

Hledání variantních slovesných tvarů od kmene prezntního

Cílem práce je nalézt na základě jednoho dotazu všechny variantní slovesné tvary určitých gramatických kategorií tak, aby součástí bylo co nejméně nerelevantních výskytů.

1. indikativu prezntu aktiva

- **-u/-i** v 1. osobě singuláru (typ *kupuji/kupuju*) – jde o:
 - některá slovesa vzoru **mazat** (= před koncovým [ui] v 1. os. sg. předchází [čřšž] nebo šl a infinitiv končí na [aá]t). Sériemi dotazů, z nichž základní byl [lc=".*[^v]([čřšž]|(šl))[ui]"&lemma=".*[aá]t"] jsem zjistil, že variantnost se netýká sloves s infinitivy na .*slat, .*lh[aá]t a početné skupiny sloves kolísající mezi I. a V. třídou (zakončení -u/-ám, viz níže). Na základě tohoto faktu jsem dospěl k závěru, že bude nejjednodušší sestavit seznam výčtem, neboť kýžených sloves je o poznání méně než zbylých sloves náležících (alespoň zčásti) ke vzoru *mazat*; výsledek tedy: [lc=".*maž[ui]|.*[kv]áž[ui]|(o|ne|do)*táž[iu]|.*(pl|sk)áč[ui]|.*píš[ui]"]
 - součástí jsou však některá substantiva se shodným zakončením (nejvýraznější homonymní tvar *pláči*), v případě lemmatizovaného korpusu můžeme pro jejich odstranění dodat &lemma=".*[aá]t")
 - slovesa vzorů **krýt** a **kupovat** (= před koncovým [ui] v 1. os. sg. předchází j). Základní dotaz [lc=".*j[ui]"&lemma=".*t"] – kromě nepravidelného slovesa *být* jsou relevantní všechna slovesa, zároveň však dostáváme spousty výskytů jiných slovních druhů, částečně je odfiltrujeme zpřesněním dotazu na [lc=".*[aeéěiu]j[ui]"], v případě lemmatizovaného korpusu ještě doplněním &lemma=".*t". Kdybychom chtěli získat pouze slovesa mohoucí mít tuto variantu, nikoliv jejich výskyt v obou tvarech, je možno dotaz zjednodušit na [lc=".*[aeéěiu]ju"], neboť zakončení -u je méně homonymní s ostatními slovy
 - výsledný dotaz:

1)

```
[lc="(.*maž[ui]|.*[kv]áž[ui]|.[oe]táž[iu]|.*(pl|sk)áč[ui]|.*píš[ui])|(.*[aeéěiu]j[ui])"]
```

2)

```
[lc="(.*maž[ui]|.*[kv]áž[ui]|.[oe]táž[iu]|.*(pl|sk)áč[ui]|.*píš[ui])|(.*[aeéěiu]ju)"]
```

3)

```
[(lc=".*maž[ui]|.*[kv]áž[ui]|.[oe]táž[iu]|.*(pl|sk)áč[ui]|.*píš[ui]"&lemma=".*[aá]t")|(lc=".*[aeéěiu]j[ui]"&lemma=".*t")]
```

1)

1	p / n	později	33638
2	p / n	raději	11088
3	p / n	kraji	8123
4	p / n	nejčastěji	6804
5	p / n	častěji	5289
6	p / n	rychleji	4020
7	p / n	okraji	3981
8	p / n	naději	3891
9	p / n	děkuji	3309
10	p / n	obličej	2941
11	p / n	potřebuji	2642
12	p / n	nedokážu	2343
13	p / n	nejraději	2340
14	p / n	přesněji	2212
15	p / n	nejrychleji	2200
16	p / n	miluji	2134
17	p / n	prodeji	2062
18	p / n	děkuju	2033
19	p / n	považuji	1805
20	p / n	pamatuju	1706

2)

1	p / n	potřebuji	2642
2	p / n	nedokážu	2343
3	p / n	miluju	2134
4	p / n	děkuju	2033
5	p / n	pamatuju	1706
6	p / n	dokážu	1681
7	p / n	píšu	1381
8	p / n	žiju	946
9	p / n	hraju	873
10	p / n	přeju	816
11	p / n	ukážu	807
12	p / n	nepamatuju	715
13	p / n	nepotřebuji	597
14	p / n	napišu	577
15	p / n	slíbju	505
16	p / n	pracuju	498
17	p / n	pochybuju	406
18	p / n	představuju	385
19	p / n	zabiju	383
20	p / n	jmenuju	382

