1. setkání

Úvod, požadavky

O interaktivní osnově – bude v ní:

* Rekapitulace setkání
* Úkoly na druhé
* Vysvětlení něčeho obtížného

**I. Nastavení korpusu SYN2020** – společná práce pomocí otázek:

1. Přejít na [www.korpus.cz](http://www.korpus.cz)
2. Otázka: Podívejte se na lištu (ukázat ji): na který odkaz kliknete? – odpověď: **Přejít na KonText**
3. Jak nastavíte korpus SYN2020?: – **Zkontrolovat obdélník s názvem korpusu – vlevo, vedle něj je hvězdička**

Pokud je tam SYN2020, je to v pořádku. Pokud není, nastavíme ho:

Kliknout na název korpusu → Všechny korpusy → Rada SYN → dole je SYN2020

Kliknout na něj.

**II. Hledání podle atributů** – společná práce pomocí otázek:

Které atributy jsou vysvětleny ve Vašem materiálu?

* Rozdíl lemma x sublemma; k čemu je dobré „sublemma“?
* Lemma – word

POZOR, POKUD NEMÁME „POKROČILÝ DOTAZ“, MUSÍME ATRIBUT NASTAVIT-

Co najdete, když zadáme *psy* a nespecifikujeme?

Korpus dotaz pochopí jako „word“ a vyhledá jen tvar *psy*.

A co najdeme, když hledáme pes a nespecifikujeme?

Korpus to pochopí jako „lemma“.

Pokud chceme skutečně jen 1. pád, musíme zadat explicitně „word“.

K čemu je „pokročilý dotaz?“

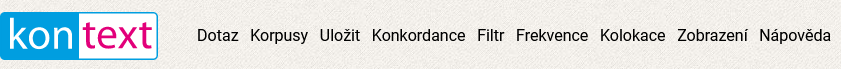
ODBOČKA – zopakování „pokročilého dotazu“.

Vysvětlení frekvence – využijeme dotaz se sufixem *-ina*.

Po vyhledání slov končících na *.\*ina* vytvoříme frekvenční seznam.

Jak?

NA LIŠTĚ NAJDEME ODKAZ **Frekvence**.



Když na něj klikneme, objeví se sloupec, v něm je na prvním místě: **lemmata**

Na tento odkaz klikneme a objeví se seznam, který máte ve studijním materiálu.

Praktické úkoly

I.

**a) Nejdříve si pro sebe zodpovězte otázky:**

V kterých případech jde o dvě varianty téhož slova?

V kterých případech jde o dvě správně napsaná slova s odlišným významem?

V kterých případech je jedno slovo napsáno správně, a jedno slovo chybně?

U každé z dvojici si poznamenejte správnou odpověď.

archiv – archív

kurs – kurz

shlédnout – zhlédnout

smoknout – zmoknout

b) Pokuste se vždy obě varianty nebo obě slova vyhledat v Českém národním korpusu SYN2020 bez ohledu na jejich správnost.

Pročtěte si studijní materiál a navrhněte, který typ dotazu využijete.

**II.**

**Připomeňte si morfologii a vzpomeňte si:**

**a) Které tvary mohou mít v 7. pádě množného čísla názvy:**

*loď*

**Které tvary mohou mít v 6. pádě množného čísla názvy:**

*jablko*

*pravomoc*

Které infinitivní tvary se užívají u slovesa, jehož jeden tvar je *peče*? (Pozor, u infinitivu nezapomeňte na záporné tvary)

b) Tvary, které se podle Vás vyskytují, zkuste vyhledat v Českém národním korpusu SYN2020 bez ohledu na to, jsou-li správné, nebo ne.

Pročtěte si studijní materiál a navrhněte, který typ dotazu využijete.

3 skupiny:

**první skupina:**

**a)** archiv – archív VARIANTY

kurs – kurz VARIANTY

shlédnout – zhlédnout DVĚ RŮZNÁ SLOVA

smoknout – zmoknout SLOVO smoknout NEEXISTUJE

**druhá skupina:**

**a) Které tvary mohou mít v 7. pádě množného čísla názvy:**

*loď – LOĎMI a LODĚMI*

**Které tvary mohou mít v 6. pádě množného čísla názvy:**

*jablko* – JABLCÍCH, JABLKÁCH

*pravomoc* – PRAVOMOCÍCH, PRAVOMOCECH

Které infinitivní tvary se užívají u slovesa, jehož jeden tvar je *peče*? (Pozor, u infinitivu nezapomeňte na záporné tvary): PÉCT/NEPÉCT PÉCI/NEPÉCI

.

