

Meeting 5:

Compiling a corpus: format, tools and selection

18 November 2021

Example I:

For this study, a corpus of speeches by politicians has been compiled. To draw on a more diverse dataset within the same language, three culturally different English-speaking countries have been selected: New Zealand, the United Kingdom, and the United States of America. Jacinda Ardern, Queen Elizabeth and Donald Trump usually employ rather different political styles, which are already well-researched and might be useful for answering the research question (on Ardern see, e.g., Mustaqim 2019; Craig 2021; Hafner and Sun 2021; Sacca and Selva 2021; on Queen Elizabeth II. see Yates 2018; Charteris-Black 2019; Vlasyan et al. 2020; Breeze 2020; on Trump see Ahmadian et al. 2017; Sclafani 2018; Panageotou and Crombez 2020; Ross and Rivers 2020; Schneiker 2020; Jones 2021).

Example II:

For this study, a corpus of speeches by politicians has been compiled. To draw on a more diverse dataset within the same language, three culturally different English-speaking countries have been selected: New Zealand, the United Kingdom, and the United States of America. Jacinda Ardern, Boris Johnson and Joe Biden Jr. usually employ rather different political styles. The corpus will initially consist of their communication in form of press statements as published on official government websites. In a second step, the Twitter statements of Boris Johnson are to be considered as well.

Example III:

To assess the success of their crisis communication, a control corpus will be created out of newspaper articles on the topic of the pandemic from the same period in order to see if their narratives were picked up or reported upon positively by the media. There were three newspapers selected for each country, based on three main criteria: daily publishing, distribution range, and accessibility. In order to better display the countries' newspaper varieties, both quality press and yellow press have been included if possible, with no differentiation of online and printed papers. For the United Kingdom, the papers selected are *The Guardian*, *The Independent*, and *The Sun*. The three newspapers for New Zealand were the *New Zealand Herald*, *Dominion Post*, and *The Press*. Lastly, the papers selected for the United States of America are the most-read newspaper *USA Today*, *The New York Times*, and *The New York Post*.

Definition

Keyword:

„A keyword refers to a lexical item which occurs with unusual frequency, either with a significantly higher or lower frequency in a target text or corpus, when compared to a reference corpus“ (Pojanapunya and Watson Todd, 2016, 1)

AntConc

<https://www.laurenceanthony.net/software/antconc/>