LingPhil research course proposal: Course content

Linguistic data search and preparation

Audience and learning goals

This course is aimed at students who will use linguistic or language-related data in their research projects. Even if great amounts of language resources are available, researchers in linguistic and philology still confront stumbling blocks in obtaining and selecting the data in the right format for the purposes of their research projects. The goal of this course is to give students knowledge and skills in locating datasets, search in datasets, and extracting, filtering, reformatting, and visualizing the data in preparation of further analysis.

Components:

1. Finding data through catalogs; metadata, usage rights and licenses.
2. Searching in corpora with Glossa.
3. Searching in corpora with Corpuscle and INESS.
4. Searching in the National Library holdings with Jupyter Notebook.
5. Filtering, converting, counting and visualizing text data with shell scripts and R.
6. Refining tabular data with Open Refine.

Teaching and learning methods

- Lectures which mainly consist of demonstrations of methods and techniques, using examples based on student needs.
- In-class exercises in which students apply techniques to example data.

Assignments

Three weeks before the course, students will be asked to submit a brief description of their projects and their specific needs for data access and handling, including examples of data.

After the course, students will be asked to send in their answer to an assignment given during the course, based on the teachers' analysis of their projects.