

General Information

Course name	Corpus Linguistics	ECTS Credits	5
		Semester	summer

Aims

The aim is to introduce corpus linguistics as a research method for descriptive and applied linguistics. To this end, the course consists of two parts: theory, which reviews the history of corpus linguistics and the basic stages of corpus building and annotation; and practice, which is a series of hands-on sessions where the main corpus tools are exercised by use of a selection of free-access corpora, dictionaries, and concordancers and concordancer-related tools. The course aims at enabling students to understand the principles of corpus-based research; successfully use the right corpus-related tools for their needs; and train in the basic stages of data collection, sampling and preparation.

Contents

Part I. Theory

1. An introduction to corpus linguistics
 - 1.1. What is corpus linguistics?
 - 1.2. Computerised corpora
 - 1.3. The state of the art in corpus linguistics
 - 1.4. The linguistic exploitation of computerised corpora
2. Corpus design and compilation
 - 2.1. Introduction
 - 2.2. Corpus design and compilation
 - 2.3. The size of corpora
 - 2.4. The Text Encoding Initiative (TEI)
3. Corpus annotation
 - 3.1. Introduction
 - 3.2. Corpus annotation
 - 3.3. Two examples of annotation: tagging and parsing
 - 3.4. Semantic disambiguation

Part II. Practice

4. (Online access to) major corpora
 - 4.1. The BNC. Access modes
 - 4.2. CREA, CORDE, SNC
 - 4.3. Other corpora or bodies of data
 - 4.4. Web as corpus
5. (Online access to) major dictionaries
 - 5.1. The OED
 - 5.2. The DRAE
 - 5.3. The SSDL
6. (Online access to) corpus tools (frequency lists)
 - 6.1. Kilgarriff (1998)
 - 6.2. Leech, Rayson & Wilson (2001)
 - 6.3. BNC Sampler-based frequency list
7. Data processing
 - 7.1. Data collection
 - 7.2. Data sampling
 - 7.3. Data preparation
8. Assignments
 - 8.1. Under-/over-/... verbal compounds and lexicalization
 - 8.2. Word-class frequency
 - 8.3. -ic/-ical

Evaluation

Bibliography

- Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.
- Dörnyei, Z. (2003). *Questionnaires in Second Language Research: Construction, Administration and Processing*. Mahwah, NJ: LEA.
- McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- McEnery, T. & Hardie, A. (2013). The history of corpus linguistics. In K. Allan (ed.), *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press.
- McEnery, A. M., Xiao, R. Z. & Tono, Y. (2006). *Corpus-Based Language Studies: an Advanced Resource Book*. London: Routledge.

Wray, A. & Bloomer, A. (1998) *Projects in Linguistics. A Practical Guide to Researching Language*. London: E. Arnold.

