

pondělí: 8.00-10.00 G13

26. 4. 2021

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 (noun)*

2. pozice: *N jako obyčejné substantivum (noun)*

3. pozice: *M jako maskulinum životné nebo I jako maskulinum neživotné (inanimatum)*

4. pozice: *P jako plurál*

5. pozice: *2 jako 2. pád (genitiv)*

a zároveň

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

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

```
[tag="NN[MI]P2.*" & lc!=".*ů"]
```

Frekvence/Vlastní/lc

Celkem: 1 157 položek (24 stránek)

	Filtr	lc [lowercase word]	Freq	
1	p / n	lidí	45314	
2	p / n	peněz	11968	
3	p / n	dní	11476	
4	p / n	obyvatel	4154	
5	p / n	přátel	3405	

Jak lze výjimky z výše uvedeného pravidla dále kategorizovat?

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

- končící na -í (typ *lidí/dní*)
- bez koncovky/ s nulovou koncovkou (typ *peněz-0/přátel-0*)
- 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

Aplikovat operaci "pozitivní filtr"

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

```
[lc=". * í"]
```

TIP Barevně zvýrazněný token s otazníkem v jednoduchém dotazu znamená možnost upřesnit jeho interpretaci. Použijte CTRL/Command a klik pro aktivaci. ([další tip](#))

 Upřesnit parametry

Rozsah hledání od: 0
do: 0

Výchozí atribut:

word

včetně KWIC

Zvýraznění shody:

první shoda



Celkem: 21 položek (1 stránka)

	Filtr	lc [lowercase word]	Freq	
1	p / n	lidí	45314	
2	p / n	dní	11476	
3	p / n	koní	2265	
4	p / n	bratří	694	
5	p / n	kněží	511	
6	p / n	hostí	433	

Chyby:

poradil . Rozbrečel jsem se a začal ječet : „

TATÍ/NNMP2-----A-----

, JÁ MUSIM ČURAT ! " Což bylo dost ponižující

. Podruhé dětský hlásek , který se ptal : "

Tatí/NNMP2-----A-----

, co ten pán dělá ? " V obou případech

Zaskočil nás , když syn ve školce prohlásil : „

Tatí/NNMP2-----A-----

, chtěl bych být holčička . " Prohrábneme si strniště

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

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

Celkem: 863 položek (18 stránek)

	Filtr	word	Freq	
1	p / n	peněz	11968	
2	p / n	obyvatel	4154	
3	p / n	přátel	3405	
4	p / n	°	3210	
5	p / n	mm	3152	
6	p / n	windows	2279	
7	p / n	§	2141	
8	p / n	times	1082	
9	p / n	m2	859	
10	p / n	nepřítel	764	
11	p / n	reuters	508	
12	p / n	klatov	450	
13	p / n	poděbrad	384	

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*, *zlatý*) a substantivizovaná adjektiva.

K dalšímu procvičení:

Mějme tvrzení 2:

Všechna slovesa končí ve 3. osobě singuláru indikativu přítomného času na *e, í, á*. Máme 5 tříd podle kmene přítomného času se 4 vzory: *nese/peče/bere/maže/umře, tiskne/mine/začne, kryje/kupuje, prosí/trpí/sází, dělá*. Máme nepravidelná 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 přítomného času sloves, tedy tvary, které mají značku (tag):

1. pozice: *V* jako sloveso (verbun)

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

4. pozice: *S* jako *singulár*

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

a zároveň

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

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

```
[tag="VB.S...3.*" & lc!=".*[eíá]" ]|
```

Mějme tvrzení 3:

Všechna slovesa končí ve 2. osobě singuláru indikativu přítomného času na *š*. Máme 5 tříd podle kmene přítomného času se 4 vzory: *neseš/pečeš/bereš/mažeš/umřeš, tiskneš/mineš/začneš, kryješ/kupuješ, prosíš/trpíš/sázíš, děláš*. Máme nepravidelná 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 přítomna sloves, tedy tvary, které mají značku (tag):

1. pozice: *V jako sloveso (verbun)*

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




4. pozice: *S jako singulár*

8. pozice: *2 jako 2. osoba*

a zároveň

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

```
[tag="VB.S...2|.*" & lc!=".*š"]
```

	Filtr	word	Freq	
1	p / n	jsi	43422	
2	p / n	nejsi	1956	
3	p / n	si	409	

Chyby v desambiguaci:

Vyskytů: 409 | l.p.m.: 3,39 (vztaženo k celému korpusu) | ARF: 211,55 | Výsledek je promíchán 1

