

# NEALT

Northern European Association for  
Language Technology

NEALT Proceedings Series Vol. 29

Proceedings of the  
21st Nordic Conference on Computational Linguistics

A nighttime photograph of a harbor, likely in Gothenburg, Sweden. In the foreground, two large gantry cranes are silhouetted against the dark sky. The water reflects the lights from the city and the cranes. In the background, a suspension bridge with two tall towers and a city skyline with various buildings and lights are visible. The overall scene is illuminated by a mix of warm city lights and cooler blue tones from the night sky and water.

**NoDaLiDa**

22-24 MAY 2017  
GOTHENBURG SWEDEN  
CONFERENCE CENTRE WALLENBERG

Editor: Jörg Tiedemann

NoDaLiDa 2017

**21st Nordic Conference of  
Computational Linguistics**

**Proceedings of the Conference**

23 – 24 May 2017  
Wallenberg Conference Center  
Gothenburg, Sweden

Cover photo: Kjell Holmner/Göteborg & Co

©2017 Linköping University Electronic Press

The proceedings are published by:

Linköping University Electronic Press  
Linköping Electronic Conference Proceedings No. 131  
<http://www.ep.liu.se/ecp/contents.asp?issue=131>

ACL Anthology  
<http://aclweb.org/anthology/W/W17/#0200>

ISSN: 1650-3686  
eISSN: 1650-3740  
ISBN: 978-91-7685-601-7



RIKSBANKENS  
JUBILEUMSFOND

STIFTELSEN FÖR HUMANISTISK OCH  
SAMHÄLLSVETENSKAPLIG FORSKNING

Språk-  
BANKEN

KIELIPANKKI  
The Language Bank of Finland

Lingsoft®  
LANGUAGE  
MANAGEMENT



FINDWISE

MLT



CLASP centre for  
linguistic theory  
and studies in probability

digital Grammars  
Language technology to rely on.

## Introduction

On behalf of the program committee, I am pleased to welcome you to the 21st Nordic Conference on Computational Linguistics (NoDaLiDa 2017), held at the Wallenberg Conference Center in the beautiful city of Gothenburg in Sweden, on May 22–24, 2017. The proceedings are published as part of the NEALT Proceedings Series by Linköping University Electronic Press and they will also be available from the ACL Anthology together with the proceedings of the co-located workshops. The NoDaLiDa conference has been organized bi-annually since 1977 and returns to this anniversary event after 40 years back to Gothenburg, where it started as a friendly gathering to discuss on-going research in the field of computational linguistics in the Nordic countries. The Northern European Association for Language Technology (NEALT) was founded later in 2006, which is now responsible for organizing NoDaLiDa among other events in the Nordic countries, the Baltic states and Northwest Russia. Since the early days, NoDaLiDa has grown into a recognized international conference and the tradition continues with the program of this year’s conference. It is a great honor for me to serve as the general chair of NoDaLiDa 2017 and I am grateful for all the support during the progress.

Before diving deeper into the acknowledgements, please, let me first briefly introduce the setup of the conference. Similar to the last edition, we included different paper categories to be presented: Regular long papers, student papers, short papers and system demonstration papers. Regular papers are presented orally during the conference and short papers received a slot in one of the two poster sessions. System demonstrations are given at the same time as the posters. We selected two student papers for an oral presentation and three student papers for poster presentations. In total, we received 78 submissions and accepted 49. The submissions included 32 regular papers (21 accepted, 65.6% acceptance rate), 8 student papers (5 accepted, 62.5% acceptance rate), 27 short papers (12 accepted, 44.4% acceptance rate) and 11 system demonstration papers (which we accepted all).

In addition to the submitted papers, NoDaLiDa 2017 also features invited keynote speakers – three distinguished international researchers: Kyunghyun Cho from New York University, Sharon Goldwater from the University of Edinburgh and Rada Mihalcea from the University of Michigan. We are excited about their contributions and grateful for their participation in the conference.

Furthermore, four workshops are connected to NoDaLiDa 2017: The First Workshop on Universal Dependencies (UDW 2017), the Joint 6th Workshop on NLP for CALL and 2nd Workshop on NLP for Research on Language Acquisition (NLP4CALL & LA), the Workshop on Processing Historical Language and the Workshop on Constraint Grammar - Methods, Tools, and Applications. We would like to thank the workshop organizers for their efforts in making these events happen enriching the whole conference and its scientific coverage.