A vyhledáváme.

**I.**

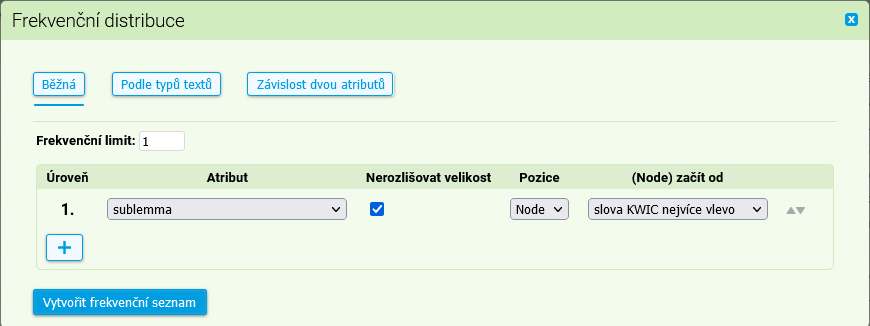
**Dotazy**

1. **možnost**

[lemma="arch.v"]

+ frekvence sublemmat

Frekvence – vlastní



Výsledek:



1. možnost – je rychlejší

[sublemma="archív"] – objeví se 163 řádků

[sublemma="archiv"] – objeví se 3744 řádků

Podobně postupujeme u variant kurs – kurz:

1)[lemma="kur."]

+ frekvence sublemmat

Frekvence – vlastní



Zajímají nás jen varianty *kurz* a *kurs*.

nebo:

[sublemma="kurz"] – objeví se 7720 řádků

[sublemma="kurs"] – objeví se 501 řádků

Slovesa:

[lemma=".hlédnout"] + Frekvence lemmat:



Zajímají nás slovesa zhlédnout a shlédnout.

nebo:

[lemma="zhlédnout"] – objeví se 840 řádků

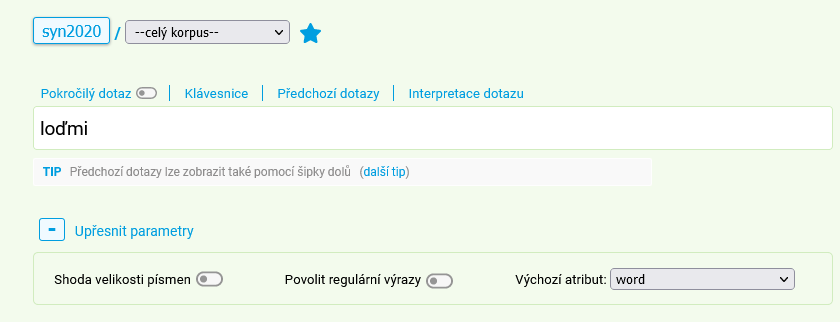
[lemma="shlédnout"] – objeví se 278 řádků

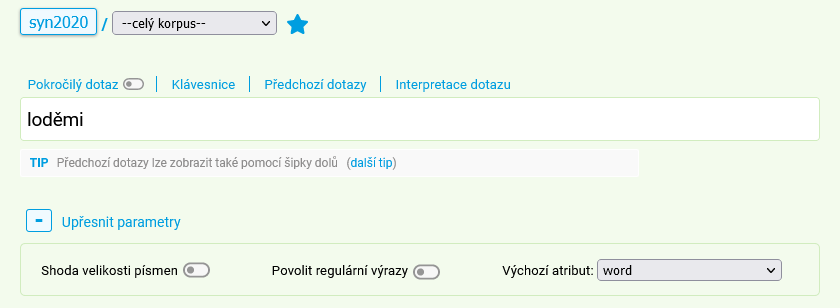
U slovesa zmoknout je postačí pouze:

[lemma=".moknout"]

**II.**

Zde budeme využívat dotaz „nepokročilý“, tj. zadání „pokročilý dotaz“ bude vyrušeno a my zvolíme jako výchozí atribut jen **word.** Nepotřebujeme ani frekvenci.





