

Preface

Orphée De Clercq*
 Ayla Rigouts Terryn*
 A. Seza Doğruöz*
 Arda Tezcan*
 Bram Vanroy*
 Sofie Labat*
 Lieve Macken*

ORPHEE.DECLERCQ@UGENT.BE
 AYLARIGOUTSTERRYN@UGENT.BE
 AS.DOGRUOZ@UGENT.BE
 ARDA.TEZCAN@UGENT.BE
 BRAM.VANROY@UGENT.BE
 SOFIE.LABAT@UGENT.BE
 LIEVE.MACKEN@UGENT.BE

* *LT3, Language and Translation Technology Team, Ghent University, Belgium*

The 11th edition of the Computational Linguistics in the Netherlands (CLIN) Journal contains some of the best work presented at the 31st edition of the CLIN Conference. The original plan was to organise this year's conference in January in the beautiful *Aula* in Ghent. Predicting that busy poster sessions would not be possible with proper social distancing, it was decided to launch a call for extended abstracts, instead of the usual short abstracts. Despite the precautions, the Covid pandemic meant that the conference had to be postponed to the 9th of July, and, ultimately, moved online. Nevertheless, CLIN2021 was highly successful. Through the *gather.town* platform (see Figure 1), the friendly atmosphere and networking opportunities could be replicated online. We were even able to take a virtual version of the traditional group picture (see Figure 2).

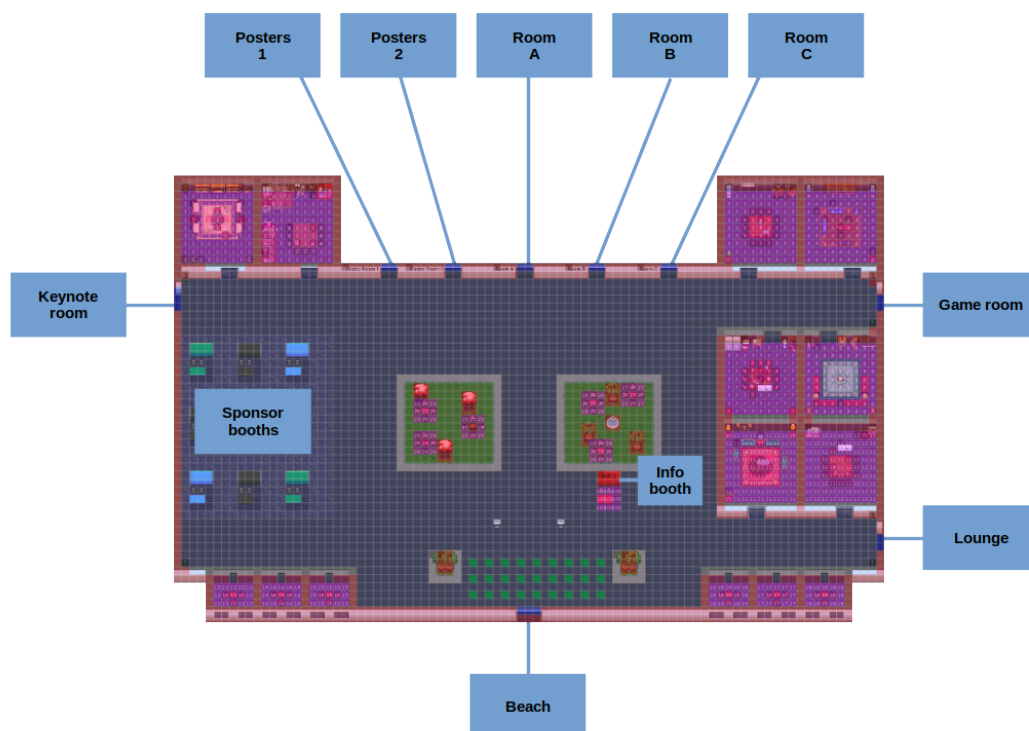


Figure 1: Map of the virtual gather.town environment



Figure 2: Virtual group picture at the "beach"

The success of the conference is reflected in the numbers. A first call for extended abstracts was launched in January, followed by a second call for short abstracts in May, when it was decided the conference would be held online. The first call yielded 39 submissions of extended abstracts, and 30 more abstracts were submitted after the second call. Out of these 69 submissions, 66 were accepted, resulting in 26 oral presentations (divided over 3 parallel sessions) and 40 poster presentations (in 2 sessions). 164 people registered for the conference and created a digital avatar to attend these presentations in the virtual rooms. Most of them came from the Netherlands (51%) and Belgium (46%), but the online event also attracted people from the UK, Ireland, Germany, and Kuwait. One of the highlights of the CLIN conference was the Keynote by Prof. Dr. Roman Klinger, senior lecturer at the Institute for Natural Language Processing (IMS) at the University of Stuttgart. He gave a riveting talk on *Show-don't-tell — how emotions are communicated in text and how psychological theories can help us in computational emotion analysis*. A recording of this interesting talk can be found at the LT3 Research YouTube channel¹. In keeping with tradition, the conference ended with a social event. While, understandably, most people were happy to leave their computers after such a full day, a few dozen brave souls stuck around for an online game of *Pictionary*, *Codenames*, or *Battle Tetris*, thus concluding this 31st edition of the CLIN Conference.

