

# Contents

<b>Contents</b>	<b>iii</b>
<b>Preface</b>	<b>ix</b>
<b>Committees</b>	<b>xi</b>
Program Committee . . . . .	xi
Local Organization Committee . . . . .	xi
Reviewers . . . . .	xii
<b>Conference program</b>	<b>xiii</b>
<b>I Invited talks</b>	<b>1</b>
DIANA F. MCCARTHY <b>Evaluating Automatic Approaches for Word Meaning Discovery and Disambiguation using Lexical Substitution</b>	<b>2</b>
WALTER DAELEMANS <b>Text Analysis and Machine Learning for Stylometrics and Stylogenetics</b>	<b>3</b>
<b>II Regular papers</b>	<b>4</b>
TANEL ALUMÄE <b>Automatic Compound Word Reconstruction for Speech Recognition of Compounding Languages</b>	<b>5</b>
GUNTIS BĀRZDIŅŠ, NORMUNDS GRŪZĪTIS, GUNTA NEŠPORE AND BAIBA SAULĪTE <b>Dependency-Based Hybrid Model of Syntactic Analysis for the Languages with a Rather Free Word Order</b>	<b>13</b>
ECKHARD BICK AND LARS NYGAARD <b>Using Danish as a CG Interlingua: A Wide-Coverage Norwegian-English Machine Translation System</b>	<b>21</b>

JANNE BONDI JOHANNESSEN, KRISTIN HAGEN, JOEL JAMES PRIESTLEY AND LARS NYGAARD	
<b>An Advanced Speech Corpus for Norwegian</b>	<b>29</b>
MARKUS BORG	
<b>Time Extraction from Real-time Generated Football Reports</b>	<b>37</b>
INGER EKMAN AND KALERVO JÄRVELIN	
<b>Spoken Document Retrieval in a Highly Inflectional Language</b>	<b>44</b>
EVA FORSBOM	
<b>Inducing Baseform Models from a Swedish Vocabulary Pool</b>	<b>51</b>
OLGA GERASSIMENKO, MARE KOIT, ANDRIELA RÄÄBIS AND KRISTA STRAND- SON	
<b>Achieving Goals in Collaboration: Analysis of Estonian Institutional Calls</b>	<b>59</b>
KÄRLIS GOBA AND ANDREJS VASIĻJEVS	
<b>Development of Text-To-Speech system for Latvian</b>	<b>67</b>
JONAS GRANFELDT AND PIERRE NUGUES	
<b>Evaluating Stages of Development in Second Language French: A Machine- Learning Approach</b>	<b>73</b>
KARIN HARBUSCH AND GERARD KEMPEN	
<b>Clausal Coordinate Ellipsis in German: The TIGER Treebank as a Source of Evidence</b>	<b>81</b>
MARTIN HASSEL AND JONAS SJÖBERGH	
<b>Widening the HolSum Search Scope</b>	<b>89</b>
HANS HJELM	
<b>Identifying Cross Language Term Equivalents Using Statistical Ma- chine Translation and Distributional Association Measures</b>	<b>97</b>
RICHARD JOHANSSON AND PIERRE NUGUES	
<b>Extended Constituent-to-Dependency Conversion for English</b>	<b>105</b>
TOOMAS KIRT AND ENE VAINIK	
<b>Comparison of the Methods of Self-Organizing Maps and Multidimen- sional Scaling in Analysis of Estonian Emotion Concepts</b>	<b>113</b>
HANJING LI, TIEJUN ZHAO, SHENG LI, JIYUAN ZHAO	
<b>The Extraction of Trajectories from Real Texts Based on Linear Clas- sification</b>	<b>121</b>