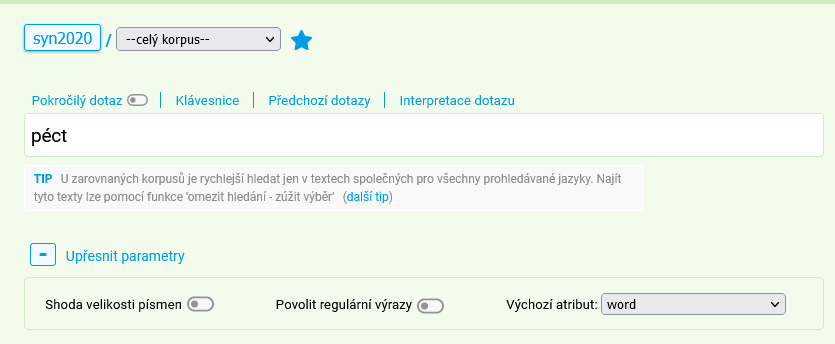
**Výsledek:**

loďmi 25 loděmi 173

jablcích 6 jablkách 35

pravomocích 8 pravomocech 5

**Sloveso:**



péct: 425 nepéct: 0

péci. 36 nepéci: 0

**třetí skupina:**

1. [lemma=”.\*yně”]

**Dotazy „word“, „lemma“, „sublemma“**

Viz přiložený materiál

**A. Kdy využijeme dotaz typu „slovní tvar“**?

Je vhodný tehdy, pokud srovnáváme např. výskyt dubletních tvarů jmen v jednom pádě nebo výskyt slovesných tvarů v jedné osobě ve stejném čísle.

**Příklad:**

balíček

hrníček

jablko

kolečko

Vzpomínáte si, v kterém pádě mají dvojtvary?

Lokál pl.

Balíčcích : balíčkách 49 : 1 (kodifikovány jsou oba)

Hrníčcích : hrníčkách 5 : 1

Jablcích : jablkách 6 : 35

Kolečcích : kolečkách 1 : 317

**Doma:**

Hříbek hříbcích : hříbkách 0 : 4

Rohlík rohlících : rohlíkách 7 : 1

Instr. Pl.

Loď loďmi : loděmi 25 : 173

Jíti – jít 42 : 17 071

Péci – péct 35 : 412

Doma

Létati – létat

Říci - říct

Sforma je zřejmě vpodstatě totéž

**Sublemma** se využívá u slov, jejichž základní tvar, tj. to, co hledáme jako „lemma“, má variantní podobu.

Příklad: diskuse

Když tento základní tvar zadáme jako lemma, vyhledá KonText jak podoby se „s“, tak podoby se „z“.

Když zvolíme výchozí atribut „lemma“ a zadáme podobu „diskuze“, nevyhledá korpus nic, protože lemmatem pro obě podoby je „diskuse“.

Chceme-li tedy zjistit, je-li častější „diskuse“, nebo „diskuze“, hledáme sublemma *diskuse* a pak sublemma *diskuze*. Ještě loni to nebylo možné, protože žádné sublemma nebylo. Museli jsme pokusem-omylem zjistit, která z dvou variant je lemmatem. Pak jsme ji zadali jako lemma, ve vyhledaných dokladech jsme zjistili frekvenci slovních tvarů a spočítali jsme varianty se „s“ a se „z“ – jinak to nešlo.

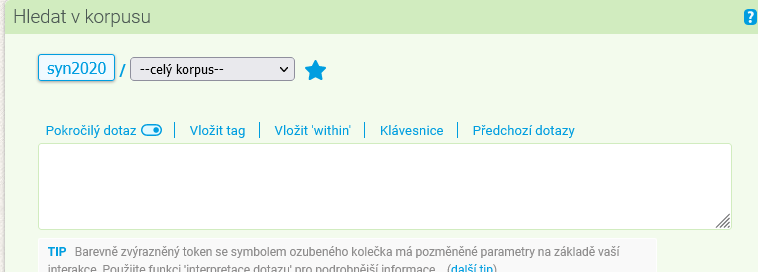
+ tag

A co když chceme vyhledat jen případy, kdy je homonymní slovo určitým slovním druhem?

