

Lecture 6. Sociolinguistics and stylistics: Individual and social variation

Aim: The lecture is based on Brezina (2018), Chapter 6. It discusses different statistical procedures available for the analysis of stylistic and sociolinguistic variation in corpora. It reviews different approaches to variation, pointing out the common connection to the notion of 'style' understood as a particular way of speaking and using language. The statistics discussed include the t-test, ANOVA, Mann-Whitney U test, Kruskal-Wallis test, correspondence analysis and mixed-effects models.

► Key terms: style, sociolect, envelope of variation, ambient variables, sociolinguistic variables, t-test, ANOVA, Mann-Whitney U test, Kruskal-Wallis test, correspondence analysis, chi-squared distance, principle of accountability, mixed-effects models

Time:

1-hour lecture.

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

1-hour individual study (readings).

Statistical tools: [Group comparison](#), [Correspondence analysis](#), [Mixed effects logistics regression](#) and [BNC64](#)

Practical exercises: [Chapter 6 Exercises and answers](#).

Data: [Chapter 6 data](#).

Readings: Brezina 2018 Chapter 6 + optionally selected Advanced readings recommended in the book.

Outline:

1. Individual style and social variation: Where does a sociolinguistic variable start?
2. Group comparison: T-test, ANOVA, Mann-Whitney U test, Kruskal-Wallis test
3. Individual style: Correspondence analysis
4. Linguistic context: Mixed-effects models
5. Things to remember

Main points – 'Things to remember':

- Sociolinguistic variation can be operationalised in different ways: Labovian meaning-preserving sociolinguistic variable (formal approach), or the functional approach looking at the distribution of linguistic features in groups of speakers.
- The t-test, ANOVA (as well as their non-parametric counterparts: Mann-Whitney U and Kruskal-Wallis) are used to investigate the effect of explanatory social variables (gender, social class) on the use of different linguistic features.
- Correspondence analysis is an exploratory analytical technique, which compares the use of multiple variables in different speakers reducing them to two factors and producing a powerful visualisation – a correspondence plot.
- Mixed-effects models is a group of sophisticated statistical techniques, which can account for multiple variables at the same time and include the effect of individual variation between speakers.