HRAFN LOFTSSON AND EIRÍKUR RÖGNVALDSSON <b>IceParser: An Incremental Finite-State Parser for Icelandic</b>	<b>128</b>
BEATA B. MEGYESI AND BENGT DAHLQVIST <b>The Swedish-Turkish Parallel Corpus and Tools for its Creation</b>	<b>136</b>
NICOLAS MORALES, JOHN H. L. HANSEN, DOROTEO T. TOLEDANO AND JAVIER GARRIDO <b>Multivariate Cepstral Feature Compensation on Band-limited Data for Robust Speech Recognition</b>	<b>144</b>
VICTORIA ROSÉN AND KOENRAAD DE SMEDT <b>Theoretically Motivated Treebank Coverage</b>	<b>152</b>
TUOMO SAARNI, JUSSI HAKOKARI, TAPIO SALAKOSKI, JOUNI ISOAHO AND OLLI AALTONEN <b>Utterance-Initial Duration of Finnish Non-Plosive Consonants</b>	<b>160</b>
INGUNA SKADIŅA, ANDREJS VASIĻJEVS, DAIGA DEKSNE, RAIVIS SKADIŅŠ AND LINDA GOLDBERGA <b>Comprehension Assistant for Languages of Baltic States</b>	<b>167</b>
RICHARD SOCHER, CHRIS BIEMANN AND RAINER OSSWALD <b>Combining Contexts in Lexicon Learning for Semantic Parsing</b>	<b>175</b>
ANDERS SØGAARD <b>Polynomial Charts For Totally Unordered Languages</b>	<b>183</b>
MARTIN VOLK AND FRIDA TIDSTRÖM <b>Comparing French PP-attachment to English, German and Swedish</b>	<b>191</b>
PONTUS WÄRNESTÅL, LARS DEGERSTEDT AND ARNE JÖNSSON <b>Interview and Delivery: Dialogue Strategies for Conversational Rec- ommender Systems</b>	<b>199</b>
<b>III Student papers</b>	<b>206</b>
BJÖRN ANDRIST AND MARTIN HASSEL <b>Linguistically Fuelled Text Similarity</b>	<b>207</b>
KONSTANTINOS CHARITAKIS <b>Using Parallel Corpora to Create a Greek-English Dictionary with Up- lug</b>	<b>212</b>

DAVE COCHRAN		
<b>Unmediated Data-Oriented Generation</b>		<b>216</b>
KARIN FRIBERG		
<b>Decomposing Swedish Compounds Using Memory-Based Learning</b>		<b>224</b>
MARIA HOLMQVIST		
<b>Memory-based Learning of Word Translation</b>		<b>231</b>
FREDRIK JØRGENSEN		
<b>Clause Boundary Detection in Transcribed Spoken Language</b>		<b>235</b>
FREDRIK JØRGENSEN		
<b>The Effects of Disfluency Detection in Parsing Spoken Language</b>		<b>240</b>
ANDERS NØKLESTAD AND ÅSHILD SØFTELAND		
<b>Tagging a Norwegian Speech Corpus</b>		<b>245</b>
ANTON RAGNI		
<b>Initial Experiments with Estonian Speech Recognition</b>		<b>249</b>
MARIANNE SANTAHOLMA		
<b>Grammar Sharing Techniques for Rule-based Multilingual NLP Systems</b>		<b>253</b>
MARIANNE STARLANDER		
<b>Using a Wizard of Oz as a Baseline to Determine which System Architecture is the Best for a Spoken Language Translation System</b>		<b>261</b>
MARGUS TREUMUTH		
<b>A Method for Recognizing Temporal Expressions in Estonian Natural Language Dialogue Systems</b>		<b>265</b>
<b>IV Posters</b>		<b>269</b>
LARS AHRENBERG		
<b>LinES: An English-Swedish Parallel Treebank</b>		<b>270</b>
DANIEL BOLANOS AND WAYNE H. WARD		
<b>Posterior Probability Based Confidence Measures Applied to a Children's Speech Reading Tracking System</b>		<b>274</b>
MARK FISHEL, HEIKI-JAAN KAALEP AND KADRI MUISCHNEK		
<b>Estonian-English Statistical Machine Translation: the First Results</b>		<b>278</b>

