

Preface

Malvina Nissim
Tommaso Caselli
Rik van Noord
Antonio Toral

M.NISSIM@RUG.NL
 T.CASELLI@RUG.NL
 R.I.K.VAN.NOORD@RUG.NL
 A.TORAL.RUIZ@RUG.NL

Center for Language of Cognition, University of Groningen, The Netherlands

This ninth issue of the Computational Linguistics in the Netherlands (CLIN) Journal contains a sample of the papers that have been presented at the 29th edition of the CLIN Conference. CLIN29 was held on January 31st 2019 in Groningen, at the Oosterpoort, a beautiful, contemporary theatre in front of the city’s canal ring.

The 35 talks featured at the conference were given in the *Kleine Zaal*, which was actually really roomy and could easily host the over 150 conference participants, coming from a total of almost 40 different institutions. The 44 posters were hanging in a large exposition area on the first floor, and were presented in two separate sessions. Besides the contributed talks and posters, we were lucky to have Ido Dagan, from Bar-Ilan University, Israel, as our keynote speaker.



CLIN29 Keynote: Ido Dagan

In his talk titled “Consolidating and Exploring Open Textual Knowledge”, Ido outlined a research program to represent information from multiple texts. The program is articulated around three under-researched pillars: a “natural” semantic representation for individual texts, based on crowd-sourcable natural language expressions via question answering; a way of consolidating information structures of different texts; and a framework for interactive exploration of multi-text information.

The conference also featured a special session on the CLIN29 Shared Task: “GxG: Cross-genre gender prediction in Dutch”. Data from various domains was provided by the organisers, which was a

joint team involving Groningen and Antwerp, and participants had to develop models able to capture gender traits from text on a domain different from that they had been trained on. Five teams were at CLIN to present their systems, out of the seven total who participated. All seven approaches are described as system papers which are published as CEUR proceedings, together with an introduction paper by the organisers (<http://ceur-ws.org/Vol-2453/>).

For this CLIN Journal issue we sent out for reviewing eight submitted papers, and selected six for publication. The variety of topics nicely reflects the breadth of computational linguistics research presented at CLIN29.

The current trend on machine translation is represented with two papers. “Predicting syntactic equivalence between source and target sentences” by Vanroy et al. shows how machine learning systems can be trained on a parallel English/Dutch corpus to predict the expected syntactic equivalence of an English source sentence without having access to its Dutch translation. In “NMT’s wonderland where people turn into rabbits. A study on the comprehensibility of newly invented words in NMT output”, Macken et al. assess the impact of non-existing words present in the output of neural machine translation systems.

Three papers have a strong linguistic flavour. “Linguistic Proxies of Readability: Comparing Easy-to-Read and regular newspaper Dutch” by Vandeghinste and Bulté, where the authors focus on

simplification and readability and explore which linguistic features define text as being easy-to-read. The contribution “From partial neural graph-based LTAG parsing towards full parsing” by Bladier et al. shows that it is possible to achieve full TAG parsing by combining supertagging with dependency parsing. A simplified coreference annotation scheme for Dutch and a rule-based coreference system, which are applied to a corpus of literary works, are described in van Cranenburgh’s paper “A Dutch coreference resolution system with an evaluation on literary fiction”.

Lastly, “How to optimize your Twitter collection” by Kreutz and Daelemans describes a method for automatically finding a list of keywords to retrieve tweets for a given language achieving both high precision and high recall.

The conference was organised in cooperation with SIKS, and could benefit from a number of sponsors which we would like to deeply thank: Bureau Taal, CLARIAH, CLCG, Crosslang, Elsevier, IvdNT, NO-TaS, Telecats, Textgain and Textkernel.

Heartfelt thanks also go to the whole organising team, namely staff and students of the Computational Linguistics group of CLCG at the University of Groningen, and to the reviewers for the papers published herein. We do hope you enjoy reading them, and we look forward to next CLIN in Utrecht!



CLIN29 Organising Team