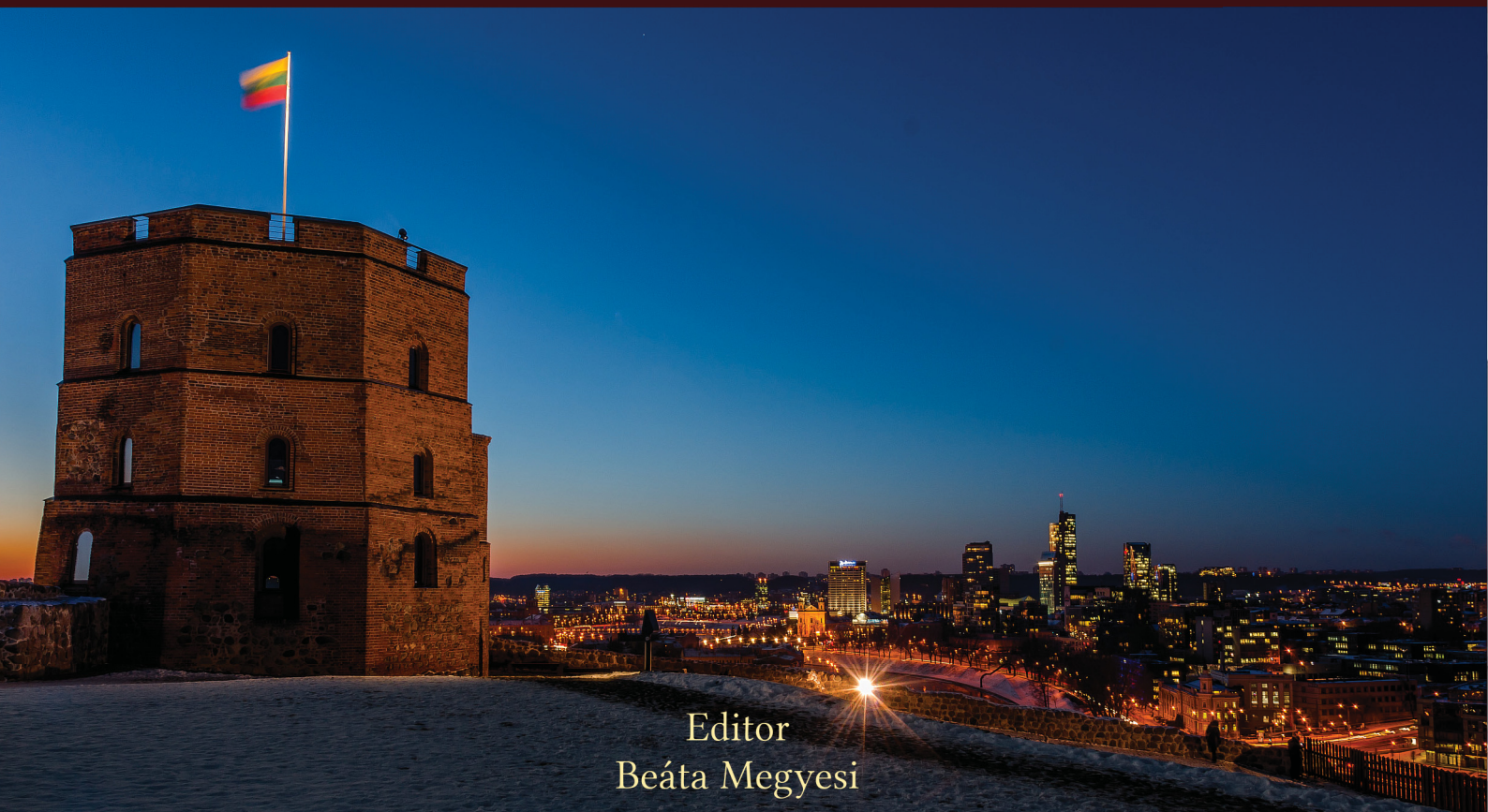


NEALT

Northern European Association for
Language Technology

NEALT Proceedings Series Vol. 23



Editor
Beáta Megyesi

Proceedings of the
20th Nordic Conference of Computational Linguistics

NODALIDA 2015

May 11-13, 2015
Institute of the Lithuanian Language
Vilnius, Lithuania

Cover photo 'Vilnius castle tower by night' by Mantas Volungevičius
<http://www.flickr.com/photos/112693323@N04/13596235485/>
Licensed under Creative Commons Attribution 2.0 Generic
See <http://creativecommons.org/licenses/by/2.0/> for full terms

Cover design Nils Blomqvist

Proceedings of the 20th Nordic Conference of Computational Linguistics NODALIDA 2015

