Základy využití korpusu v praxi **cjbb75**

**Jak lze v korpusech hledat doklady pro výzkum morfologie?**

**Mějme tvrzení 1:**

**Všechna česká maskulina v genitivu plurálu mají koncovku na *–ů*.**

**Máme 6 vzorů: *pán/pánů, hrad/hradů, muž/mužů, stroj/strojů, předseda/předsedů, soudce/soudců.***

**Existují z tohoto pravidla nějaké výjimky?**

**Vzpomenete si na ně?**

**Jak jsou obvyklé (jsou to slova z centra/periferie slovní zásoby)?**

**Jak vyhledat data v korpusu:**

**Co hledáme?**

**1) tvary genitivu plurálu substantiv maskulin, tedy tvary, které mají značku (tag):**

**1. pozice: *N*** *jako substantivum (****n****oun)*

**2. pozice: *N*** *jako obyčejné substantivum (****n****oun)*

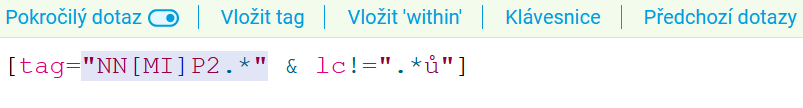
**3. pozice: *M*** *jako* ***m****askulinum životné* **nebo *I*** *jako maskulinum neživotné (****i****nanimatum)*

**4. pozice: *P*** *jako* ***p****lurál*

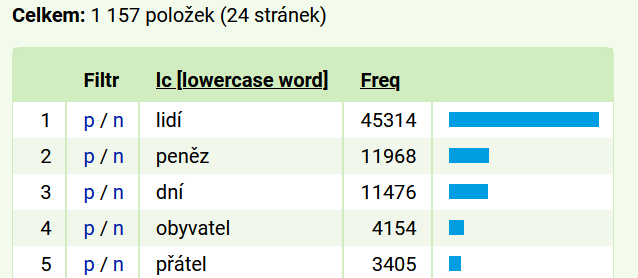
**5. pozice: *2*** *jako* ***2.*** *pád (genitiv)*

a zároveň

**2) tvary (bez ohledu na velikost písma - lc), které nekončí na *ů***



**Frekvence/Vlastní/lc**



**Jak lze výjimky z výše uvedeného pravidla dále kategorizovat?**

**Jednoduše je lze rozdělit podle tří zastoupených koncovek u prvních 20:**

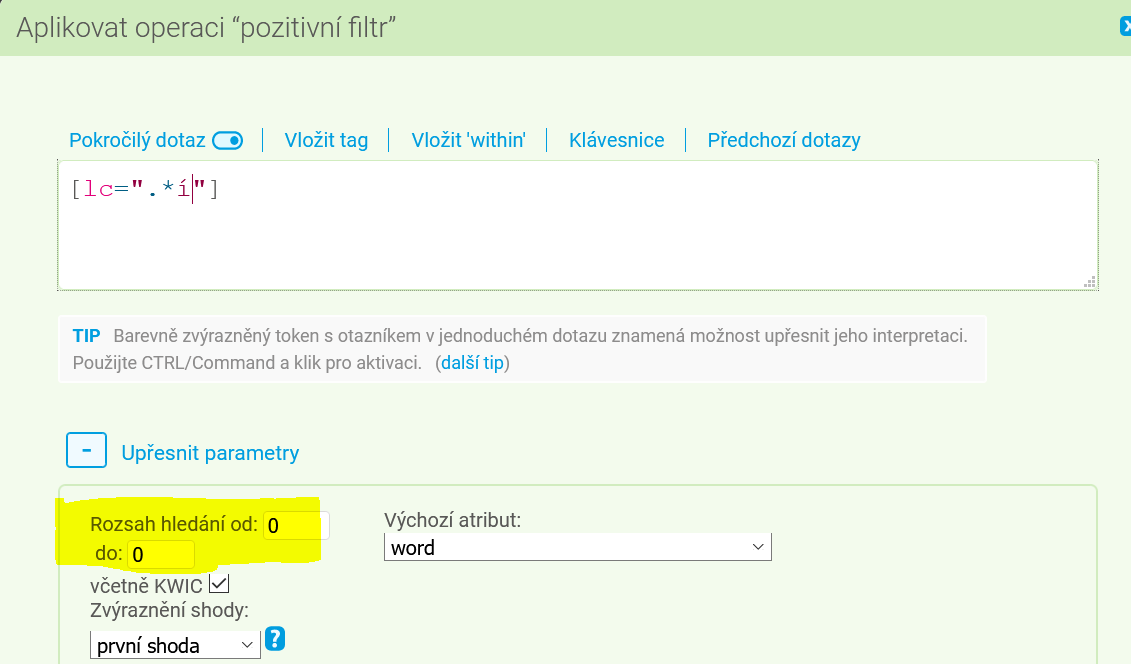
1. **končící na –í (typ *lidí/dní*)**
2. **bez koncovky/ s nulovou koncovkou (typ *peněz-0/přátel-0*)**
3. **s adjektivní flexí (typ *dospělých/zlatých*)**

