# *Introduction to Corpus Linguistics: Theoretical and methodological basics*

## **Julien Perrez**
### ***University of Liege, Belgium***

Since the 1980's and the rise of computer-assisted technologies, Corpus Linguistics (CL) has become a mainstream methodology in linguistics, making it possible to analyze 'very extensive collections of transcribed utterances or written texts' (McEnery & Hardie, 2012: i). This workshop will be devoted to main theoretical and methodological basics of Corpus Linguistics. It will be composed of three main parts. Firstly, we will address the process of corpus construction, with a focus on data collection, balance and representativeness. Secondly, we will discuss essential notions of CL, such as tokens, types, concordances, collocations, corpus annotation and the distinction between corpus-based and corpus-driven approaches. Finally, we will present various types of specialized corpora (for example, monolingual and bilingual corpora, learner corpora and political corpora) to give an overview of the research questions that can be addressed thanks to Corpus Linguistics in a variety of disciplines.

This workshop will also include a hands-on session during which the participants will have the opportunity to apply the notions that have been discussed to their own corpus, using the free corpus processing softwares AntConc (http://www.laurenceanthony.net/software/antconc, Anthony, 2019) and Unitex (https://unitexgramlab.org, Paumier, 2020).

## References

Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Available from https://www.laurenceanthony.net/software

McEnery, T., & Hardie, A. (2012). Corpus Linguistics: Method, Theory and Practice. Cambridge University Press.

Paumier, S. (2020). Unitex 3.2 User Manual. https://unitexgramlab.org/releases/3.2/man/Unitex-GramLab-3.2-usermanual-en.pdf