Editor
Beáta Megyesi

May 11-13, 2015
Institute of the Lithuanian Language
Vilnius, Lithuania

Published by

ACL Anthology
<https://www.aclweb.org/anthology/>

Linköping University Electronic Press, Sweden
Linköping Electronic Conference Proceedings #109
ISSN: 1650-3638
eISSN: 1650-3740
ISBN: 978-91-7519-098-3

Sponsors



Preface: Program Chair

We are very pleased to introduce the proceedings of the 20th Nordic Conference on Computational Linguistics (NODALIDA 2015), held at the Institute of the Lithuanian Language in Vilnius, Lithuania, between May 11 and May 13, 2015. The proceedings is published as part of the NEALT Proceedings Series by Linköping University Electronic Press, and is also publicly available in the ACL Anthology for the first time.

NODALIDA have been held bi-annually since 1977, first organized as a friendly gathering in Gothenburg, Sweden to discuss on-going research in computational linguistics in the Nordic countries. Nearly 30 years later, in 2006, the Northern European Association for Language Technology (NEALT) was founded to organize NODALIDA and other events in the Nordic countries, the Baltic states, and Northwest Russia to promote research, cooperation and information exchange in the field of language technology in a wide sense. Today, NODALIDA addresses all aspects of computational linguistics, natural language processing, and speech technology, including work in closely related neighboring disciplines. The conference has been internationally recognized outside the Nordic regions and submissions are received from all over the world. It is a great honor to serve as the Program Chair for NODALIDA 2015, to be held in Lithuania for the first time.

Following the pattern of previous years, the Program Committee invited paper submissions in four distinct tracks: *regular papers* on substantial, original, and unpublished research, including empirical evaluation results, where appropriate; *student papers* on completed or ongoing work, where at least the first author is a student; *short papers* on smaller, focused contributions, work in progress, negative results, surveys, or opinion pieces; and *demonstration papers* summarizing a software system or language resource, to be accompanied by a live demonstration at the conference.

The conference received 68 submissions from all over Europe as well as from Canada, India, Japan, and the US. We followed the standards of recent NODALIDAS—emerged since 2007—with high quality technical track with peer-review of all papers, and an acceptance rate of 61%. All submitted papers went through a rigorous review process. The regular, student and short papers were reviewed by three experts in the field while the demonstration papers were reviewed by two experts. The final selection was made by the Program Committee, which was not an easy task due to many submissions with high scores and overall positive reviews, and the time and space constraints of the two day long main conference. We aimed at achieving balance between regular and short papers, and was more lenient in the student and demo categories. Our goal was to include papers dealing with a wide variety of topics from various regions while maintaining NODALIDA’s regional character as the major conference for Nordic research. 42 submissions were accepted for presentation either as long talks, poster presentations, lightning talks with poster presentations, or demos. In the final program, there are 22 regular, 5 student, 8 short, and 7 demonstration papers, all collected in this volume.

In addition to the accepted papers, we are proud to present three invited keynote speakers, distinguished researchers from France, Great Britain, and the US, to cover different areas of the conference.

Kevin Knight (University of Southern California) presents work on the use of redundancy information occurring in natural language by humans to improve automatic language processing applications, such as summarization or machine translation. Catherine Pelachaud (CNRS-LTICI, TELECOM-ParisTech) talks about how to model virtual agents with socio-emotional capabilities, in particular how we can animate laughter and virtual agents who can laugh when interacting with humans. Sebastian Riedel (University College London) presents his work on teaching machines to read and to reason about what was read. In addition, we followed the quite recent tradition of organizing a local language tutorial on Lithuanian.

The conference program also includes four workshops: i) NLP for Computer Assisted Language Learning, ii) Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities, iii) Constraint Grammar - Methods, Tools and Applications and iv) Innovative Corpus Query and Visualization Tools. Each workshop has been organized by their own committees, and produced their own proceedings published in the same series.

Organizing a conference with a good program is complex and relies on the goodwill of many researchers involved in the field. I would like to express my gratitude and appreciation to my fellows on the Program Committee for their hard and invaluable work for sharing the effort of creating the program. A special thanks goes to Stephan Oepen, the program chair of NODALIDA 2013, and NEALT's president, Bolette Sandford Pedersen, for generous advice. Wholehearted thanks go to the 62 reviewers for their time and effort to contribute to the reviewing and selection of papers. I am also grateful to the three keynote speakers, the presenter of the local language tutorial, and the workshop organizers! And of course, all the authors who submitted papers deserve special thanks. Without you, this conference would not take place! In addition, I would also like to acknowledge and thank Nils Blomqvist for professionally serving as the proceedings co-manager. I am also indebted to Lars Ahrenberg, the editor-in-chief of NEALT, for helping the publication of the proceedings in ACL anthology, in parallel with Linköping University Electronic Press, come true. My greatest debt goes to the Institute of the Lithuanian Language, Jolanta Zabarskaitė, and in particular Violeta Meiliūnaitė for carrying the heavy burden of the local organization, and for being a great host in the picturesque city of Vilnius. Lastly, I am grateful to my colleagues at Department of Linguistics and Philology at Uppsala University for their patience with me during the last year and generously letting me hide from time to time while organizing this conference, and to my nearest and dearest—my twins and friends—for generously giving me the space to disappear into our world of language technology.

