NEALT PROCEEDINGS SERIES

VOL. 4

Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009

May 14-16, 2009

Odense, Denmark

Editors

Kristiina Jokinen and Eckhard Bick

NORTHERN EUROPEAN ASSOCIATION FOR LANGUAGE TECHNOLOGY

Proceedings of the NODALIDA 2009

NEALT Proceedings Series, Vol. 4

ISSN 1736-6305

Published by
Northern European Association for Language
Technology (NEALT)
http://omilia.uio.no/nealt

Electronically published at
Tartu University Library (Estonia)
http://dspace.utlib.ee/dspace/handle/10062/9206

Volume Editors
Kristiina Jokinen and Eckhard Bick

Contents

Preface

We are pleased to present the Proceedings of NODALIDA 2009, the 17th Nordic Conference of Computational Linguistics, held 14-16 May 2009 in Odense, Denmark.

The NODALIDA conference has always been an important meeting for the Nordic computational linguistics and language technology community. In recent years, especially through the establishment of the Northern European Association of Language Technology (NEALT), it has emerged as a major conference covering the geographical area of the Nordic countries as well as the Baltic countries and Northwest Russia. The previous NODALIDA conference was a success along the new dimension of being both a regional and an international character, and the current NODALIDA conference follows these lines. Although smaller in numbers, it shows similar recognition on the international level, as witnessed by a fair amount of submissions from outside the core geographical areas in Europe, and also from the US, India, and Japan.

We received altogether 82 submissions from 24 countries in the five categories of regular full papers, regular short papers, student papers, demos and workshops. The review process was rigorous and aimed at high scientific standards: each submission received three reviews and borderline cases were further subjected to discussion among the reviewers and the Programme Committee members. This resulted in the acceptance of 43 high-quality papers which appear in these proceedings, as well as five workshops which will produce their own proceedings. Of the accepted papers in the main conference, nine are short papers, three are student papers, and five are demos. The low number of student papers was disappointing, and we hope this will improve in future conferences.

The conference also features two distinguished invited speakers. Their talks concern language research and technological applications that allow us to address challenges encountered in the multilingual and multimodal contexts. Jean Carletta (University of Edinburgh) talks about interdisciplinary work on corpus collection, analysis of group dynamics, and interaction management in her keynote talk "Developing Meeting Support Technologies: From Data to Demonstration (and Beyond)". Ralf Steinberger (EC - Joint Research Centre) presents the cross-lingual functionality of a news analysis system and highlights various language technology topics in a rich multilingual environment (between 19 and 43 languages) in his keynote talk "Linking News Content Across Languages".

Besides presenting novel research, another important goal of the NODALIDA conference is to establish a series of tutorials concerning state-of-the-art language technology and computational linguistics research. In this conference, Graham Wilcock (University of Helsinki) presents an overview of linguistic annotation using open source tools in his tutorial "Text Annotation with OpenNLP and UIMA".

The conference programme also includes five workshops as specialised meetings on various relevant topics. We are proud to offer the following workshops, held immediately before the main conference:

W1: Nordic Perspectives on the CLARIN Infrastructure of Common Language Resources

W2: Multimodal Communication: from Human Behaviour to Computational Models

W3: Lexical Semantic Resources for NLP Purposes - the Interplay between Lexical Semantics, Lexicography, Terminology and Formal Ontologies

W4: Extraction and Use of Constructions in NLP

W5: Constraint Grammar and Robust Parsing

The conference has also attracted two satellite events, held before the workshops: the student and board meetings of the NGSLT (The Nordic Graduate School of Language Technology), and the project-related workshop "Linguistic Theory and Raw Sound" organised by Peter Juel Henrichsen (Copenhagen Business School). Moreover, during the conference there will be the second NEALT business meeting.

The organisation of a conference of this size is not possible without the efforts of several people working together in a friendly yet efficient manner. We would first like to thank our international Review Committee for their wonderful work on reviewing. Their prompt and constructive judgments greatly assisted us in putting together the current, exciting programme. We also wish to thank the Program Committee for their insightful comments, inviting the reviewers, and in general sharing their views on many complicated issues dealing with the structure and format of the conference. A big thank you goes to the Local Organisation Committee at the University of Southern Denmark for all their hard work concerning conference logistics and practical issues for the conference, and to the Institute of Language and Communication for financial and logistic support. Special thanks go to Mare Koit, Editor-in-Chief of the NEALT Publication Series at University of Tartu, for her kind help in the production of the electronic proceedings.

