

15th Nordic Conference of Computational Linguistics (NODALIDA 2005)

Joensuu, Finland
20 – 21 May 2005

Editor:

Stefan Werner

ISBN: 978-1-5108-3518-4

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© (2005) 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

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
Web: www.proceedings.com

Contents

Preface	v
Marita Ailomaa, Vladimír Kadlec, Martin Rajman, Jean-Cédric Chappelier: <i>Robust stochastic parsing: Comparing and combining two approaches for processing extra-grammatical sentences</i>	1
Ergin Altintas, Elif Karşlıgil, Vedat Coskun: <i>A new semantic similarity measure evaluated in word sense disambiguation</i>	8
Lili Aunimo, Reeta Kuuskoski: <i>Reformulations of Finnish questions for question answering</i>	12
Chris Biemann, Uwe Quasthoff: <i>Dictionary acquisition using parallel text and co-occurrence statistics</i>	22
Heidi Brøseth: <i>The influence of written task descriptions in Wizard of Oz experiments</i>	30
Hercules Dalianis: <i>Improving search engine retrieval using a compound splitter for Swedish</i>	38
Kristin Hagen, Lars Nygaard: <i>Improbable morphological forms in a computational lexicon</i>	43
Johan Hall, Joakim Nivre: <i>A generic architecture for data-driven dependency parsing</i>	47
Lars Hellan, Dorothee Beermann: <i>The ‘specifier’ in an HPSG grammar implementation of Norwegian</i>	57
Hans Hjelm, Christoph Schwarz: <i>LiSa - morphological analysis for information retrieval</i>	65
Jon Holmlund, Magnus Sahlgren, Jussi Karlgren: <i>Creating bilingual lexica using reference wordlists for alignment of monolingual semantic vector spaces</i>	71

Johan Hovold: <i>Naive Bayes spam filtering using word-position-based attributes and length-sensitive classification thresholds</i>	78
Christer Johansson, Lars G. Johnsen: <i>An analytical relation between analogical modeling and memory based learning</i>	88
Tuomo Kakkonen: <i>Dependency treebanks: methods, annotation schemes and tools</i>	94
Viggo Kann, Magnus Rosell: <i>Free construction of a free Swedish dictionary of synonyms</i>	105
Jussi Karlgren: <i>Compound terms and their constituent elements in information retrieval</i>	111
Bodil Kyst, Peter Juel Henrichsen: <i>Synthetic regional Danish</i>	116
Hanne Moa: <i>Compounds and other oddities in machine translation</i>	124
Jyrki Niemi, Lauri Carlson: <i>Towards modeling the semantics of calendar expressions as extended regular expressions</i>	133
Kristina Nilsson, Aisha Malmgren: <i>Towards automatic recognition of product names: an exploratory study of brand names in economic texts</i>	146
Kristina Nilsson, Hans Hjelm, Henrik Oxhammar: <i>SUiS - cross-language ontology-driven information retrieval in a restricted domain</i>	139
Anders Nøklestad, Christer Johansson: <i>Detecting reference chains in Norwegian</i>	156
Henrik Oxhammar: <i>Mapping product descriptions to a large ontology</i>	163
Magnus Rosell, Sumithra Velupillai: <i>The impact of phrases in document clustering for Swedish</i>	173
Jonas Sjöbergh: <i>Chunking: an unsupervised method to find errors in text</i>	180
Peter Rossen Skadhauge, Peter Juel Henrichsen: <i>DanPO - a transcription-based dictionary for Danish speech technology</i>	186
Anders Søgaard, Petter Haugereid: <i>Functionality in grammar design</i>	193

Heli Uibo: <i>Optimizing the finite-state description of Estonian morphology</i>	203
Hans Friedrich Witschel, Chris Biemann: <i>Rigorous dimensionality reduction through linguistically motivated feature selection for text categorization</i>	210