I wish you all a fruitful conference and hope you will enjoy NODALIDA 2015!

Beáta Megyesi (Program Chair)

Preface: Local Organizer Chair

I would like to extend a warm welcome to the participants and guests at NODALIDA 2015.

The fact that this conference, with its highly acclaimed status and prominence in the academia, is taking place in Lithuania in 2015 is significant in several ways.

We find it very important that such a high-level event, which attracts a great many scientists and graduate students from all over the world, is taking place in a country that has its language classed as one of the less-used. It shows that NEALT understands and supports the involvement of the minor European languages that do not have big technological markets in the processes of supporting multilingualism and multiculturalism both in Europe and on a world-wide scale, which processes are in fact aimed at developing and upgrading language technologies. Needless to say, this kind of involvement is critical to the Lithuanian language.

What is more, in the digital age language acquires many new functions, evolving from a tool of communication and a persuader into something that creates added value in the society of knowledge and creative process. Figuratively speaking, the impalpability of the digital world brings forth perfectly tangible things. As John Searle, one of the most prominent contemporary linguistic philosophers, once said – words make things. And it is the rapidly developing language technologies that make it happen in the first place. Language technologies facilitate the retrieval of information, management of different things, communication, and exchange of creative ideas that are rooted in the unique nature of each and every language. Ideas are the backbone of innovation as the key precondition for global development. As a researcher of the Lithuanian language, I believe that eventually there will be no more minor and major languages in the world, as they all have equal opportunities. And all that thanks to language technologies alone.

Another important thing is that NODALIDA 2015 is taking place in a country whose language is considered one of the languages that have preserved the structure of the Indo-European parent language the best, and for a good reason. The structure of the Lithuanian language is indeed very complicated, and its digitalisation poses a significant challenge to scientists. I truly hope that as a result of this conference more researchers will discover a passion for tackling difficult problems, such as the Lithuanian language and its next-of-kin, the Latvian language.

The future of languages is in the hands of language technologists. Sharing scientific expertise, discovering new contacts, meeting old friends, setting up and updating scientific networks, developing and presenting new ideas is what we all expect from NODALIDA 2015.

Many thanks to all who have gathered here in Vilnius.

Jolanta Zabarskaitė (Local Organizer Chair)

Program Committee

- Filip Ginter, University of Turku, Finland
- Kristiina Jokinen, University of Helsinki, Finland
- Arne Jönsson, Linköping University, Sweden
- Violeta Meiliūnaitė, Institute of the Lithuanian Language, Lithuania
- Beáta Megyesi (Program Chair), Uppsala University, Sweden
- Costanza Navarretta, University of Copenhagen, Denmark
- Stephan Oepen, University of Oslo, Norway
- Oscar Täckström, Google Inc.

Organizing Committee

- Jolanta Zabarskaitė
- Violeta Meiliūnaitė

Reviewers

- Yvonne Adesam, University of Gothenburg, Sweden
- Lars Ahrenberg, Linköping University, Sweden
- Krasimir Angelov, University of Gothenburg, Sweden
- Bernd Bohnet, University of Stuttgart, Germany
- Janne Bondi Johannessen, University of Oslo, Norway
- Johan Boye, Royal Technical Institute (KTH), Sweden
- Diego Ceccarelli, National Research Council, Italy
- Lin Chen, University of Illinois, USA
- Koenraad De Smedt, University of Bergen, Norway
- Rodolfo Delmonte, Ca'Foscari University Venice, Italy
- Jens Edlund, Royal Technical Institute (KTH), Sweden
- Anna Esposito, Second University of Naples, Italy
- Björn Gambäck, Norwegian University of Science and Technology, Norway