3)

1	p / n	děkuji	3309
2	p / n	potřebuji	2642
3	p / n	nedokážu	2343
4	p / n	miluju	2134
5	p / n	děkuju	2033
6	p / n	považuji	1805
7	p / n	pamatuju	1706
8	p / n	dokážu	1681
9	p / n	pamatuji	1413
10	p / n	píšu	1319
11	p / n	potřebuji	1299
12	p / n	miluji	1120
13	p / n	doporučuji	988
14	p / n	přeji	955
15	p / n	žiju	946
16	p / n	hraju	873
17	p / n	přeju	816
18	p / n	ukážu	807
19	p / n	pracuji	739
20	p / n	nepamatuju	715

- + **-í/-ou** (typ *kupují/kupujou*): v předchozím dotazu bude [ui] nahrazeno (í | ou), sem se ale dostane nadměrné množství výsledků náležící ke vzoru *dělat*, bude proto lepší odfiltrovat zakončení *. *aj (í | ou)* a poté jen výčtem postihnout slovesa III. třídy mající takové zakončení (pozorováním jsem objevil pouze [hz]raj (í | ou), [tsk]aj (í | ou) a v[aj] (í | ou) – nutno však počítat s obtížným začleněním prefigovaných variant u [tsk]aj (í | ou), neboť totožně končí množství sloves podle *dělat* (*říkají, čekají, chystají, klesají, ...*). Součástí jsou rovněž nerelevantní slovesa IV. třídy, ta lze odstranit použitím atributu lemma (2).

- výsledný dotaz:

1)

```
[lc="(.*maž(í|ou)|.*[kv]áž(í|ou)|.[oe]táž(í|ou)|.*(pl|sk)áč(í|ou)|.*píš(í|ou))|.*( [eéiuy]j(í|ou)|.*[hz]raj(í|ou)|[tsk]aj(í|ou)|.*v[aj](í|ou)"]
```

2)

```
[lc="(.*maž(í|ou)|.*[kv]áž(í|ou)|.[oe]táž(í|ou)|.*(pl|sk)áč(í|ou)|.*píš(í|ou))|.*( [eéiuy]j(í|ou)|.*[hz]raj(í|ou)|[tsk]aj(í|ou)|.*v[aj](í|ou)"&lemma=".*[ět"]
```

1)

1	p / n	její	87053
2	p / n	chtějí	10789
3	p / n	existují	6351
4	p / n	žijí	4944
5	p / n	musejí	4306
6	p / n	představují	4096
7	p / n	hrají	4047
8	p / n	potřebují	4017
9	p / n	nabízejí	3644
10	p / n	objevují	3358
11	p / n	dokážou	3235
12	p / n	pracují	3066
13	p / n	ukazují	2888
14	p / n	obsahují	2738
15	p / n	umožňují	2730
16	p / n	nechtějí	2632
17	p / n	přicházejí	2591
18	p / n	vyskytují	2551
19	p / n	vytvářejí	2541
20	p / n	nacházejí	2450

2)

1	p / n	chtějí	10789
2	p / n	existují	6351
3	p / n	žijí	4944
4	p / n	představují	4096
5	p / n	hrají	4047
6	p / n	potřebují	4017
7	p / n	objevují	3358
8	p / n	dokážou	3235
9	p / n	pracují	3066
10	p / n	ukazují	2888
11	p / n	obsahují	2738
12	p / n	umožňují	2730
13	p / n	nechtějí	2632
14	p / n	vyskytují	2551
15	p / n	pohybují	2152
16	p / n	považují	2096
17	p / n	fungují	1964
18	p / n	poskytují	1793
19	p / n	váží	1733
20	p / n	vyžadují	1710

