

pondělí: 8.00-10.00 G13

27. 4. 2020

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

### Mějme tvrzení:

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:

Hledat v korpusu

Korpus:

Typ dotazu: CQL [?](#)

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

Dotaz: `[tag="NN[MI]P2.*"]`

Předchozí dotazy lze zobrazit také pomocí šipky dolů [\(další tip\)](#)

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

► [Specifikovat kontext](#)

► [Omezit hledání](#)

Frekvenční distribuce ✕

---

Frekvenční limit: 1

Úroveň	Atribut	Nerozlišovat velikost	Pozice	(Node) začít od
1.	lc [lowercase word]	<input type="checkbox"/>	Node	slova KWIC nejvíce vlevo

## Frekvenční seznam

1 ▶

Minimální frekvence:

**Celkem:** 16 684 položek (334 stránek)

	Filter	<u>lc [lowercase word]</u>	<u>Freq</u>	
1	p / n	lidí	45314	<div style="width: 100%; height: 10px; background-color: #007bff;"></div>
2	p / n	milionů	19697	<div style="width: 43%; height: 10px; background-color: #007bff;"></div>
3	p / n	metrů	13055	<div style="width: 29%; height: 10px; background-color: #007bff;"></div>
4	p / n	peněz	11968	<div style="width: 26%; height: 10px; background-color: #007bff;"></div>
5	p / n	dní	11476	<div style="width: 25%; height: 10px; background-color: #007bff;"></div>
6	p / n	měsíců	9780	<div style="width: 22%; height: 10px; background-color: #007bff;"></div>
7	p / n	dolarů	9700	<div style="width: 21%; height: 10px; background-color: #007bff;"></div>
8	p / n	mužů	8855	<div style="width: 19%; height: 10px; background-color: #007bff;"></div>
9	p / n	kilometrů	8550	<div style="width: 19%; height: 10px; background-color: #007bff;"></div>
10	p / n	důvodů	7973	<div style="width: 18%; height: 10px; background-color: #007bff;"></div>
11	p / n	druhů	7828	<div style="width: 17%; height: 10px; background-color: #007bff;"></div>
12	p / n	bodů	7279	<div style="width: 16%; height: 10px; background-color: #007bff;"></div>
13	p / n	dnů	7181	<div style="width: 16%; height: 10px; background-color: #007bff;"></div>
14	p / n	problémů	7062	<div style="width: 16%; height: 10px; background-color: #007bff;"></div>
15	p / n	zdrojů	6837	<div style="width: 15%; height: 10px; background-color: #007bff;"></div>
16	p / n	států	5616	<div style="width: 12%; height: 10px; background-color: #007bff;"></div>
17	p / n	stromů	5576	<div style="width: 12%; height: 10px; background-color: #007bff;"></div>
18	p / n	rodičů	5418	<div style="width: 12%; height: 10px; background-color: #007bff;"></div>
19	p / n	projektů	5376	<div style="width: 12%; height: 10px; background-color: #007bff;"></div>
20	p / n	členů	5321	<div style="width: 12%; height: 10px; background-color: #007bff;"></div>

## Aplikovat operaci "filtr"



Filtr:

negativní ▾

Rozsah hledání: od: 0 do: 0 , včetně KWIC

Typ dotazu:

Slovní tvar ▾ ?

[Klávesnice](#) | [Předchozí dotazy](#)

.\*|

Dotaz:

Předchozí dotazy lze zobrazit také pomocí šipky dolů [\(další tip\)](#)

Slovní druh: Nespecifikováno ▾ Shoda velikosti písmen:

Hledat

Celkem: 1 157 položek (24 stránek)

	Filter	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	
6	p / n	°	3210	
7	p / n	mm	3152	
8	p / n	windows	2279	
9	p / n	koní	2265	
10	p / n	§	2141	
11	p / n	dospělých	1234	
12	p / n	times	1082	
13	p / n	cestujících	1076	
14	p / n	nemocných	931	
15	p / n	známých	921	
16	p / n	příbuzných	907	
17	p / n	m2	859	
18	p / n	nepřátel	764	
19	p / n	bratří	694	
20	p / n	zlatých	576	
21	p / n	kněží	511	

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:

► Negativní filtr: .\*ů (118 236 výskytů) ► Pozitivní filtr: .\*í (60 797 výskytů)

## Frekvenční seznam

1

Minimální frekvence:

[Použít](#)