- Gintare Grigonyte, University of Stockholm, Sweden
- Christian Hardmeier, Uppsala universitet, Sweden
- Petter Haugereid, NTNU, Trondheim, Norway
- Katarina Heimann Mühlenbock, University of Gothenburg, Sweden
- Anna Hjalmarsson, Royal Technical Institute (KTH), Sweden
- Angelina Ivanova, University of Oslo, Norway
- Richard Johansson, University of Gothenburg, Sweden
- Sofie Johansson Kokkinakis, University of Gothenburg, Sweden
- Jussi Karlgren, Gavagai and KTH, Sweden
- Mare Koit, University of Tartu, Estonia
- Marco Kuhlmann, Linköping University, Sweden
- Alessandro Lenci, University of Pisa, Italy
- Krister Lindén, University of Helsinki, Finland
- Christian Lindgren, Lund University, Sweden
- Diego Marcheggiani, Italian National Research Council, Italy
- Antonia Martí, University of Barcelona, Spain
- Mattias Nilsson, Karolinska Institutet, Sweden
- Joakim Nivre, Uppsala University, Sweden
- Pierre Nugues, Lund University, Sweden
- Gustav Öqvist Seimyr, Karolinska Institutet, Sweden
- Constantin Orasan, University of Wolverhampton, UK
- Robert Östling, Stockholm University, Sweden
- Lilja Øvrelid, University of Oslo, Norway
- Patrizia Paggio, University of Copenhagen, Denmark
- Eva Pettersson, Uppsala University, Sweden
- Tommi A Pirinen, Dublin City University, Ireland
- Barbara Plank, University of Copenhagen, Denmark
- Sampo Pyysalo, University of Turku, Finland

- Siva Reddy, University of Edinburgh, UK
- Fabio Rinaldi, University of Zurich, Switzerland
- Eiríkur Rognvaldsson, School of Humanities, Iceland
- Rune Sætre, NTNU, Trondheim, Norway
- Giampiero Salvi, Royal Technical Institute (KTH), Sweden
- Marina Santini, Nordstedts, Sweden
- Arne Skjærholt, University of Oslo, Norway
- Anders Søgaard, University of Copenhagen, Denmark
- Sara Stymne, Uppsala University, Sweden
- Torbjørn Svendsen, Norwegian University of Science and Technology, Norway
- Nina Tahmasebi, University of Gothenburg, Sweden
- Joel Tetreault, Yahoo Labs, USA
- Jörg Tiedemann, Uppsala University, Sweden
- Olga Uryupina, University of Trento, Italy
- Erik Velldal, University of Oslo, Norway
- Sumithra Velupillai, Stockholm University, Sweden
- Martin Volk, University of Zurich, Switzerland
- Jürgen Wedekind, University of Copenhagen, Denmark
- Mats Wirén, Stockholm University, Sweden
- Roman Yangarber, University of Helsinki, Finland
- Heike Zinsmeister, University of Hamburg, Germany

INVITED TALK: How Much Information Does a Human Translator Add to the Original?

Kevin Knight

ISI, University of Southern California, USA

knight@isi.edu

Abstract

It is well-known that natural language has built-in redundancy. By using context, we can often guess the next word or character in a text. Two practical communities have independently exploited this fact. First, automatic speech and translation researchers build language models to distinguish fluent from non-fluent outputs. Second, text compression researchers convert predictions into short encodings, to save disk space and bandwidth. I will explore what these two communities can learn from each others' (interestingly different) solutions. Then I will look at the less-studied question of redundancy in bilingual text, addressing questions like "How well can we predict human translator behavior?" and "How much information does a human translator add to the original?" (This is joint work with Barret Zoph and Marjan Ghazvininejad.)

Bio

Kevin Knight is Director of Natural Language Technologies at the Information Sciences Institute (ISI) of the University of Southern California (USC), and a Professor in the USC Computer Science Department. He received a PhD in computer science from Carnegie Mellon University and a bachelor's degree from Harvard University. Prof. Knight's research interests include machine translation, automata theory, and decipherment of historical manuscripts. Prof. Knight co-wrote the textbook "Artificial Intelligence", served as President of the Association for Computational Linguistics, and was a co-founder of the machine translation company Language Weaver, Inc. He is a Fellow of the Association for the Advancement of Artificial Intelligence (AAAI), the Association for Computational Linguistics (ACL), and the Information Sciences Institute (ISI).

INVITED TALK: Modeling Socio-Emotional Humanoid Agent

Catherine Pelachaud

CNRS-LTCl, TELECOM-ParisTech, France

`catherine.pelachaud@telecom-paristech.fr`

Abstract

In this talk, I will present our current work toward endowing virtual agents with socio-emotional capabilities. I will start describing an interactive system of an agent dialoging with human users in an emotionally colored manner. Through its behaviors, the agent can sustain a conversation as well as show various attitudes and levels of engagement. I will present our latest work on laughter. I will address several issues such as: how to animate laughter in a virtual agent looking particularly at rhythmic movements; how to laugh with human participant and how laughing agent is perceived.

