automatic summarisation techniques to formative feedback</i> Debora Field, Stephen Pulman, Nicolas Van Labeke, Denise Whitelock and John Richardson .	277
<i>Matching sets of parse trees for answering multi-sentence questions</i> Boris Galitsky, Dmitry Ilvovsky, Sergei O. Kuznetsov and Fedor Strok	285
<i>Realization of common statistical methods in computational linguistics with functional automata</i> Stefan Gerdjikov, Petar Mitankin and Vladislav Nenchev	294
<i>Mining Fine-grained Opinion Expressions with Shallow Parsing</i> Sucheta Ghosh, Sara Tonelli and Richard Johansson	302
<i>Justifying Corpus-Based Choices in Referring Expression Generation</i> Helmut Horacek	311
<i>A Boosting-based Algorithm for Classification of Semi-Structured Text using the Frequency of Substructures</i> Tomoya Iwakura	319
<i>Headerless, Quoteless, but not Hopeless? Using Pairwise Email Classification to Disentangle Email Threads</i> Emily Jamison and Iryna Gurevych	327
<i>Using Parallel Corpora for Word Sense Disambiguation</i> Dimitar Kazakov and Ahmad R. Shahid	336
<i>Recognizing semantic relations within Polish noun phrase: A rule-based approach</i> Paweł Kędzia and Marek Maziarz	342
<i>Unsupervised Induction of Arabic Root and Pattern Lexicons using Machine Learning</i> Bilal Khaliq and John Carroll	350
<i>Towards Domain Adaptation for Parsing Web Data</i> Mohammad Khan, Markus Dickinson and Sandra Kübler	357
<i>Capturing Anomalies in the Choice of Content Words in Compositional Distributional Semantic Space</i> Ekaterina Kochmar and Ted Briscoe	365
<i>Incremental and Predictive Dependency Parsing under Real-Time Conditions</i> Arne Köhn and Wolfgang Menzel	373
<i>Rationale, Concepts, and Current Outcome of the Unit Graphs Framework</i> Maxime Lefrançois and Fabien Gandon	382
<i>The Unit Graphs Framework: Foundational Concepts and Semantic Consequence</i> Maxime Lefrançois and Fabien Gandon	389
<i>Confidence Estimation for Knowledge Base Population</i> Xiang Li and Ralph Grishman	396

<i>Towards Fine-grained Citation Function Classification</i> Xiang Li, Yifan He, Adam Meyers and Ralph Grishman	402
<i>Supervised Morphology Generation Using Parallel Corpus</i> Alireza Mahmoudi, Mohsen Arabsorkhi and Hesham Faili	408
<i>Sentiment Analysis of Reviews: Should we analyze writer intentions or reader perceptions?</i> Isa Maks and Piek Vossen	415
<i>Revisiting the Old Kitchen Sink: Do we Need Sentiment Domain Adaptation?</i> Riham Mansour, Nesma Refaei, Michael Gamon, Ahmed Abdul-Hamid and Khaled Sami	420
<i>Evaluation of baseline information retrieval for Polish open-domain Question Answering system</i> Michał Marcińczuk, Adam Radziszewski, Maciej Piasecki, Dominik Piasecki and Marcin Ptak	428
<i>WCCL Relation — a Toolset for Rule-based Recognition of Semantic Relations Between Named Entities</i> Michał Marcińczuk	436
<i>Beyond the Transfer-and-Merge Wordnet Construction: plWordNet and a Comparison with WordNet</i> Marek Maziarz, Maciej Piasecki, Ewa Rudnicka and Stan Szpakowicz	443
<i>History Based Unsupervised Data Oriented Parsing</i> Mohsen Mesgar and Gholamreza Ghasem-Sani	453
<i>Contrasting and Corroborating Citations in Journal Articles</i> Adam Meyers	460
<i>CCG Categories for Distributional Semantic Models</i> Paramita Mirza and Raffaella Bernardi	467
<i>Discourse-aware Statistical Machine Translation as a Context-sensitive Spell Checker</i> Behzad Mirzababaei, Hesham Faili and Nava Ehsan	475
<i>Cross-Lingual Information Retrieval and Semantic Interoperability for Cultural Heritage Repositories</i> Johanna Monti, Mario Monteleone, Maria Pia di Buono and Federica Marano	483
<i>Improving Web 2.0 Opinion Mining Systems Using Text Normalisation Techniques</i> Alejandro Mosquera and Paloma Moreda Pozo	491
<i>Identifying Social and Expressive Factors in Request Texts Using Transaction/Sequence Model</i> Dasa Munkova, Michal Munk and Zuzana Fraterova	496
<i>Parameter Optimization for Statistical Machine Translation: It Pays to Learn from Hard Examples</i> Preslav Nakov, Fahad Al Obaidli, Francisco Guzman and Stephan Vogel	504
<i>Automatic Cloze-Questions Generation</i> Annamaneni Narendra, Manish Agarwal and Rakshit shah	511
<i>High-Accuracy Phrase Translation Acquisition Through Battle-Royale Selection</i> Lionel Nicolas, Egon W. Stemle, Klara Kranebitter and Verena Lyding	516
<i>Enriching Patent Search with External Keywords: a Feasibility Study</i> Ivelina Nikolova, Irina Temnikova and Galia Angelova	525

