

Preface

Jan Odijk
Tejaswini Deoskar
Martijn van der Klis
Marijn Schraagen

J.ODIJK@UU.NL
 T.DEOSKAR@UU.NL
 M.H.VANDERKLIS@UU.NL
 M.P.SCHRAAGEN@UU.NL

Utrecht University, Utrecht, the Netherlands

The tenth issue of the Computational Linguistics in the Netherlands (CLIN) Journal contains a selection of the papers that have been presented at the 30th edition of the CLIN Conference. CLIN30 was held on January 30, 2020, at the place where it all started in 1990: Utrecht University’s Drift complex. If you want to know why CLIN started in 1990, and why in Utrecht, we refer you to the history of computational linguistics in the Netherlands by Leonoor van der Beek, at least, if you read Dutch.¹

CLIN30 had a record number of 228 registered participants. Of the 104 submissions to the conference, 99 were accepted (the rejected papers were considered out of the scope of CLIN), and with 6 withdrawals we ended up with 93 contributions to the conference. The oral presentations (39), posters (52) and demos (2) were held in parallel sessions in different rooms of the Drift complex.

We were very happy to welcome Joakim Nivre (Uppsala University, Sweden), the main proponent of the Universal Dependencies framework, as our keynote speaker. There was just enough space in the largest auditorium available for the keynote speech to accommodate the large number of participants (as well as for



Figure 1: The Atrium at Drift 21 ©Khiet Truong

a unexpected intruder, a mouse). Fortunately, CLIN30 took place just before the restrictions to address the COVID-19 pandemic made such large gatherings of people impossible.

In his talk titled “Multilingual Dependency Parsing: From Universal Dependencies to Sesame Street”, Joakim Nivre sketched the framework and resources of Universal Dependencies, and discussed advances in multilingual dependency parsing enabled by these resources in combination with deep learning techniques, ranging from traditional word and character embeddings to deep contextualized word representations like ELMo and BERT.

1. <https://www.let.rug.nl/vannoord/TST-Geschiedenis/>.

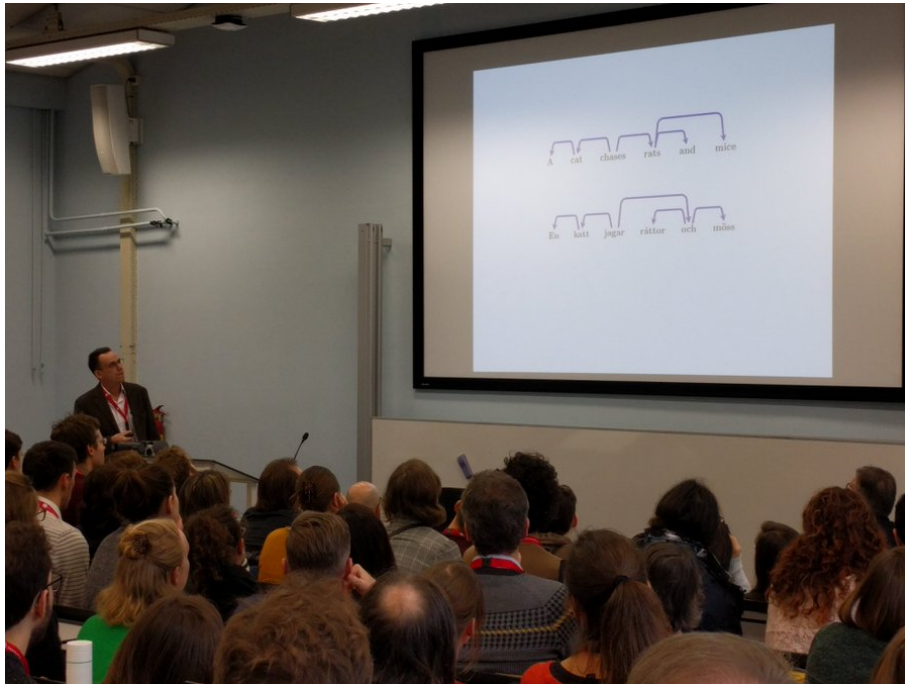


Figure 2: A full house at Joakim Nivre’s keynote lecture! ©Martijn van der Klis

The conference also featured a special session on the CLIN30 Shared Task: “Small data for predicting perfect doubling”, a task in the domain of historical computational linguistics, where the amount of available data is by definition finite, and in most cases relatively limited. The Shared Task deals with *have*-doubling constructions in historical varieties of Dutch, which is an example of a linguistic phenomenon that is very sparsely distributed, but nevertheless exhibits interesting linguistic properties. For more information, we refer to Schraagen *et al.* in this volume.

For this CLIN Journal issue we received 17 submissions. Each submission was reviewed by three independent reviewers. These reviews formed the basis for selecting 10 papers for publication.

The papers in this volume cover a wide range of topics, reflecting the wide range of topics of the CLIN30 conference. We classify these papers here by their most salient topic.

Syntax plays a prominent role in many papers. Allein *et al.* analyze Dutch sentences to identify and correct the incorrect use of relative pronouns. Van Noord *et al.* describe a syntactic profiler for Dutch. De Kok & Pütz use self-distillation to obtain better dependency parsers for Dutch and German. Kroon *et al.* attempt to detect syntactic differences between languages automatically using the Minimum Description Length principle.

Research on historical Dutch is well represented. Creten *et al.* report on research on automatic part-of-speech tagging and lemmatisation for historical varieties of Dutch. Schraagen *et al.* describe the CLIN30 shared task on predicting the occurrence of *have*-doubling in historical varieties of Dutch.

Several papers analyze texts in various ways. Van de Poel & Speelman present a tool for keyword analysis using stable lexical marker analysis, De Clercq *et al.* carry out topic classification of news text as a first step towards diverse news recommendations, and Gatti & Van Stegeren describe research on improving Dutch sentiment analysis.

From a methodological point of view, it comes as no surprise that several papers use various forms of deep learning (Allein *et al.*, De Kok & Pütz, Creten *et al.*, De Clercq *et al.*, Schraagen *et al.*).

The conference was organized in cooperation with the Netherlands Research School for Information and Knowledge Systems (SIKS) and we are very grateful for the support of our sponsors: Institute for the Dutch Language (INT), Utrecht University, Stichting Taaltechnologie, ReadSpeaker, TextKernel, CLARIN, CLARIAH, Elsevier, NOTaS, TeleCats, Textgain, and Yoast.

Heartfelt thanks go to the whole organising team, in particular to Yvonne van Adrichem and Sylvia Pascalis from the UiL-OTS reception, to Paul van der Lugt and his facility management team at the Drift complex, to a large group of student volunteers from linguistics and artificial intelligence, and to the reviewers of the papers published in this issue.

We hope that you will enjoy reading the papers in this volume. We also hope to see you all at CLIN31, which is organised by the University of Ghent, Belgium!