

**Thematic issue:  
'Corpus Phonology'**

Projected to appear as one of the issues of *Phonology* 40 (2023)

*Corpus Phonology* emerged around the turn of the 21<sup>st</sup> century as a relatively autonomous field, at the interface between phonology and corpus linguistics. While there is a plurality of approaches within Corpus Phonology, it appears that all scholars working in this field share the belief that phonology is an empirical science, and that phonological generalizations must be drawn from large collections of data instead of isolated examples. As such, Corpus Phonology shares many concerns and methods with other empirical subfields of linguistics such as variationist sociolinguistics, Laboratory Phonology as well as natural language processing.

The vitality of this research program is illustrated by the fact that there exists a dedicated handbook, *The Oxford Handbook of Corpus Phonology* (2014). In addition, a number of large projects, targeting a growing variety of languages, have been developed to provide reliable, primary data sources that can serve as a basis for phonological analysis and theorizing. Corpora most often consist of (annotated) samples of recorded spoken or sign language, but may also take the form of lexical databases containing phonetic or phonological transcriptions. The availability of large corpora has been accompanied by the development of a wealth of tools for the annotation and analysis of phonological data.

Among other contributions to the empirical basis of phonology, corpora have led to more accurate descriptions of phonological patterns that differ, sometimes markedly, from traditional accounts. Furthermore, they have allowed phonologists to integrate frequency and different types of context-related effects to the understanding of sound phenomena, often taking advantage of new quantitative methods.

This thematic issue seeks to offer an overview of the current state of the field, show how corpora shed new light on old problems or uncover new phenomena, and highlight potential avenues and challenges going forward. Contributions that try to articulate the link between phonological data and phonological theory are particularly welcome, provided that phonological corpora play a central role in the analysis. Topics addressed include (but are not limited to) the following:

- the use of corpora to test (or challenge) formal models in phonology
- the advantages and/or limitations of (different types of) corpora for phonological research, including comparisons with other empirical approaches, such as experimental methods
- corpus-based analyses of phonological phenomena that use state-of-the-art quantitative methods (e.g., mixed-effects models, Information Theory, Bayesian methods)
- corpus-based analyses of lesser documented languages
- data-driven, inductive approaches to phonological structure
- corpus-based analyses of sign language phonology

This thematic issue, which will be edited by Marie-Hélène Côté (University of Lausanne) and Julien Eychenne (University of Sherbrooke), is open to all potential contributors, from all theoretical persuasions, and is projected to appear as one of the issues of *Phonology* 40 (2023).

**The deadline for submissions is 1 August 2022.**

General information on the submission of manuscripts can be found on the Phonology website (<http://journals.cambridge.org/pho>). Submissions should be uploaded in PDF format to [ScholarOne](#). Contributors should feel free to contact the editors directly with questions at any time in the submission process, at [marie-helene.cote@unil.ch](mailto:marie-helene.cote@unil.ch), [julien.eychenne@usherbrooke.ca](mailto:julien.eychenne@usherbrooke.ca), and [submissions@phonology.uk](mailto:submissions@phonology.uk). An abstract (no longer than 150 words) should be included. Preference will be given to papers which will occupy no more than 20 printed pages in the journal (around

8000 words). Submissions will be read by at least two reviewers and by the editors of the thematic issue.