

# **Využití korpusů**

**při výuce češtiny jako cizího jazyka**

**PLIN022**

**1./2.**

**Jak začít s korpusem**

# Zasina, A. J. (2023)

## Úvod. Jak začít s korpusem

Aby byl uživatel schopen plně využívat všechny dostupné funkce portálu <https://www.korpus.cz>, je nutná registrace, která je bezplatná. Registrace je velmi jednoduchá; návod je k dispozici zde: <https://wiki.korpus.cz/doku.php/kurz:zaciname?rev=1533556421>. Potřebujete-li se blíže seznámit se způsobem vyhledávání v korpuse, můžete projít on-line kurzem práce s korpusem s rozhraním v sedmi lekcích (<https://wiki.korpus.cz/doku.php/kurz:uvod>). Nemusíte se bát, pokud něčemu nerozumíte – vždy se můžete obrátit na podporu a položit libovolný dotaz (<https://podpora.korpus.cz/>). K dispozici je rovněž repozitář korpusem cvičení Pro školy (<https://korpus.cz/proskoly>), který je sice zaměřen na rodilé mluvčí, nicméně najdete zde i některé pracovní listy vhodné pro cizince.

ADRIAN JAN ZASINA

**KORPUSOVÁ CVIČEBNICE  
PRO STUDENTY ČEŠTINY  
JAKO CIZÍHO JAZYKA**

**Zadání:** Seznamte se s vyhledáváním v korpusu v první lekci on-line kurzu ([https://wiki.korpus.cz/doku.php/kurz:prvni\\_dotaz?rev=1615195905](https://wiki.korpus.cz/doku.php/kurz:prvni_dotaz?rev=1615195905)). Kurz vás provede základním vyhledáváním a typy dotazů v nástroji KonText.

**Zadání:** Nyní zkuste v rozhraní KonText (<https://www.korpus.cz/kontext>) vyhledat seznam kolokací doporučeným způsobem. K substantivům *věk*, *muž*, *nemocnice* najděte jejich tři nejčastější adjektivní kolokáty na základě dat z korpusu SYN2020.

\_\_\_\_\_ věk                      \_\_\_\_\_ muž                      \_\_\_\_\_ nemocnice

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 (klíčového slova v kontextu) – a označíte přídavné jméno. Následně stiskněte *Enter*, nebo klikněte na *Hledat*.

Specifikovat kontext

## Lemmata v kontextu

Rozsah kontextu:

... -5 -4 -3 -2 -1 KWIC 1 2 3 4 5 ...

Seznam lemmat oddělených mezerami:

musí být přítomny všechny položky

## Slovní druhy v kontextu

Rozsah kontextu:

... -5 -4 -3 -2 -1 KWIC 1 2 3 4 5 ...

PoS list:

- částice
- číslovka
- citoslovce
- cizí slovo
- interpunkce
- neznámý
- podstatné jméno
- předložka
- přídavné jméno
- příslovce

musí být přítomny všechny položky

Abyste získali seznam nejčastějších adjektiv, vyberte v dalším kroku z nabídky *Frekvence* > *Vlastní...* Zobrazí se vám okno, v němž zvolíte atribut *lemma* (základní slovníkový tvar) a pozici *1L* (adjektivum se nachází na první pozici vlevo).

Frekvenční distribuce x

Běžná Podle typů textů Závislost dvou atributů

Frekvenční limit:

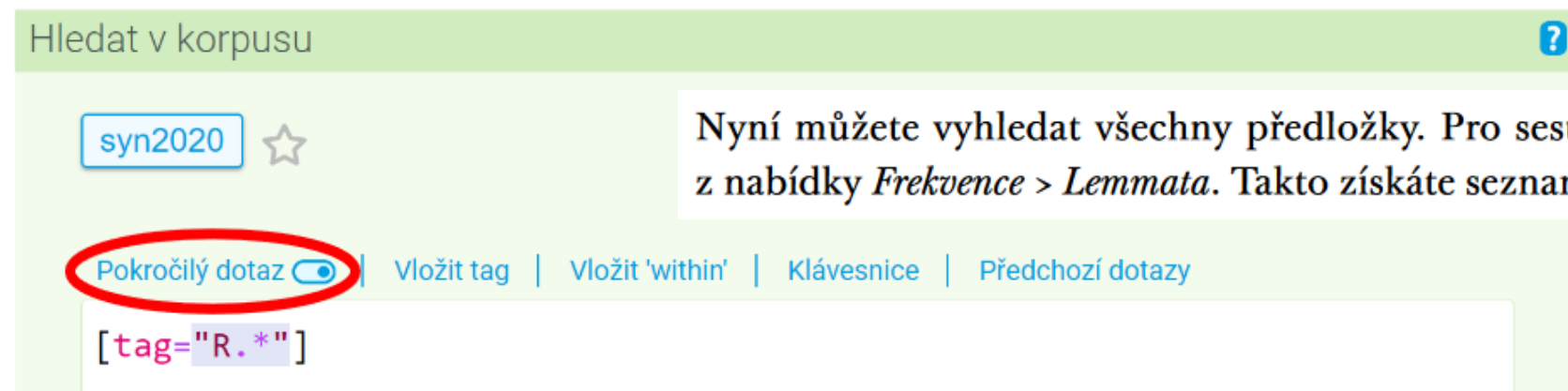
Úroveň	Atribut	Nerozlišovat velikost	Pozice	(Node) začít od
1.	<input type="text" value="lemma"/>	<input type="checkbox"/>	<input type="text" value="1L"/>	<input type="text" value="slova KWIC nejvíce vlevo"/>

