

## Kováříková presentation 15.6.2023

The Czech National Corpus (CNC) webpage is [www.korpus.cz](http://www.korpus.cz)

### 1. CNC CORPORA

- parallel corpus **InterCorp** with [41 different languages](https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze15#corpus_size_in_thousands_of_words) ([https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze15#corpus\\_size\\_in\\_thousands\\_of\\_words](https://wiki.korpus.cz/doku.php/en:cnk:intercorp:verze15#corpus_size_in_thousands_of_words))
- the newest version is InterCorp\_v15
- however, if you want to use the Universal Dependencies tagset, use version 13
- you can align two or more languages, just remember, that not all texts are available for all language pairs
- you can also work with one language only, in that case, I recommend using the online tool **Alpha** ([alpha.korpus.cz](http://alpha.korpus.cz)) that searches the InterCorp version 13
  - if you want to search for a specific lemma or a part of a word, always use quotation marks
  - otherwise, you can form the corpus query in normal language
  - here are some examples
    - find lemma "red", "blue", "green" in English
    - find all adjectives starting with "anti" in English corpus
    - end of sentence, then adverb, then comma in French

### 2. KONTEXT

- the primary tool for working with CNC corpora
- if you want to learn to use KonText, here is the manual in English: <https://wiki.korpus.cz/doku.php/en:manualy:kontext:index>

### 3. Some examples of corpus searches that can be used not only in linguistics, but in humanities and social sciences (using the Alpha tool)

- negative/positive spin of lemma *ideology* or multiword unit *because of*
- lemma "green" not only the colour, but also ecology-related
- concept of mansplaining:
  - query in Alpha: masculine subject dependent on lemma "explain"
  - query in Alpha: feminine subject dependent on lemma "explain"
- compare *fascist* vs. *Nazi*

### 4. TREQ for everyone who needs a dictionary

- Treq ([treq.korpus.cz](http://treq.korpus.cz)) is a translation equivalent database, automatically generated from InterCorp data
- you can find equivalents in any of the 41 languages
- you can click on the translation equivalent and examine it closer in contexts
- disadvantages:
  - fully automatically processed so there are some error
  - intended for experienced speakers – the user needs to pick one of the offered alternatives
- advantages:
  - number of languages
  - you can click on the translation equivalent and examine it closer in contexts

- after clicking on the equivalent, you can examine collocations (menu> collocations)

## 5. QuitaUp

- QuitaUp application ([korpus.cz/quitaup](http://korpus.cz/quitaup)) is a simple tool for calculating selected stylometric indicators that quantitatively express some properties of the text
- the manual is available in the application itself (on the first page)
- upload any text in one of the offered languages and find the results in no time