Finally, I would also like to thank the entire team behind the conference. Organizing such an event is a complex process and would not be possible without the help of many people. I would like to thank all members of the program committee, especially Beáta Megyesi for a smooth transition from the previous NoDaLiDa and all her valuable input coming from the organization of that event, Inguna Skadiņa for taking care of the submission system EasyChair, Lilja Øvrelid for the publicity and calls that we need to send out. My greatest relieve from organisational pain came from the professional local committee in Gothenburg. It is a pleasure to work together with their team and without the hard work of the local organizers we could not run the event in any way. Thank you very much and especially thanks to Nina Tahmasebi for leading the local team. All names are properly listed below and I am grateful to all of you and your efforts. I would also like to acknowledge the large number of reviewers and sub-reviewers for their assessment of the submissions, NEALT for backing up the conference, Linköping University Press for publishing the proceedings as well as people behind the ACL Anthology for offering the space for storing our publications. And, last but not least, I would also thank our sponsors for all the financial

support, which really helped us to organize a pleasant and affordable meeting.

With all of these acknowledgments, and with my apologies for forgetting to mention many names that should be listed here, I would like to wish you all, once again, a fruitful conference and a nice stay in Gothenburg. And I wish you a lot of pleasure with reading the contributions in this volume especially if you, for whatever reason, happen to read this welcome address after the conference has already ended.

*Jörg Tiedemann (general chair of NoDaLiDa 2017)*

## **Welcome Message by the Local Organizers**

40 years ago, the first NoDaLiDa conference took place here in Gothenburg, arranged by what is now Språkbanken, the Swedish Language Bank. The aim of the conference was to bring together researchers from the five Nordic countries to discuss all aspects of language technology. Experiences and results with respect to the Nordic languages would benefit from being shared, compared and discussed. The community has grown and expanded since; today, NoDaLiDa brings together researchers and practitioners from 23 countries and a diverse field of studies. The first meeting attracted over 60 people and the current installment has more than doubled that amount. Though the field has developed, many of the topics raised in 1977 are still highly relevant and interesting today. For example, the creation of lexical resources, sense disambiguation and syntactic analysis, and analysis of large amounts of text. To celebrate the 40th anniversary of NoDaLiDa, we are happy to welcome you back to Gothenburg and this 21st edition of the conference. We hope that Gothenburg will show its most beautiful side, offering you sunshine, vast views of the sea and good food. We hope that you will enjoy interesting talks, posters and workshops during these three days of NoDaLiDa, which will provide inspiration in the 40 years to come. Welcome!

*Nina Tahmasebi*

*Yvonne Adesam*

*Martin Kaså*

*on behalf of the local organizers*

**General Chair:**

Jörg Tiedemann, University of Helsinki, Finland

**Program Committee:**

Jón Guðnason, Reykjavik University, Iceland  
Beáta Megyesi, Uppsala University, Sweden  
Kadri Muischnek, University of Tartu, Estonia  
Inguna Skadiņa, University of Latvia, Latvia  
Anders Søgaard, University of Copenhagen, Denmark  
Andrius Utka, Vytautas Magnus University, Lithuania  
Lilja Øvrelid, University of Oslo, Norway

**Organizing Committee:**

Nina Tahmasebi (chair), University of Gothenburg, Sweden  
Yvonne Adesam (co-chair), University of Gothenburg, Sweden  
Martin Kaså (co-chair), University of Gothenburg, Sweden  
Sven Lindström (publication chair), University of Gothenburg, Sweden  
Elena Volodina (sponsor chair), University of Gothenburg, Sweden  
Lars Borin, University of Gothenburg, Sweden  
Dana Dannélls, University of Gothenburg, Sweden  
Ildikó Pilán, University of Gothenburg, Sweden

**Invited Speaker:**

Rada Mihalcea, University of Michigan  
Kyunghyun Cho, New York University  
Sharon Goldwater, University of Edinburgh

