



Contribution ID: 112

Type: **Presenters (Oral Presentation)** – Live ZOOM Presentation

## A Diachronic Corpus Compilation Process and Principles

*Friday, 29 April 2022 12:10 (20 minutes)*

In the current report, a corpus of texts on political economy published in 1841-1850 is presented as a case study. It was compiled within the framework of an international project “LEXECON. The Economic Teacher: A transnational and diachronic study of treatises and textbooks of economics (18th to 20th century). Intra- and interlingual corpus-driven and corpus-based analysis with a focus on lexicon and argumentation”, realized by the University of Pisa, the University of Padova and the University of Palermo, and funded by the Italian Ministry of University and Research for the period 2021-2023 as research of national interest. The goal of this report is to explore the process and principles of corpus formation, answering the following research questions: how does one comply with the corpus criteria; what are the challenges of each corpus formation stage and what are the functionalities of corpus-based approach? The research explores authenticity, representativeness, balance and sampling, and size as corpus criteria and describes bibliographical research, corpus sampling, editing, and structuring as stages of corpus creation. To illustrate the functionalities of corpus-based approach, the use of the first-person singular pronoun across four genres - essay, academic lecture, textbook and treatise - is examined, using such corpus analysis tools as Sketch Engine and Hyperbase 10. The preliminary findings suggest that balance is the most challenging corpus criterion to fulfil, corpus editing is the most time-consuming corpus creation stage, and the context and surplus-deficiency extraction contribute to the research results no less than relative frequency data.

Keywords: corpus linguistics, corpus compilation, political economy, diachronic corpus, genre

### Biographical note(s) of the author(s)

Kristina Korneliusa holds a BA degree in Humanities (2020). Her BA thesis, “Modifiers and Verb forms in “Mayor of Casterbridge” by T. Hardy” employed such functionalities of corpus linguistics as frequency count and semantic corpus tagging. The report to be presented is based on the empirical part of her upcoming MA Thesis, which involves a larger variety of tools and more complex manual corpus compilation. She has presented several reports in literary studies and translation studies, which involved the creation of diachronic multilingual corpora. She intends to continue applying the corpus linguistics method for texts of different topics and genres.

### Recommendation (for student section)

Zigrīda Vinčela, [zigrida.vincela@lu.lv](mailto:zigrida.vincela@lu.lv), 67034916

### Contact e-mail address

[kristina.korneliusa@gmail.com](mailto:kristina.korneliusa@gmail.com)

### Affiliation of the author(s)

University of Latvia

**Primary author:** KORNELIUSA, Kristina (University of Latvia)

**Co-author:** VINČELA, Zigrīda (University of Latvia)

**Presenter:** KORNELIUSA, Kristina (University of Latvia)

**Session Classification:** Student Session

**Track Classification:** Student Section: All topics