

NEALT PROCEEDINGS SERIES  
VOL. 11

Proceedings of the  
18th Nordic Conference of Computational Linguistics  
NODALIDA 2011

May 11-13, 2011  
Riga, Latvia

*Editors*

Bolette Sandford Pedersen, Gunta Nešpore and Inguna Skadiņa

Proceedings of the NODALIDA 2011

NEALT Proceedings Series, Vol. 11

© 2011 The editors and contributors.

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/10062/16955>

*Volume Editors*

Bolette Sandford Pedersen, Gunta Nešpore and Inguna Skadiņa

*Series Editor-in-Chief*

Mare Koit

*Series Editorial Board*

Lars Ahrenberg  
Koenraad De Smedt  
Kristiina Jokinen  
Joakim Nivre  
Patrizia Paggio  
Vytautas Rudžionis

*Supported by*

Institute of Mathematics and Computer Science, University of Latvia (ERAF project,  
agreement No. 2010/0206/2DP/2.1.1.2.0/10/APIA/VIAA/011)

# Contents

Preface	viii
Committees	x
Conference Program	xii
<b>I Invited Papers</b>	<b>1</b>
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
<b>II Regular papers</b>	<b>17</b>
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 Hjälmsjö, 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
<b>III Regular short papers</b>	<b>270</b>
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

## Preface

The computational linguistics and language technology communities in the Nordic and Baltic countries have always considered the NODALIDA conference as one of the important events for meeting and interchanging new research in the field. Through the establishment of the Northern European Association of Language Technology (NEALT) in 2006, the NODALIDA conference has increased its importance and is now recognized outside the Nordic regions, as can be seen by the fact that we have received several European submissions from outside the Nordic and Baltic countries, as well as submissions from outside Europe such as the US, India, and Pakistan. We are very pleased to hereby present the Proceedings of NODALIDA 2011, the 18th Nordic Conference of Computational Linguistics, held 11-13 May 2011 in Riga, Latvia. We hope that these proceedings will serve as a useful and comprehensive repository of information, will facilitate research in language technology and will encourage the development of further language resources for the Nordic and Baltic languages!

According to the reviews provided by the review committee, a vast majority of the papers submitted for the conference this year were of very good quality. This is a positive sign of the fact that language technology in the Nordic and Baltic countries is striving. However, maintaining the tradition of the NODALIDA conference running over two days plus a workshop day, time scarcity has enforced us to accept only a limited number of papers. This means that even with an acceptance rate above 60%, several quality papers have been rejected. To sum up in figures, we received altogether 85 submissions from 20 countries in the four categories of full papers, short /demo papers, student papers, and workshops. Each submission received three reviews and borderline cases were further subjected to discussion among the Program Committee members. For the conference, we have accepted 52 papers which appear in these proceedings, as well as three workshops which will produce their own proceedings. Of the accepted papers in the main conference, 33 are long papers presented as talk or poster, 14 are short papers presented as poster or demo and five are student papers of which three are presented as talk and two as poster. It should be pointed out that most of the submissions are from the Nordic countries and only a limited number of papers are from the Baltic region. This may be because the Baltic HLT conference was held only recently. The papers selected for the conference represent a wide range of topics of research, including corpus linguistics, lexicography, morphological and syntactic processing, machine translation, speech technologies, semantics, and other areas of language technology.

We also have the pleasure of presenting three invited speakers at NODALIDA 2011, one of which is invited to present ongoing research in the host country, Latvia, and two others to present ongoing research in Sweden and Scotland, respectively. The invited talks concern central aspects of language technology such as discourse analysis, dependency parsing, and controlled natural languages. Bonnie Webber from University of Edinburgh talks about discourse structures and language technology and discusses how discourse structures can help to improve language technologies, and further, how language technologies can help to induce and model discourse structures. Joakim Nivre from Uppsala University gives a survey of recent advances in so-called bare-bones dependency parsing; focusing in particular on transition-based methods for highly efficient parsing. Guntis Bārzdiņš from University of Latvia talks about a new kind of rich controlled natural language which allows to narrow the gap with true natural language.