Finally, on behalf of the organisers, we would like to thank all the conference speakers and participants. Your interactions and enthusiasm will make the actual conference into what it aims to be: a forum for fruitful conversations and discussions which contribute to connections and work for years to come.

We wish you inspiring, useful, and enjoyable conference days at NODALIDA 2009.

Kristiina Jokinen

Programme Chair

NODALIDA 2009

Eckhard Bick

Local Chair

NODALIDA 2009

Committees

PROGRAM COMMITTEE

Kristiina Jokinen (CHAIR), University of Helsinki and University of Tartu
Robin Cooper, University of Göteborg
Anna Korhonen, University of Cambridge
Kaili Müürisep, University of Tartu
Joakim Nivre, Uppsala University
Patrizia Paggio, University of Copenhagen
Koenraad de Smedt, University of Bergen
Roman Yangarber, University of Helsinki

LOCAL ORGANIZATION COMMITTEE

Eckhard Bick (CHAIR), University of Southern Denmark
Poul Søren Kjærsgaard, University of Southern Denmark
Klaus Robering, University of Southern Denmark
Anette Wulff, University of Southern Denmark

REVIEWERS

Helena Ahonen-Myka, University of Helsinki, Finland
Lars Ahrenberg, University of Linköping, Sweden
Tanel Alumäe, Technical University of Tallinn, Estonia
Antti Arppe, University of Helsinki, Finland
Gemma Boleda, Universitat Politècnica de Catalunya, Spain
Francis Bond, NICT, Japan
Lars Borin, University of Gothenburg, Sweden
Rolf Carlson, KTH, Sweden
Mathias Creutz, Helsinki University of Technology, Finland
Antoine Doucet, University of Caen, France
Elisabet Engdahl, University of Gothenburg, Sweden
Stefan Evert, University of Osnabrück, Germany
Björn Gambäck, SICS, Sweden
Barbara Gawronska, University of Skövde, Sweden
Jeroen Geertzen, University of Cambridge, UK
Janne Bondi Johannessen, University of Oslo, Norway
Christer Johansson, University of Bergen, Norway
Heiki-Jaan Kaalep, University of Tartu, Estonia
Kaarel Kaljurand, University of Tartu, Estonia
Viggo Kann, KTH, Sweden
Jussi Karlgren, SICS, Sweden
Mare Koit, University of Tartu, Estonia
Dimitrios Kokkinakis, University of Gothenburg, Sweden
Kimmo Koskenniemi, University of Helsinki, Finland
Udo Kruschwitz, University of Sussex, UK
Yuval Krymolowski, Bar-Ilan University, Israel
Marco Kuhlmann, Uppsala University, Sweden
Krista Lagus, Helsinki University of Technology, Finland
Miro Lehtonen, University of Helsinki, Finland

Ian Lewin, EBI, UK
Krister Lindén, University of Helsinki, Finland
Ramón López-Cózar Delgado, University of Granada, Spain
Jan Tore Lønning, University of Oslo, Norway
Bente Maegaard, University of Copenhagen, Denmark
Beáta Megyesi, Uppsala University, Sweden
Kadri Muischnek, University of Tartu, Estonia
Costanza Navarretta, University of Copenhagen, Denmark
Pierre Nugues, University of Lund, Sweden
Jussi Piitulainen, University of Helsinki, Finland
Ari Pirkola, University of Tampere, Finland
Gábor Prószéky, Pázmány University, Hungary
Eiríkur Rögnvaldsson, University of Reykyavik, Iceland
Bolette Sandford Pedersen, University of Copenhagen, Denmark
Inguna Skadiņa, Tilde, Latvia
Torbjørn Svendsen, Norwegian University of Science and Technology, Norway
Anders Søgaard, University of Copenhagen, Denmark
Jürgen Wedekind, University of Copenhagen, Denmark
Aline Villavicencio, University of Rio Grande do Sul, Brazil
Martin Volk, University of Stockholm, Sweden
Atro Voutilainen, Connexor, Finland
Michael Zock, LIF, CNRS, Marseille, France

Conference program
NODALIDA-2009

13 May
Satellite events

9.15-12:20 **Workshop:** Linguistic Theory and Raw Sound
Organiser: Peter Juel Henrichsen (Copenhagen Business School)
15.00-18.00 **NGSLT board meeting**

