Lecture 7. Change over time: Working with diachronic data

Aim: The lecture is based on Brezina (2018), Chapter 7. It discusses statistical procedures that can be used to explore historical or diachronic data. First, specific features of diachronic studies are outlined and techniques that provide effective visualizations of diachronic change are introduced. Second, the lecture focuses on the statistical comparison of two time periods using a procedure called bootstrapping. Next, the diachronic application of the cluster analysis is discussed. Finally, the chapter presents a method for statistical identification of peaks and troughs in diachronic data and an extension called Usage Fluctuation Analysis (UFA).

- Key terms: longitudinal study | diachronic corpora | line graph | candlestick plot | bootstrapping test | VNC | peaks and troughs | UFA

Time:

1-hour lecture.

2-hour computer lab session with exercises and Lancaster Stats Tools online (optional).

1-hour individual study (readings).

Statistical tools: Bootstrapping test, Neighbour clusters, Peaks and troughs and UFA

Practical exercises: Chapter 7 Exercises and answers.

Data: Chapter 7 data.

Readings: Brezina 2018 Chapter 7 + optionally selected Advanced readings recommended in the book.

Outline:

1. Time as a variable: Measuring and visualizing time
2. Finding and interpreting differences: Percentage change and the Bootstrap test
3. Grouping time periods: Neighbouring cluster analysis
4. Modelling changes in discourse: peaks and troughs and UFA
5. Things to remember

Main points – 'Things to remember':

- Historical analyses, because they use available and imperfect data, require critical consideration of i) diachronic representativeness of corpora, ii) alternative interpretations of linguistic development and iii) fluctuation of the meaning of linguistic forms.
- Visualization options include line graphs, boxplots and error bars, sparklines and candlestick plots.
- The bootstrapping test is used to compare two corpora (representing different points in time); it makes use of a technique of multiple resampling of corpus data.
- Peaks and troughs is a technique which fits a non-linear regression to historical data, producing a graph which highlights significant outliers in the process of historical development of language and discourse.
- MFA (Meaning Fluctuation Analysis) is a complex procedure combining automatic collocation comparison in a given historical period and the peaks and troughs technique.