In addition, the conference program includes three workshops; two on the specialized topics terminology and Constraint Grammar, and one with the broader focus on visibility of language resources.

Moreover, the conference has attracted a satellite event, held before the workshops: The project-related meeting in META-NET/META-NORD which is the Nordic and Baltic branch of a Network of Excellence dedicated to building the technological foundations of a multilingual European information society. Finally, during the conference there will be the third NEALT business meeting.

The organization of a conference of this size is a joint effort between several organizational units. We would first like to thank our reviewers for their conscientious work in reviewing all the submitted contributions. We also wish to thank the Program Committee for inviting the reviewers as well as for the fruitful discussions regarding how to ensure a conference of high quality. A big thank you goes to the Local Organization Committee at the Institute of Mathematics and Computer Science of University of Latvia for their work concerning practical issues for the conference. Special thanks go to Mare Koit, Editor-in-Chief of the NEALT Publication Series at University of Tartu, for producing the electronic proceedings.

We wish you an inspiring conference!

Bolette Sandford Pedersen  
Program Chair  
NODALIDA 2011

Inguna Skadiņa  
Local Chair  
NODALIDA 2011

## Committees

### PROGRAM COMMITTEE

Bolette Sandford Pedersen (Program Chair), University of Copenhagen, Denmark  
Kristiina Jokinen, University of Helsinki, Finland  
Jussi Karlgren, Swedish Institute of Computer Science, Sweden  
Ruta Marcinkeviciene, Vytautas Magnus University, Lithuania  
Meelis Mihkila, Institute of the Estonian Language, Estonia  
Costanza Navarretta, University of Copenhagen, Denmark  
Anders Nøklestad, University of Oslo, Norway  
Eirikur Rögnvaldsson, University of Iceland, Iceland

### LOCAL ORGANIZATION COMMITTEE

Inguna Skadiņa (Local Chair), Institute of Mathematics and Computer Science,  
University of Latvia  
Rihards Balodis, Institute of Mathematics and Computer Science, University of Latvia  
Gunta Nešpore, Institute of Mathematics and Computer Science, University of Latvia  
Gunta Plataiskalna, Institute of Mathematics and Computer Science, University of Latvia  
Ilmārs Poikāns, Institute of Mathematics and Computer Science, University of Latvia  
Baiba Saulīte, Institute of Mathematics and Computer Science, University of Latvia  
Andrejs Spektors, Institute of Mathematics and Computer Science, University of Latvia

### REVIEWERS

Toomas Altosaar, Helsinki University of Technology, Finland  
Tanel Alumäe, Tallinn University of Technology, Estonia  
Ilze Auziņa, University of Latvia, Latvia  
Eckhard Bick, Syddansk Universitet, Denmark  
Kristín Bjarnadóttir, Árni Magnússon Institute, Iceland  
Anne Bjerre, Syddansk Universitet, Denmark  
Anna Braach, University of Copenhagen, Denmark  
Hanne Fersøe, University of Copenhagen, Denmark  
Jody Foo, Linköping University, Sweden  
Björn Gambäck, Norwegian University of Science and Technology, Norway & Swedish  
Institute of Computer Science, Sweden  
Tatiana Gornostay, Tilde, Latvia  
Gintare Grigonyte, Vytautas Magnus University, Lithuania  
Joakim Gustafson, Kungliga Tekniska Högskolan, Sweden  
Kristin Hagen, University of Oslo, Norway  
Daniel Hardt, Copenhagen Business School, Denmark  
Sigrún Helgadóttir, Árni Magnússon Institute, Iceland  
Janne Bondi Johannessen, University of Oslo, Norway  
Lars G. Johnsen, University of Bergen, Norway  
Heikki-Jaan Kaalep, University of Tartu, Estonia  
Mari-Liis Kalvik, Institute of the Estonian Language, Estonia  
Sabine Kirchmeier-Andersen, Danish Language Council, Denmark  
Krista Lagus, Aalto University, Finland  
Yves Lepage, Waseda University, Japan

