

# **17th Nordic Conference of Computational Linguistics (NODALIDA 2009)**

NEALT Proceedings Series Volume 4

Odense, Denmark  
14 – 16 May 2009

**Editors:**

**Kristiina Jokinen  
Eckhard Bick**

ISBN: 978-1-5108-3465-1

**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© (2009) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Contents

<b>Contents</b>	<b>iii</b>
<b>Preface</b>	<b>vii</b>
<b>Committees</b>	<b>ix</b>
<b>Conference Program</b>	<b>xi</b>
<b>I Invited Papers</b>	<b>1</b>
<b>JEAN CARLETTA</b> Developing Meeting Support Technologies: From Data to Demonstration (and Beyond)	2
<b>RALF STEINBERGER</b> Linking News Content Across Languages	4
<b>II Tutorial</b>	<b>6</b>
<b>GRAHAM WILCOCK</b> Text Annotation with OpenNLP and UIMA	7
<b>III Regular papers</b>	<b>9</b>
<b>LENE ANTONSEN, SAARA HUHMARNIEMI AND TROND TROSTERUD</b> Interactive pedagogical programs based on constraint grammar	10
<b>JARI BJÖRNE, FILIP GINTER, JUHO HEIMONEN, SAMPO PYYSALO AND TAPIO SALAKOSKI</b> Learning to Extract Biological Event and Relation Graphs	18
<b>HERCULES DALIANIS, MARTIN RIMKA AND VIGGO KANN</b> Using Uplug and SiteSeeker to construct a cross language search engine for Scandinavian languages	26
<b>EVA FORSBOM</b> Extending the View: Explorations in Bootstrapping a Swedish PoS Tagger	34
<b>TATIANA GORNOSTAY AND INGUNA SKADIŅA</b> Pattern-based English-Latvian Toponym Translation	41
<b>NATHAN GREEN, PAUL BREIMYER, VINAY KUMAR AND NAGIZA F SAMATOVA</b> WebBANC: Building Semantically-Rich Annotated Corpora from Web User Annotations of Minority Languages	48
<b>CHRISTIAN HARDMEIER AND MARTIN VOLK</b> Using Linguistic Annotations in Statistical Machine Translation of Film Subtitles	57

<b>KATRI HAVERINEN, FILIP GINTER, VERONIKA LAIPPALA AND TAPIO SALAKOSKI</b> <b>Parsing Clinical Finnish: Experiments with Rule-Based and Statistical Dependency Parsers</b>	<b>65</b>
<b>JANNE BONDI JOHANNESSEN, JOEL PRIESTLEY, KRISTIN HAGEN, TOR ANDERS ÅFARLI AND ØYSTEIN ALEXANDER VANGSNES</b> <b>The Nordic Dialect Corpus — an advanced research tool</b>	<b>73</b>
<b>PETER KOLB</b> <b>Experiments on the difference between semantic similarity and relatedness</b>	<b>81</b>
<b>KRISTER LINDÉN AND TOMMI PIRINEN</b> <b>Weighted Finite-State Morphological Analysis of Finnish Compounding with HFST-LEXC</b>	<b>89</b>
<b>KRISTER LINDÉN AND JUSSI TUOVILA</b> <b>Corpus-based Paradigm Selection for Morphological Entries</b>	<b>96</b>
<b>HRAFN LOFTSSON, IDA KRAMARCZYK, SIGRÚN HELGADÓTTIR AND EIRÍKUR RÖGNVALDSSON</b> <b>Improving the PoS tagging accuracy of Icelandic text</b>	<b>103</b>
<b>OLGA LASHEVSKAJA AND OLGA MITROFANOVA</b> <b>Disambiguation of Taxonomy Markers in Context: Russian Nouns</b>	<b>111</b>
<b>YVES LEPAGE AND CHOOI LING GOH</b> <b>Towards automatic acquisition of linguistic features</b>	<b>118</b>
<b>MIGUEL A. MOLINERO, BENOÎT SAGOT AND LIONEL NICOLAS</b> <b>Building a morphological and syntactic lexicon by merging various linguistic resources</b>	<b>126</b>
<b>KRISTINA NILSSON AND HANS HJELM</b> <b>Using Semantic Features Derived from Word-Space Models for Swedish Coreference Resolution</b>	<b>134</b>
<b>JACOB PERSSON, RICHARD JOHANSSON AND PIERRE NUGUES</b> <b>Text Categorization Using Predicate–Argument Structures</b>	<b>142</b>
<b>MAGNUS ROSELL</b> <b>Part of Speech Tagging for Text Clustering in Swedish</b>	<b>150</b>
<b>BOLETTE SANDFORD PEDERSEN AND ANNA BRAASCH</b> <b>What do we need to know about humans? A view into the DanNet database</b>	<b>158</b>
<b>NATALIE SCHLUTER AND JOSEF VAN GENABITH</b> <b>Dependency Parsing Resources for French: Converting Acquired Lexical Functional Grammar F-Structure Annotations and Parsing F-Structures Directly</b>	<b>166</b>
<b>MIIKKA SILFVERBERG AND KRISTER LINDÉN</b> <b>Conflict Resolution Using Weighted Rules in HFST-TWOLC</b>	<b>174</b>
<b>ANDERS SØGAARD</b> <b>A linear time extension of deterministic pushdown automata</b>	<b>182</b>
<b>ANDERS SØGAARD</b> <b>Verifying context-sensitive treebanks and heuristic parses in polynomial time</b>	<b>190</b>
<b>MICHAEL WIEGAND AND DIETRICH KLAKOW</b> <b>Predictive Features in Semi-Supervised Learning for Polarity Classification and the Role of Adjectives</b>	<b>198</b>