Bio

Catherine Pelachaud is a Director of Research at CNRS in the laboratory LTCl, TELECOM ParisTech. Her research interest includes embodied conversational agent, nonverbal communication (face, gaze, and gesture), expressive behaviors and socio-emotional agents. She is associate editor of several journals among which IEEE Transactions on Affective Computing, ACM Transactions on Interactive Intelligent Systems and Journal on Multimodal User Interfaces. She has co-edited several books on virtual agents and emotion-oriented systems.

INVITED TALK: Embedding Probabilistic Logic for Machine Reading

Sebastian Riedel

University College London, UK

sebastian.riedel@gmail.com

Abstract

We want to build machines that read, and make inferences based on what was read. A long line of the work in the field has focussed on approaches where language is converted (possibly using machine learning) into a symbolic and relational representation. A reasoning algorithm (such as a theorem prover) then derives new knowledge from this representation. This allows for rich knowledge to be captured, but generally suffers from two problems: acquiring sufficient symbolic background knowledge and coping with noise and uncertainty in data. Probabilistic logics (such as Markov Logic) offer a solution, but are known to often scale poorly.

In recent years a third alternative emerged: latent variable models in which entities and relations are embedded in vector spaces (and represented "distributionally"). Such approaches scale well and are robust to noise, but they raise their own set of questions: What type of inferences do they support? What is a proof in embeddings? How can explicit background knowledge be injected into embeddings? In this talk I first present our work on latent variable models for machine reading, using ideas from matrix factorisation as well as both closed and open information extraction. Then I will present recent work we conducted to address the questions of injecting and extracting symbolic knowledge into/from models based on embeddings. In particular, I will show how one can rapidly build accurate relation extractors through combining logic and embeddings.

Bio

Dr. Riedel is a Senior Lecturer in the Department of Computer Science at University College London, leading the Machine Reading lab. He received his MSc and PhD (in 2009) in Computer Science from the University of Edinburgh. He was a researcher at the University of Tokyo, and a postdoc with Andrew McCallum at the University of Massachusetts Amherst. He is an Allen Distinguished Investigator, a Marie Curie CIG fellow, was a finalist for the Microsoft Research Faculty Award in 2013 and recently received a Google Focused Research award. Sebastian is generally interested in the intersection of NLP and machine learning, and particularly interested in teaching machines to read, and to reason with what was read.

Table of Contents

Invited Keynotes

<i>Kevin Knight</i> How Much Information Does a Human Translator Add to the Original?	xi
<i>Catherine Pelachaud</i> Modeling Socio-Emotional Humanoid Agent	xiii
<i>Sebastian Riedel</i> Embedding Probabilistic Logic for Machine Reading	xv

Regular Papers

<i>Yvonne Adesam, Gerlof Bouma and Richard Johansson</i> Defining the Eukalyptus forest – the Koala treebank of Swedish	1
<i>Malin Ahlberg, Peter Andersson, Markus Forsberg and Nina Tahmasebi</i> A case study on supervised classification of Swedish pseudo-coordination	11
<i>Héctor Martínez Alonso, Anders Johannsen, Sussi Olsen, Sanni Nimb, Nicolai Hartvig Sørensen, Anna Braasch, Anders Søgaard and Bolette Sandford Pedersen</i> Supersense tagging for Danish	21
<i>Eckhard Bick and Tino Didriksen</i> CG-3 — Beyond Classical Constraint Grammar	31
<i>Loïc Boizou, Jolanta Kovalevskaitė and Erika Rimkutė</i> Automatic Lemmatisation of Lithuanian MWEs	41
<i>Alexandr Chernov, Volha Petukhova and Dietrich Klakow</i> Linguistically Motivated Question Classification	51
<i>Jana Götze and Johan Boye</i> Resolving Spatial References using Crowdsourced Geographical Data	61
<i>Richard Johansson and Luis Nieto Piña</i> Combining Relational and Distributional Knowledge for Word Sense Disambiguation	69
<i>Peter Juel Henriksen</i> Talebob - an Interactive Speech Trainer for Danish	79
<i>Jurgita Kapočiūtė-Dzikiėnė, Ligita Šarkutė and Andrius Utkas</i> The Effect of Author Set Size in Authorship Attribution for Lithuanian	87
<i>Sigrid Klerke, Héctor Martínez Alonso and Anders Søgaard</i> Looking hard: Eye tracking for detecting grammaticality of automatically compressed sentences	97
<i>Veronika Laippala, Jenna Kanerva, Anna Missilä, Sampo Pyysalo, Tapio Salakoski and Filip Ginter</i> Towards the Classification of the Finnish Internet Parsebank: Detecting Translations and Informality	107
<i>Jostein Lien, Erik Vellidal and Liĳa Øvrelid</i> Improving cross-domain dependency parsing with dependency-derived clusters	117
<i>Mihai Lintean and Vasile Rus</i> An Optimal Quadratic Approach to Monolingual Paraphrase Alignment	127
<i>Juhani Luotolahti and Filip Ginter</i> Sentence Compression For Automatic Subtitling	135

