

A Quantitative Analysis of Rival Forms

MA Course: Linguistic data: quantitative analysis and visualization
Final project

In the project, the students are supposed to explore the use of the rival forms in the written or oral speech. The rival forms can be:

- two or more synonyms
- two or more grammatical forms
- rival word formation models
- rival syntactic constructions
- rival pronunciation models, etc.

Just to give an idea, the choice of the rival units can be driven by certain contextual factors (e.g. words, syntactic patterns), features of the rival units themselves (e.g. the gender of the noun, the tense of the verb), genre and register of the text, sociolinguistic parameters (e.g. age, sex, profession, education, place of birth). We expect students to explore at least three factors of any kind in their case study. For inspiration, you can look at some examples of the data annotation at the open repository Trolling (REF).

During the project, the students formulate their initial hypothesis, collect data (either corpus-based or experimental), annotate data and do the preliminary descriptive, exploratory and inferential statistical analysis. After that, they can update their hypothesis, include more or exclude some factors, and collect more data/annotate more parameters in order to improve their analysis. The amount of the data collected should be enough to support the statistical analysis. It is by no means evident that it is strongly prohibited to exclude data that contradict the hypothesis or make any other sort of the hypothesis-biased fraud.

The students are expected to prepare the final project in a written form as an electronic document that includes the following parts:

- Research objectives and hypothesis to be tested.
- Description of input data: features and values, descriptive statistics.
- Discussion of the methods of analysis and their applicability.
- Obtained results and their interpretation.
- All the code used (as an attachment).
- Annotated data (as an attachment).
- Experimental survey questionnaires, if applicable (as an attachment).

Language under analysis: any natural language

Language of the project paper: English (Language Theory program), Russian (Computational Linguistics program)

Type of analysis: multi-factor

Type of the project: individual