JOHAN HALL, JOAKIM NIVRE AND JENS NILSSON <b>A Hybrid Constituency-Dependency Parser for Swedish</b>	<b>284</b>
ERLA HALLSTEINSDÓTTIR, THOMAS ECKART, CHRIS BIEMANN, UWE QUASTHOFF AND MATTHIAS RICHTER <b>Íslenskur Orðasjóður — Building a Large Icelandic Corpus</b>	<b>288</b>
HARALD HAMMARSTRÖM <b>A Survey and Classification of Methods for (Mostly) Unsupervised Learning</b>	<b>292</b>
OLE HARTVIGSEN, ERIK HARBORG, TORE AMBLE AND MAGNE H. JOHNSEN <b>Marvina — A Norwegian Speech-Centric, Multimodal Visitors' Guide</b>	<b>297</b>
PETER JUEL HENRICHSEN <b>A Norwegian Letter-to-Sound Engine with Danish as a Catalyst</b>	<b>305</b>
SIMON KEIZER AND ROSER MORANTE <b>Dialogue Simulation and Context Dynamics for Dialogue Management</b>	<b>310</b>
KIMMO KETTUNEN <b>Managing Keyword Variation with Frequency Based Generation of Word Forms in IR</b>	<b>318</b>
WANWISA KHANARAKSOMBAT AND JONAS SJÖBERGH <b>Developing and Evaluating a Searchable Swedish-Thai Lexicon</b>	<b>324</b>
DIMITRIOS KOKKINAKIS AND ANDERS THURIN <b>Identification of Entity References in Hospital Discharge Letters</b>	<b>329</b>
DIMITRIOS KOKKINAKIS, MARIA TOPOROWSKA GRONOSTAJ, CATALINA HALLETT AND DAVID HARDCASTLE <b>Lexical Parameters, Based on Corpus Analysis of English and Swedish Cancer Data, of Relevance for NLG</b>	<b>333</b>
MIKKO LOUNELA <b>Anatomy of an XML-based Text Corpus Server</b>	<b>337</b>
LYA MEISTER <b>Perceptual Assessment of the Degree of Russian Accent</b>	<b>345</b>
MAGNUS MERKEL AND JODY FOO <b>Terminology Extraction and Term Ranking for Standardizing Term Banks</b>	<b>349</b>

JYRKI NIEMI AND KIMMO KOSKENNIEMI <b>Representing Calendar Expressions with Finite-State Transducers that Bracket Periods of Time on a Hierarchical Timeline</b>	<b>355</b>
HELEN NIGOL <b>Parsing Manually Detected and Normalized Disfluencies in Spoken Es- tonian</b>	<b>363</b>
LIISI PIITS, MEELIS MIHKLA, TÕNIS NURK AND INDREK KIISSEL <b>Designing a Speech Corpus for Estonian Unit Selection Synthesis</b>	<b>367</b>
INES REHBEIN AND JOSEF VAN GENABITH <b>Evaluating Evaluation Measures</b>	<b>372</b>
JÜRGEN RIEDLER AND SERGIOS KATSIKAS <b>Development of a Modern Greek Broadcast-News Corpus and Speech Recognition System</b>	<b>380</b>
JANNE SAVELA, STINA OJALA, OLLI AALTONEN AND TAPIO SALAKOSKI <b>Role of Different Spectral Attributes in Vowel Categorization: the Case of Udmurt</b>	<b>384</b>
JONAS SJÖBERGH AND KENJI ARAKI <b>Recreating Humorous Split Compound Errors in Swedish by Using Grammaticality</b>	<b>389</b>
HÅKAN SUNDBLAD <b>A Re-examination of Question Classification</b>	<b>394</b>
TARMO TRUU, HALDUR ÕIM AND MARE KOIT <b>Interpretation of Yes/No Questions as Metaphor Recognition</b>	<b>398</b>
CENNY WENNER <b>Rule-based Logical Forms Extraction</b>	<b>402</b>
<b>Author Index</b>	<b>410</b>