

Contents

Preface	viii
Committees	x
Conference Program	xii
I Invited Papers	1
When FrameNet meets a Controlled Natural Language <i>Guntis Bārzdīņš</i>	2
Bare-Bones Dependency Parsing — A Case for Occam’s Razor? <i>Joakim Nivre</i>	6
Discourse Structures and Language Technologies <i>Bonnie Webber</i>	12
II Regular papers	17
Identification of sense selection in regular polysemy using shallow features <i>Hector Martinez Alonso, Núria Bel and Bolette Sandford Pedersen</i>	18
Decision Strategies for Incremental POS Tagging <i>Niels Beuck, Arne Köhn and Wolfgang Menzel</i>	26
A FrameNet for Danish <i>Eckhard Bick</i>	34
Extraction from relative and embedded interrogative clauses in Danish <i>Anne Bjerre</i>	42
The Formal Patterns of the Lithuanian Verb Forms <i>Loïc Boizou</i>	50
Semantic search in literature as an e-Humanities research tool: CONPLISIT — Consumption patterns and life-style in 19th century Swedish literature <i>Lars Borin, Markus Forsberg and Christer Ahlberger</i>	58

Evaluation of terminologies acquired from comparable corpora: an application perspective <i>Estelle Delpech</i>	66
A quantitative and qualitative analysis of Nordic surnames <i>Eirini Florou and Stasinou Konstantopoulos</i>	74
Experiments on Lithuanian Term Extraction <i>Gintarė Grigonytė, Erika Rimkutė, Andrius Utkas and Loïc Boizou</i>	82
Fishing in a speech stream, angling for a lexicon <i>Peter Juel Henriksen</i>	90
The Impact of Part-of-Speech Filtering on Generation of a Swedish-Japanese Dictionary Using English as Pivot Language <i>Ingemar Hjälmstad, Martin Hassel and Maria Skeppstedt</i>	98
A Gold Standard for English–Swedish Word Alignment <i>Maria Holmqvist and Lars Ahrenberg</i>	106
Relevance Prediction in Information Extraction using Discourse and Lexical Features <i>Silja Huttunen, Arto Vihavainen and Roman Yangarber</i>	114
What kind of corpus is a web corpus? <i>Janne Bondi Johannessen and Emiliano Raúl Guevara</i>	122
Morphological analysis of a non-standard language variety <i>Heiki-Jaan Kaalep and Kadri Muischnek</i>	130
Editing Syntax Trees on the Surface <i>Peter Ljunglöf</i>	138
Do wordnets also improve human performance on NLP tasks? <i>Kristiina Muhonen and Krister Lindén</i>	146
Creating Comparable Multimodal Corpora for Nordic Languages <i>Costanza Navarretta, Elisabeth Ahlsén, Jens Allwood, Kristiina Jokinen and Patrizia Paggio</i>	153
Estimating language relationships from a parallel corpus. A study of the Europarl corpus <i>Taraka Rama and Lars Borin</i>	161
Improving Sentence-level Subjectivity Classification through Readability Measurement <i>Robert Remus</i>	168
Iterative, MT-based Sentence Alignment of Parallel Texts <i>Rico Sennrich and Martin Volk</i>	175
Combining Statistical Models for POS Tagging using Finite-State Calculus <i>Miikka Silfverberg and Krister Lindén</i>	183

Toponym Disambiguation in English-Lithuanian SMT System with Spatial Knowledge <i>Raivis Skadiņš, Tatiana Gornostay and Valters Šics</i>	191
Automatic summarization as means of simplifying texts, an evaluation for Swedish <i>Christian Smith and Arne Jönsson</i>	198
Using graphical models for PP attachment <i>Anders Søgaard</i>	206
Corrective re-synthesis of deviant speech using unit selection <i>Sofia Strömbergsson</i>	214
Psycho-acoustically motivated formant feature extraction <i>Bea Valkenier, Dirkjan Krijnders, Ronald Van Elburg and Tjeerd Andringa</i>	218
Random Indexing Re-Hashed <i>Erik Velldal</i>	224
Evaluating the effect of word frequencies in a probabilistic generative model of morphology <i>Sami Virpioja, Oskar Kohonen and Krista Lagus</i>	230
Disambiguation of English Contractions for Machine Translation of TV Subtitles <i>Martin Volk and Rico Sennrich</i>	238
Probabilistic Models for Alignment of Etymological Data <i>Hannes Wettig and Roman Yangarber</i>	246
Convolution Kernels for Subjectivity Detection <i>Michael Wiegand and Dietrich Klakow</i>	254
Explorations on Positionwise Flag Diacritics in Finite-State Morphology <i>Anssi Yli-Jyrä</i>	262
III Regular short papers	270
Experiments to investigate the utility of nearest neighbour metrics based on linguistically informed features for detecting textual plagiarism <i>Per Almquist and Jussi Karlgren</i>	271
CFG based grammar checker for Latvian <i>Daiga Deksnė and Raivis Skadiņš</i>	275
Query Constraining Aspects of Knowledge <i>Ann-Marie Eklund</i>	279
A categorization scheme for analyzing rules from a handbook of Swedish writing rules <i>Jody Foo</i>	283

Something Old, Something New — Applying a Pre-trained Parsing Model to Clinical Swedish <i>Martin Hassel, Aron Henriksson and Sumithra Velupillai</i>	287
Knowledge-free Verb Detection through Sentence Sequence Alignment <i>Christian Hänig</i>	291
”Andre ord” — a wordnet browser for the Danish wordnet, DanNet (DEMO) <i>Anders Johannsen and Bolette Sandford Pedersen</i>	295
Modularisation of Finnish Finite-State Language Description — Towards Wide Collaboration in Open Source Development of a Morphological Analyser <i>Tommi Pirinen</i>	299
A Prague Markup Language profile for the SemTi-Kamols grammar model <i>Lauma Pretkalniņa, Gunta Nešpore, Kristīne Levāne-Petrova and Baiba Saulīte</i>	303
Dialect classification in the Himalayas: a computational approach <i>Anju Saxena and Lars Borin</i>	307
Extraction of Knowledge-Rich Contexts in Russian – A Study in the Automotive Domain <i>Anne-Kathrin Schumann</i>	311
Iterative reordering and word alignment for statistical MT <i>Sara Stymne</i>	315
A double-blind experiment on interannotator agreement: the case of dependency syntax and Finnish <i>Atro Voutilainen and Tanja Purtonen</i>	319
Automatic Question Generation from Swedish Documents as a Tool for Information Extraction <i>Kenneth Wilhelmsson</i>	323
IV Student papers	327
Linguistic Motivation in Automatic Sentence Alignment of Parallel Corpora: the Case of Danish-Bulgarian and English-Bulgarian <i>Angel Genov and Georgi Iliev</i>	328
Finding statistically motivated features influencing subtree alignment performance <i>Gideon Kotzé</i>	332
Evaluating the speech quality of the Norwegian synthetic voice Brage <i>Marius Olaussen</i>	336
A Statistical Part-of-Speech Tagger for Persian <i>Mojgan Seraji</i>	340

Identification of context markers for Russian nouns <i>Anastasia Shimorina and Maria Grachkova</i>	344
Author Index	348