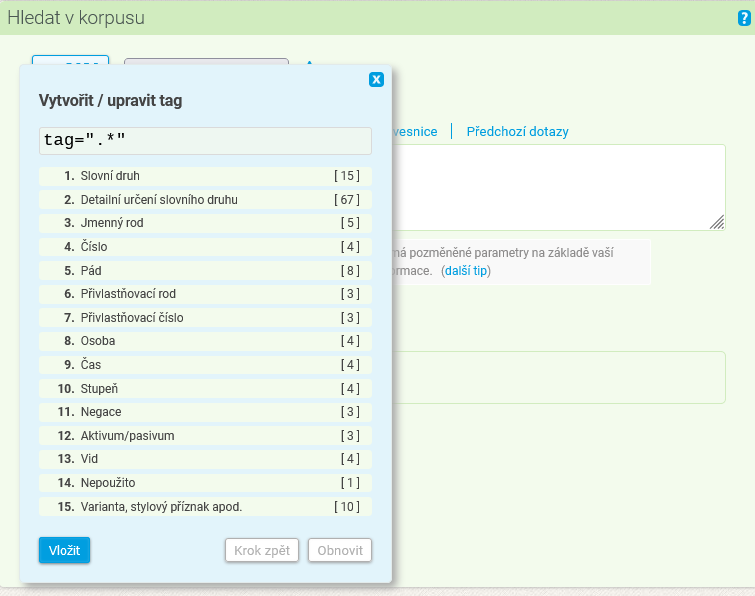
**IV. Které tvary mají substantiva *hotel*, *psycholog* v 6. pádě plurálu?**

Budeme využívat „Pokročilý dotaz“ a atribut tag = morfologické vlastnosti.

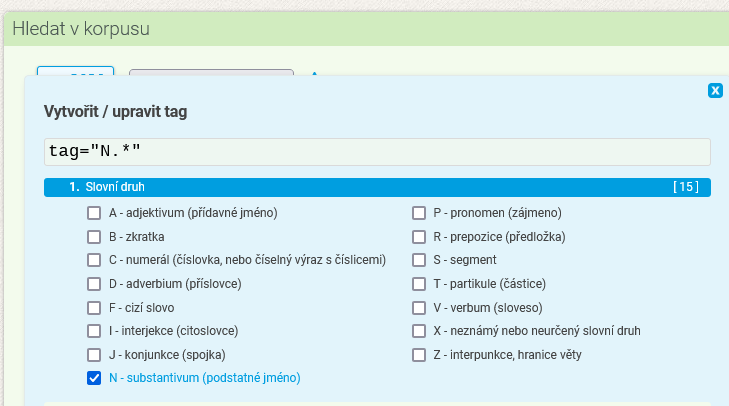
1. Odklikneme *Vložit tag*.



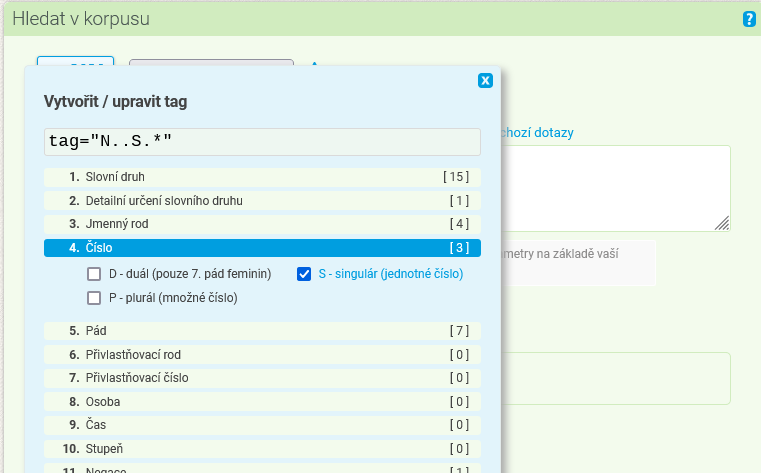
1. Objeví se nabídka, z níž využijeme Slovní druh (1), Číslo (4) a Pád (5)



