

ČEŠTINA PRO CIZINCE V 21. STOLETÍ
ANEB JAK UPLATNIT KORPUSY V HODINĚ
X. MEZINÁRODNÍ SYMPOZIUM
O ČEŠTINĚ JAKO CIZÍM JAZYKU
10.-14. 8. 2021

Seminář bude zacílen na korpusový přístup ve výuce ČCJ a bude se skládat z následujících částí:

1. Práce s korpusy (existující korpusy, nástroj KonText)
 2. Volně dostupné korpusové nástroje a jejich využití ve výuce (WaG, SyD, Morfio)
 3. Jak vytvářet korpusová cvičení – ukázka příkladových cvičení
-

Seminář **nezajišťuje** základní informace, které si účastníci budou muset doplnit formou samostudia:

1. Vstupní informace o korpusové lingvistice a jazykovém korpusu
2. Registrace
3. Úvod do práce s korpusy (základní vyhledávání, frekvenční distribuce)

K jednotlivým bodům zpracujte úkoly obsažené v pracovním listu na další straně.

Čas věnovaný samostudiu: 90 minut

PRACOVNÍ LIST

1. ZÁKLADNÍ POJMY

Zadání: V rámci samostudia si přečtete definice korpusové lingvistiky (<https://www.czechency.org/slovník/KORPUSOV%C3%81%20LINGVISTIKA>) a jazykového korpusu (<https://www.czechency.org/slovník/KORPUS>). Odpovězte na níže uvedené otázky.

V čem spočívá odlišnost korpusové lingvistiky ve zkoumání dat ve srovnání s přístupem N. Chomského?

.....
.....
.....

Jaké problémy musí řešit korpusová lingvistika v současnosti?

.....
.....
.....

Čím se liší korpus od běžného elektronického archivu?

.....
.....
.....

Co je to *hapax*?

.....
.....

2. REGISTRACE

Abyste mohli plně využívat všechny dostupné funkce v nástroji KontText, je nutná registrace, která je zcela zadarmo.

Registrace je jednoduchá. Její návod najdete zde:

[https://wiki.korpus.cz/doku.php/kurz:zaciname?s\[\]=registrace](https://wiki.korpus.cz/doku.php/kurz:zaciname?s[]=registrace).

3. ÚVOD DO PRÁCE S KORPUSY

Zadání: Seznamte se s vyhledáváním v korpusu v první lekci online kurzu (https://wiki.korpus.cz/doku.php/kurz:prvni_dotaz). Kurz vás provede základním vyhledáváním a typy dotazů v nástroji KonText.

Zadání: Nyní zkuste vyhledat seznam kolokací doporučeným způsobem. K níže uvedeným substantivům (*věk, muž, nemocnice*) najděte jejich tři nejčastější adjektivní kolokáty na základě dat z korpusu SYN2020.

K tomu postačí zadat dotaz a následně specifikovat kontext (viz obrázek), kde v rámci slovního druhu zvolíte rozsah kontextu **-1** – adjektivum se nachází na první pozici vlevo od KWIC – a označíte *přídavné jméno*. Následně stiskněte *Enter*, nebo klikněte na *Hledat*.

The screenshot shows a search interface with two main sections: "Lemmata v kontextu" and "Slovní druhy v kontextu".

Lemmata v kontextu:

- Section: "Lemmata v kontextu"
- Label: "Rozsah kontextu:"
- Context range: A row of buttons from -5 to 5, with "KWIC" in the center. The "-1" button is highlighted with a red circle.
- Label: "Seznam lemmat oddělených mezerami:"
- Input field: Empty
- Dropdown: "musí být přítomny všechny položky"

Slovní druhy v kontextu:

- Section: "Slovní druhy v kontextu"
- Label: "Rozsah kontextu:"
- Context range: A row of buttons from -5 to 5, with "KWIC" in the center. The "-1" button is highlighted with a red circle.
- Label: "PoS list:"
- Radio buttons for parts of speech: "podstatné jméno", "přídavné jméno" (highlighted with a red circle), "zajmeno", "číslovka", "sloveso", "příslovce", "předložka", "spojka", "částice", "citoslovce", "interpunkce", "neznámý".
- Dropdown: "musí být přítomny všechny položky"

V dalším kroku, abyste získali seznam nejčastějších adjektiv, vyberte z nabídky *Frekvence > Vlastní....* Zobrazí se vám okno, v němž vyberte atribut „lemma“ a pozici „1L“ (adjektivum se nachází na první pozici vlevo).

The screenshot shows the "Frekvenční distribuce" window with the following settings:

- Buttons: "Běžná", "Podle typů textů", "Závislost dvou atributů"
- Label: "Frekvenční limit:"
- Input field: "1"
- Table with columns: "Úroveň", "Atribut", "Nerozlišovat velikost", "Pozice", "(Node) začít od"
- Row 1: "1.", "lemma" (highlighted with a red circle), a checkbox, "1L" (highlighted with a red circle), "slova KWIC nejvíce vlevo"
- Button: "+"
- Button: "Vytvořit frekvenční seznam"

..... věk

..... muž

..... nemocnice