**Computer-assisted translation**

Course scenario

# CLASS 3

|  |  |  |
| --- | --- | --- |
| **TOPIC** | Corpora and Corpus Analysis Tools | |
| **LEARNING CONTENT - DETAILED CHARACTERISTICS** | Topics to be covered: What is a corpus, Representativeness, Types of corpora, Types of corpora according to the number of languages: monolingual corpora, bilingual corpora, bilingual comparable corpora, Basic corpus analysis tools: frequency lists and concordance programs, KWIC concordance, English language corpora, Online corpora | |
| **KEY WORDS** | Corpus, concordance, representativeness, frequency list, wordlist, types, tokens | |
| **SUGGESTED TOOLS** | MS PowerPoint / Canvas (or similar)  Access to online corpora (e. g. <https://www.english-corpora.org/>)  Corpus analysis tools (e. g. AntConc) | |
| **TIPS / METHODOLOGICAL REMARKS**  **(if applicable)** | * Renew and test information about demonstrated corpora and links prior to the class. * Multilingual lecture material for practical tasks needs to be prepared according to the student’s knowledge of languages. You need to learn it during the previous classes. * Compile small toy-corpora for the practical tasks and make them accessible to the students (e.g. via Moodle). * Choose the words for the analysis that are present in the selected toy-corpora. * Minimum CEFR Level required: B1 (Reading, Writing, Listening and Speaking). | |
| **IMPLEMENTATION OF THE CLASSES** | **STEP 1** | Give a self-assessment quiz from the previous class. |
| **STEP 2** | Ask students prior questions about the knowledge of the topic. E.g. do they already know, what is *corpus, representativeness, different types of corpora, corpus analysis tools.* Some philology students are well familiar with the topics and the repetition of the material would not be effective. |
| **STEP 3** | Based on the student’s answers present the explanation of the above concepts. If students are well familiar with the above concepts, concentrate more on practical tasks. If the concepts are new to students, firstly concentrate on theoretical explanations and presentation. |
| **STEP 4** | Present students with practical tasks and give them reasonable time to complete the tasks. |
| **STEP 5** | Review and compare results of the task. |

**ADITIONAL MATERIAL 1 (PICTURE1)**

Picture exemplifies the representativeness of a corpus.

**A blue oval with white border

Description automatically generated**

Source: Original

**ADITIONAL MATERIAL 2 (PRACTICAL TASK AND WORK CARD)**

1. Download *AntConc* software (**version 3.4.4w**) (<https://www.laurenceanthony.net/software/antconc/releases/AntConc344/>).
2. Download the texts from *Moodle*:
   1. EN-text
   2. DE-text (German, French or some other language)
3. Prepare the texts for the analysis (*AntConc* can only process (plain text));
4. Analyse the text with the *AntConc* tool and fill in the table
5. Upload the results into Moodle folder “Corpora tasks”

|  |  |  |
| --- | --- | --- |
|  | EN | DE |
| What‘s the size of the text? (number of word tokens) |  |  |
| How many different words there are in the text? (types) |  |  |
| How many times the wordform “connection” occurs? |  |  |
| What are left collocates of the wordform “centre” in English text when the span is equal 5. |  |  |
| How many wordforms there are of the word “inspektor” |  |  |
| what are the 3 most frequent collocates of the wordform *freiheit* in the right position when the span is equal to 3 |  |  |