

3. lekce: Hledání v Českém Národním Korpusu

Cíly těchto dvou posledních lekcí

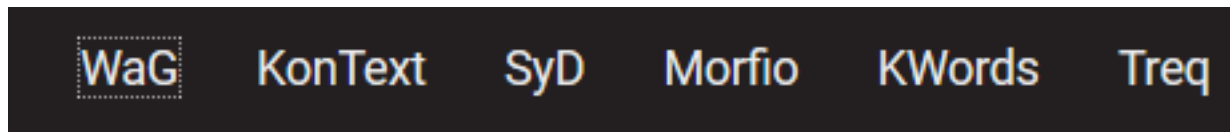
- Zkoumat valenci českých sloves pomocí korpusy a porovnávat ji s valencí sloves neznámého slovanského jazyka: slovinština.
- Seznamovat se s několika aplikacemi Českého Národního Korpusu (ČNK)

1. úkol

- Hledat ve Vallexu valenci těchto sloves komunikace a kopírovat jejich rámeček do wordu (Stačí jejich 1. význam, tj. 1. rámeček):
 - lhát
 - mlčet
 - mluvit
 - odpovědět
 - omlouvat
 - zavolat

Seznámení se s aplikacemi Českého Narodního Korpusu (ČNK)

- Jděte na stránky ČNK: <https://www.korpus.cz/>
- Není to nutné v rámci této lekce, ale je lépe se tam registrovat, aby byl přístup k aplikacím snadnější
- Nahoře máte přístup k hlavními aplikacemi



- V rámci této lekce pracovat s aplikací KonText, ale můžete se dívat nejdříve na ostatní aplikace a je krátce otestovat.

2. úkol - Jednoduché hledání v KonTextu

Typ dotazu: Základní ?

Klávesnice | Předchozí dotazy

Dotaz: Ihát
Na hodnotu tagu lze kliknout s př

► Specifikovat kontext

► Omezit hledání

Hledat

- Jděte na aplikaci KonText a napište dotaz na sloveso „Ihát“, pak klikněte na „hledat“.
- Mezi výsledky najděte a kopírujte do wordu 2 příklady, které odpovídají rámci tohoto slovesa ve Vallexu (obsahující alespoň jeden ze funktorů označených s „opt“). Příklady musejí být snadno čitelné, občas stačí kus věty.
- Opakujte to samé se slovesem „mlčet“.

Komplexní hledání v KonTextu

- Vyberte tentokrát „CQL“ jako typ dotazu.

Hledat v korpusu

Korpus: ☆

Typ dotazu: ?

Dotaz: (popis morf. značek [🔗](#))