Krister Linden, University of Helsinki, Finland  
Hrafn Loftsson, Reykjavik University, Iceland  
Jan Tore Lønning, University of Oslo, Norway  
Bente Maegaard, University of Copenhagen, Denmark  
Sanni Nimb, Danish Society for Language and Literature, Denmark  
Joakim Nivre, Uppsala University, Sweden  
Stephan Oepen, University of Oslo, Norway  
Fredrik Olsson, Gavagai, Sweden  
Patrizia Paggio, University of Copenhagen, Denmark  
Hille Pajupuu, Institute of the Estonian Language, Estonia  
Ari Pirkola, Tampere, University of Tampere, Finland  
Gailius Raskinis, Vytautas Magnus University, Lithuania  
Anders Søgaard, University of Copenhagen, Denmark  
Hanne Erdman Thomsen, Copenhagen Business School, Denmark  
Trond Trosterud, University of Tromsø, Norway  
Oscar Täckström, Swedish Institute of Computer Science & Uppsala University, Sweden  
Andrius Utkas, Vytautas Magnus University, Lithuania  
Martti Vainio, University of Helsinki, Finland  
Erik Velldal, University of Oslo, Norway  
Sumithra Velupillai, Stockholm University, Sweden  
Carl Vogel, Trinity College Dublin, Ireland  
Joel Wallenberg, University of Iceland, Iceland  
Jürgen Wedekind, University of Copenhagen, Denmark  
Matthew Whelpton, University of Iceland, Iceland  
Atro Voutilainen, University of Helsinki, Finland  
Mats Wirén, Stockholm University, Sweden  
Roman Yangarber, University of Helsinki, Finland  
Robert Östling, Stockholm University  
Lilja Øvreliid, University of Oslo, Norway

# Conference program

NODALIDA-2011

## 11 May

### Satellite events

#### Workshops

- Workshop on Creation, Harmonization and Application of Terminology Resources
- Workshop in Constraint Grammar Applications
- Workshop on Visibility and Availability of LT resources

19.00            *Welcome reception*

## 12 May

9.00–9.30        **Opening**

Mārcis Auziņš (Rector of the University of Latvia)  
Janne Bondi Johannessen (President of NEALT)  
Inguna Skadiņa (Chair of the Local Organizing Committee)  
Bolette Sandford Pedersen (Chair of the Program Committee)

9.30–10.30      **Invited Talk** (Chair: Costanza Navarretta)

Prof. Bonnie Webber (University of Edinburgh). *Discourse Structures and Language Technologies*

10.30–11.00     *Coffee*

11.00–13.00 **3 parallel sessions: REGULAR papers**

<b>Corpus creation, annotation and use</b> (Chair: Eiríkur Rögnvaldsson)	
11.00–11.30	Costanza Navarretta, Elisabeth Ahlsén, Jens Allwood, Kristiina Jokinen and Patrizia Paggio. <i>Creating Comparable Multimodal Corpora for Nordic Languages</i>
11.30–12.00	Rico Sennrich and Martin Volk. <i>Iterative, MT-based Sentence Alignment of Parallel Texts</i>
12.00–12.30	Estelle Delpech. <i>Evaluation of Terminologies Acquired from Comparable Corpora: an Application Perspective</i>
12.30–13.00	Janne Bondi Johannessen and Emiliano Raúl Guevara. <i>What Kind of Corpus is a Web Corpus?</i>
<b>Text and language classification</b> (Chair: Hanne Fersøe)	
11.00–11.30	Taraka Rama and Lars Borin. <i>Estimating Language Relationships from a Parallel Corpus. A Study of the Europarl Corpus</i>
11.30–12.00	Robert Remus. <i>Improving Sentence-level Subjectivity Classification through Readability Measurement</i>
12.00–12.30	Michael Wiegand and Dietrich Klakow. <i>Convolution Kernels for Subjectivity Detection</i>
<b>Morphology and POS tagging</b> (Chair: Janne Bondi Johannessen)	
11.00–11.30	Miikka Silfverberg and Krister Lindén. <i>Combining Statistical Models for POS Tagging using Finite-State Calculus</i>
11.30–12.00	Niels Beuck, Arne Köhn and Wolfgang Menzel. <i>Decision Strategies for Incremental POS Tagging</i>
12.00–12.30	Anssi Yli-Jyrä. <i>Explorations on Positionwise Flag Diacritics in Finite-State Morphology</i>
12.30–13.00	Heiki-Jaan Kaalep and Kadri Muischnek. <i>Morphological Analysis of a Non-Standard Language Variety</i>