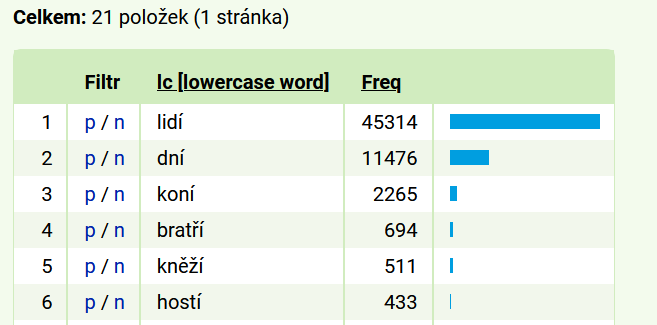
**Vidíme, že jsou zastoupeny oba rody, jak maskulina životná, tak neživotná.**

**Pozorujme jednotlivé typy z hlediska frekvence:**

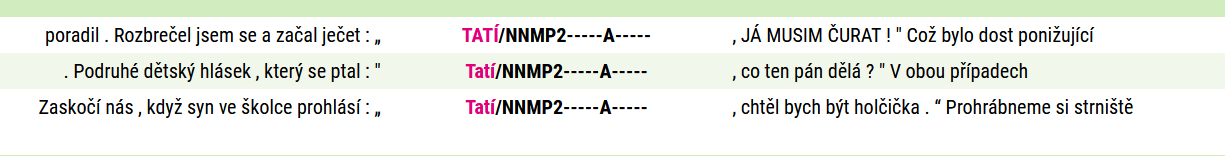
**Jak?**

**P-filtr**





**Chyby:**



Jak budeme formulovat dotaz, pokud chceme získat doklady, které ani nekončí na *í*, ani se neskloňují adjektivně?

[tag="NN[MI]P2.\*" & lc!=".\*([ůí]|[íý]ch)"]



**Co můžeme na základě pozorovaných dat říci o slovech s výjimečnou koncovkou u maskulin v gen. pl.?**

1. **Do skupiny s koncovkou –*í* patří omezený seznam slov, mezi nimi jsou ovšem buď domácí slova velmi frekventovaná, nebo jejich deriváty. Nefrekventovanou výjimkou je nesklonný název měny/vlastní jméno *a/Afghání* (adaptace).**
2. **Do skupiny bez koncovky patří z domácí slovní zásoby pouze substantiva *peníze, sudety* a životná *(ne)přítel, obyvatel* a propria – plurália tantum (*Klatovy, Poděbrady, Karpaty)*. Dále se objevují nesklonná přejatá/adaptovaná substantiva, zkratky, atd.**
3. **Adjektivní flexi mají substantiva s adjektivní flexí (*radní, mluvčí, zlotý*) a substantivizovaná adjektiva.**

**K dalšímu procvičení:**

**Mějme tvrzení 2:**

**Všechna slovesa končí ve 3. osobě singuláru indikativu prézentu aktiva na *e, í, á.* Máme 5 tříd podle kmene přítomného se 4 vzory: *nese/peče/bere/maže/umře, tiskne/mine/začne, kryje/kupuje, prosí/trpí/sází, dělá.* Máme nepravidélná slovesa: *být/je, mít/má, chtít/chce …***