- **-m/-me** v 1. osobě plurálu (typ *nesem/neseme*)
 - jde o slovesa I., II. a III. třídy: pro III. třídu využijeme dotaz z předešlého bodu, tedy `[lc=".*[aeéiuy]jem(|e)"]`. Zbylá slovesa mohou před koncovým `em(|e)` obsahovat různé souhlásky, avšak pozorováním jsem zjistil, že mezi nimi nikdy nejsou `[ghkqwx]`, c mimo `.*chcem(|e)`, `[zs]` jen s předcházejícím `e(+.*pasem(|e))`
 - Stejně jako v prvním případě můžeme hledat pouze charakterističtější zakončení (*-me*, 2), nebo použít atribut lemma (3).
 - výsledný dotaz:

1)

```
[lc=".*[aeéiuy]jem(|e)|.*[^ááééěííoóúúůýýjghkqwxcsz]em(|e)|.*e[sz]em(|e)|.*pasem(|e)|.*chcem(|e)|.*stůněm(|e)"]
```

2)

```
[lc=".*[aeéiuy]jeme|.*[^ááééěííoóúúůýýjghkqwxcsz]eme|.*e[sz]eme|.*paseme|.*chceme|.*stůněme"]
```

3)

```
[lc=".*[aeéiuy]jem(|e)|.*[^ááééěííoóúúůýýjghkqwxcsz]em(|e)|.*e[sz]em(|e)|.*pasem(|e)|.*chcem(|e)|.*stůněm(|e) "&lemma=".*[cszaáééěííoóúúůý]t|. *out"]
```

1)

1	p/n	kolem	56791
2	p/n	ovšem	39233
3	p/n	můžeme	24371
4	p/n	způsobem	20930
5	p/n	budeme	17825
6	p/n	všem	15671
7	p/n	směrem	15200
8	p/n	vzhledem	13836
9	p/n	lidem	10940
10	p/n	cílem	10054
11	p/n	čem	9378
12	p/n	chceme	7879
13	p/n	pohledem	7848
14	p/n	jménem	7481
15	p/n	firem	7366
16	p/n	našem	7237
17	p/n	předem	6450
18	p/n	dětem	6058
19	p/n	názvem	6043
20	p/n	málem	5738

2)

1	p/n	můžeme	24371
2	p/n	budeme	17825
3	p/n	chceme	7879
4	p/n	nemůžeme	4879
5	p/n	najdeme	3560
6	p/n	potřebujeme	3368
7	p/n	nebudeme	3163
8	p/n	jdeme	2562
9	p/n	dostaneme	2301
10	p/n	žijeme	1863
11	p/n	nechceme	1722
12	p/n	půjdeme	1533
13	p/n	jedeme	1426
14	p/n	začneme	1272
15	p/n	dokážeme	1222
16	p/n	považujeme	1184
17	p/n	vezmeme	1125
18	p/n	doporučujeme	1120
19	p/n	hrájeme	1067
20	p/n	děkujeme	999

3)

1	p/n	budeme	17825
2	p/n	chceme	7879
3	p/n	najdeme	3560
4	p/n	potřebujeme	3368
5	p/n	nebudeme	3163
6	p/n	světem	2910
7	p/n	předmětem	2764
8	p/n	jdeme	2562
9	p/n	prstem	2349
10	p/n	počtem	2326
11	p/n	dostaneme	2301
12	p/n	žijeme	1863
13	p/n	nechceme	1722
14	p/n	pocitem	1679
15	p/n	státem	1550
16	p/n	půjdeme	1533
17	p/n	jedeme	1426
18	p/n	začneme	1272
19	p/n	dokážeme	1222
20	p/n	považujeme	1184

- **-u/-ám** v 1. osobě singuláru (typ *kopu/kopám*)
 - jde o některá slovesa vzorů **mazat** a **brát**
 - vzor *mazat*: nejpřesnější výsledky dostaneme opět výčtem, jehož docílíme pozorováním dat:


```
[lc=".*(ře|vr|klou|hry|lí)(žu|zám)|.*(o|pá)(řu
```

| rám) | . * (kou | če | klu) (šu | sám) "] . Součástí jsou některé nerelevantní výskyty mající stejné zakončení (především na . *oru a . *lížu), ty by se daly odstranit doplněním všech možných předpon těchto sloves