## Reviewers:

Lars Ahrenberg, Linköping University  
Tanel Alumäe, Tallinn University of Technology  
Isabelle Augenstein, University College London  
Ilze Auziņa, University of Latvia  
Joachim Bingel, University of Copenhagen  
Johannes Bjerva, University of Groningen  
Bernd Bohnet, Google  
Chloé Braud, University of Copenhagen  
Lin Chen, UIC  
Koenraad De Smedt, University of Bergen  
Rodolfo Delmonte, Università Ca' Foscari  
Leon Derczynski, University of Sheffield  
Jens Edlund, Royal Technical Institute (KTH), Stockholm  
Murhaf Fares, University of Oslo  
Björn Gambäck, Norwegian University of Science and Technology  
Filip Ginter, University of Turku  
Gintarė Grigonytė, University of Stockholm  
Normunds Grūzītis, University of Latvia  
Angelina Ivanova, University of Oslo  
Richard Johansson, University of Gothenburg  
Sofie Johansson, Institutionen för svenska språket  
Arne Jönsson, Linöping University  
Heiki-Jaan Kaalep, University of Tartu  
Jussi Karlgren, Gavagai & KTH Stockholm  
Sigrid Klerke, University of Copenhagen  
Mare Koit, University of Tartu  
Andrey Kutuzov, University of Oslo  
Ophélie Lacroix, University of Copenhagen  
Emanuele Lapponi, University of Oslo  
Krister Lindén, University of Helsinki  
Pierre Lison, Norwegian Computing Center  
Kaili Müürisep, University of Tartu  
Costanza Navarretta, University of Copenhagen  
Joakim Nivre, Uppsala University  
Pierre Nugues, Lund University  
Stephan Oepen, Universitetet i Oslo  
Heili Orav, University of Tartu  
Patrizia Paggio, University of Copenhagen  
Eva Pettersson, Uppsala University  
Mārcis Pinnis, Tilde  
Tommi Pirinen, Universität Hamburg  
Fabio Rinaldi, IFI, University of Zurich  
Eiríkur Rögnvaldsson, University of Iceland  
Marina Santini, SICS East ICT  
Kairit Sirts, Tallinn University of Technology  
Raivis Skadiņš, Tilde  
Sara Stymne, Uppsala University  
Torbjørn Svendsen, Norwegian University of Science and Technology  
Andrius Utkas, Vytautas Magnus University  
Erik Velldal, University of Oslo



Sumithra Velupillai, KTH Royal Institute of Technology, Stockholm  
Martin Volk, University of Zurich  
Jürgen Wedekind, University of Copenhagen  
Mats Wirén, Stockholm University  
Roman Yangarber, University of Helsinki  
Gustaf Öqvist Seimyr, Karolinska Institutet  
Robert Östling, Stockholm University

**Additional Reviewers:**

Amrhein, Chantal  
Fahlborg, Daniel  
Klang, Marcus  
Mascarell, Laura  
Mitchell, Jeff  
Pretkalniņa, Lauma  
Vare, Kadri  
Waseem, Zeerak