14 May
9-17 Workshops

W1: Nordic perspectives on the CLARIN infrastructure of common language resource
W2: Multimodal Communication: from Human Behaviour to Computational Models
W3: WordNets and other Lexical Semantic Resources
W4: Extraction and Use of Constructions in NLP
W5: Constraint Grammar and robust parsing

19.00 **Reception at Odense Town Hall**

Main conference

15 May
9.00- 9.30 **Opening**
Kimmo Koskenniemi (President of NEALT)
Eckhard Bick (Chair of the local Organizing Committee)
Kristiina Jokinen (Chair of the Program Committee)

9.30-10.30 **Invited Talk** (Chair: Patrizia Paggio)
Jean Carletta (University of Edinburgh). *Developing Meeting Support Technologies: From Data to Demonstration (and Beyond)*

10.30-11.00 **Coffee**

11.00-13.00 **Regular papers** (3 parallel sessions)

Corpus, annotation, and their use (Chair: Rickard Domeij)	
11.00-11.30	Janne Bondi Johannessen, Joel Priestley, Kristin Hagen, Tor Anders Åfarli and Øystein Alexander Vangsnes. *The Nordic Dialect Corpus - an advanced research tool*
11.30-12.00	Nathan Green, Paul Breimyer, Vinay Kumar and Nagiza F. Samatova. *WebBANC: Building Semantically-Rich Annotated Corpora from Web User Annotations of Minority Languages*
12.00-12.30	Olga Lashevskaja and Olga Mitrofanova. *Disambiguation of Taxonomy Markers in Context: Russian Nouns*
12.30-13.00	Krister Lindén and Jussi Tuovila. *Corpus-based Paradigm Selection for Morphological Entries*

Morphology and Syntax (Chair: Koenraad de Smedt)	
11.00-11.30	Krister Lindén and Tommi Pirinen. *Weighted Finite-State Morphological Analysis of Finnish Compounding with HFST-LEXC*
11.30-12.00	Miikka Silfverberg and Krister Lindén. *Conflict Resolution Using Weighted Rules in HFST-TWOLC*
12.00-12.30	Hrafn Loftsson, Ida Kramarczyk, Sigrún Helgadóttir and Eiríkur Rögnvaldsson. I*mproving the PoS tagging accuracy of Icelandic text*
12.30-13.00	Katri Haverinen, Filip Ginter, Veronika Laippala and Tapio Salakoski. *Parsing Clinical Finnish: Experiments with Rule-Based and Statistical Dependency Parsers*

Semantic Classification (Chair: Robin Cooper)	
11.00-11.30	Michael Wiegand and Dietrich Klakow. *Predictive Features in Semi-Supervised Learning for Polarity Classification and the Role of Adjectives*
11.30-12.00	Jari Björne, Filip Ginter, Juho Heimonen, Sampo Pyysalo and Tapio Salakoski. *Learning to Extract Biological Event and Relation Graphs*
12.00-12.30	Jacob Persson, Richard Johansson and Pierre Nugues. *Text Categorization Using Predicate-Argument Structures*
12.30-13.00	Peter Kolb. *Experiments on the difference between distributional similarity and relatedness*

13.00-14.00 **Lunch**

14.00-14.45 **Demos** (Chair: Eiríkur Rögnvaldsson)

Eckhard Bick. *DeepDict - A Graphical Corpus-based Dictionary of Word Relations*

Sandra Derbring, Peter Ljunglöf and Maria Olsson. *SubTTS: Light-weight automatic reading of subtitles*

Peter Ljunglöf, Staffan Larsson, Katarina Mühlenbock and Gunilla Thunberg. *TRIK: A talking and drawing robot for children with communication disabilities*

Arne Martinus Lindstad, Anders Nøklestad, Janne Bondi Johannessen and Øystein A. Vangsnes. *The Nordic Dialect Database: Mapping Microsyntactic Variation in the Scandinavian Languages*

Bodil Nistrup Madsen and Hanne Erdman Thomsen. *CAOS - A Tool for the Construction of Terminological Ontologies*

14.45-15.30 **Student posters** (Chair: Kaili Müürisep)

Maria Eskevich. *Prominence detected by listeners for future speech synthesis application*

Okko Räsänen and Joris Driesen. *A comparison and combination of segmental and fixed-frame signal representations in NMF-based word recognition*

