16th Nordic Conference of Computational Linguistics (NODALIDA 2007)

Tartu, Estonia
25 – 26 May 2007

Editors:

Joakim Nivre  Kadri Muischnek
Heiki-Jaan Kaalep  Mare Koit

ISBN: 978-1-5108-3519-1
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 Meaning Discovery and Disambiguation using Lexical Substitution .................................................. 2
   WALTER DAELMANS Text Analysis and Machine Learning for Stylometrics and Stylogenetics ........................................... 3

II Regular papers ................................. 4
   TANEL ALUMÄE
      Automatic Compound Word Reconstruction for Speech Recognition of Compounding Languages ........................................... 5
   GUNTIS BĀRZDINŠ, NORMUNDS GRŪZĪTIS, GUNTA NEŠPORE AND BAIBA SAULĪTE
      Dependency-Based Hybrid Model of Syntactic Analysis for the Languages 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 JOHANNESSSEN, 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 STRANDSON
Achieving Goals in Collaboration: Analysis of Estonian Institutional Calls 59

KÄRLIS GOBA AND ANDREJIS VASILJEVS
Development of Text-To-Speech system for Latvian 67

JONAS GRANFELDT AND PIERRE NUGUES

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 Machine 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 Multidimensional 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 Classification 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 RAÍNER 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 Recom-
mender 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 Cochrane  
Unmediated Data-Oriented Generation  

Karin Friberg  
Decomposing Swedish Compounds Using Memory-Based Learning  

Maria Holmqvist  
Memory-based Learning of Word Translation  

Fredrik Jørgensen  
Clause Boundary Detection in Transcribed Spoken Language  

Fredrik Jørgensen  
The Effects of Disfluency Detection in Parsing Spoken Language  

Anders Nøklestad and Åshild Søfteland  
Tagging a Norwegian Speech Corpus  

Anton Ragni  
Initial Experiments with Estonian Speech Recognition  

Marianne Santaholma  
Grammar Sharing Techniques for Rule-based Multilingual NLP Systems  

Marianne Starlander  
Using a Wizard of Oz as a Baseline to Determine which System Architecture is the Best for a Spoken Language Translation System  

Margus Treumuth  
A Method for Recognizing Temporal Expressions in Estonian Natural Language Dialogue Systems  

IV Posters  

Lars Ahrenberg  
LinES: An English-Swedish Parallel Treebank  

Daniel Bolanos and Wayne H. Ward  
Posterior Probability Based Confidence Measures Applied to a Children’s Speech Reading Tracking System  

Mark Fishel, Heiki-Jaan Kaalep and Kadri Muischnek  
Estonian-English Statistical Machine Translation: the First Results
JOHAN HALL, JOAKIM NIVRE AND JENS NILSSON
A Hybrid Constituency-Dependency Parser for Swedish  

ERLÁ HALLSTEINSDÓTTIR, THOMAS ECKART, CHRIS BIEMANN, UWE QUASTHOFF AND MATTHIAS RICHTER
Íslenskur Orðasjóður — Building a Large Icelandic Corpus  

HARALD HAMMARSTRÖM
A Survey and Classification of Methods for (Mostly) Unsupervised Learning  

OLE HARTVIGSEN, ERIK HARBORG, TORE AMBLE AND MAGNE H. JOHNSEN
Marvina — A Norwegian Speech-Centric, Multimodal Visitors’ Guide  

PETER JUEL HENRICHSEN
A Norwegian Letter-to-Sound Engine with Danish as a Catalyst  

SIMON KEIZER AND ROSER MORANTE
Dialogue Simulation and Context Dynamics for Dialogue Management  

KIMMO KETTUNEN
Managing Keyword Variation with Frequency Based Generation of Word Forms in IR  

WANWISA KHANARAKSOMBAT AND JONAS SJÖBERGH
Developing and Evaluating a Searchable Swedish-Thai Lexicon  

DIMITRIOS KOKKINAKIS AND ANDERS THURIN
Identification of Entity References in Hospital Discharge Letters  

DIMITRIOS KOKKINAKIS, MARIA TOPOROWSKA GRONOSTAJ, CATALINA HALLETT AND DAVID HARDCASTLE
Lexical Parameters, Based on Corpus Analysis of English and Swedish Cancer Data, of Relevance for NLG  

MIKKO LOUNELA
Anatomy of an XML-based Text Corpus Server  

LYA MEISTER
Perceptual Assessment of the Degree of Russian Accent  

MAGNUS MERKEL AND JODY FOO
Terminology Extraction and Term Ranking for Standardizing Term Banks  

vii
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 Estonian 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