

Required DigiLing coursework prior to arrival

Variability of Languages in Time and Space (abbrev. Variab)

- Please complete the following online course units:
 - Unit 1 and Unit 2 (Introduction)
 - Unit 4 (Phonemic inventories, Stress, Syllable)
 - Unit 5 (Morphological typology)
 - Unit 8 (Writing systems)
- Examine linguistic resources on the world's languages described in Unit 1.3.
 - Explore databases Ethnologue and WALS. Using these databases do the Exercise 1-3 in Unit 1.3.

Introduction to Text Processing and Analysis (IntroProc)

- Participants are required to:
 - Create a free account at <https://www.korpus.cz/signup>
- Please go through the following online course units:
 - Unit 1 (File formats related to textual data)
 - Unit 2 (Plain text: encoding, data and metadata)
 - Unit 3 (XML)
 - Unit 4 (Regular expressions)
 - Unit 5 (Tokenization, segmentation, corpus data formats)
 - Unit 6 (Morphological analysis)
- Please prepare the following materials:
 - any richly formatted text in Word (1-2 pages)
 - an article of your own choice in English (to be later lemmatized and tagged)
 - possibly any text with unknown encoding you had trouble opening
- Participants will need to bring their laptops to each session.

Post-Editing Machine Translation (Postedit)

- Please complete the following Units: Units 0, Unit 1, Unit 2, Unit 3
We wrap up and discuss the contents in the first session.
- Laptops would be great, but are not mandatory

Computational Lexicology and Lexicography (Lexico)

- Please complete the following online course units:
 - Unit 1 (Introduction to Computational lexicology and lexicography)
 - Unit 3 (Morphological lexicons)
 - Unit 5 (Lexical relations and lexical databases)
 - Unit 8 (Semantic lexicons)
- Complete exercises in Unit 1, 3, 5 and 8
- Complete activity in Unit 5 (hands-on use of Wordnet) and activity in Unit 8 (hands-on use of semantic lexicons)

Introduction to Python for Linguists (Python)

- Participant will need to install Python version 2 on their laptops.
- Participants will need to bring their own laptops to each session.
- Please complete the following online course units:
 - Unit 1 - Introduction to Python
 - Unit 2 - Basic data types
 - Unit 3 - Variables
 - Unit 4 - Basic operators
 - Unit 5 - Basic built-in functions
- Please read the following online course unit:
 - Unit 12 - Final assignment
- Create a text document of any genre in your native language. The document should be at least 100,000 characters long (including white space). Save the document as a .txt file with UTF-8 encoding.

Localization Tools and Workflows (Localiz)

- All participants will need to complete the following tasks and upload the relevant files:
 - Create free accounts with the recommended CAT tools, Wordfast Anywhere, Memsource and SmartCAT (links to the associated websites can be found in Unit 5 Practice Project Instructions)
 - Complete the online course units (Units 0-10)
 - ensuring that the files required in each Unit are submitted via the Moodle. Transcripts are also available to accompany the videos for each Unit.
- Participants will need to bring their own laptops to each session.

Mining and Managing Multilingual Terminology (TermMine)

- Participants are required to complete the following tasks:
 - Complete the course units 1, 2, 3 including assignments, units 4 and 6 (assignment optional)
 - Test access to the SketchEngine via Institutional login
 - Study the basics of CQL and regular expressions (see links in Unit 8)
 - Locate or collect a small parallel corpus English-XX for a domain of your choice (a few specialised texts and their translations), align and convert to TMX using any available tool (LF aligner - free, memoQ / SDL / ... - paid but trial versions available). Note: This last task can be completed in pairs or groups of several students.
- Participants will need to bring their laptops to each session.

Introduction to Sketch Engine (SketchEng)

- No prior requirements.