**Existují z tohoto pravidla nějaké výjimky?**

**Vzpomenete si na ně?**

**Jak jsou obvyklé (jsou to slova z centra/periferie slovní zásoby)?**

**Jak vyhledat data v korpusu:**

**Co hledáme?**

**1) tvary 3. osoby singuláru indikativu prézenta sloves, tedy tvary, které mají značku (tag):**

**1. pozice: *V*** *jako sloveso (****v****erbun)*

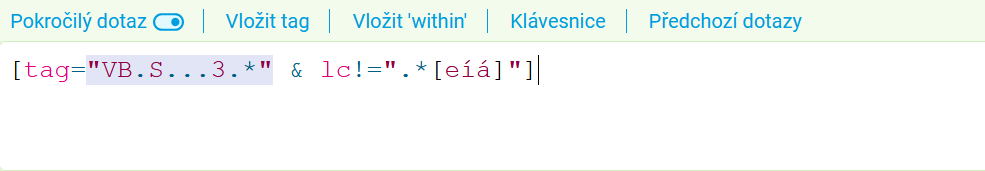
**2. pozice: *B*** *jako prézens* (žádná mnemotechnika)

**4. pozice: *S*** *jako* ***s****ingulár*

**8. pozice: *3*** *jako* ***3.*** *osoba*

a zároveň

**2) tvary (bez ohledu na velikost písma - lc), které nekončí na *e, í, á***



**Mějme tvrzení 3:**

**Všechna slovesa končí ve 2. osobě singuláru indikativu prézentu aktiva na *š.* Máme 5 tříd podle kmene přítomného se 4 vzory: *neseš/pečeš/bereš/mažeš/umřeš, tiskneš/mineš/začneš, kryješ/kupuješ, prosíš/trpíš/sázíš, děláš.* Máme nepravidélná slovesa: *být/jsi, mít/máš, chtít/chceš …***

**Existují z tohoto pravidla ještě nějaké výjimky?**

**Vzpomenete si na ně?**

**Jak jsou obvyklé (jsou to slova z centra/periferie slovní zásoby)?**

**Jak vyhledat data v korpusu:**

**Co hledáme?**

**1) tvary 3. osoby singuláru indikativu prézenta sloves, tedy tvary, které mají značku (tag):**

**1. pozice: *V*** *jako sloveso (****v****erbun)*

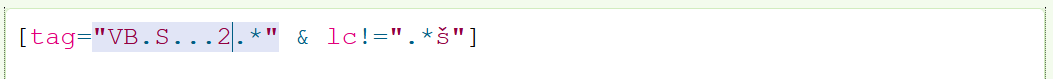
**2. pozice: *B*** *jako prézens* (žádná mnemotechnika)

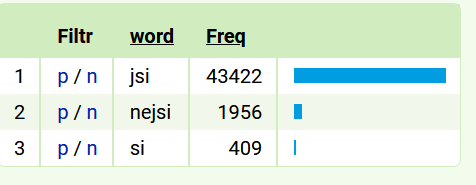
**4. pozice: *S*** *jako* ***s****ingulár*