# Table of Contents

## Regular Papers

<i>Joint UD Parsing of Norwegian Bokmål and Nynorsk</i> Erik Velldal, Lilja Øvrelid and Petter Hohle .....	1
<i>Replacing OOV Words For Dependency Parsing With Distributional Semantics</i> Prasanth Kolachina, Martin Riedl and Chris Biemann .....	11
<i>Real-valued Syntactic Word Vectors (RSV) for Greedy Neural Dependency Parsing</i> Ali Basirat and Joakim Nivre .....	20
<i>Tagging Named Entities in 19th Century and Modern Finnish Newspaper Material with a Finnish Semantic Tagger</i> Kimmo Kettunen and Laura Löfberg .....	29
<i>Machine Learning for Rhetorical Figure Detection: More Chiasmus with Less Annotation</i> Marie Dubremetz and Joakim Nivre .....	37
<i>Coreference Resolution for Swedish and German using Distant Supervision</i> Alexander Wallin and Pierre Nugues .....	46
<i>Aligning phonemes using finite-state methods</i> Kimmo Koskenniemi .....	56
<i>Twitter Topic Modeling by Tweet Aggregation</i> Asbjørn Steinskog, Jonas Therkelsen and Björn Gambäck .....	77
<i>A Multilingual Entity Linker Using PageRank and Semantic Graphs</i> Anton Södergren and Pierre Nugues .....	87
<i>Linear Ensembles of Word Embedding Models</i> Avo Muromägi, Kairit Sirts and Sven Laur .....	96
<i>Using Pseudowords for Algorithm Comparison: An Evaluation Framework for Graph-based Word Sense Induction</i> Flavio Massimiliano Cecchini, Chris Biemann and Martin Riedl .....	105
<i>North-Sámi to Finnish rule-based machine translation system</i> Tommi Pirinen, Francis M. Tyers, Trond Trosterud, Ryan Johnson, Kevin Unhammer and Tiina Puolakainen .....	115
<i>Machine translation with North Saami as a pivot language</i> Lene Antonsen, Ciprian Gerstenberger, Maja Kappfjell, Sandra Nystø Ráhka, Marja-Liisa Olthuis, Trond Trosterud and Francis M. Tyers .....	123
<i>SWEGRAM –A Web-Based Tool for Automatic Annotation and Analysis of Swedish Texts</i> Jesper Näsman, Beáta Megyesi and Anne Palmér .....	132
<i>Optimizing a PoS Tagset for Norwegian Dependency Parsing</i> Petter Hohle, Lilja Øvrelid and Erik Velldal .....	142

<i>Creating register sub-corpora for the Finnish Internet Parsebank</i> Veronika Laipala, Juhani Luotolahti, Aki-Juhani Kyröläinen, Tapio Salakoski and Filip Ginter	152
<i>KILLE: a Framework for Situated Agents for Learning Language Through Interaction</i> Simon Dobnik and Erik de Graaf	162
<i>Data Collection from Persons with Mild Forms of Cognitive Impairment and Healthy Controls - Infrastructure for Classification and Prediction of Dementia</i> Dimitrios Kokkinakis, Kristina Lundholm Fors, Eva Björkner and Arto Nordlund	172
<i>Evaluation of language identification methods using 285 languages</i> Tommi Jauhiainen, Krister Lindén and Heidi Jauhiainen	183
<i>Can We Create a Tool for General Domain Event Analysis?</i> Siim Orasmaa and Heiki-Jaan Kaalep	192
<i>From Treebank to Propbank: A Semantic-Role and VerbNet Corpus for Danish</i> Eckhard Bick	202

## Student Papers

<i>Acoustic Model Compression with MAP adaptation</i> Katri Leino and Mikko Kurimo	65
<i>OCR and post-correction of historical Finnish texts</i> Senka Drobac, Pekka Kauppinen and Krister Lindén	70
<i>Mainstreaming August Strindberg with Text Normalization</i> Adam Ek and Sofia Knuutinen	266
<i>Improving Optical Character Recognition of Finnish Historical Newspapers with a Combination of Fraktur &amp; Antiqua Models and Image Preprocessing</i> Mika Koistinen, Kimmo Kettunen and Tuula Pääkkönen	277
<i>The Effect of Excluding Out of Domain Training Data from Supervised Named-Entity Recognition</i> Adam Persson	289

## Short Papers

<i>Cross-lingual Learning of Semantic Textual Similarity with Multilingual Word Representations</i> Johannes Bjerva and Robert Östling	211
<i>Will my auxiliary tagging task help? Estimating Auxiliary Tasks Effectivity in Multi-Task Learning</i> Johannes Bjerva	216
<i>Iconic Locations in Swedish Sign Language: Mapping Form to Meaning with Lexical Databases</i> Carl Börstell and Robert Östling	221
<i>Docforia: A Multilayer Document Model</i> Marcus Klang and Pierre Nugues	226