within' | Klávesnice | Předchozí dotazy

👉 můžete vložit další řádek stisknutím klávesy Shift+ENTER ▶

CQL

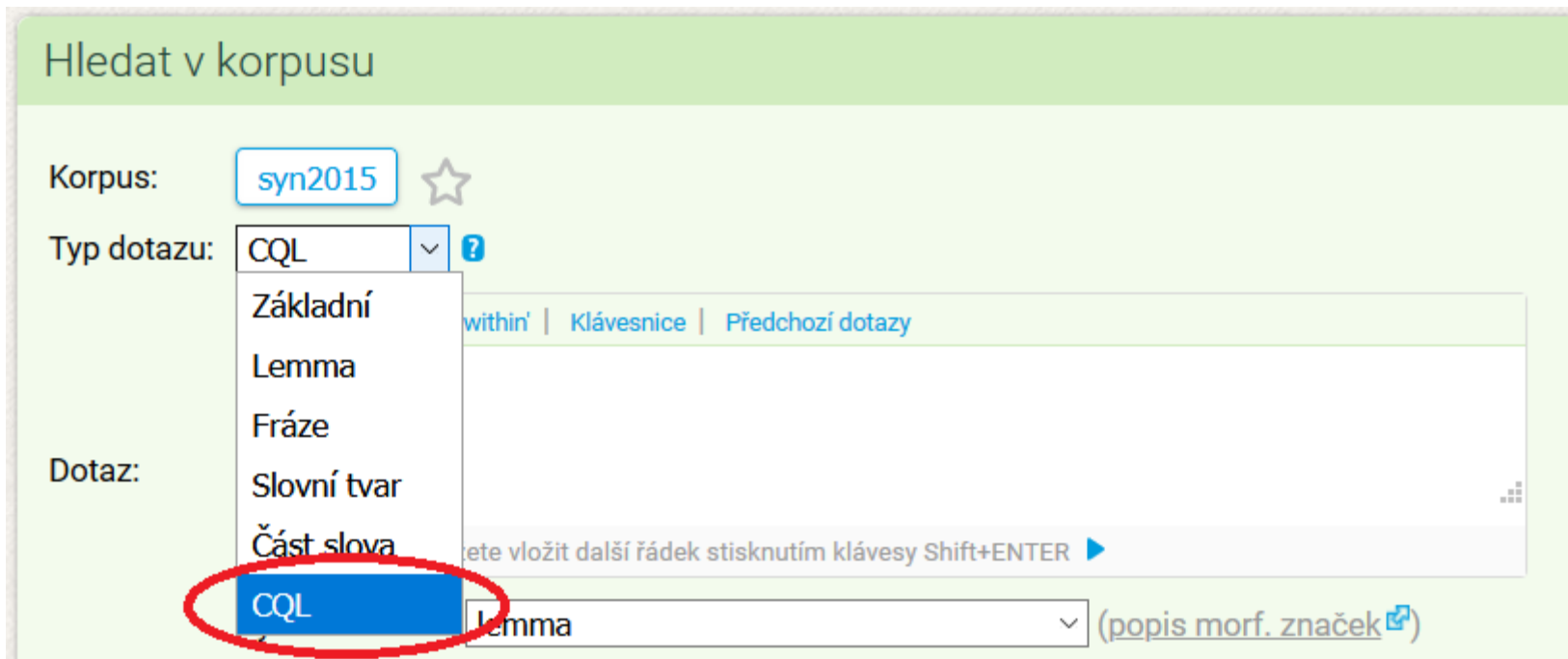
Základní

Lemma

Fráze

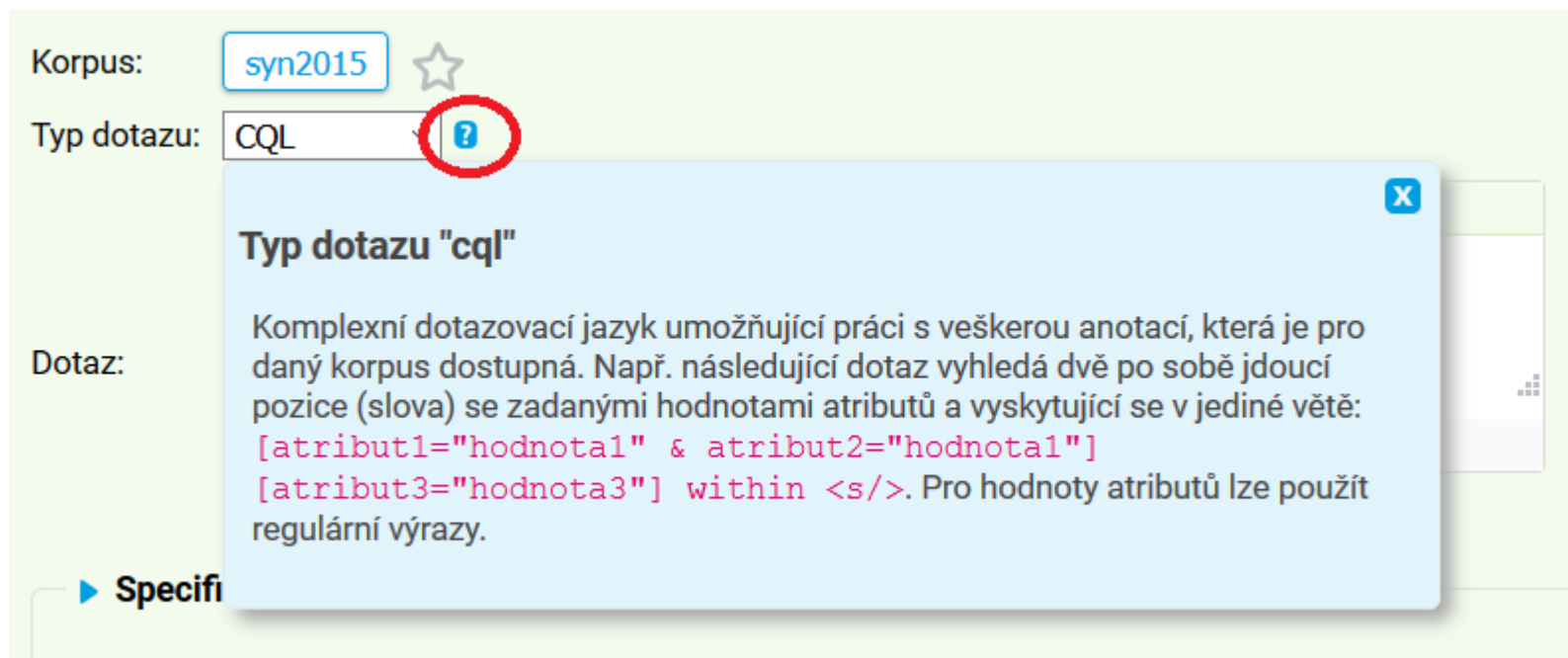
Slovní tvar


Část slova




Komplexní hledání v KonTextu

- Můžete se cestou získat vysvětlení o různých typech dotazů pomocí otazníku



Korpus: 

Typ dotazu: 

Dotaz:

Typ dotazu "cql"

Komplexní dotazovací jazyk umožňující práci s veškerou anotací, která je pro daný korpus dostupná. Např. následující dotaz vyhledá dvě po sobě jdoucí pozice (slova) se zadanými hodnotami atributů a vyskytující se v jediné větě:
`[atribut1="hodnota1" & atribut2="hodnota1"]
[atribut3="hodnota3"] within <s/>`. Pro hodnoty atributů lze použít regulární výrazy.

► Specifi

Komplexní hledání v KonTextu

- Díky takovému hledání můžete lépe cílit výsledky, které potřebujete.
- Ve Vallexu má většina sloves několik rámců, tady chcete mít výsledky, které jsou lépe cílené na 1. rámeček každého slovesa.
- Nejdříve se podívejte na následující příklad

příklad se slovesem „dělat“

- „dělat“ má celkem 23 rámců ve Vallexu. Tady nás zajímá jenom 1. rámeček, ve smyslu „pracovat, provádět, věnovat se něčemu“.:
dělat (si)^{impf}, dělávat (si)^{iter}