Bálint Sass. *Verb Argument Browser for Danish*

Regular poster

Anders Søgaard. *A linear time extension of deterministic pushdown automata*

15.30-16.00 **Coffee**

16.00-17.00 **Tutorial** (Chair: Joakim Nivre)
Graham Wilcock (University of Helsinki). *Text Annotation with OpenNLP and UIMA*

19.00 **Conference dinner**

16 May

9.00-10.00 **Invited Talk** (Chair: Kristiina Jokinen)
Ralf Steinberger (EC - Joint Research Centre). *Linking News Content Across Languages*

10.00-10.30 **Coffee**

10.30-11.30 **Regular papers** (3 parallel sessions)

Semantics (Chair: Costanza Navarretta)	
10.30-11.00	Bolette Sandford Pedersen and Anna Braasch. *What do we need to know about humans? A view into the DanNet database*
11.00-11.30	Kristina Nilsson and Hans Hjelm. *Using Semantic Features Derived from Word-Space Models for Swedish Coreference Resolution*

Parallel Corpora and Translation (Chair: Inguna Skadiņa)	
10.30-11.00	Hercules Dalianis, Martin Rimka and Viggo Kann. *Using Uplug and SiteSeeker to construct a cross language search engine for Scandinavian languages*
11.00-11.30	Christian Hardmeier and Martin Volk. *Using Linguistic Annotations in Statistical Machine Translation of Film Subtitles*

Algorithms (Chair: Kimmo Koskenniemi)	
10.30-11.00	Anders Søgaard. V*erifying context-sensitive treebanks and heuristic parses in polynomial time*
11.00-11.30	Anssi Yli-Jyrä. *An Efficient Double Complementation Algorithm for Superposition-Based Finite-State Morphology*

11.30-12.30 **Posters** (Chair: Eckhard Bick)

Lene Antonsen, Trond Trosterud and Saara Huhmarniemi. *Interactive pedagogical programs based on constraint grammar*

Eva Forsbom. *Extending the View: Explorations in Bootstrapping a Swedish PoS Tagger*
Tatiana Gornostay and Inguna Skadina. Pattern-based English-Latvian Toponym Translation

Yves Lepage and Chooi Ling Goh. *Towards automatic acquisition of linguistic features*

Miguel A. Molinero, Benoit Sagot and Lionel Nicolas. *Building a morphological and syntactic lexicon by merging various linguistic resources*

Magnus Rosell. *Part of Speech Tagging for Text Clustering in Swedish*

Natalie Schluter and Josef van Genabith. *Dependency Parsing Resources for French: Converting Acquired Lexical Functional Grammar F-Structure Annotations and Parsing F-Structures Directly*

12.30-13.30 **Lunch**

13.30-14.30 **Short regular papers** (3 parallel sessions)

Parsing and Tagging (Chair: Janne Bondi Johannessen)	
13.30-13.50	Mark Fishel and Joakim Nivre. *Voting and Stacking in Data-Driven Dependency Parsing*
13.50-14.10	Rashmi Gangadharaiah, Ralf Brown and Jaime Carbonell. *Active Learning in Example-Based Machine Translation*
14.10-14.30	Beata B. Megyesi. *The Open Source Tagger HunPoS for Swedish*

Semantic Analysis (Chair: Poul Søren Kjærsgaard)	
13.30-13.50	Eckhard Bick and M. Pilar Valverde Ibáñez. *Automatic Semantic Role Annotation for Spanish*
13.50-14.10	Manfred Klenner, Angela Fahrni and Stefanos Petrakis. *PolArt: A Robust Tool for Sentiment Analysis*
14.10-14.30	Lilja Øvrelid. *Cross-lingual porting of distributional semantic classification*

Applications (Chair: Klaus Robering)	
13.30-13.50	Inguna Skadiņa and Edgars Bralitis. *English-Latvian SMT: knowledge or data?*
13.50-14.10	Karin Friberg Heppin. *MedEval - Six Test Collections in One*
14.10-14.30	Anton Karl Ingason, Skúli Bernhard Jóhannsson, Eiríkur Rögnvaldsson, Hrafn Loftsson and Sigrún Helgadóttir. *Context-Sensitive Spelling Correction and Rich Morphology*

14.30-15.30 **NEALT Business meeting**

15.30-16.00 **Closing**

16.00-16.30 **Coffee**