

Towards an accessible and sound corpus methodology in research and teaching

This special issue broadens the scope of the *Kalbotyra* journal beyond linguistic analyses to encompass broader uses of corpora within German philology, or *Germanistik*, including literary analysis and second language teaching. A particular emphasis is placed on the use of corpora within an academic setting where the corpus language is studied as a second language, while recognising the challenges faced by learners at various levels of proficiency. To reflect the rapidly evolving digital landscape, the scope of this issue is also expanded by including studies examining the role of generative artificial intelligence, and specifically evaluating the application of large language models for research and teaching in philology, as well as studies addressing other languages such as English, Lithuanian, or Latvian.

For linguists, literary scholars, and experts of German as a foreign language, working with data in the form of empirical or literary corpora constitutes an indispensable resource for both research and teaching. Corpora offer structured access to language data across different topics and text types, including literary texts. Their language data can be enriched by linguistic and other interpretative annotation at multiple levels, which makes latent characteristics retrievable, thereby enabling generalisations and insights beyond the text surface. However, the practical use of corpora is far from straightforward. Researchers, teachers and students alike are often confronted with the challenge of mastering corpus access in terms of interfaces and query languages, while pedagogical resources for users with limited prior knowledge remain scarce. Another challenge they have to master is critical data literacy, for example, the capacity of understanding that every corpus compilation process has inherent limitations that are reflected in the findings derived from the resulting corpora.

This special issue addresses these challenges by presenting contributions that explore the use of corpora in research and teaching, reflect on the necessary corpus literacy, and consequences for corpus didactics. It brings together case studies demonstrating the application of specific corpora and tools to concrete research questions, teaching

scenarios that integrate corpus methods either in the preparation of materials or in direct student engagement, as well as discussions of the methodological and technical foundations required for corpus-based inquiry. Furthermore, it includes reflections on challenges in compiling and annotating corpora for specialised purposes, and considers the potential role of generative artificial intelligence in supporting linguistic and literary analyses and corpus development.

Although initiated by an open call, a majority of the contributions in this volume originate from talks or practical sessions at workshops, research and teaching stays organised within the framework of the project “Corpus competence for formulaic language / Korpusdidaktik für formelhafte Sprache” (KoDi-FS), an institutional partnership project between Vilnius University and the University of Hamburg. Each paper has undergone thorough anonymous peer reviewing by at least two experts. We would like to thank the reviewers for their valuable time and constructive feedback, which has greatly contributed to the scientific quality of this issue. We also gratefully acknowledge the support of the German Academic Exchange Service (DAAD) with funds of the German Federal Foreign Office, which has funded the project KoDi-FS since 2022, and has also supported this special issue. Funding to this publication has also been provided by the Research Council of Lithuania under the Lithuanian Studies Programme 2025–2030 (Agreement No. P-LISs-25-62).

Heike Zinsmeister, Vaiva Žeimantienė, Skaistė Volungevičienė, and Carla Sökefeld
Hamburg and Vilnius, 2025