**Celkem:** 21 položek (1 stránka)

	Filter	lc [lowercase word]	Freq	
1	p / n	lidí	45314	<div style="width: 100%;"></div>
2	p / n	dní	11476	<div style="width: 25%;"></div>
3	p / n	koní	2265	<div style="width: 5%;"></div>
4	p / n	bratří	694	<div style="width: 1%;"></div>
5	p / n	kněží	511	<div style="width: 1%;"></div>
6	p / n	hostí	433	<div style="width: 1%;"></div>
7	p / n	podlidí	26	<div style="width: 0.5%;"></div>
8	p / n	spolubratří	14	<div style="width: 0.3%;"></div>
9	p / n	nadlidí	10	<div style="width: 0.2%;"></div>
10	p / n	bří	10	<div style="width: 0.2%;"></div>
11	p / n	pralidí	9	<div style="width: 0.2%;"></div>
12	p / n	pakoní	7	<div style="width: 0.15%;"></div>
13	p / n	kachlí	6	<div style="width: 0.13%;"></div>
14	p / n	velekněží	5	<div style="width: 0.11%;"></div>
15	p / n	opolidí	3	<div style="width: 0.06%;"></div>
16	p / n	tatí	3	<div style="width: 0.06%;"></div>
17	p / n	reagencí	3	<div style="width: 0.06%;"></div>
18	p / n	novokněží	2	<div style="width: 0.04%;"></div>
19	p / n	člověkodní	2	<div style="width: 0.04%;"></div>
20	p / n	afghání	2	<div style="width: 0.04%;"></div>
21	p / n	pololidí	2	<div style="width: 0.04%;"></div>

### Chyby:

poradil . Rozbrečel jsem se a začal ječet : „ **TATÍ** , JÁ MUSIM ČURAT ! ” Což bylo dost ponižující

. Podruhé dětský hlásek , který se ptal : “ **Tatí** , co ten pán dělá ? ” V obou případech

Zaskočí nás , když syn ve školce prohlásí : „ **Tatí** , chtěl bych být holčička . ” Prohrábeme si strniště

necháme reagovat s jednou ze štěpících látek . První sada **reagencií** chemicky modifikuje ty nukleotidy , pro něž jsou specifické ,  
mikrozkumavce jednoduše tak , že se DNA smíchá se souborem **reagencií** a mikrozkumavka se vloží do termálního cykléru ( termocyklér )  
dobře táhnoucí digestoři nebo na volném prostranství . I zbytky **reagencií** k jeho výrobě likvidujeme v digestoři . Navážíme dvojnásobek teoreticky

Odstraníme tvary na *-í* a lemmata na *.\*[ýí]*

1

Minimální frekvence:  Použít

Celkem: 857 položek (18 stránek)






	Filter	lc [lowercase 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	nepřátel	764	
11	p / n	reuters	508	
12	p / n	klatov	450	
13	p / n	poděbrad	384	
14	p / n	km2	346	
15	p / n	m3	308	
16	p / n	cm3	302	
17	p / n	karpát	284	
18	p / n	beatles	278	
19	p / n	drobných	251	
20	p / n	jo	233	

## Frekvenční seznam

1 ▶

Minimální frekvence:

Celkem: 279 položek (6 stránek)

	Filter	lc [lowercase word]	Freq	
1	p / n	dospělých	1234	
2	p / n	cestujících	1076	
3	p / n	nemocných	931	
4	p / n	známých	921	
5	p / n	příbuzných	907	
6	p / n	zlatých	576	
7	p / n	pracujících	484	
8	p / n	rozhodčích	451	
9	p / n	nezaměstnaných	333	
10	p / n	radních	271	
11	p / n	nadřízených	190	
12	p / n	podřízených	179	
13	p / n	poddaných	160	
14	p / n	obžalovaných	147	
15	p / n	kinských	143	
16	p / n	neznámých	137	
17	p / n	starých	134	
18	p / n	mluvčích	117	
19	p / n	bližních	113	
20	p / n	zlotých	111	

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 – plurálie 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.**

**ÚKOL na 4. 5. 2020**

**Porovnáme-li maskulina v pozorovaném tvaru s femininy a neutry, zjistíme, že dvě výjimečné koncovky (-í a -o) figurují v genitivu plurálu právě u vzorů feminin a neuter:**

*růže/růží, píseň/písni, kost/kostí, moře/moří, stavení/stavení a žena/žen, město/měst, kuře/kuřat.*

**Platí i zde nějaké výjimky?**

**Vzpomenete si na ně?**

**Použijte k jejich zjištění korpus.**

**Jakého rodu jsou *Tatry, Krkonoše, Beskydy*?**

**Všimněte si, jak výjimečné koncovky u maskulin plurálií tantum působí na nejistotu při určení rodu u rodilých mluvčí.**