**Zadání:** V korpusovém rozhraní KonText (<https://www.korpus.cz/kontext>) existují dva typy dotazu: jednoduchý a pokročilý. V jednoduchém typu dotazu můžete vyhledávat jednotlivá slova nebo slovní spojení. V pokročilém typu dotazu můžete např. vyhledávat podle specifikace gramatických kategorií. Zkuste vyhledat v korpusu SYN2020 předložky a vypsat pět nejčastějších na základě frekvenčního seznamu.

Nečastější předložky v korpusu SYN2020

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_
4. \_\_\_\_\_
5. \_\_\_\_\_

Ve vyhledávacím formuláři přepněte dotaz na pokročilý. Hned vedle se vám zobrazí funkce *Vložit tag*, kterou použijete pro vložení tagu označujícího předložku. Stačí vybrat z nabídky slovní druh a následně kliknout na *Vložit*.



Hledat v korpusu ?

syn2020 ☆

Nyní můžete vyhledat všechny předložky. Pro sestavení frekvenčního seznamu vyberte z nabídky *Frekvence > Lemmata*. Takto získáte seznam nejčastějších předložek v korpusu.

Pokročilý dotaz  | Vložit tag | Vložit 'within' | Klávesnice | Předchozí dotazy

[tag="R.\*"]

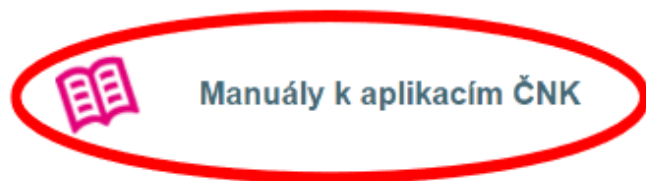
# KORPUSOVÉ NÁSTROJE

Pro práci s korpusovými daty lze využít několik různých nástrojů. Každý nástroj je vhodný pro jiné účely. V rámci Českého národního korpusu existuje několik aplikací, jejichž nabídka stále roste. Aktuální přehled všech nástrojů najdete na stránkách korpusové Wiki <https://wiki.korpus.cz> (viz obrázek). Aplikace jsou také aktualizovány, proto je vždy nejlepší seznámit se s aktuální verzí na webových stránkách.

## Jaké informace zde najdete?

---

Tato wiki slouží uživatelům ČNK nejen jako zdroj informací o ČNK (popis zveřejněných korpusů a dokumentace k nim, manuály k aplikacím), ale také jako průběžně doplňovaná báze znalostí o korpusové lingvistice. Hlavní části wiki tvoří tyto oddíly:



Přehled korpusů dostupných v rámci ČNK



Kurz práce s korpusem v 7 lekcích



Slovníček základních pojmů z korpusové lingvistiky



Seznamy značek, slov, nástrojů a zdrojů

S korpusovým vyhledávačem KonText jste se seznámili v předchozí části. Nyní se zaměříme na další aplikace, jež jsou vhodné pro využití ve výuce. Krátká zadání vám přiblíží základní funkcionalitu. Všechny aplikace jsou dostupné na portále korpusu z horní lišty aplikací; stačí kliknout na název příslušné aplikace.

**Slovo v kostce (WaG):** přehled informací o slově  
(<https://wiki.korpus.cz/doku.php/manualy:wag>)

**Zadání:** Zadejte do nástroje **Slovo v kostce** dotaz *oko* a zodpovězte níže uvedené otázky.

Jaký je nejčastější slovní tvar lexému *oko*? \_\_\_\_\_

V jakém typu textu se lexém *oko* nejčastěji vyskytuje? \_\_\_\_\_

Jaká je nejčastější kolokace lexému *oko* v psaném jazyce? \_\_\_\_\_

**SyD:** porovnávání variant slov či slovních spojení v psané i mluvené češtině  
(<https://wiki.korpus.cz/doku.php/manualy:syd>)

**Zadání:** Podívejte se do nástroje **SyD**, která varianta *sýra* × *sýru* je častější v psaném a mluveném jazyce.



**Treq:** překladové ekvivalenty pro ČJ/AJ a další jazyky

(<https://wiki.korpus.cz/doku.php/manualy:treq>)

**Zadání:** Představte si, že studujete francouzštinu a potřebujete přeložit slovíčka. V nástroji

**Treq** najděte francouzské překladové ekvivalenty pro následující lexémy:

bouřka \_\_\_\_\_ prázdniny \_\_\_\_\_  
stáž \_\_\_\_\_ chřipka \_\_\_\_\_

**Morfio:** vyhledávání slovotvorných vztahů

(<https://wiki.korpus.cz/doku.php/manualy:morfio>)

**Zadání:** S pomocí nástroje **Morfio** zjistěte, která slova se společnou částí mohou končit jak na -ovitý, tak i -ný.

společná část	odlišná část
.+	ovitý
	ný