<i>Finnish resources for evaluating language model semantics</i> Viljami Venekoski and Jouko Vankka .....	231
<i>Málrómur: A Manually Verified Corpus of Recorded Icelandic Speech</i> Steinþór Steingrímsson, Jón Guðnason, Sigrún Helgadóttir and Eiríkur Rögnvaldsson .....	237
<i>The Effect of Translationese on Tuning for Statistical Machine Translation</i> Sara Stymne .....	241
<i>Word vectors, reuse, and replicability: Towards a community repository of large-text resources</i> Murhaf Fares, Andrey Kutuzov, Stephan Oepen and Erik Velldal .....	271
<i>Redefining Context Windows for Word Embedding Models: An Experimental Study</i> Pierre Lison and Andrey Kutuzov .....	284
<i>Quote Extraction and Attribution from Norwegian Newspapers</i> Andrew Salway, Paul Meurer, Knut Hofland and Øystein Reigem .....	293
<i>Wordnet extension via word embeddings: Experiments on the Norwegian Wordnet</i> Heidi Sand, Erik Velldal and Lilja Øvrelid .....	298
<i>Universal Dependencies for Swedish Sign Language</i> Robert Östling, Carl Börstell, Moa Gärdenfors and Mats Wirén .....	303

## **System Demonstration Papers**

<i>Multilingwis2 –Explore Your Parallel Corpus</i> Johannes Graën, Dominique Sandoz and Martin Volk .....	247
<i>A modernised version of the Glossa corpus search system</i> Anders Nøklestad, Kristin Hagen, Janne Bondi Johannessen, Michał Kosek and Joel Priestley .....	251
<i>Dep_search: Efficient Search Tool for Large Dependency Parsebanks</i> Juhani Luotolahti, Jenna Kanerva and Filip Ginter .....	255
<i>Proto-Indo-European Lexicon: The Generative Etymological Dictionary of Indo-European Languages</i> Jouna Pyysalo .....	259
<i>Tilde MODEL - Multilingual Open Data for EU Languages</i> Roberts Rozis and Raivis Skadiņš .....	263
<i>Services for text simplification and analysis</i> Johan Falkenjack, Evelina Rennes, Daniel Fahlborg, Vida Johansson and Arne Jönsson .....	309
<i>Exploring Properties of Intralingual and Interlingual Association Measures Visually</i> Johannes Graën and Christof Bless .....	314
<i>TALERUM - Learning Danish by Doing Danish</i> Peter Juel Henriksen .....	318
<i>Cross-Lingual Syntax: Relating Grammatical Framework with Universal Dependencies</i> Aarne Ranta, Prasanth Kolachina and Thomas Hallgren .....	322

<i>Exploring Treebanks with INESS Search</i>	
Victoria Rosén, Helge Dyvik, Paul Meurer and Koenraad De Smedt .....	326
<i>A System for Identifying and Exploring Text Repetition in Large Historical Document Corpora</i>	
Aleksi Vesanto, Filip Ginter, Hannu Salmi, Asko Nivala and Tapio Salakoski .....	330

# Conference Program

**Tuesday, May 23, 2017**

9:10–9:30      Opening Remarks

9:30–10:30    Invited Talk by Kyunghyun Cho

## **Session 1: Syntax**

11:00–11:30    *Joint UD Parsing of Norwegian Bokmål and Nynorsk*  
Erik Velldal, Lilja Øvrelid and Petter Hohle

11:30–12:00    *Replacing OOV Words For Dependency Parsing With Distributional Semantics*  
Prasanth Kolachina, Martin Riedl and Chris Biemann

12:00–12:30    *Real-valued Syntactic Word Vectors (RSV) for Greedy Neural Dependency Parsing*  
Ali Basirat and Joakim Nivre

## **Session 2: Semantics and Discourse**

11:00–11:30    *Tagging Named Entities in 19th Century and Modern Finnish Newspaper Material with a Finnish Semantic Tagger*  
Kimmo Kettunen and Laura Löfberg

11:30–12:00    *Machine Learning for Rhetorical Figure Detection: More Chiasmus with Less Annotation*  
Marie Dubremetz and Joakim Nivre

12:00–12:30    *Coreference Resolution for Swedish and German using Distant Supervision*  
Alexander Wallin and Pierre Nugues

**Tuesday, May 23, 2017 (continued)**

**Session 3: Mixed Topics**

- 11:00–11:30 *Aligning phonemes using finite-state methods*  
Kimmo Koskenniemi
- 11:30–12:00 *Acoustic Model Compression with MAP adaptation*  
Katri Leino and Mikko Kurimo
- 12:00–12:30 *OCR and post-correction of historical Finnish texts*  
Senka Drobac, Pekka Kauppinen and Krister Lindén
- 12:30–13:30 Lunch

