

20th Nordic Conference of Computational Linguistics (NODALIDA 2015)

NEALT Proceedings Series Volume 23

**Vilnius, Lithuania
11 – 13 May 2015**

Editor:

B. Megyesi

ISBN: 978-1-5108-0299-5

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© (2015) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2015)

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

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
Web: www.proceedings.com

Table of Contents

Invited Keynotes

<i>Kevin Knight</i> How Much Information Does a Human Translator Add to the Original?	xi
<i>Catherine Pelachaud</i> Modeling Socio-Emotional Humanoid Agent	xiii
<i>Sebastian Riedel</i> Embedding Probabilistic Logic for Machine Reading	xv

Regular Papers

<i>Yvonne Adesam, Gerlof Bouma and Richard Johansson</i> Defining the Eukalyptus forest – the Koala treebank of Swedish	1
<i>Malin Ahlberg, Peter Andersson, Markus Forsberg and Nina Tahmasebi</i> A case study on supervised classification of Swedish pseudo-coordination	11
<i>Héctor Martínez Alonso, Anders Johannsen, Sussi Olsen, Sanni Nimb, Nicolai Hartvig Sørensen, Anna Braasch, Anders Søgaard and Bolette Sandford Pedersen</i> Supersense tagging for Danish	21
<i>Eckhard Bick and Tino Didriksen</i> CG-3 — Beyond Classical Constraint Grammar	31
<i>Loïc Boizou, Jolanta Kovalevskaitė and Erika Rimkutė</i> Automatic Lemmatisation of Lithuanian MWEs	41
<i>Alexandr Chernov, Volha Petukhova and Dietrich Klakow</i> Linguistically Motivated Question Classification	51
<i>Jana Götze and Johan Boye</i> Resolving Spatial References using Crowdsourced Geographical Data	61
<i>Richard Johansson and Luis Nieto Piña</i> Combining Relational and Distributional Knowledge for Word Sense Disambiguation	69
<i>Peter Juel Henriksen</i> Talebob - an Interactive Speech Trainer for Danish	79
<i>Jurgita Kapočiūtė-Dzikiėnė, Ligita Šarkutė and Andrius Utkas</i> The Effect of Author Set Size in Authorship Attribution for Lithuanian	87
<i>Sigrid Klerke, Héctor Martínez Alonso and Anders Søgaard</i> Looking hard: Eye tracking for detecting grammaticality of automatically compressed sentences	97
<i>Veronika Laippala, Jenna Kanerva, Anna Missilä, Sampo Pyysalo, Tapio Salakoski and Filip Ginter</i> Towards the Classification of the Finnish Internet Parsebank: Detecting Translations and Informality	107
<i>Jostein Lien, Erik Velldal and Lilja Øvrelid</i> Improving cross-domain dependency parsing with dependency-derived clusters	117
<i>Mihai Lintean and Vasile Rus</i> An Optimal Quadratic Approach to Monolingual Paraphrase Alignment	127
<i>Juhani Luotolahti and Filip Ginter</i> Sentence Compression For Automatic Subtitling	135

<i>Michael Nokel and Natalia Loukachevitch</i> Topic Models: Accounting Component Structure of Bigrams	145
<i>Eva Pettersson and Joakim Nivre</i> Improving Verb Phrase Extraction from Historical Text by use of Verb Valency Frames	153
<i>Sampo Pyysalo, Jenna Kanerva, Anna Missilä, Veronika Laippala and Filip Ginter</i> Universal Dependencies for Finnish	163
<i>Robert Reynolds and Francis Tyers</i> Automatic word stress annotation of Russian unrestricted text	173
<i>Dominik Sacha, Yuki Asano, Christian Rohrdantz, Felix Hamborg, Daniel Keim, Bettina Braun and Miriam Butt</i> Self Organizing Maps for the Visual Analysis of Pitch Contours	181
<i>Jörg Tiedemann</i> Improving the Cross-Lingual Projection of Syntactic Dependencies	191
<i>Julián Zapata and Andreas Søeborg Kirkedal</i> Assessing the Performance of Automatic Speech Recognition Systems When Used by Native and Non-Native Speakers of Three Major Languages in Dictation Workflows	201
Student Papers	
<i>Max Berggren, Jussi Karlgren, Robert Östling and Mikael Parkvall</i> Inferring the location of authors from words in their texts	211
<i>Inari Listenmaa and Francis M. Tyers</i> Automatic conversion of colloquial Finnish to standard Finnish	219
<i>Vytautas Mickevičius, Tomas Krilavičius, Vaidas Morkevičius and Aušra Mackutė-Varoneckienė</i> Automatic Thematic Classification of the Titles of the Seimas Votes	225
<i>Birgitta Ojamaa, Päivi Kristiina Jokinen and Kadri Muischenk</i> Sentiment analysis on conversational texts	233
<i>Scharolta Katharina Sienčnik</i> Adapting word2vec to Named Entity Recognition	239
Short Papers	
<i>Héctor Martínez Alonso, Anders Johannsen, Barbara Plank and Anders Søgaard</i> Active learning for sense annotation	245
<i>Johan Bos and Malvina Nissim</i> Uncovering Noun-Noun Compound Relations by Gamification	251
<i>Katarina Heimann Mühlenbock, Sofie Johansson Kokkinakis, Caroline Liberg, Åsa af Geijerstam, Jenny Wiksten Folkeryd, Arne Jönsson, Erik Kanebrant and Johan Falkenjack</i> A multivariate model for classifying texts' readability	257
<i>Robert Östling, Carl Börstell and Lars Wallin</i> Enriching the Swedish Sign Language Corpus with Part of Speech Tags Using Joint Bayesian Word Alignment and Annotation Transfer	263
<i>Gustavo Henrique Paetzold</i> Using Positional Suffix Trees to Perform Agile Tree Kernel Calculation	269
<i>Ildikó Pilán</i> Helping Swedish words come to their senses: word-sense disambiguation based on sense associations from the SALDO lexicon	275

<i>Askars Salimbajevs and Jevgenijs Strigins</i> Using sub-word n-gram models for dealing with OOV in large vocabulary speech recognition for Latvian	281
<i>Steinþór Steingrímsson, Sigrún Helgadóttir and Eiríkur Rögnvaldsson</i> Analysing Inconsistencies and Errors in PoS Tagging in two Icelandic Gold Standards	287
<i>Demonstration Papers</i>	
<i>Magnus Birkenes, Lars Johnsen, Arne Lindstad and Johanne Ostad</i> From digital library to n-grams: NB N-gram	293
<i>Janne Bondi Johannessen</i> The Corpus of American Norwegian Speech (CANS)	297
<i>Johan Bos</i> Open-Domain Semantic Parsing with Boxer	301
<i>Sam Hardwick, Miikka Silfverberg and Krister Lindén</i> Extracting Semantic Frames using hfst-pmatch	305
<i>Arne Neumann</i> discoursegraphs: A graph-based merging tool and converter for multilayer annotated corpora	309
<i>Tommi A Pirinen</i> Omorfi-Free and Open Source Morphological Lexical Database of Finnish	313
<i>Evelina Rennes and Arne Jönsson</i> A Tool for Automatic Simplification of Swedish Texts	317