**8. pozice: *2*** *jako* ***2.*** *osoba*

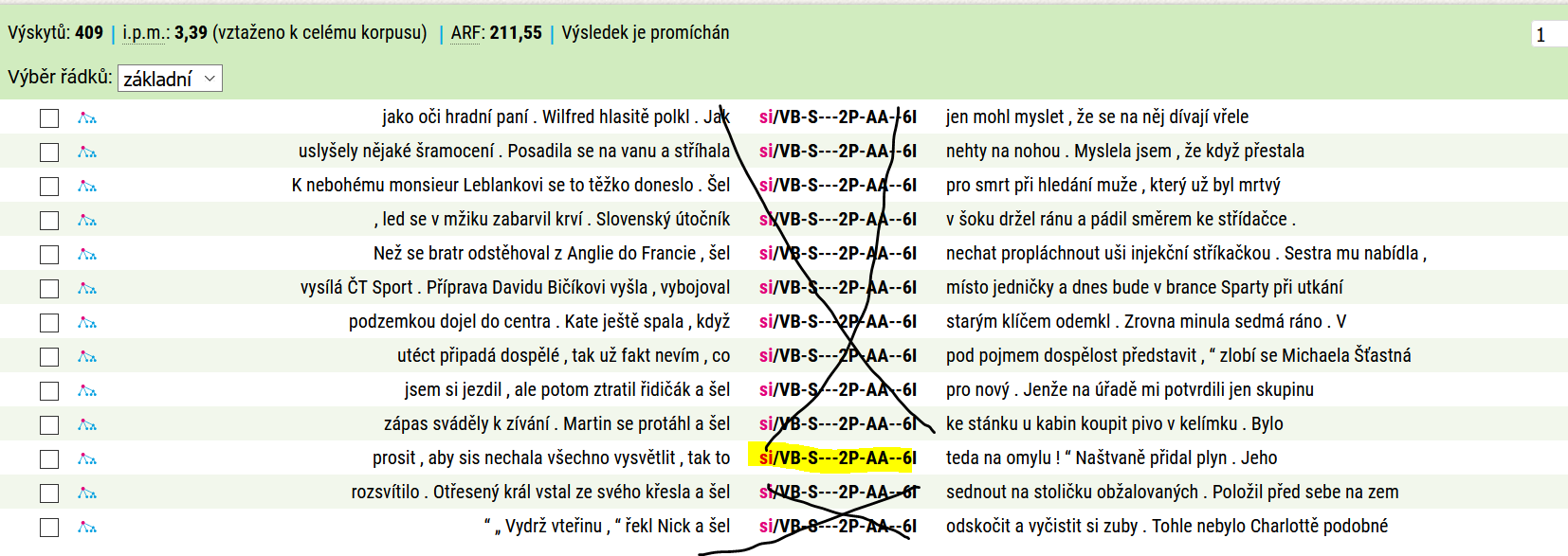
a zároveň

**2) tvary (bez ohledu na velikost písma - lc), které nekončí na *š***





**Chyby v desambiguaci:**



**Mějme tvrzení 4:**

**Substantiva typu *muž* mají ve vokativu singuláru koncovku i (*muži*).**

**Mohou končit i jinak?**

**Vzpomenete si na nějaký příklad?**

**Jak vyhledat data v korpusu:**

**Co hledáme?**

**1) tvary vokativu singuláru maskulin životných, která mají značku (tag):**

**1. pozice: *N*** *jako substantivum (****n****oun)*

**2. pozice: *N*** *jako obyčejné substantivum (****n****oun)*

**3. pozice: *M*** *jako* ***m****askulinum životné*

**4. pozice: *S*** *jako* ***s****ingulár*

**5. pozice: *5*** *jako* ***5.*** *pád (vokativ)*

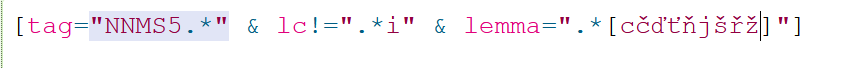
a zároveň

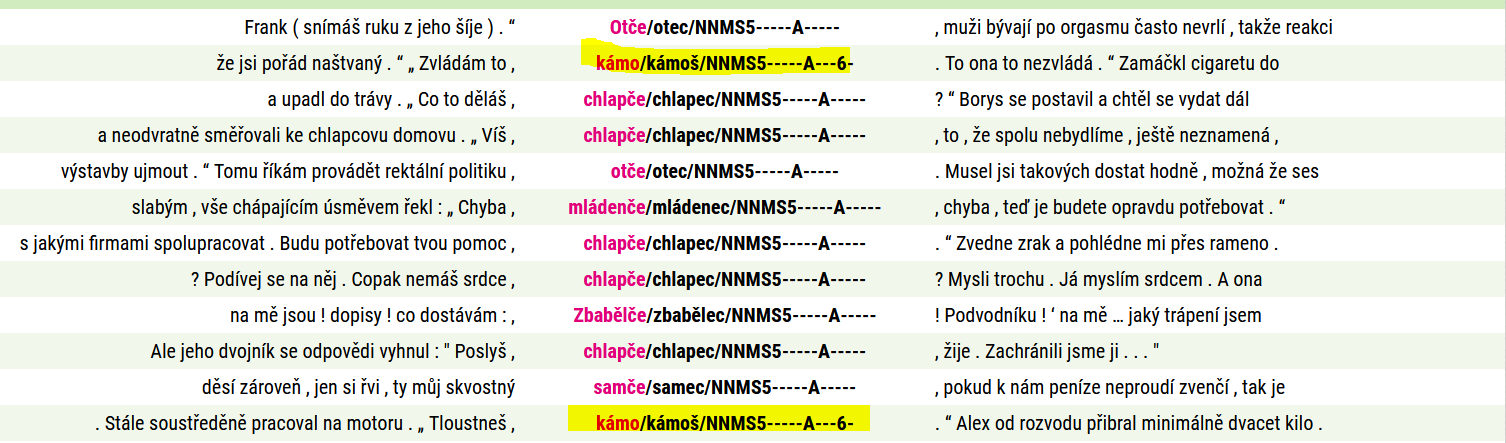
**2) tvary (bez ohledu na velikost písma - lc), které nekončí na *i***

a zároveň

**3)** Skloňují se podle typu *muž,* což znamená, že **mají lemma končící na souhlásku „měkkou“ (?nebo obojetnou).**

Pro jednoduchost necháme obojetné souhlásky stranou:





Má tvar *kámo* koncovku *–o*?