Výběr řádků: základní

<input type="checkbox"/>	Λα	jako oči hradní paní . Wilfred hlasitě polkl . Jak	si/VB-S---2P-AA--6I	jen mohl myslet , že se na něj dívají vřele
<input type="checkbox"/>	Λα	uslyšely nějaké šramocení . Posadila se na vanu a stříhala	si/VB-S---2P-AA--6I	nehty na nohou . Myslela jsem , že když přestala
<input type="checkbox"/>	Λα	K nebohému monsieur Leblankovi se to těžko doneslo . Šel	si/VB-S---2P-AA--6I	pro smrt při hledání muže , který už byl mrtvý
<input type="checkbox"/>	Λα	, led se v mžiku zabarvil krví . Slovenský útočník	si/VB-S---2P-AA--6I	v šoku držel ránu a pádil směrem ke střídačce .
<input type="checkbox"/>	Λα	Než se bratr odstěhoval z Anglie do Francie , šel	si/VB-S---2P-AA--6I	nechat propláchnout uši injekční stříkačkou . Sestra mu nabídla ,
<input type="checkbox"/>	Λα	vysílá ČT Sport . Příprava Davidu Bičíkovi vyšla , vybojoval	si/VB-S---2P-AA--6I	místo jedničky a dnes bude v brance Sparty při utkání
<input type="checkbox"/>	Λα	podzemkou dojel do centra . Kate ještě spala , když	si/VB-S---2P-AA--6I	starým klíčem odemkl . Zrovna minula sedmá ráno . V
<input type="checkbox"/>	Λα	utěct připadá dospělé , tak už fakt nevím , co	si/VB-S---2P-AA--6I	pod pojmem dospělost představit , " zlobí se Michaela Šťastná
<input type="checkbox"/>	Λα	jsem si jezdil , ale potom ztratil řidičák a šel	si/VB-S---2P-AA--6I	pro nový . Jenže na úřadě mi potvrdili jen skupinu
<input type="checkbox"/>	Λα	zápas sváděly k zívání . Martin se protáhl a šel	si/VB-S---2P-AA--6I	ke stánku u kabin koupit pivo v kelímku . Bylo
<input type="checkbox"/>	Λα	prosít , aby sis nechala všechno vysvětlit , tak to	si/VB-S---2P-AA--6I	teda na omylu ! " Naštvaně přidal plyn . Jeho
<input type="checkbox"/>	Λα	rozsvítilo . Otržený král vstal ze svého křesla a šel	si/VB-S---2P-AA--6I	sednout na stoličku obžalovaných . Položil před sebe na zem
<input type="checkbox"/>	Λα	" , Vydřz vterinu , " řekl Nick a šel	si/VB-S---2P-AA--6I	odskočit a vyčistit si zuby . Tohle nebylo Charlottě podobné

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 (noun)*

2. pozice: *N jako obyčejné substantivum (noun)*

3. pozice: *M jako maskulinum životné*

4. pozice: *S jako singulár*

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:







```
[tag="NNMS5.*" & lc!=".*i" & lemma=".*[cčďtňjšřž]" ]
```

Frank (snímáš ruku z jeho šije). "	Otče/otec/NNMS5-----A-----	, muži bývají po orgasmu často nevrli , takže reakci
že jsi pořád naštvaný . " . Zvládám to ,	kámo/kámoš/NNMS5-----A---6-	. To ona to nezvládá . " Zamáčkli cigaretu do
a upadl do trávy . „ Co to děláš ,	chlapče/chlapec/NNMS5-----A-----	? " Borys se postavil a chtěl se vydat dál
a neodvratně směřovali ke chlapcovu domovu . „ Víš ,	chlapče/chlapec/NNMS5-----A-----	, to , že spolu nebydlíme , ještě neznamená ,
výstavby ujmout . " Tomu říkám provádět rektální politiku ,	otče/otec/NNMS5-----A-----	. Musel jsi takových dostat hodně , možná že ses
slabým , vše chápajícím úsměvem řekl : „ Chyba ,	mládenče/mládenc/NNMS5-----A-----	, chyba , teď je budete opravdu potřebovat . "
s jakými firmami spolupracovat . Budu potřebovat tvou pomoc ,	chlapče/chlapec/NNMS5-----A-----	. " Zvedne zrak a pohlédne mi přes rameno .
? Podívej se na něj . Copak nemáš srdce ,	chlapče/chlapec/NNMS5-----A-----	? Mysli trochu . Já myslím srdcem . A ona
na mě jsou ! dopisy ! co dostávám : ,	Zbabělče/zbabělec/NNMS5-----A-----	! Podvodníku ! ' na mě ... jaký trápení jsem
Ale jeho dvojník se odpovědi vyhnul : " Poslyš ,	chlapče/chlapec/NNMS5-----A-----	, žije . Zachránili jsme ji ... "
děsí zároveň , jen si řvi , ty můj skvostný	samče/samec/NNMS5-----A-----	, pokud k nám peníze neproudí zvenčí , tak je
. Stále soustředěně pracoval na motoru . „ Tlousteš ,	kámo/kámoš/NNMS5-----A---6-	. " Alex od rozvodu přibral minimálně dvacet kilo .

Má tvar *kámo* koncovku *-o*?

Co mají společného ostatní substantiva?

Celkem: 126 položek (3 stránky)

	Filtr	word	Freq	
1	p / n	otče	1066	
2	p / n	chlapče	1008	
3	p / n	kámo	405	
4	p / n	pitomče	81	
5	p / n	blbče	71	
6	p / n	starče	43	
7	p / n	mládenče	35	
8	n / n	bratranče	35	

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 na 3. 5. 2021

3. 5. Dů:

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 genitivu 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 morřio, 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?