- vzor *brát*: těžko rozlišitelná podoba od sloves časovaných výhradně dle *dělat*, přesto jsem při jejich pozorování zjistil některé zákonitosti:
 - variantní tvary mají slovesa zakončená na **p** (u | ám) mimo . *čerpám, . *stoupám, . *chápu, . *cpu
 - variantní tvary nemají slovesa zakončená na [bmlvf] ám mimo . *škráb(u | ám), . *kýv(u | ám), . *plav(u | ám), . *raf(u | ám), . *lám(u | ám), hýb(u | ám)
- Opět můžeme hledat pouze charakterističtější zakončení (-ám, 2), nebo použít atribut lemma (3).
- výsledný dotaz:

1)

```
[lc=".*(ře|vr|klou|hry|lí)(žu|zám)|.*(o|pá)(řu|rám)|
.*(kou|če|klu)(šu|sám)|.*p(u|ám)|.*škráb(u|ám)|.*kýv
(u|ám)|.*plav(u|ám)|.*raf(u|ám)|.*lám(u|ám)|(n|)(e|)
hýb(u|ám)"&lc!=".*(čer|stou)pám|.*cpu|.*chápu"]
```

2)

```
[lc=".*(ře|vr|klou|hry|lí)zám|.*(o|pá)rám|.*(kou|če|
klu)sám|.*pám|.*škrábám|.*kývám|.*plavám|.*rafám|.*l
ámám|(n|)(e|)hýbám"&lc!=".*(čer|stou)pám"]
```

3)

```
[lc=".*(ře|vr|klou|hry|lí)(žu|zám)|.*(o|pá)(řu|rám)|
.*(kou|če|klu)(šu|sám)|.*p(u|ám)|.*škráb(u|ám)|.*kýv
(u|ám)|.*plav(u|ám)|.*raf(u|ám)|.*lám(u|ám)|(n|)(e|)
hýb(u|ám)"&lc!=".*(čer|stou)pám|.*cpu|.*chápu"&lemma
=".*[cszaáééěííóúůý]t|.*out"]
```

1)

1	p / n	typu	12127
2	p / n	přístupu	3526
3	p / n	vstupu	3160
4	p / n	postupu	3011
5	p / n	principu	2995
6	p / n	stropu	2417
7	p / n	nákupu	2051
8	p / n	evropu	1736
9	p / n	nástupu	1669
10	p / n	stopu	1532
11	p / n	mapu	1333
12	p / n	výstupu	1172
13	p / n	cupu	1012
14	p / n	islámu	879
15	p / n	přikopu	764
16	p / n	chalupu	752
17	p / n	tempu	697
18	p / n	kopu	667
19	p / n	trupu	646
20	p / n	lampu	625

2)

1	p / n	úsporám	590
2	p / n	horám	177
3	p / n	kývám	101
4	p / n	záplavám	57
5	p / n	bramborám	54
6	p / n	oporám	50
7	p / n	prostorám	40
8	p / n	lapám	33
9	p / n	stopám	30
10	p / n	mapám	30
11	p / n	zácpám	26
12	p / n	lampám	24
13	p / n	útrapám	21
14	p / n	etapám	19
15	p / n	komorám	19
16	p / n	vrávorám	18
17	p / n	zalapám	13
18	p / n	alpám	12
19	p / n	načapám	12
20	p / n	chalupám	11

3)

1	p / n	kývám	101
2	p / n	plavu	86
3	p / n	šlapu	74
4	p / n	klepu	70
5	p / n	lámú	52
6	p / n	koušu	40
7	p / n	zaklepu	39
8	p / n	nasypu	34
9	p / n	škrábu	32
10	p / n	tápu	32
11	p / n	lapám	32
12	p / n	sypu	28
13	p / n	hýbu	27
14	p / n	řežu	27
15	p / n	koupu	26
16	p / n	kloužu	25
17	p / n	vykoupu	23
18	p / n	vysypu	19
19	p / n	poklepu	18
20	p / n	vrávorám	18

- **-í/-ejí** ve 3. osobě plurálu (typ *sází/sázejí*)
 - jde o slovesa vzoru **sázet**, jejichž kratší tvar je shodný se 3. osobou singuláru + koncovka **í** je velmi homonymní, takže hledáme pouze delší tvary podle dotazu [lc=".*[zžcčsšlřj]ejí"]
 - součástí i některá slovesa III. třídy, ta by šla podobně jako v analýze **-í/-ou** odstranit přidáním atributu lemma, nebo výčtem (není jich mnoho)