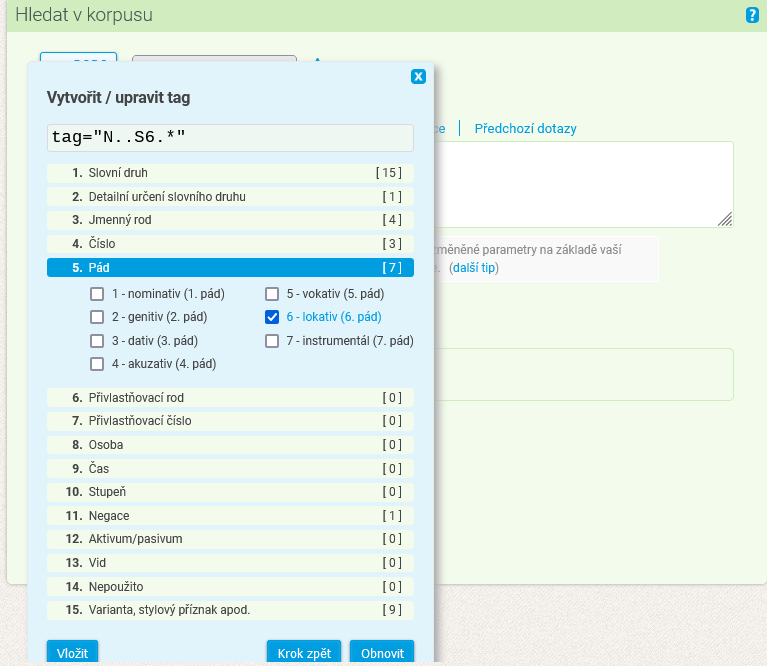
a)



b)

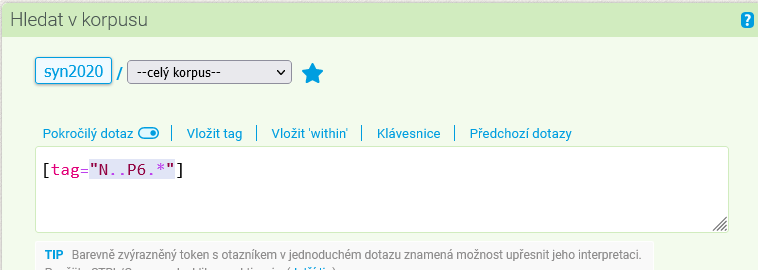


c)



Po zadání všech „tagových“ vlastností klikneme na odkaz Vložit.

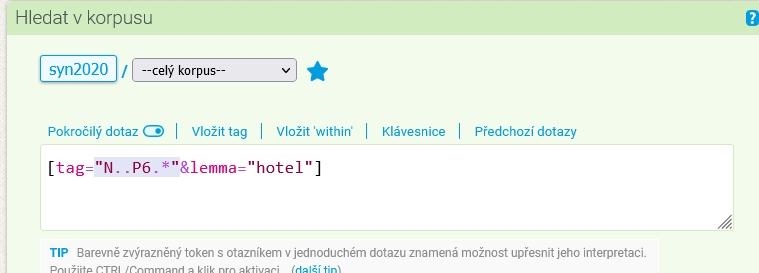
Dotazový řádek vypadá takto:



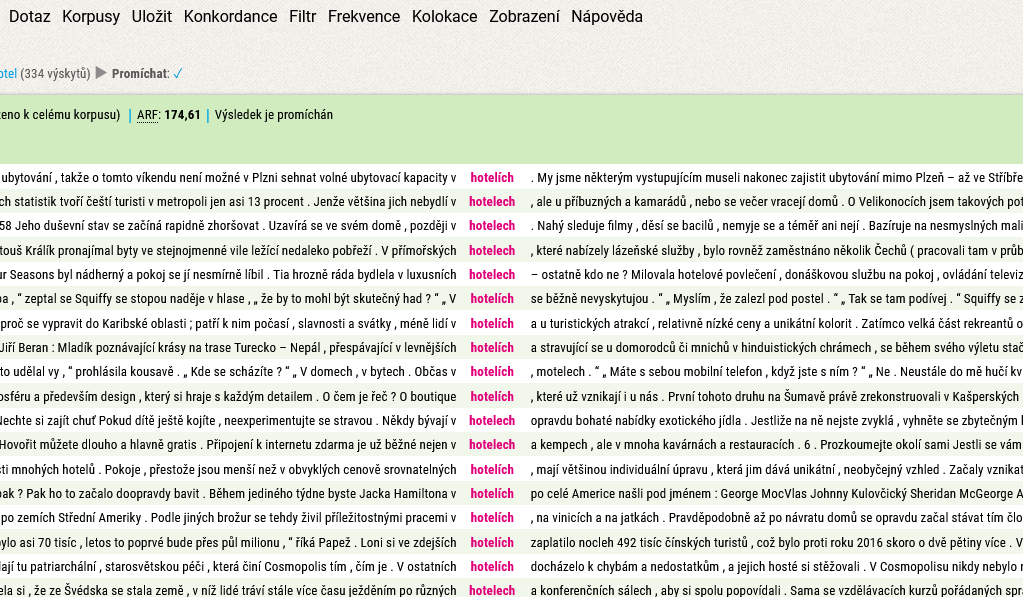
Vedle zakončujících uvozovek vložíme znak &, který vytvoříme klávesovou zkratkou Alt038.

