

16th Nordic Conference of Computational Linguistics (NODALIDA 2007)

Tartu, Estonia
25 – 26 May 2007

Editors:

**Joakim Nivre
Heiki-Jaan Kaalep**

**Kadri Muischnek
Mare Koit**

ISBN: 978-1-5108-3519-1

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© (2007) 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

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

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

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

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

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

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