**Session 4: Social Media**

- 13:30–14:00 *Twitter Topic Modeling by Tweet Aggregation*  
Asbjørn Steinskog, Jonas Therkelsen and Björn Gambäck
- 14:00–14:30 *A Multilingual Entity Linker Using PageRank and Semantic Graphs*  
Anton Södergren and Pierre Nugues

**Session 5: Lexical Semantics**

- 13:30–14:00 *Linear Ensembles of Word Embedding Models*  
Avo Muromägi, Kairit Sirts and Sven Laur
- 14:00–14:30 *Using Pseudowords for Algorithm Comparison: An Evaluation Framework for Graph-based Word Sense Induction*  
Flavio Massimiliano Cecchini, Chris Biemann and Martin Riedl



**Tuesday, May 23, 2017 (continued)**

**Session 6: Machine Translation**

- 13:30–14:00 *North-Sámi to Finnish rule-based machine translation system*  
Tommi Pirinen, Francis M. Tyers, Trond Trosterud, Ryan Johnson, Kevin Unhammer and Tiina Puolakainen
- 14:00–14:30 *Machine translation with North Saami as a pivot language*  
Lene Antonsen, Ciprian Gerstenberger, Maja Kappfjell, Sandra Nystø Ráhka, Marja-Liisa Olthuis, Trond Trosterud and Francis M. Tyers
- 14:30–15:00 Lightning Talks
- 15:30–16:30 Posters and System Demonstrations
- 16:30–17:15 Tutorial on Swedish, Maia Andreasson
- 17:15–18:00 SIG Meetings

**Wednesday, May 24, 2017**

- 9:00–10:00 Invited Talk by Sharon Goldwater
- 10:00–10:30 Lightning Talks
- 10:30–11:30 Posters and System Demonstrations
- 11:30–12:30 NEALT Business Meeting
- 12:30–13:30 Lunch

**Wednesday, May 24, 2017 (continued)**

**Session 7: Syntax**

- 13:30–14:00 *SWEGRAM –A Web-Based Tool for Automatic Annotation and Analysis of Swedish Texts*  
Jesper Näsman, Beáta Megyesi and Anne Palmér
- 14:00–14:30 *Optimizing a PoS Tagset for Norwegian Dependency Parsing*  
Petter Hohle, Lilja Øvrelid and Erik Velldal
- 14:30–15:00 *Creating register sub-corpora for the Finnish Internet Parsebank*  
Veronika Laippala, Juhani Luotolahti, Aki-Juhani Kyröläinen, Tapio Salakoski and Filip Ginter

**Session 8: NLP Applications**

- 13:30–14:00 *KILLE: a Framework for Situated Agents for Learning Language Through Interaction*  
Simon Dobnik and Erik de Graaf
- 14:00–14:30 *Data Collection from Persons with Mild Forms of Cognitive Impairment and Healthy Controls - Infrastructure for Classification and Prediction of Dementia*  
Dimitrios Kokkinakis, Kristina Lundholm Fors, Eva Björkner and Arto Nordlund
- 14:30–15:00 *Evaluation of language identification methods using 285 languages*  
Tommi Jauhiainen, Krister Lindén and Heidi Jauhiainen

**Session 9: NLP Applications**

- 13:30–14:00 *Can We Create a Tool for General Domain Event Analysis?*  
Siim Orasmaa and Heiki-Jaan Kaalep
- 14:00–14:30 *From Treebank to Propbank: A Semantic-Role and VerbNet Corpus for Danish*  
Eckhard Bick
- 15:30–16:30 Invited Talk by Rada Mihalcea
- 16:30–17:00 Best Paper Awards and Closing Remarks

