

Editorial

Opportunities in the new decade

The beginning of a new decade is always a good moment for self-reflection. As corpus linguists, we generally have been ready to think about the contribution we make to the study of language through our focus on evidence of real language use. We have spent much time considering how corpus methods challenge traditional ways of describing language. We also have not shied away from internal debates over where we see corpus linguistics in terms of methodology versus theory (at the beginning of the last decade, *IJCL* published a special issue on such a debate; issue 15(3)).

Now, in 2020, we are at a point where a key question is: are we outward-looking enough? Or are we spending too much time talking among ourselves, defining smaller and smaller boxes for subtypes, specialisms and areas of corpus linguistics? As a field grows, such developments are not unnatural. We still might miss an opportunity.

In today's digital world, a lot of big data is textual data. Data scientists often analyse linguistic data. Algorithms are employed to shape communication on social media and AI approaches open up new dimensions of computer-human interaction. Amidst all the technology, language is still key. It is part of the data, as well as part of the means to make sense of the data.

Corpus linguists have produced pioneering language research for more than half a century. We have compiled a range of synchronic and diachronic corpora and developed ground-breaking software and tools. Still the impact of corpus linguistics has not been fully realised yet. The term 'corpus linguistics' in itself is a term which very few people outside of academia will have heard of. And even within academia, communication across disciplines can be limited. Looking across the digital humanities, for instance, there is much research that addresses related problems or uses similar methods as we do in corpus linguistics, but terminological and conceptual barriers unnecessarily veil the potential for cross-fertilisation.

Our aim is not to paint a picture of pessimism. On the contrary, we want to highlight three areas where corpus linguistics has the potential to make a significant difference in the present decade.

i. Applied humanities

Much is being said about the current crisis of the humanities, where attempts are made to express the value of humanities research in simplified monetary terms. At the same time, societal challenges that come with globalisation, diversification, mobility, and developments in artificial intelligence and machine-learning highlight a growing complexity and interdependence of matter, people and objects. The application of humanities knowledge is crucial to gaining a better understanding of this complexity and language plays a central role. Language is the substance of any kind of global discourse. Corpus linguistics has the potential to provide methods and approaches for applied humanities at scale.

ii. Interdisciplinarity

To address complex and global challenges interdisciplinarity is crucial. Interdisciplinarity is multi-dimensional: it enables research on and solutions to the problems which cannot be solved by one discipline alone. This includes tackling the grand societal challenges and specifically the UN Sustainability Goals. No matter what the focus is, language and discourses – how they are used, analysed, and explained – play a crucial role. Generally, science communication has been moved up the ladder of transferrable skills for researchers. In any discipline, findings will be described through language so that these findings can be sensibly communicated and applied. Scientific publications, archives of primary sources, reports of experiments, questionnaire and survey data, fieldnotes, etc. are all potential data sets for corpus linguistic analysis. Corpus linguistics offers vast opportunities to better understand how disciplines communicate and to consider how cross-disciplinary discourse might work.

iii. Open research

Across all disciplines there is an increasing need to make data open and ensure the reproducibility of methods and approaches. There is also a need for data to reflect diversity. Data can be linguistic, semiotic, digital. The choice of corpus is a choice of a particular point of view. Our approach to data needs to move away from the Western-centred epistemes and dichotomies of standard and variety. We need to broaden our view beyond more easily retrievable data sets, such as newspaper articles or canonical literary texts, in favour of an inclusive and diverse approach to data. Corpora provide an opportunity to address biases in terms of

gender, race, cultural and literary history. Through the experience with corpus design and compilation, and the development of corpus methods and software, corpus linguistics has an important contribution to make to current initiatives and the definition of standards beyond our own discipline. Especially for humanities research, corpus linguistics has the potential to create significant momentum for the open research agenda.

There is still much for us to do – thinking outside the box, looking beyond our own horizons, and contributing to tackling the big problems of society. *IJCL* is keen to provide a forum for this kind of diversity and innovative thinking.

We are looking forward to your submissions!

Michaela Mahlberg (Editor)
Beatrix Busse (Reviews Editor)