13.00–14.00 *Lunch*

14.00–15.30

**12 Posters and Demos** (Chair: Anders Nøklestad)

<b>Wordnets and lexical issues</b>
<p>Kristiina Muhonen and Krister Lindén. <i>Do Wordnets also Improve Human Performance on NLP Tasks?</i></p> <p>Loïc Boizou. <i>The Formal Patterns of the Lithuanian Verb Forms</i></p> <p>Hector Martinez Alonso, Núria Bel and Bolette Sandford Pedersen. <i>Identification of Sense Selection in Regular Polysemy using Shallow Features</i></p> <p>Anders Johannsen and Bolette Sandford Pedersen. “<i>Andre ord</i>” — <i>a Wordnet Browser for the Danish Wordnet, DanNet</i> (DEMO)</p>
<b>Syntax</b>
<p>Anne Bjerre. <i>Extraction from Relative and Embedded Interrogative Clauses in Danish</i></p> <p>Martin Hassel, Aron Henriksson and Sumithra Velupillai. <i>Something Old, Something New — Applying a Pre-trained Parsing Model to Clinical Swedish</i></p> <p>Atro Voutilainen and Tanja Purtonen. <i>A Double-blind Experiment on Interannotator Agreement: the Case of Dependency Syntax and Finnish</i></p> <p>Lauma Pretkalniņa, Gunta Nešpore, Kristīne Levāne-Petrova and Baiba Saulīte. <i>A Prague Markup Language Profile for the SemTi-Kamols Grammar Model</i></p> <p>Daiga Dekšne and Raivis Skadiņš. <i>CFG Based Grammar Checker for Latvian</i></p>
<b>Morphology</b>
<p>Sami Virpioja, Oskar Kohonen and Krista Lagus. <i>Evaluating the Effect of word Frequencies in a Probabilistic Generative Model of Morphology</i></p> <p>Tommi Pirinen. <i>Modularisation of Finnish Finite-State Language Description — Towards Wide Collaboration in Open Source Development of a Morphological Analyser</i></p>
<b>Machine translation</b>
<p>Sara Stymne. <i>Iterative Reordering and Word Alignment for Statistical MT</i></p>

15.30–15.45

*Coffee*

15.45–17.15      **3 parallel sessions: REGULAR papers**

<b>Speech</b> (Chair: Meelis Mihkla)	
15.45–16.15	Sofia Strömbergsson. <i>Corrective Re-synthesis of Deviant Speech Using Unit Selection</i>
16.15–16.45	Peter Juel Henriksen. <i>Fishing in a Speech Stream, Angling for a Lexicon</i>
16.45–17.15	Bea Valkenier, Dirkjan Krijnders, Ronald van Elburg and Tjeerd Andringa. <i>Psycho-Acoustically Motivated Formant Feature Extraction</i>
<b>Search and information extraction</b> (Chair: Costanza Navarretta)	
15.45–16.15	Lars Borin, Markus Forsberg and Christer Ahlberger. <i>Semantic Search in Literature as an e-Humanities Research Tool: CONPLISIT — Consumption Patterns and Life-Style in 19th Century Swedish Literature</i>
16.15–16.45	Silja Huttunen, Arto Vihavainen and Roman Yangarber. <i>Relevance Prediction in Information Extraction Using Discourse and Lexical Features</i>
16.45–17.15	Gintarė Grigonytė, Erika Rimkutė, Andrius Utkā and Loïc Boizou. <i>Experiments on Lithuanian Term Extraction</i>
<b>Syntax, indexing</b> (Chair: Jussi Karlgren)	
15.45–16.15	Peter Ljunglöf. <i>Editing Syntax Trees on the Surface</i>
16.15–16.45	Anders Søgaard. <i>Using Graphical Models for PP Attachment</i>
16.45–17.15	Erik Velldal. <i>Random Indexing Re-Hashed</i>