Vedle znaku & napíšeme: lemma = „hotel“

Dotaz pak vypadá takto:



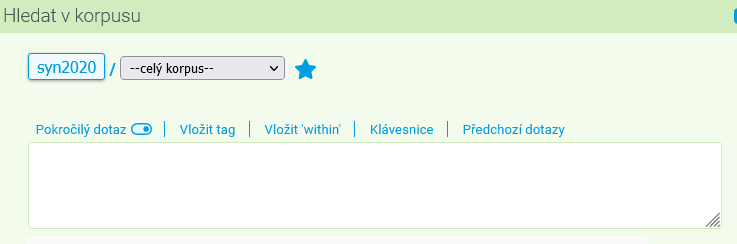
Po vyhledání dokladů využijeme Frekvenci slovních tvarů:

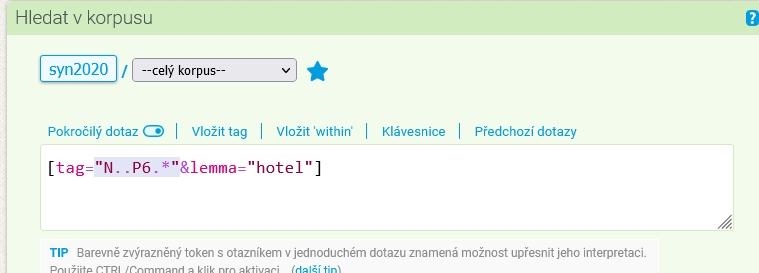




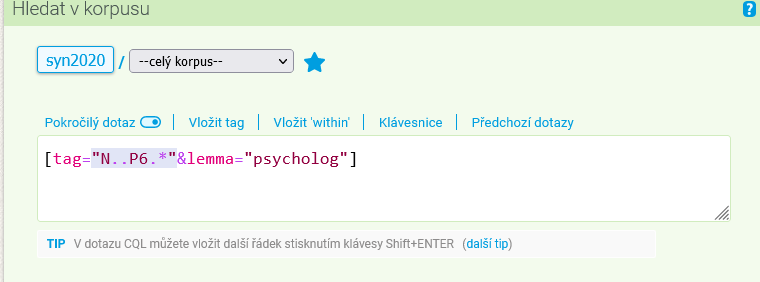
Zjistili jsme tedy, že v 6. p. pl. jsou tvary *hotelech* a *hotelích*, mezi jejichž výskytem není velký rozdíl.

Když budeme zjišťovat tvary slova psycholog, využijeme Předchozí dotazy. Zobrazí se vždy 10 posledních dotazů, vybereme si ten, který je nejblíže našemu dotazu.





Místo slova hotel napíšeme *psycholog*.



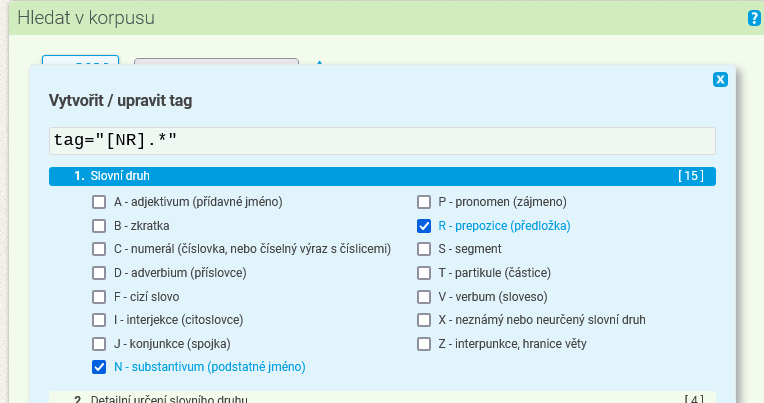
Po vyhledání dokladů zjistíme, že v 6. p. pl. je pouze tvar psycholozích a že se vyskytuje velmi málo (7x).

