Northern European Journal of Language Technology, 2019, Vol. 6, Article 1, pp 1–3
DOI 10.3384/nejlt.2000-1533.1961

Special Issue of Selected Contributions from the Seventh Swedish Language Technology Conference (SLTC 2018)

Hercules Dalianis[1], Robert Östling[2], Rebecka Weegar[1] and Mats Wirén[2]
[1]Department of Computer and Systems Science
[2]Department of Linguistics
Stockholm University
hercules@dsv.su.se robert@ling.su.se
rebeckaw@dsv.su.se mats.wiren@ling.su.se

Introduction

This Special Issue contains three papers that are extended versions of abstracts presented at the Seventh Swedish Language Technology Conference (SLTC 2018), held at Stockholm University 8–9 November 2018.[1] SLTC 2018 received 34 submissions, of which 31 were accepted for presentation. The number of registered participants was 113, including both attendees at SLTC 2018 and two co-located workshops that took place on 7 November. 32 participants were internationally affiliated, of which 14 were from outside the Nordic countries. Overall participation was thus on a par with previous editions of SLTC, but international participation was higher.

Based on the SLTC 2018 reviews (three for each submission) and the quality of the presentations, we invited 12 groups of authors to submit extended versions of their abstracts to this Special Issue. Three groups accepted and eventually submitted. Below are brief presentations of their papers.

Adesam and Bouma (2019) describe the Koala part-of-speech tagset and annotation scheme, developed in the infrastructure project with the same name, funded by Riksbankens Jubileumsfond 2014–17. The aim of the Koala project was to improve the quality and relevance of the linguistic annotations of the corpus-analysis pipeline at Språkbanken Text, Department of Swedish at the University of Gothenburg. The main point of reference for the Koala tagset is the Swedish Academy Grammar, SAG (Teleman et al., 1999), but the authors also compare it to MAMBA, used in the Talbanken project in the 1970s (Teleman, 1974), the SUC tagset used for the Stockholm–Umeå Corpus (Ejerhed et al., 1992), and the Swedish part of the recent Universal Dependencies project (Nivre, 2014). The Koala tagset will be applied to the entire range of Swedish corpora available through Språkbanken Text's corpus tool Korp. This will add to its great interest to the

[1] For more information, see https://sltc2018.su.se/, where also the abstracts are available.

language technology community, since, as far as we know, Korp provides access to the largest annotated and freely searchable corpus collection for any language in the world (about 16 billion tokens as of November 2019[2]). Also, in related work which was presented separately at SLTC 2018 (Adesam et al., 2018), the Koala guidelines were used for annotating the Eukalyptus treebank, containing 100 000 tokens of contemporary Swedish public domain texts.

Kurtz and Kuhlmann (2019) provide new insights into the training of neural networks for dependency parsing, by examining the interplay between constraints imposed on the parsers and the selection of a loss function. During training of a neural network, the loss function provides a measure of how close the predicted dependency output is to the true output, and this article highlights the importance of selecting a suitable loss function for this task. Specifically, the authors investigate how structural constraints, such as not allowing for crossing arcs in the dependency graphs, interact with the structural hinge loss function. They show that training a network with constraints and structural hinge loss can lead to a suboptimal model, and further suggest modifications for the hinge loss function, leading to improved performance of the parsers.

Volodina et al. (2019) summarise the results of the SweLL project, whose goal is to produce an annotated corpus of learner Swedish along with the necessary infrastructure for anonymisation, normalisation and annotation of the corrections made during normalisation. As one of the first major corpus creation projects since the introduction of the General Data Protection Regulation (GDPR), the legal and technical solutions provided by the authors will provide a template for similar projects in the years to come. Furthermore, the SweLL corpus itself with its liberal licence and rich annotations will be of tremendous use for language learning research, as well as for research on educational applications in language technology.

References

Adesam, Yvonne and Gerlof Bouma. 2019. The Koala Part-of-Speech Tagset. *Northern European Journal of Language Technology (NEJLT)* 6, Special Issue of Selected Contributions from the Seventh Swedish Language Technology Conference (SLTC 2018) [this issue].

Adesam, Yvonne, Gerlof Bouma, Richard Johansson, Lars Borin, and Markus Forsberg. 2018. The Eukalyptus Treebank of Written Swedish. In Swedish Language Technology Conference (SLTC). Available at `https://sltc2018.blogs.dsv.su.se/files/2019/11/Final-compilation-of-abstracts-SLTC-2018-Nov-19.pdf/`. Stockholm University.

Ejerhed, Eva, Gunnel Källgren, Ola Wennstedt, and Magnus Åström. 1992. The Linguistic Annotation System of the Stockholm–Umeå Corpus Project. Department of General Linguistics, University of Umeå.

Kurtz, Robin and Marco Kuhlmann. 2019. The Interplay Between Loss Functions and Structural Constraints in Dependency Parsing. *Northern European Journal of Lan-*

[2] Lars Borin, personal communication, 10 November 2019.

guage Technology (NEJLT) 6, Special Issue of Selected Contributions from the Seventh Swedish Language Technology Conference (SLTC 2018) [this issue].

Nivre, Joakim. 2014. Universal Dependencies for Swedish. In Swedish Language Technology Conference (SLTC). Available at `https://www2.lingfil.uu.se/SLTC2014/abstracts/sltc2014_submission_7.pdf`. Uppsala University.

Teleman, Ulf. 1974. *Manual för grammatisk beskrivning av talad och skriven svenska.* ISBN 91-44-10721-8. Lund: Studentlitteratur.

Teleman, Ulf, Staffan Hellberg, and Erik Andersson. 1999. *Svenska Akademiens Grammatik.* Stockholm: Svenska Akademien.

Volodina, Elena, Lena Granstedt, Arild Matsson, Beáta Megyesi, Ildikó Pilán, Julia Prentice, Dan Rosén, Lisa Rudebeck, Carl-Johan Schenström, Gunlög Sundberg, and Mats Wirén. 2019. The SweLL Language Learner Corpus: From Design to Annotation. *Northern European Journal of Language Technology (NEJLT)* 6, Special Issue of Selected Contributions from the Seventh Swedish Language Technology Conference (SLTC 2018) [this issue].