Co mají společného ostatní substantiva?



Doplňující otázka: Vidíme, že vokativ substantiva ***stařec*** je ***starče***. Je alternace ***ř/r*** v pravidelná? Jak lze použít korpus?

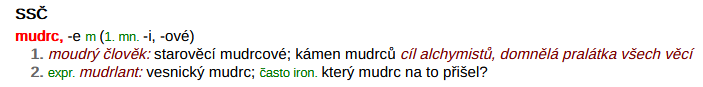
Vyhledáme lemmata životných maskulin zakončená na *řec*: [lemma=".\*řec" & tag="NNM.\*"]. Dále by bylo možné pomocí pFiltru vybrat pouze vokativy sg.: [tag="NNMS5.\*"]. Tento postup nám ale bohužel žádná data nepřinese. Můžeme zapojit lingvistickou fantazii a ptát se, jak bychom oslovili pána, který se jmenuje *Tomáš Kuřec, Jan Mendřec, Adam Šářec,* jak bychom oslovili prvoka, který se jmenuje *bachořec* a konečně jak bychom si poradili se slovem *mudřec*.

Obsah obrázku text

Popis byl vytvořen automaticky

Obsah obrázku text

Popis byl vytvořen automaticky



Obsah obrázku stůl

Popis byl vytvořen automaticky

Obsah obrázku stůl

Popis byl vytvořen automaticky

**Mějme tvrzení 5:**

**Kmenotvorná přípona kmene infinitivního sloves typu *dělat, mazat, brát* může být buď *–a-* (*maz-a-t, děl-a-t*) nebo *–á-* (*br-á-t*).**

**Volba *–a-/-á-* podléhá pravidlu.**

**Z pravidla existují výjimky.**

**Vyhledejme v korpusu podklady, o které lze opírat pravidlo a v nichž lze nalézt výjimky.**

[lemma=".\*át" & tag="Vf.\*"]

[lemma=".\*at" & tag="Vf.\*"]

**K jaké třídě a vzoru mohou patřit slovesa, jejichž lemma=".\*át".**

**Existuje nějaký postup, jak odfiltrovat slovesa typu *krýt*?**

**ÚKOL**

1. V dativu singuláru mají feminina typu *žena* koncovku *–e* realizovanou jako *[eě]*, kdežto feminina typu *růže* mají v dativu singuláru (stejně jako feminina typu *píseň, kost*) koncovku *–i*. Vzpomenete si na nějakou výjimku z tohoto pravidla? Svůj jazykový cit ověřte v korpusu a popište, jak byste doklady výjimky z tohoto pravidla hledali v korpusu.

2. Neutra typu *město* mají v lokálu singuláru koncovku koncovku *–e* realizovanou jako *[eě]*, ale mohou mít i koncovku *u.* Formulujte dotaz v jazyce cql tak, abyste získali seznam lemmat substantiv.

3. Neutra typu *kuře*  mají v singuláru koncovky shodné s neutry typu *moře* a v plurálu koncovky shodné s koncovkami typu *město*. Existují další skupiny substantiv neuter, která ve flexi kolísají mezi uvedenými vzory, takže např. mají v genituvu sg. koncovku *–a* (jako bez *města*) a v genitivu pl. *–í* (jako bez *moří*). Pokud si na žádné nevzpomenete, jak budete formulovat dotaz v nástroji morfio, abyste našli odpověď.

4. V instrumentálu plurálu mají koncovku *–mi* substantiva skloňovaná podle typu *kost*. Koncovku *–mi* (tedy nikoli *–ami, -emi, -ěmi, -ími*, *-ými*) mají také substantiva, která se neskloňují podle typu *kost*. Vzpomenete si na nějaké? Jak je najdete v korpusu?