Of the 66 accepted abstracts for the CLIN Conference, 24 were elaborated into full journal papers and submitted to the CLIN Journal. Of these 24, 16 were accepted for publication and can be found in the current edition. This makes the current edition the largest volume of the CLIN journal to date, with contributions by no fewer than 44 (co-)authors. While it is not surprising, it is still noteworthy that 11 of the 16 papers specifically focus on the Dutch language, illustrating the role of CLIN to stimulate this research and to provide a platform for it. Of the remaining 5 papers, 1 is applied to Dutch in combination with other languages, and the other 4 work with English.

A first example of research that focuses on Dutch is the work of Anthe Sevenants and Dirk Speelman, who present *An agent-based simulation of the divergence of the standard Dutch pronunciations in the Netherlands and Belgium*. Bob Van Dyck, Bagher Babaali, and Dirk Van Compernelle researched Dutch speech as well and developed *A hybrid ASR system for Southern Dutch*. Of course, not all papers that focus on the Dutch language relate to speech, as illustrated by Lasha Abzianidze

1. <https://www.youtube.com/watch?v=XGZ6uh0HKjs>

and Konstantinos Kogkalidis who present their work on *A logic-based framework for natural language inference in Dutch*. One of their proposed pipelines is based on Dutch syntactic parsing, which is something we also see in the work of Gosse Bouma, who relies on a syntactically parsed corpus to investigate *Probing for Dutch Relative Pronoun Choice*. Another recurring theme this paper shares with others is the use of BERT-based language models. For instance, Jens Van Nooten, Iliia Markov, and Walter Daelemans use pre-trained contextual word embedding for their project on *Evaluating the impact of word classes on cross-domain age detection models' performance*, and BERT models are a crucial component of Lore De Greve, Gunther Martens, Cynthia Van Hee, Pranaydeep Singh, and Els Lefever's investigation of *Aspect-based sentiment analysis for German: Aalyzing "Talk of Literature" surrounding literary prizes on social media*. BERT models are the main topic of three other papers. Mohamed Barbouch, Suzan Verberne, and Tessa Verhoef present *WN-BERT: integrating WordNet and BERT for lexical semantics in natural Language Understanding*. Pieter Delobelle, Thomas Winters, and Bettina Berendt wrote about *RobBERTje: a Distilled Dutch BERT Model*, and Stella Verkijk and Piek Vossen about *MedRoBERTa.nl: A Language Model for Dutch Electronic Health Records*. Given the current context, it is not surprising to find other projects with a medical, or, more specifically, Covid-related topic, such as *Measuring Shifts in Attitudes Towards COVID-19 Measures in Belgium* by Kristen M. Scott, Pieter Delobelle, and Bettina Berendt and *Vaccinpraat: monitoring vaccine skepticism in Dutch Twitter and Facebook comments* by Jens Lemmens, Tess Dejeaghère, Tim Kreutz, Jens Van Nooten, Iliia Markov, and Walter Daelemans. Another project that investigates the potential for an NLP application to have a positive societal impact is that by Bram Bulté, Vincent Vandeghinste, Leen Sevens, Ineke Schuurman, and Frank Van Eynde, who try to answer the question: *Can pictograph translation technologies facilitate communication and integration in migration settings?*. The final four papers that have not yet been mentioned all revolve around data. Javad Pourmostafa Roshan Sharami, Dimitar Shterionov, and Pieter Spronck describe their project on *Selecting parallel in-domain sentences for neural machine translation using monolingual texts*. Luna De Bruyne, Orphée De Clercq, and Véronique Hoste illustrate how valuable annotated datasets can be in *Prospects for Dutch emotion detection: insights from the new EmotioNL dataset*. Annotation is also the subject of Liqin Zhang, Howard Spoelstra, and Marco Kalz's work on the *Annotation of a Dutch essay corpus with argument structures and quality indicators*. Finally, Wieke Harmsen, Catia Cucchiari, and Helmer Strik discuss *Automatic detection and annotation of spelling errors and orthographic properties in the Dutch BasiScript corpus*.

To conclude this introduction we want to express our sincerest gratitude to everyone who made this year's CLIN Conference and CLIN Journal a success. We extend our thanks to our sponsors: Clariah, CrossLang, De Taalsector, Instituut voor de Nederlandse Taal, Nederlandse Organisatie voor Taal- en Spraaktechnologie, Taalunie, Telecats, Textgain, the Faculty of Arts & Humanities and the Department of Translation, Interpreting and Communication of Ghent University. We are also deeply grateful to the organising committee of the CLIN Conference, and to everyone who reviewed for the journal.

We hope you enjoy the contributions in this journal and look forward to seeing you at next year's CLIN Conference, organised by the Tilburg School of Humanities and Digital Sciences (TSHD), Tilburg University.