① limit dělat^{impf}, dělávat^{iter}
pracovat; provádět; věnovat se něčemu
frame **ACT₁^{obl} PAT₄^{obl} MANN^{typ} BEN_{3,pro+4}^{typ}**
example dělal to, co měl dělat, poradně; dělal svou práci rychle; dělal test / zkoušky; dělat film / divadlo / noviny

② limit dělat^{impf}, dělávat^{iter}
přetvářet; připravovat; vytvořit
frame ACT₁^{obl} PAT₄^{obl} ORIG_{z+2}^{opt} BEN_{3,pro+4}^{typ}
example dělal dětem loutku z papíru

③ limit dělat^{impf}, dělávat^{iter}
působit změnu; stanovovat
frame ACT₁^{obl} PAT₄^{obl} EFF_{7,adj-7}^{obl}
example dělali ho učitelem

④ limit dělat^{impf}, dělávat^{iter}
pracovat (zaměstnání)

příklad se slovesem „dělat“


- V tom rámci, sloveso „dělat“ má vždy objekt v akuzativu a může typicky mít doplnění označující způsob a příjemce.
- Pokud používáme základní dotaz, bude mezi výsledaty těžko najít zajímavé příklady kvůli tomu, že má sloveso „dělat“ tolik možných valencí.


ovat; provádět; věnovat se něčemu
e **ACT^{obl}₁ PAT^{obl}₄ MANN^{typ} BEN^{typ}_{3,pro+4}**
anle dělal to, co měl dělat, napřed dělal svou práci

příklad se slovesem „dělat“

Abychom dostali omezenější a jasnější výsledky, můžeme s typem dotazů CQL hledat přesný typ slov, který chceme najít např. po slovesu „dělat“ .