**Tuesday, May 23, 2017**

**Poster Session 1**

- 15:30–16:30 *Cross-lingual Learning of Semantic Textual Similarity with Multilingual Word Representations*  
Johannes Bjerva and Robert Östling
- 15:30–16:30 *Will my auxiliary tagging task help? Estimating Auxiliary Tasks Effectivity in Multi-Task Learning*  
Johannes Bjerva
- 15:30–16:30 *Iconic Locations in Swedish Sign Language: Mapping Form to Meaning with Lexical Databases*  
Carl Börstell and Robert Östling
- 15:30–16:30 *Docforia: A Multilayer Document Model*  
Marcus Klang and Pierre Nugues
- 15:30–16:30 *Finnish resources for evaluating language model semantics*  
Viljami Venekoski and Jouko Vankka
- 15:30–16:30 *Málrómur: A Manually Verified Corpus of Recorded Icelandic Speech*  
Steinþór Steingrímsson, Jón Guðnason, Sigrún Helgadóttir and Eiríkur Rögnvaldsson
- 15:30–16:30 *The Effect of Translationese on Tuning for Statistical Machine Translation*  
Sara Stymne

**System Demonstrations 1**

- 15:30–16:30 *Multilingwis2 –Explore Your Parallel Corpus*  
Johannes Graën, Dominique Sandoz and Martin Volk
- 15:30–16:30 *A modernised version of the Glossa corpus search system*  
Anders Nøklestad, Kristin Hagen, Janne Bondi Johannessen, Michał Kosek and Joel Priestley
- 15:30–16:30 *Dep\_search: Efficient Search Tool for Large Dependency Parsebanks*  
Juhani Luotolahti, Jenna Kanerva and Filip Ginter

**Tuesday, May 23, 2017 (continued)**

- 15:30–16:30 *Proto-Indo-European Lexicon: The Generative Etymological Dictionary of Indo-European Languages*  
Jouna Pyysalo
- 15:30–16:30 *Tilde MODEL - Multilingual Open Data for EU Languages*  
Roberts Rozis and Raivis Skadiņš

**Wednesday, May 24, 2017**

**Poster Session 2**

- 10:30–11:30 *Mainstreaming August Strindberg with Text Normalization*  
Adam Ek and Sofia Knuutinen
- 10:30–11:30 *Word vectors, reuse, and replicability: Towards a community repository of large-text resources*  
Murhaf Fares, Andrey Kutuzov, Stephan Oepen and Erik Velldal
- 10:30–11:30 *Improving Optical Character Recognition of Finnish Historical Newspapers with a Combination of Fraktur & Antiqua Models and Image Preprocessing*  
Mika Koistinen, Kimmo Kettunen and Tuula Pääkkönen
- 10:30–11:30 *Redefining Context Windows for Word Embedding Models: An Experimental Study*  
Pierre Lison and Andrey Kutuzov
- 10:30–11:30 *The Effect of Excluding Out of Domain Training Data from Supervised Named-Entity Recognition*  
Adam Persson
- 10:30–11:30 *Quote Extraction and Attribution from Norwegian Newspapers*  
Andrew Salway, Paul Meurer, Knut Hofland and Øystein Reigem
- 10:30–11:30 *Wordnet extension via word embeddings: Experiments on the Norwegian Wordnet*  
Heidi Sand, Erik Velldal and Lilja Øvrelid
- 10:30–11:30 *Universal Dependencies for Swedish Sign Language*  
Robert Östling, Carl Börstell, Moa Gärdenfors and Mats Wirén

Wednesday, May 24, 2017 (continued)

**System Demonstrations 2**

- 10:30–11:30 *Services for text simplification and analysis*  
Johan Falkenjack, Evelina Rennes, Daniel Fahlborg, Vida Johansson and Arne Jönsson
- 10:30–11:30 *Exploring Properties of Intralingual and Interlingual Association Measures Visually*  
Johannes Graën and Christof Bless
- 10:30–11:30 *TALERUM - Learning Danish by Doing Danish*  
Peter Juel Henriksen
- 10:30–11:30 *Cross-Lingual Syntax: Relating Grammatical Framework with Universal Dependencies*  
Aarne Ranta, Prasanth Kolachina and Thomas Hallgren
- 10:30–11:30 *Exploring Treebanks with INESS Search*  
Victoria Rosén, Helge Dyvik, Paul Meurer and Koenraad De Smedt
- 10:30–11:30 *A System for Identifying and Exploring Text Repetition in Large Historical Document Corpora*  
Aleksi Vesanto, Filip Ginter, Hannu Salmi, Asko Nivala and Tapio Salakoski