<i>Michael Nokel and Natalia Loukachevitch</i> Topic Models: Accounting Component Structure of Bigrams	145
<i>Eva Pettersson and Joakim Nivre</i> Improving Verb Phrase Extraction from Historical Text by use of Verb Valency Frames	153
<i>Sampo Pyysalo, Jenna Kanerva, Anna Missilä, Veronika Laippala and Filip Ginter</i> Universal Dependencies for Finnish	163
<i>Robert Reynolds and Francis Tyers</i> Automatic word stress annotation of Russian unrestricted text	173
<i>Dominik Sacha, Yuki Asano, Christian Rohrdantz, Felix Hamborg, Daniel Keim, Bettina Braun and Miriam Butt</i> Self Organizing Maps for the Visual Analysis of Pitch Contours	181
<i>Jörg Tiedemann</i> Improving the Cross-Lingual Projection of Syntactic Dependencies	191
<i>Julián Zapata and Andreas Søeborg Kirkedal</i> Assessing the Performance of Automatic Speech Recognition Systems When Used by Native and Non-Native Speakers of Three Major Languages in Dictation Workflows	201
Student Papers	
<i>Max Berggren, Jussi Karlgren, Robert Östling and Mikael Parkvall</i> Inferring the location of authors from words in their texts	211
<i>Inari Listenmaa and Francis M. Tyers</i> Automatic conversion of colloquial Finnish to standard Finnish	219
<i>Vytautas Mickevičius, Tomas Krilavičius, Vaidas Morkevičius and Aušra Mackutė-Varoneckienė</i> Automatic Thematic Classification of the Titles of the Seimas Votes	225
<i>Birgitta Ojamaa, Päivi Kristiina Jokinen and Kadri Muischenk</i> Sentiment analysis on conversational texts	233
<i>Scharolta Katharina Sienčnik</i> Adapting word2vec to Named Entity Recognition	239
Short Papers	
<i>Héctor Martínez Alonso, Anders Johannsen, Barbara Plank and Anders Søgaard</i> Active learning for sense annotation	245
<i>Johan Bos and Malvina Nissim</i> Uncovering Noun-Noun Compound Relations by Gamification	251
<i>Katarina Heimann Mühlenbock, Sofie Johansson Kokkinakis, Caroline Liberg, Åsa af Geijerstam, Jenny Wiksten Folkeryd, Arne Jönsson, Erik Kanebrant and Johan Falkenjack</i> A multivariate model for classifying texts' readability	257
<i>Robert Östling, Carl Börstell and Lars Wallin</i> Enriching the Swedish Sign Language Corpus with Part of Speech Tags Using Joint Bayesian Word Alignment and Annotation Transfer	263
<i>Gustavo Henrique Paetzold</i> Using Positional Suffix Trees to Perform Agile Tree Kernel Calculation	269
<i>Ildikó Pilán</i> Helping Swedish words come to their senses: word-sense disambiguation based on sense associations from the SALDO lexicon	275

<i>Askars Salimbajevs and Jevgenijs Strigins</i> Using sub-word n-gram models for dealing with OOV in large vocabulary speech recognition for Latvian	281
<i>Steinþór Steingrímsson, Sigrún Helgadóttir and Eiríkur Rögnvaldsson</i> Analysing Inconsistencies and Errors in PoS Tagging in two Icelandic Gold Standards	287
<i>Demonstration Papers</i>	
<i>Magnus Birkenes, Lars Johnsen, Arne Lindstad and Johanne Ostad</i> From digital library to n-grams: NB N-gram	293
<i>Janne Bondi Johannessen</i> The Corpus of American Norwegian Speech (CANS)	297
<i>Johan Bos</i> Open-Domain Semantic Parsing with Boxer	301
<i>Sam Hardwick, Miikka Silfverberg and Krister Lindén</i> Extracting Semantic Frames using hfst-pmatch	305
<i>Arne Neumann</i> discoursegraphs: A graph-based merging tool and converter for multilayer annotated corpora	309
<i>Tommi A Pirinen</i> Omorfi-Free and Open Source Morphological Lexical Database of Finnish	313
<i>Evelina Rennes and Arne Jönsson</i> A Tool for Automatic Simplification of Swedish Texts	317

CONFERENCE PROGRAM

Monday, May 11, 2015 NODALIDA WORKSHOPS

- 09:00-13:00 **Constraint Grammar - Methods, Tools and Applications**
- 09:00-18:00 **4th Workshop on NLP for Computer Assisted Language Learning**
- 13:00-18:00 **Innovative Corpus Query and Visualization Tools**
- 13:00-17:30 **Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities**
- 19:30 **Welcome reception**

Tuesday, May 12, 2015 MAIN CONFERENCE

- 09:10 – 09:30 **Opening**
- 09:30 – 10:30 **Keynote**
Kevin Knight
How Much Information Does a Human Translator Add to the Original?

