

- Paul Baker. 2010. Corpus methods in linguistics. Litosseliti, ed., 93-113.
- Stefan Th. Gries & John Newman. 2013. Creating and using corpora. Podesva & Sharma, eds., 257-287.

Structure of the chapters

| Baker 2010 | Gries & Newman 2013 |
|--------------------------------------|---|
| 1. Introduction | 1. Introduction |
| 2. Theoretical concepts | 2. Creating corpora |
| 3. Building and annotating corpora | 2.1 The notion of “a corpus”: a prototype and dimensions of variation |
| 4. Types and applications of corpora | 2.2 Collecting the corpus data |
| 5. Corpus software and analysis | 2.3 Preparing the corpus data |
| 6. Critical considerations | 2.3.1 Cleaning up and standardizing |
| | 2.3.2 Marking up and annotating |
| | 2.4 Several widely used corpora |
| | 3. Using corpora |
| | 3.1 Analytical tools of corpus linguistics |
| | 3.1.1 Frequency lists and dispersion |
| | 3.1.2 Collocations |
| | 3.1.3 Concordances |
| | 3.2 Tools for analysis in corpus linguistics |
| | 3.3 Programming tools for corpus linguistics |

Questions the answer to which can be found in the chapters

- What is a corpus? What is typical for a modern corpus used in linguistics?
- What is the advantage of corpora/corpus studies over other kinds of linguistic analysis?
- What are disadvantages?
- Name different types of corpora (or “dimensions” along which corpora are distinguished).
- How large is a “large” corpus?
- How large must a good corpus be? (or What determines the size of a good corpus?)
- Where to get data for a corpus of written language? Where to get data for a corpus of spoken language?
- What should we have in mind when collecting texts for a corpus? (what are categories for sampling)
- What has to be done with the raw texts (in order to prepare a corpus useful to linguistic research)?
- How can we use corpora for linguistic research? What are typical research questions and findings?
- Where can the results of corpus studies be used further? or: Which fields of (applied) linguistics can profit from corpus studies?
- What is this:
collocation, concordance, KWIC, tagging, regular expressions