<b>ANSSI YLI-JYRÄ</b> An Efficient Double Complementation Algorithm for Superposition-Based Finite-State Morphology	206
<b>IV Regular short paper</b>	<b>214</b>
<b>ECKHARD BICK AND M. PILAR VALVERDE IBÁÑEZ</b> Automatic Semantic Role Annotation for Spanish	215
<b>MARK FISHEL AND JOAKIM NIVRE</b> Voting and Stacking in Data-Driven Dependency Parsing	219
<b>KARIN FRIBERG HEPPIN</b> MedEval Six Test Collections in One	223
<b>RASHMI GANGADHARAI AH, RALF D. BROWN AND JAIME CARBONELL</b> Active Learning in Example-Based Machine Translation	227
<b>ANTON K. INGASON, SKÚLI B. JÓHANSSON, EIRÍKUR RÖGNVALDSSON, HRAFN LOFTSSON AND SIGRÚN HELGADÓTTIR</b> Context-Sensitive Spelling Correction and Rich Morphology	231
<b>MANFRED KLENNER, ANGELA FAHRNI AND STEFANOS PETRAKIS</b> PolArt: A Robust Tool for Sentiment Analysis	235
<b>BEÁTA B. MEGYESI</b> The Open Source Tagger HunPoS for Swedish	239
<b>INGUNA SKADIŅA AND EDGARS BRĀLĪTIS</b> English-Latvian SMT: knowledge or data?	242
<b>LILJA ØVRELID</b> Cross-lingual porting of distributional semantic classification	246
<b>V Student papers</b>	<b>250</b>
<b>MARIA ESKEVICH</b> Prominence detected by listeners for future speech synthesis application	251
<b>OKKO RÄSÄNEN AND JORIS DRIESEN</b> A comparison and combination of segmental and fixed-frame signal representations in NMF-based word recognition	255
<b>BÁLINT SASS</b> Verb Argument Browser for Danish	263
<b>VI Demos</b>	<b>267</b>
<b>ECKHARD BICK</b> DeepDict — A Graphical Corpus-based Dictionary of Word Relations	268
<b>SANDRA DERBRING, PETER LJUNGLÖF AND MARIA OLSSON</b> SubTTS: Light-weight automatic reading of subtitles	272
<b>PETER LJUNGLÖF, STAFFAN LARSSON, KATARINA MÜHLENBOCK AND GUNILLA THUNBERG</b> TRIK: A Talking and Drawing Robot for Children with Communication Disabilities	275

<b>BODIL NISTRUP MADSEN AND HANNE ERDMAN THOMSEN</b> <b>CAOS — A tool for the Construction of Terminological Ontologies</b>	<b>279</b>
<b>ARNE MARTINUS LINDSTAD, ANDERS NØKLESTAD, JANNE BONDI JOHANNESSEN AND ØYSTEIN A. VANGSNES</b> <b>The Nordic Dialect Database: Mapping Microsyntactic Variation in the Scandinavian Languages</b>	<b>283</b>
<b>Author Index</b>	<b>287</b>