5. Summary

Research in CLCG continues to be of a high quality. Situated in the North, a special focus of CLCG concerns local language varieties (e.g. Dutch, Frisian, Low Saxon), combined with a focus on multilingualism and multilingual practices such as in education and health care. Successful areas of research include computational linguistics, discourse and communication, psycholinguistics, neuro- and clinical linguistics, applied linguistics, theoretical and empirical linguistics and multilingualism. The number and impact of academic articles appear to be increasing, the articles are cited by peers, and CLCG researchers are successful in academic grant applications.

The nature of research in several linguistics disciplines appears to be moving in a direction in which empirical research methods are expanding, leading to an increased need for lab infrastructure - not only lab space and hardware, but also support for advanced statistical and computational modelling, and support for research participants and (open) research data management. Improvement of this infrastructure therefore continues to be an important goal for the next few years.

Most of CLCG's research is driven by societal problems and is thus highly societally relevant, with application areas in Healthy Aging (e.g. clinical linguistics, health communication) and Sustainable Society (e.g. multilingualism, language education and second-language learning). Despite the fact that CLCG researchers have obtained several grants financed by non-academic partners, the societal relevance of our research is not always as visible to the general public or the relevant societal partners as we would like. We aim to further improve the visibility of our research in the next few years.

Case studies

Case study 1
Speech and language technology for low-resource languages

The first case study combines the CLCG focus on regional languages with advanced research in computational methods, in particular with recently developed deep neural networks, applied to text and speech. All research described in this case study is partly or fully funded by the Province of Groningen which provides the research funding for Centrum Groninger Taal & Cultuur in which this research line is embedded.

The Low Saxon regional language is one of the officially recognized minority languages spoken in the north-eastern part of the Netherlands. In contrast to majority languages such as English or Dutch, speech and language technology is often not available for low-resource languages. Together with supervisors Martijn Wieling and Mark Liberman (University of Pennsylvania), PhD student Martijn Bartelds has developed techniques to quantify pronunciation differences using acoustic recordings. Currently, the benchmark technique to quantify pronunciation differences is the edit (or Levenshtein) distance, applied to phonetic transcriptions of speech. This technique closely matches human perceptual judgements of pronunciation differences (with a correlation of $|r| = 0.77$). Unfortunately, transcribing one minute of speech takes about one hour, and these transcriptions are often not available for minority languages. Consequently, Bartelds has developed several acoustic-only methods that do not require transcriptions (e.g. Bartelds et al., 2020 – Frontiers in Artificial Intelligence). His most recent method is based on extracting information from deep learning models of speech that clearly outperformed the transcription-based approach (i.e., $|r| = 0.87$; Bartelds et al., 2022 – Journal of Phonetics). Moreover, Bartelds showed that such models can be successfully used for (related) minority languages. This work has been further developed as part of a collaboration between Bartelds and Dan Jurafsky at Stanford University (San et al., 2021 – IEEE Workshop on Automatic Speech Recognition and Understanding).

Together with supervisors Malvina Nissim and Martijn Wieling, Wietse de Vries focuses on making large (deep learning) language models suitable for processing text and speech in minority languages. While language models are generally developed for majority languages, de Vries has shown how these models can be adjusted so they achieve adequate performance when applied to minority languages (de Vries et al., 2021 – ACL Findings). Importantly, this adjustment requires a minimal amount of computational resources and only a very small amount of data, which is realistic and beneficial for low-resource languages. Furthermore, funded by a Google Data Center grant (2019) as well as a grant from the University Center for Information Technology (2021; with Prof. van Doorn at the Faculty of Economics and Business as co-applicant), de Vries has developed a text-to-speech system for Gronings (one of the regional variants of Low Saxon). This system is currently being integrated in the WoordWaark website (which aims to provide an overview of all texts written in Gronings), so that users of the website will not only be able to read texts in Gronings, but also to listen to them. This line of research has also gained interest from the Frisian provincial government, who have committed to partly fund (225 k€) a PhD student (starting at the end of 2022) who will focus on speech technology for Frisian (including text-to-speech).

Case study 2
Verb processing in children with paediatric posterior fossa tumours

Clinical linguistics is traditionally a strong CLCG research area of expertise. Earlier studies have led to improved clinical testing of aphasia patients by including verbs, in addition to nouns, in diagnostic tests. In subsequent CLCG work, verbs were also found to be important for language testing during awake brain surgery. During such tests, electric stimuli are administered to the areas of the brain thought to control language. A wrong answer while the stimulus is being given indicates that the surgeon is encroaching on a language area and preferably must not remove that tissue. CLCG researchers produced a MOOC on this topic which, with 5,740 course enrolments around the globe, has been very successful. The NWO VIDI project by Vânia de Aguiar fits nicely in this tradition and is the topic of our second case study.

Cancer treatments continue to undergo exceptional technical advances. In the field of radiotherapy, one such advance is the progressive replacement of photon therapy with proton therapy. In essence, conventional radiotherapy of cancer has photons passing through healthy tissue to reach the target region such a brain tumour. Along the way, they damage not only tumour cells but also healthy brain tissue. Conversely, proton therapy allows for a more targeted approach, in which less healthy tissue is affected. The UMCG proton therapy centre provides this treatment to paediatric cancer patients from across the entire country, who most often receive brain surgery in Utrecht, at the Princess Máxima Center (PMC).

Proton therapy increases survival rates for children with brain tumours considerably. However, the tumour itself, the surgical procedures and the post-surgical treatments may all lead to long-term cognitive problems. In 2021, the Dutch Research Council (NWO) awarded a VIDI grant to Vânia de Aguiar to study (1) the nature of language processing difficulties in child-survivors of posterior fossa tumours, (2) the relation between brain damage and language impairment, and (3) the effects of proton therapy on brain and language development.

It is well documented that children who survive brain tumours have difficulties with vocabulary development, often requiring special education. De Aguiar's research will shed renewed light on the exact nature of the vocabulary difficulties, leading to more targeted treatments. In addition, de Aguiar is interested in mapping the regions of the brain that contribute to vocabulary development in children. This is a broader research aim, which is relevant not only in this clinical context, but also more broadly to the neurobiology of language. The new knowledge obtained in this project thus helps improve long-term outcomes for children who survive brain tumours. Clinicians will have access to the outcomes, which can empower them to tailor their treatments in a way that improves the children's development.

Thanks to the NWO funding, De Aguiar works on this project together with two PhD candidates and several research assistants. De Aguiar collaborates closely with neuropsychologists Annet Kingma (UMCG) and Marita Partanen (PMC), paediatric radiation oncologists John Maduro and Agata Bannink-Gawryszuk (UMCG), and physicist Branislava Curcic-Blake (Neuroimaging Centre - UMCG). Overseas, De Aguiar cooperates with medical radiologist Andreia Faria (Johns Hopkins University) and linguist Sudha Arunachalam (New York University). This network provides a unique mosaic of expertise, which is specifically tailored and instrumental to address the aims of this exciting research.