# Department of Cultural Studies

| General information | |
|---|---|
| **Name of the course** | Digital linguistics:  Tools and methods |
| **Course instructor** | Benedikt Perak |
| **Study programme** | Graduate study programme in Cultural studies |
| **Status of the course** | Elective |
| **Year of study** | 3rd |
| **ECTS credits and manner of instruction** | 5 |
| | Number of class hours (Lectures + Exercises + Seminars) — 30+0+15 |

## 1. COURSE DESCRIPTION

### 1.1. Course objectives

The objectives of the course are focused on the development of empirical methods, knowledge of language resources and the use of computer tools and resources that are an essential part of modern research in linguistics, socio-linguistic, discourse studies, literature and cultural communication. The students will:

1. Get acquainted with the theoretical background and possibilities of application of tools for computer word processing and linguistic data.
2. Define the processes of collection, storage, processing, analysis, presentation and dissemination of language data.
3. Apply modern digital tools, formats and services for collection, storage, processing and presentation of language data.
4. Get acquainted with the implemented methods of researching language structures and communication using the corpus tool Sketch Engine.
5. Describe and implement methods of researching language structures and communication using language, text and data science related Python packages
6. Describe methods and implement language, text and communication research using machine learning algorithms.

### 1.2. Course enrolment requirements and entry competences required for the course

No prior knowledge is necessary

### 1.3. Expected course learning outcomes

1. Develop theoretical understanding and practical skills related to the design and application of scientific methods of linguistic research in the digital environment
2. Introduce and apply tools and formats for digital preparation, storage and processing of linguistic data: txt, csv, json, xml, conllu, as well as for linguistic tagging, such as Recogito.
3. Know how to use the Sketch Engine service for exploratory morphosyntactic analysis and corpus creation, including: corpus selection, lexical summarization, frequency research, keyword identification, concordance analysis, thesaurus analysis, word sketch analysis, language pattern analysis using CQL (corpus query language), analysis of parallel corpora

4. Develop theoretical understanding and practical skills for retrieving, preparing, morphosyntactic analysis, named entity analysis and publishing linguistic data, and creating linguistic databases
5. Use NLTK library for various language analysis tasks, including semantic labelling, sentiment analysis.
6. Develop theoretical understanding and practical skills of using Python NLP and data science packages for processing and visualization of linguistic data, including: pandas, scipy, nltk, udpipe, spacy, as well as other Python libraries.

## 1.4. Course content

- Theoretical grounding and practical applications of text processing.
- Theoretical foundation and connection with linguistic systems: cognitive linguistics, corpus linguistics, computational linguistics, discourse analysis.
- Overview of applications in scientific research, digital communication, commercial applications.
- An overview of important research approaches and research groups.
- Practical application of corpus analysis tools, resources and methods
- Application of basic tools, resources and methods of natural language processing (NLP).

## 1.5. Manner of instruction

✓ Lectures
✓ Exercises
  Fieldwork
✓ Individual assignments
✓ Multimedia and network
✓ Mentorship
✓ Other: consultations