Jiný úkol:

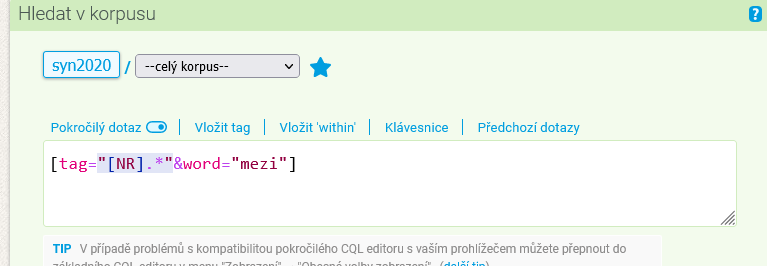
**Zjistěte, zda je slovo *mezi* častěji podstatným jménem, nebo předložkou.**

Postup: Využijeme „pokročilý dotaz“ a zkusíme hledat současně předložku i podstatné jméno.

1. Využijeme odkaz „Vložit tag“ a zvolíme slovní druh. Zaklikneme oba hledané slovní druhy současně – N (podstatné jméno) i R (předložka).



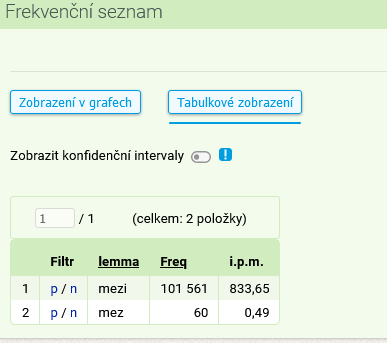
1. Tag vložíme, vedle něj napíšeme & a poté *word=“mezi“* – POZOR, MUSÍ TO BÝT word. U neohebných slov nezáleží na tom, zda volíme word nebo lemma, ale u podstatných jmen ano. A *mezi* je pouze jedním tvarem podstatného jména, jehož lemma je jiné.



Po odeslání dotazu získáme 101 621 dokladů.

POZOR, TEĎ VYUŽIJEME FREKVENCI **LEMMAT**. Tak poznáme, kdy je *mezi* wordem a současně lemmatem – to jsou ty doklady, v nichž je předložkou.

U dokladů, v nichž je mezi tvarem podstatného jména, je lemmatem *mez*.



předložka (lemmatem je *mezi*) podstatné jméno (lemmatem je *mez*, mezi je jeho slovní tvar)

**Jak můžeme zadávat syntaktické funkce (větné členy)**

To je atribut, který se zadává stejně, jako word, lemma, sublemma, tag, např. **afun=“Sb“, afun=“Obj“.**

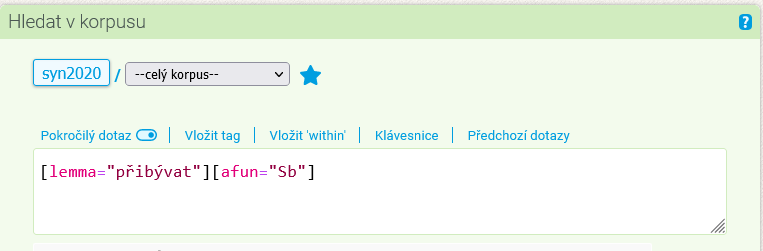
| **Afun** | **Definice** | **Poznámka** |
| --- | --- | --- |
| [Pred](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03.html#X_4_2_1_prisudek_predikat_pred_pnom_auxv_) | Predikát (přísudek) | pouze v hlavní větě |
| [Sb](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03s02.html) | Subjekt (podmět) |  |
| [Obj](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03s04.html) | Objekt (předmět) | včetně určení činitele děje u opisného pasiva; vč. infinitivu u modálních a fázových sloves aj. |
| [Adv](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03s05.html) | Adverbiale (příslovečné určení) | bez dalšího rozlišení, vč. předložkových pádů u slovesa *být*, některých volných dativů aj. |
| [Atv](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03s06.html) | Doplněk | jen tzv. určující; technicky zavěšen na neslovesném členu |
| [AtvV](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03s06.html) | Doplněk | jen tzv. určující; visící na slovese (ve větě není vyjádřen druhý řídící člen) |
| [Atr](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03s03.html) | Atribut (přívlastek) | včetně spojení typu ***Jan****Novák*, genitivu po číslovkách aj. |
| [Pnom](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03.html#X_4_2_1_prisudek_predikat_pred_pnom_auxv_) | Jmenná část verbonom. přísudku | se sponou *být, bývat*; mimo předložkové pády |
| [AuxV](http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s03.html#X_4_2_1_prisudek_predikat_pred_pnom_auxv_) | Pomocné sloveso *být* | (Auxiliary Verb) |

Spojovací znak, znamená „a současně“.