10:30-11:00 Coffee Break

11:00-12:30 **Regular papers**

Parallel session 1 *Syntax*

- 11:00-11:30 *Eckhard Bick and Tino Didriksen.*
CG-3 – Beyond Classical Constraint Grammar
- 11:30-12:00 *Jostein Lien, Erik Velldal and Lilja Øvrelid.*
Improving Cross-Domain Dependency Parsing with Dependency-Derived Clusters
- 12:00-12:30 *Jörg Tiedemann.*
Improving the Cross-Lingual Projection of Syntactic Dependencies

Parallel Session 2 *Annotation, Lithuanian NLP*

- 11:00-11:30 *Yvonne Adesam, Gerlof Bouma and Richard Johansson.*
Defining the Eukalyptus forest – the Koala treebank of Swedish
- 11:30-12:00 *Loïc Boizou, Jolanta Kovalevskaitė and Erika Rimkutė.*
Automatic Lemmatisation of Lithuanian MWEs

12:00-12:30 *Jurgita Kapočiūtė-Dzikienė, Ligita Šarkutė and Andrius Utkā.*
The Effect of Author Set Size in Authorship Attribution for Lithuanian

Parallel session 3 *Speech*

11:00-11:30 *Robert Reynolds and Francis Tyers.*
Automatic Word Stress Annotation of Russian Unrestricted Text

11:30-12:00 *Dominik Sacha, Yuki Asano, Christian Rohrdantz, Felix Hamborg, Daniel Keim, Bettina Braun and Miriam Butt.*
Self Organizing Maps for the Visual Analysis of Pitch Contours

12:00-12:30 *Julián Zapata and Andreas Sjøeborg Kirkedal.*
Assessing the Performance of Automatic Speech Recognition Systems When Used by Native and Non-Native Speakers of Three Major Languages in Dictation Workflows

12:30-13:30 Lunch

13:30-14:30 **Lightning talks**

Parallel Session 1 *Syntax and Semantics*

13:30-13:50 *Johan Bos and Malvina Nissim.*
Uncovering Noun-Noun Compound Relations by Gamification

13:50-14:10 *Héctor Martínez Alonso, Anders Johannsen, Barbara Plank and Anders Sjøgaard.*
Active Learning for Sense Annotation

14:10-14:30 *Ildikó Pilán.*
Helping Swedish words Come to their Senses: Word-Sense Disambiguation Based on Sense Associations from the SALDO Lexicon

Parallel Session 2 *Sign language and Speech*

13:30-13:50 *Robert Östling, Carl Börstell and Lars Wallin.*
Enriching the Swedish Sign Language Corpus with Part of Speech Tags Using Joint Bayesian Word Alignment and Annotation Transfer

13:50-14:10 *Peter Juel Henriksen.*
Talebob - an Interactive Speech Trainer for Danish

14:10-14:30 *Askars Salimbajevs and Jevgenijs Strigins.*
Using Sub-Word N-Gram Models for Dealing with OOV in Large Vocabulary Speech Recognition for Latvian

14:30-15:00 Coffee break

14:30-16:00 **Posters and demos**

Posters:

Johan Bos and Malvina Nissim.

Uncovering Noun-Noun Compound Relations by Gamification

Peter Juel Henriksen.

Talebob - an Interactive Speech Trainer for Danish

Héctor Martínez Alonso, Anders Johannsen, Barbara Plank and Anders Søgaard.

Active Learning for Sense Annotation

Ildikó Pilán.

Helping Swedish words Come to their Senses: Word-Sense Disambiguation Based on Sense Associations from the SALDO Lexicon

Askars Salimbajevs and Jevgenijs Strigins.

Using Sub-Word N-Gram Models for Dealing with OOV in Large Vocabulary Speech Recognition for Latvian

Robert Östling, Carl Börstell and Lars Wallin.

Enriching the Swedish Sign Language Corpus with Part of Speech Tags Using Joint Bayesian Word Alignment and Annotation Transfer

Demos:

Magnus Birkenes, Lars Johnsen, Arne Lindstad and Johanne Ostad.

From Digital Library to N-Grams: NB N-gram

Janne Bondi Johannessen.

The Corpus of American Norwegian Speech (CANS)

Johan Bos.