<i>A clustering approach for translationese identification</i> Sergiu Nisioi and Liviu P. Dinu	532
<i>PurePos 2.0: a hybrid tool for morphological disambiguation</i> György Orosz and Attila Novak	539
<i>More than Bag-of-Words: Sentence-based Document Representation for Sentiment Analysis</i> Georgios Paltoglou and Mike Thelwall	546
<i>Information Spreading in Expanding Wordnet Hypernymy Structure</i> Maciej Piasecki, Radosław Ramocki and Michał Kaliński	553
<i>Context Independent Term Mapper for European Languages</i> Mārcis Pinnis	562
<i>Semi-supervised vs. Cross-domain Graphs for Sentiment Analysis</i> Natalia Ponomareva and Mike Thelwall	571
<i>Towards a Hybrid Rule-based and Statistical Arabic-French Machine Translation System</i> fatihā sadat	579
<i>Segmenting vs. Chunking Rules: Unsupervised ITG Induction via Minimum Conditional Description Length</i> Markus Saers, Karteek Addanki and Dekai Wu	584
<i>A Combined Pattern-based and Distributional Approach for Automatic Hypernym Detection in Dutch.</i> Gwendolijn Schropp, Els Lefever and Veronique Hoste	593
<i>Exploiting Synergies Between Open Resources for German Dependency Parsing, POS-tagging, and Morphological Analysis</i> Rico Sennrich, Martin Volk and Gerold Schneider	601
<i>Using a Weighted Semantic Network for Lexical Semantic Relatedness</i> Reda Sibli and Leila Kosseim	610
<i>A New Approach to the POS Tagging Problem Using Evolutionary Computation</i> Ana Paula Silva, Arlindo Silva and Irene Rodrigues	619
<i>How Joe and Jane Tweet about Their Health: Mining for Personal Health Information on Twitter</i> Marina Sokolova, Stan Matwin, Yasser Jafer and David Schramm	626
<i>What Sentiments Can Be Found in Medical Forums?</i> Marina Sokolova and Victoria Bobicev	633
<i>Automated learning of everyday patients' language for medical blogs analytics</i> Giovanni Stilo, Moreno De Vincenzi, Alberto E. Tozzi and Paola Velardi	640
<i>How Symbolic Learning Can Help Statistical Learning (and vice versa)</i> Isabelle Tellier and Yoann Dupont	649
<i>Measuring Closure Properties of Patent Sublanguages</i> Irina Temnikova, Negacy Hailu, Galia Angelova and K. Bretonnel Cohen	659
<i>Closure Properties of Bulgarian Clinical Text</i> Irina Temnikova, Ivelina Nikolova, William A. Baumgartner, Galia Angelova and K. Bretonnel Cohen	667

<i>Analyzing the Use of Character-Level Translation with Sparse and Noisy Datasets</i> Jörg Tiedemann and Preslav Nakov	676
<i>A Feature Induction Algorithm with Application to Named Entity Disambiguation</i> Laura Tolosi, Valentin Zhikov, Georgi Georgiev and Borislav Popov	685
<i>Introducing a Corpus of Human-Authored Dialogue Summaries in Portuguese</i> Norton Trevisan Roman, Paul Piwek, Ariadne M. B. Rizzoni Carvalho and Alexandre Rossi Alvares	692
<i>Wikipedia as an SMT Training Corpus</i> Dan Tufiş, Radu Ion, Stefan Dumitrescu and Dan Stefanescu	702
<i>DutchSemCor: in quest of the ideal sense-tagged corpus</i> Piek Vossen, Ruben Izquierdo and Attila Görög	710
<i>Towards detecting anomalies in the content of standardized LMF dictionaries</i> Wafa WALI, Bilel Gargouri and Abdelmajid BEN HAMADOU	719
<i>Edit Distance: A New Data Selection Criterion for Domain Adaptation in SMT</i> Longyue Wang, Derek F. Wong, Lidia S. Chao, Junwen Xing, Yi Lu and Isabel Trancoso	727
<i>Automatic Enhancement of LTAG Treebank</i> Farzaneh Zarei, Ali Basirat, Hesham Faili and Maryam S.Mirian	733
<i>Inductive and deductive inferences in a Crowdsourced Lexical-Semantic Network</i> Manel Zarrouk, Mathieu Lafourcade and Alain Joubert	740
<i>Machine Learning for Mention Head Detection in Multilingual Coreference Resolution</i> Desislava Zhekova and Sandra Kübler	747
<i>Combining POS Tagging, Dependency Parsing and Coreferential Resolution for Bulgarian</i> Valentin Zhikov, Georgi Georgiev, Kiril Simov and Petya Osenova	755
<i>magyarlanc: A Tool for Morphological and Dependency Parsing of Hungarian</i> János Zsibrita, Veronika Vincze and Richárd Farkas	763