1. píšeme v uvozovkách lemma, které nás tady zajímá, tj. sloveso „dělat“


Korpus: 

Typ dotazu: 

[Vložit tag](#) | [Vložit 'within'](#) | [Klávesnice](#) | [Předchozí dotazy](#)

Dotaz:

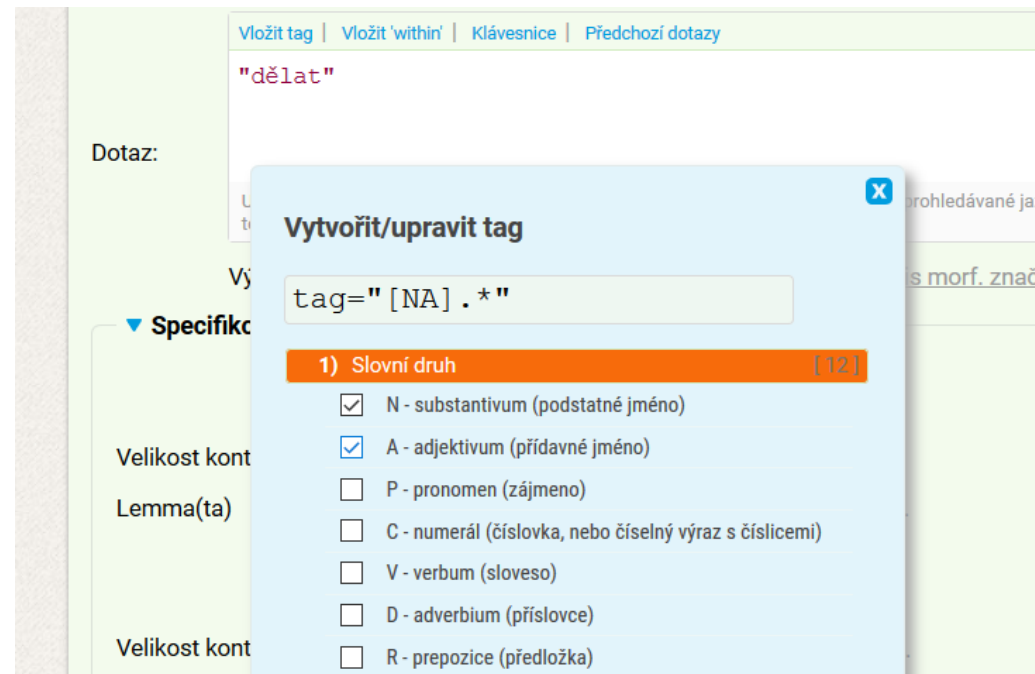
U zarovnaných korpusů je rychlejší hledat jen v textech společných pro všechny prohledávané jazyky. Najít tyto texty lze pomocí funkce 'omezit hledání - zúžit výběr' 

Výchozí atribut:  [\(popis morf. značek !\[\]\(f7111f351a527d2f0cf6f01bbd71c182_img.jpg\)\)](#)

příklad se slovesem „dělat“

2. Klikneme na „vložit tag“. Tím budeme postupně vybírat vlastnosti následujícího slova. V našem případě chceme podstatní nebo přídavné jméno v akuzativu:

Klikneme tedy na „slovní druh“, a pak „substantivum“ a „adjektivum“



The screenshot shows a web application interface with a search bar containing the word "dělat". Below the search bar, there is a dropdown menu with the option "Vložit tag" selected. A dialog box titled "Vytvořit/upravit tag" is open, showing a text input field with the tag "tag=" [NA] . * ". Below the input field, there is a list of word classes (Slovní druh) with checkboxes:

- N - substantivum (podstatné jméno)
- A - adjektivum (přídavné jméno)
- P - pronomen (zájmeno)
- C - numerál (číslovka, nebo číselný výraz s číslicemi)
- V - verbum (sloveso)
- D - adverbium (příslovce)
- R - prepozice (předložka)

příklad se slovesem „dělat“

Níže ještě na „pád“,
a pak „akuzativ“.