Open-Domain Semantic Parsing with Boxer

Sam Hardwick, Miikka Silfverberg and Krister Lindén.

Extracting Semantic Frames using hfst-pmatch

Arne Neumann.

discoursegraphs: A Graph-Based Merging Tool and Converter for Multilayer Annotated Corpora

Tommi A Pirinen.

Omorfi-Free and Open Source Morphological Lexical Database of Finnish

Evelina Rennes and Arne Jönsson.

A Tool for Automatic Simplification of Swedish Texts

16:00-17:00 **Keynote**
Catherine Pelachaud
Modeling socio-emotional humanoid agent

19:00 Conference Dinner

Wednesday, May 13, 2015 MAIN CONFERENCE

09:00 -10:00 **Keynote**
Sebastian Riedel
Embedding Probabilistic Logic for Machine Reading

10:00 -10:30 Coffee break

10:30-12:00 **Regular and student papers**

Parallel Session 1 *Semantics*

10:30-11:00 *Juhani Luotolahti and Filip Ginter.*
Sentence Compression For Automatic Subtitling

11:00-11:30 *Sigrid Klerke, Héctor Martínez Alonso and Anders Sjøgaard.*
Looking hard: Eye Tracking for Detecting Grammaticality of Automatically Compressed Sentences

11:30-12:00 *Richard Johansson and Luis Nieto Piña.*
Combining Relational and Distributional Knowledge for Word Sense Disambiguation

Parallel Session 2 *Syntax and Semantics*

10:30-11:00 *Malin Ahlberg, Peter Andersson, Markus Forsberg and Nina Tahmasebi.*
A case study on supervised classification of Swedish pseudo-coordination

11:00-11:30 *Eva Pettersson and Joakim Nivre.*
Improving Verb Phrase Extraction from Historical Text by use of Verb Valency Frames

11:30-12:00 *Sampo Pyysalo, Jenna Kanerva, Anna Missilä, Veronika Laippala and Filip Ginter.*
Universal Dependencies for Finnish

Parallel Session 3 *Student Session*

10:30-11:00 *Max Berggren, Jussi Karlgren, Robert Östling and Mikael Parkvall.*
Inferring the Location of Authors from Words in their Texts

11:00-11:30	<i>Inari Listenmaa and Francis M. Tyers.</i> Automatic Conversion of Colloquial Finnish to Standard Finnish
11:30-12:00	<i>Scharolta Sienčnik.</i> Adapting word2vec to Named Entity Recognition
12:00-13:00	Lunch
13:00-14:00	Business meeting
14:00-14.30	Coffee break
14:00-15:30	Poster session

Long papers:

Héctor Martínez Alonso, Anders Johannsen, Sussi Olsen, Sanni Nimb, Nicolai Hartvig Sørensen, Anna Braasch, Anders Sjøgaard and Bolette Sandford Pedersen.
Supersense tagging for Danish

Alexandr Chernov, Volha Petukhova and Dietrich Klakow.
Linguistically Motivated Question Classification

Jana Götze and Johan Boye.
Resolving Spatial References using Crowdsourced Geographical Data

Veronika Laippala, Jenna Kanerva, Anna Missilä, Sampo Pyysalo, Tapio Salakoski and Filip Ginter.
Towards the Classification of the Finnish Internet Parsebank: Detecting Translations and Informality

Mihai Lintean and Vasile Rus.
An Optimal Quadratic Approach to Monolingual Paraphrase Alignment

Michael Nokel and Natalia Loukachevitch.
Topic Models: Accounting Component Structure of Bigrams

Short papers:

Katarina Heimann Mühlenbock, Sofie Johansson Kokkinakis, Caroline Liberg, Åsa af Geijerstam, Jenny Folkeryd, Arne Jönsson, Erik Kanebrant and Johan Falkenjack.
A Multivariate Model for Classifying Texts' Readability

Gustavo Henrique Paetzold.
Using Positional Suffix Trees to Perform Agile Tree Kernel Calculation

Steinþór Steingrímsson, Sigrún Helgadóttir and Eiríkur Rögnvaldsson.
Analysing Inconsistencies and Errors in PoS Tagging in two Icelandic Gold Standards

Student papers:

Vytautas Mickevičius, Tomas Krilavičius Vaidas Morkevičius and Aušra Mackutė-Varoneckienė.

Automatic Thematic Classification of the Titles of the Seimas Votes

Birgitta Ojamaa, Päivi Kristiina Jokinen and Kadri Muischenk.

Sentiment Analysis on Conversational Texts

15:30-16:30

Tutorial on Lithuanian

Algirdas Saugdargas.

Lithuanian Language in the Architecture of Knowledge

16:30-16:40

Closing