22nd Nordic Conference of Computational Linguistics (NoDaLiDa 2019)

NEALT Proceedings Series Volume 42

Turku, Finland
30 September – 2 October 2019

Editors:

Mareike Hartmann
Barbara Plank

ISBN: 978-1-5108-9444-0

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2019) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2019)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Long Papers

Comparison between NMT and PBSMT Performance for Translating Noisy User-Generated Content 2
José Carlos Rosales Nuñez, Djamé Seddah and Guillaume Wisniewski

Bootstrapping UD treebanks for Delexicalized Parsing 15
Prasanth Kolachina and Aarne Ranta

Lexical Resources for Low-Resource PoS Tagging in Neural Times 25
Barbara Plank and Sigrid Klerke

Gender Bias in Pretrained Swedish Embeddings 35
Magnus Sahlgren and Fredrik Olsson

A larger-scale evaluation resource of terms and their shift direction for diachronic lexical semantics 44
Astrid van Aggelen, Antske Fokkens, Laura Hollink and Jacco van Ossenbruggen

Some steps towards the generation of diachronic WordNets 55
Yuri Bizzoni, Marius Mosbach, Dietrich Klakow and Stefania Degaetano-Ortlieb

An evaluation of Czech word embeddings 65
Karolína Hořeňovská

Language Modeling with Syntactic and Semantic Representation for Sentence Acceptability Predictions 76
Adam Ek, Jean-Philippe Bernardy and Shalom Lappin

Comparing linear and neural models for competitive MWE identification 86
Hazem Al Saied, Marie Candito and Mathieu Constant

Syntax-based identification of light-verb constructions 97
Silvio Ricardo Cordeiro and Marie Candito

Comparing the Performance of Feature Representations for the Categorization of the Easy-to-Read Variety vs Standard Language 105
Marina Santini, Benjamin Danielsson and Arne Jönsson

Unsupervised Inference of Object Affordance from Text Corpora 115
Michele Persiani and Thomas Hellström

Annotating evaluative sentences for sentiment analysis: a dataset for Norwegian 121
Petter Mæhlum, Jeremy Barnes, Lilja Øvrelid and Erik Velldal

An Unsupervised Query Rewriting Approach Using N-gram Co-occurrence Statistics to Find Similar Phrases in Large Text Corpora 131
Hans Moen, Laura-Maria Peltonen, Henry Suhonen, Hanna-Maria Matinolli, Riitta Mieronkoski, Kirsi Telen, Kirsi Terho, Tapio Salakoski and Sanna Salanterä

Compiling and Filtering ParIce: An English-Icelandic Parallel Corpus 140
Starkaður Barkarson and Steinþór Steingrímsson

DIM: The Database of Icelandic Morphology 146

Kristín Bjarnadóttir, Kristín Ingibjörg Hlynsdóttir and Steinþór Steingrímsson

Tools for supporting language learning for Sakha 155
Sardana Ivanova, Anisia Katinskaia and Roman Yangarber

Inferring morphological rules from small examples using 0/1 linear programming 164
Ann Lilliestr̈om, Koen Claessen and Nicholas Smallbone

Lexicon information in neural sentiment analysis: a multi-task learning approach 175
Jeremy Barnes, Samia Touileb, Lilja Øvrelid and Erik Velldal

Aspect-Based Sentiment Analysis using BERT 187
Mickel Hoang, Oskar Alija Bihorac and Jacobo Rouces

Political Stance in Danish 197
Rasmus Lehmann and Leon Derczynski

Joint Rumour Stance and Veracity Prediction 208
Anders Edelbo Lillie, Emil Refsgaard Middelboe and Leon Derczynski

Named-Entity Recognition for Norwegian 222
Bjarte Johansen

Projecting named entity recognizers without annotated or parallel corpora 232
Jue Hou, Maximilian Koppatz, José María Hoya Quecedo and Roman Yangarber

Template-free Data-to-Text Generation of Finnish Sports News 242
Jenna Kanerva, Samuel Rönnqvist, Riina Kekki, Tapio Salakoski and Filip Ginter

Matching Keys and Encrypted Manuscripts 253
Eva Pettersson and Beata Megyesi

Perceptual and acoustic analysis of voice similarities between parents and young children 262
Evgeniia Rykova and Stefan Werner

Enhancing Natural Language Understanding through Cross-Modal Interaction: Meaning Recovery from Acoustically Noisy Speech 272
Ozge Alacam

Predicting Prosodic Prominence from Text with Pre-trained Contextualized Word Representations 281
Aarne Talman, Antti Suni, Hande Celikkanat, Sofoklis Kakouros, Jörg Tiedemann and Martti Vainio

Short Papers

Toward Multilingual Identification of Online Registers 292
Veronika Laippala, Roosa Kyllönen, Jesse Egbert, Douglas Biber and Sampo Pyysalo

A Wide-Coverage Symbolic Natural Language Inference System 298
Stergios Chatzikyriakidis and Jean-Philippe Bernardy

Ensembles of Neural Morphological Inflection Models 304
Ilmari Kylliäinen and Miikka Silfverberg

Nefnir: A high accuracy lemmatizer for Icelandic 310
Svanhvít Lilja Ingólfsdóttir, Hrafn Loftsson, Jón Friðrik Daðason and Kristín Bjarnadóttir

Natural Language Processing in Policy Evaluation: Extracting Policy Conditions from IMF Loan Agreements . . . 316
Joakim Åkerström, Adel Daoud and Richard Johansson

Interconnecting lexical resources and word alignment: How do learners get on with particle verbs? 321
David Alfter and Johannes Graën

May I Check Again? —A simple but efficient way to generate and use contextual dictionaries for Named Entity Recognition. Application to French Legal Texts. . . . 327
Valentin Barriere and Amaury Fouret

Predicates as Boxes in Bayesian Semantics for Natural Language . . . 333
Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, Shalom Lappin and Aleksandre Maskharashvili

Bornholmsk Natural Language Processing: Resources and Tools . . . 338
Leon Derczynski and Alex Speed Kjeldsen

Morphosyntactic Disambiguation in an Endangered Language Setting . . . 345
Jeff Ens, Mika Hämäläinen, Jack Rueter and Philippe Pasquier

Tagging a Norwegian Dialect Corpus . . . 350
Andre Kåsen, Anders Nøklestad, Kristin Hagen and Joel Priestley

The Lacunae of Danish Natural Language Processing . . . 356
Andreas Kirkedal, Barbara Plank, Leon Derczynski and Natalie Schluter

Docria: Processing and Storing Linguistic Data with Wikipedia . . . 363
Marcus Klang and Pierre Nugues

Towards High Accuracy Named Entity Recognition for Icelandic . . . 369
Svanhvít Lilja Ingólfsdóttir, Sigurjón Þorsteinsson and Hrafn Loftsson

Neural Cross-Lingual Transfer and Limited Annotated Data for Named Entity Recognition in Danish 376
Barbara Plank

The Seemingly (Un)systematic Linking Element in Danish . . . 382
Sidsel Boldsen and Manex Agirrezabal

Demo Papers

LEGATO: A flexible lexicographic annotation tool . . . 388
David Alfter, Therese Lindström Tiedemann and Elena Volodina

The OPUS Resource Repository: An Open Package for Creating Parallel Corpora and Machine Translation Services . . . 395
Mikko Aulamo and Jörg Tiedemann

Garnishing a phonetic dictionary for ASR intake . . . 401
Iben Nyholm Debess, Sandra Saxov Lamhauge and Peter Juel Henrichsen

UniParse: A universal graph-based parsing toolkit . . . 406
Daniel Varab and Natalie Schluter