Máme tím všechny potřebné
specifikace, můžeme na konci
ještě kliknout na „vložit“, úplně
dolů

tag=" [NA] . . . 4 . * "

1) Slovní druh [12]

2) Detailní určení slovního druhu [6]

3) Jmenný rod [4]

4) Číslo [2]

5) Pád [8]

1 - nominativ (1. pád)

2 - genitiv (2. pád)

3 - dativ (3. pád)

4 - akuzativ (4. pád)

5 - vokativ (5. pád)

6 - lokativ (6. pád)

7 - instrumentál (7. pád)

Nevyplněno

6) Přivlastňovací rod [3]

7) Přivlastňovací číslo [0]

8) Osoba [0]

9) Čas [0]

10) Stupeň [4]

11) Negace [3]

12) Aktivum/pasivum [0]

13) Nepoužito [0]

14) Nepoužito [0]

15) Varianta, stylový příznak apod. [5]

16) Vid [0]

Vložit

Krok zpět

Obnovit

příklad se slovesem „dělat“

Jestli chceme ještě specifikovat kontext kolem slovesa „dělat“, můžeme přidat k hledání volnější kritéria, která se mohou nacházet před slovesem nebo po něm více nebo méně daleko od něj.

▼ Specifikovat kontext

Filtrovat podle lemmatu

Velikost kontextu: vlevo i vpravo ▾ 5 ▾ tokenů.

Lemma(ta) pro všechny ▾ z těchto položek.

Slovní druh

Velikost kontextu: vlevo i vpravo ▾ 3 ▾ tokenů.

Slovní druh:
(použijte Ctrl+klik pro vícenásobný výběr)

- číslovka
- sloveso
- příslovce**
- předložka
- spojka

všechny ▾ z těchto položek.

příklad se slovesem „dělat“

Tady chceme najít rezultat obsahující předložku „pro“, a to 5 slov před nebo po slovesu „dělat“. Dále chceme i mít v resultátu příslovce, 3 slova před nebo po slovesu „dělat“. Tím zvětšujeme šanci, abychom našli lepší příklad obsahující i způsob pomocí příslovce i příjemce, pomocí předložky „pro“.

Filtrovat podle lemmatu

Velikost kontextu: tokenů.

Lemma(ta) z těchto položek.

Slovní druh

Velikost kontextu: tokenů.

Slovní druh:
(použijte Ctrl+klik pro vícenásobný výběr)

- číslovka
- sloveso
- příslovce**
- předložka
- spoika

z těchto položek.

příklad se slovesem „dělat“

1. Tímto způsobem najdeme několik příkladů odpovídajících nejkompletnějšímu rámci slovesa dělat:

vědomě se snaží

dělat maximum

pro to , aby se okolí skutečnou pravdu nedozvědělo .

Společně s Diandrou

dělal rozhovor

s Michaelem Grossem pro New York Times

Nedokážu si představit , že bych

dělal film

pouze pro sebe

Komplexní hledání v KonTextu

- Tímto způsobem najděte pro každé zbyvající sloveso 3 příklady odpovídající jejich kompletním rámcem. Můžete vymýšlet a zkoušet různé funkce a kombinace, může být více než jeden správný způsob, jak je hledat.
- Pomocí funkce „Print screen“ (tlačítko na klavesnici většinou označeno s „PrtSc“), kopírujte do souboru celý dotaz, který jste k hledání použily.
- Kopírujte vybrané příklady z korpusu do souboru.

Shrnutí úkolů

Ve finální verzi souboru máte:

1. Valenční rámce 6 sloves
2. 2 věty z korpusu pro sloveso lhát, a 2 věty pro sloveso mlčet (výsledek jednoduchého hledání)
3. 3 věty z korpusu pro každé ostatní sloveso + dotazy, které jste vymysleli, abyste našli tyto věty

Věty musejí odpovídat 1. rámci každého slovesa

Ještě pár otázek

- Jste už dříve použili korpus? Který? V rámci jaké práce?
- Z ostatních aplikací dostupných na ČNK (WaG, SyD, Morfio, Kwords, TreQ), která se vám zdá nejzajímavější? Proč?

- Tento úkol pošlete ve formátu mojeJméno-tretilekce.doc(/.odt...) na adrese gregoire.labbe@ff.cuni.cz nejpozději do 22.05.2020 včetně.
- Neváhejte se ptát v případě, že máte otázku!