1	p / n	její	87053
2	p / n	musejí	4306
3	p / n	nabízejí	3644
4	p / n	přicházejí	2591
5	p / n	vytvářejí	2541
6	p / n	nacházejí	2450
7	p / n	pocházejí	2013
8	p / n	vycházejí	1667
9	p / n	přinášejí	1287
10	p / n	vracejí	1233
11	p / n	ztrácejí	1023
12	p / n	nemusejí	1018
13	p / n	procházejí	921
14	p / n	podílejí	909
15	p / n	kolejí	841
16	p / n	odcházejí	767
17	p / n	souvisejí	694
18	p / n	vyvíjejí	646
19	p / n	přejí	596
20	p / n	odrážejí	575

2. imperativ

- **-0/-ej** (singulár), **-te/-ejte** (plurál)
 - Jelikož se potýkáme s nulovou koncovkou, nelze ji formálně graficky postihnout, navíc se tyto varianty týkají jen malé skupiny slov těžko odlišitelné od ostatních; jde o některá slovesa vzorů *mazat* a *brát*, která přejímají variantní koncovku podle vzoru *dělat*, někdy navíc s alternujícími hláskami nebo kvantitou (*šlap(|ej)*, *plav(|ej)*, *pla(č|kej)*, *lam|lámej*, ...) a ojedinele slovesa kolísající mezi vzory *prosít* a *sázet* (*stav(|ěj)*, ...).
- **-0/-i** (typ *oprost/oprostiti*)
 - Tyto varianty se týkají některých sloves se základem zakončeným na [sš] ě, t [ř] a v menší míře zd, dč a č
 - Dotaz [lc=".*[sš](ť|ti)] dává příliš mnoho nerelevantních výsledků, zvláště kvůli homonymním tvarům zakončeným na *.*sti*, lepší je tedy hledat alternace s nulovou koncovkou (*stě, ště*), i zde ale dostaneme jen málo kýžených výsledků, navíc ne všechna slovesa se chovají stejně: *oprost, oprostiti – odpust, *odpusti*. Specifičtější zakončení je *.*tř(|i)*, *.*+z(d|di)*, *.*dč(|i)*, výsledků však v žádné skupině není mnoho, i proto, že imperativ (těchto sloves) není tolik užívaný.
 - Příklad hledání: [lc=".*tř(|i)"]

1	p / n	tři	52725
2	p / n	uvnitř	11876
3	p / n	dovnitř	9139
4	p / n	bratři	2258
5	p / n	zevnitř	1466
6	p / n	ministři	795
7	p / n	mistři	641
8	p / n	vevnitř	461
9	p / n	tř	188
10	p / n	psychiatři	142
11	p / n	kmotři	123
12	p / n	ušetři	69
13	p / n	lotři	45
14	p / n	šetři	35
15	p / n	utři	31
16	p / n	pediatři	28
17	p / n	arbitři	27
18	p / n	sestři	24
19	p / n	nešetři	23
20	p / n	vicemistři	23

3. přechodník přítomný

- Podle IJP mají variantní tvary slovesa „typu ,třít‘ (dřít, přít se, vřít)“. Kromě jmenovaných jsem našel ještě sloveso *mřít*, všechna tedy pojmem dotazem:

[lc=" [tdpvm] ř (e | íc) (| e) | [tdpvm] r (a | ouc) (| e) "]

- tvary na -e jsou však homonymní s 3. os. sg. prézentu a celkově je dat velmi malé množství

1	p / n	pře	282
2	p / n	vře	198
3	p / n	dře	139
4	p / n	tře	122
5	p / n	pra	66
6	p / n	tra	29
7	p / n	dra	21
8	p / n	mře	11
9	p / n	vra	8
10	p / n	dříc	1

Zdroje:

Křen, M. – Cvrček, V. – Čapka, T. – Čermáková, A. – Hnátková, M. – Chlumská, L. – Jelínek, T. – Kovářiková, D. – Petkevič, V. – Procházka, P. – Skoumalová, H. – Škrabal, M. – Truneček, P. – Vondříčka, P. – Zasina, A.: SYN2015: reprezentativní korpus psané češtiny. Ústav Českého národního korpusu FF UK, Praha 2015. Dostupný z WWW: <http://www.korpus.cz>

<http://prirucka.ujc.cas.cz/>