V. Vyhledejte věty se slovesem *přibývat* a podmětem. Je podmět častěji ve tvaru 1. pádu, nebo 2. pádu?

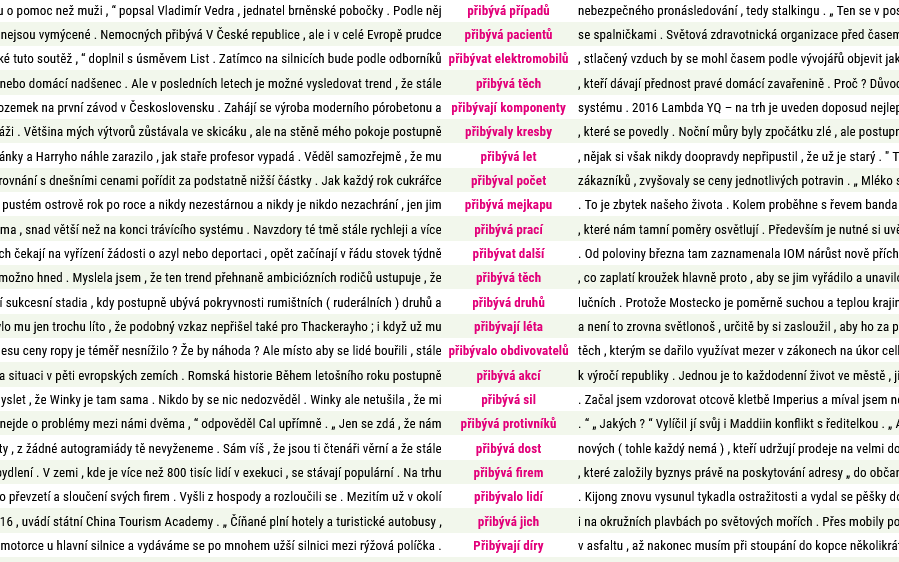
Zadáme dva atributy vedle sebe, každý bude v hranatých závorkách.

lemma=“přibývat“ afun=“Sb“.

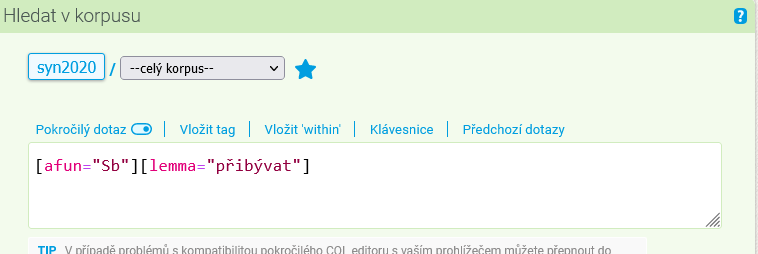


Dvě slova následující po sobě ve větě v tomto pořadí:

Část výsledku: převažuje podmět ve 2. pádě v 1. pádě je méně často



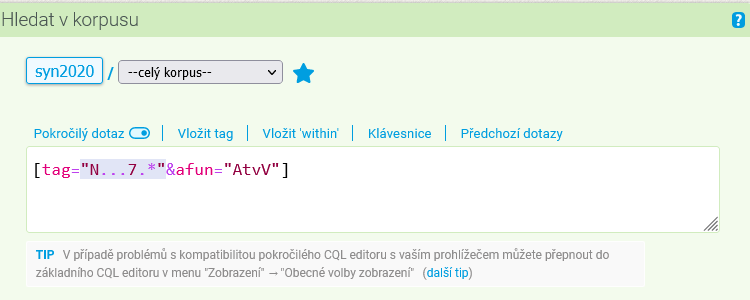
Můžeme podmět a sloveso přibývat zadat v opačném pořadí.





Jiný dotaz: Která podstatná jména bývají nejčastěji doplňkem v 7. pádě?

Dotaz:



Pozor, **tag** vytvoříme pokynem Vložit tag, ale **afun** musíme napsat, ten se takto vložit nedá.

Výsledek bohužel není přesný. Jen v několika případech jde skutečně o doplněk.



Často jde spíše o příslodečné určení způsobu (srovnání) – viz *obrůstáni jako břečtanem*, *jako zázrakem*, *jako mávnutím kouzelného proutku*.