17.15–18.15      **Invited Talk** (Chair: Inguna Skadiņa)

Prof. Guntis Bārzdīņš (University of Latvia). *When FrameNet Meets a Controlled Natural Language*

19.30              **Conference dinner**

## 13 May

9.00–10.00      **Invited Talk** (Chair: Kristiina Jokinen)  
Prof. Joakim Nivre (Uppsala University). *Bare-Bones Dependency Parsing — A Case for Occam's Razor?*

10.00–10.30      *Coffee*

10.30–12.00      **3 parallel sessions: REGULAR papers and STUDENT papers**

<b>Lexicon, etymology</b> (Chair: Bolette Sandford Pedersen)	
10.30–11.00	Eckhard Bick. <i>A FrameNet for Danish</i>
11.00–11.30	Ingemar Hjälmsstad, Martin Hassel and Maria Skeppstedt. <i>The Impact of Part-of-Speech Filtering on Generation of a Swedish-Japanese Dictionary using English as Pivot Language</i>
11.30–12.00	Hannes Wettig and Roman Yangarber. <i>Probabilistic Models for Alignment of Etymological Data</i>
<b>Machine translation; classification</b> (Chair: Andrejs Vasiļjevs)	
10.30–11.00	Martin Volk and Rico Sennrich. <i>Disambiguation of English Contractions for Machine Translation of TV Subtitles</i>
11.00–11.30	Raivis Skadiņš, Tatiana Gornostay and Valters Šics. <i>Toponym Disambiguation in English-Lithuanian SMT System with Spatial Knowledge</i>
11.30–12.00	Eirini Florou and Stasinou Konstantopoulos. <i>A Quantitative and Qualitative Analysis of Nordic Surnames</i>
<b>Student papers</b> (Chair: Normunds Grūzītis)	
10.30–11.00	Marius Olausen. <i>Evaluating the Speech Quality of the Norwegian Synthetic Voice Brage</i>
11.00–11.30	Mojgan Seraji. <i>A Statistical Part-of-Speech Tagger for Persian</i>
11.30–12.00	Angel Genov and Georgi Iliev. <i>Linguistic Motivation in Automatic Sentence Alignment of Parallel Corpora: the Case of Danish-Bulgarian and English-Bulgarian</i>

12.00–13.00      *Lunch*

13.00–14.30      **12 Posters/demos** (Chair: Kristiina Jokinen)

<b>Classification &amp; summarization</b>
<p>Christian Smith and Arne Jönsson. <i>Automatic Summarization as Means of Simplifying Texts, an Evaluation for Swedish</i></p> <p>Per Almquist and Jussi Karlgren. <i>Experiments to Investigate the Utility of Nearest Neighbor Metrics Based on Linguistically Informed Features for Detecting Textual Plagiarism</i></p> <p>Jody Foo. <i>A Categorization Scheme for Analyzing Rules from a Handbook of Swedish Writing Rules</i></p> <p>Anju Saxena and Lars Borin. <i>Dialect Classification in the Himalayas: a Computational Approach</i></p>
<b>Knowledge systems</b>
<p>Ann-Marie Eklund. <i>Query Constraining Aspects of Knowledge</i></p> <p>Kenneth Wilhelmsson. <i>Automatic Question Generation from Swedish Documents as a Tool for Information Extraction</i></p>
<b>Corpus creation, annotation and use</b>
<p>Christian Hänig. <i>Knowledge-free Verb Detection through Sentence Sequence Alignment</i></p> <p>Maria Holmqvist and Lars Ahrenberg. <i>A Gold Standard for English-Swedish Word Alignment</i></p> <p>Anne-Kathrin Schumann. <i>Corpus-based Terminology: Detection, Description and Representation of Knowledge-rich Contexts in Russian</i></p>
<b>Student posters</b>
<p>Anastasia Shimorina and Maria Grachkova. <i>Identification of Context Markers for Russian Nouns</i></p> <p>Gideon Kotzé. <i>Finding Statistically Motivated Features Influencing Subtree Alignment Performance</i></p>

14.30–15.30      **NEALT Business meeting**

15.30–16.00      **Closing**

16.00-16.30      *Coffee*