

CJBB84

Morfologie a korpus

8.00-9.30 G13

v.

Morfologické vlastnosti sloves na základě analýzy korpusových dat

- Morfologické značkování neobsahuje informace o třídě/vzoru
- Je možné vyhledat v korpusech lemmata podle třídy?
- Triviální a netriviální vyhledávání sloves podle třídy a vzoru
- K čemu je takové cvičení dobré?

Morfologické značkování neobsahuje informace o třídě/vzoru

- Jaký je tvar slovesného lemmatu?
- Morfologické značkování neobsahuje informace o třídě/vzoru
- Které kategorie zachycuje morfologická značka?

Je možné vyhledat v korpusech lemmata podle třídy? **Opakování**

třída		vzor
-e	nese	nést
	peče	péci/péct
	bere	brát, vybrat
	maže	mazat
	umře	umřít
-ne	tiskne	tisknout
	mine	minout
	začne	začít
-je	kryje	krýt, pít
	kupuje	kupovat
-í	prosí	prosit, ctít
	trpí	trpět
	sází	sázet

Lze hledat kombinace tvar+tag

- Výhoda: přesnost
- Nevýhoda: pokrytí (dotaz je zúžen na tvary, nicméně k hledané třídě mohou patřit i další slovesa, se která se v použitém korpusu vyskytují, ale nikoliv ve hledaném tvaru)
- přechody sloves mezi třídami





















Dotazy





















- [lc=".*e" & lc!=".*[nj]e" & tag= "VB.S...3.*"]
- [lc=".*ne" & tag= "VB.S...3.*"]
- [lc=".*je" & tag= "VB.S...3.*"]
- [lc=".*í" & tag= "VB.S...3.*"]
- [lc=".*á" & tag= "VB.S...3.*"]

Co opomíjejí první dva dotazy?

	Filtr	lemma	Freq	
1	p / n	být	170285	
2	p / n	moci	109465	
3	p / n	jít	67782	
4	p / n	chtít	35553	
5	p / n	vést	15314	
6	p / n	přijít	14125	
7	p / n	dokázat	13969	
8	p / n	dojít	7517	
9	p / n	psát	7459	
10	p / n	růst	6473	
11	p / n	pomoci	5829	
12	p / n	najít	4801	
13	p / n	nést	4798	
14	p / n	brát	4525	
15	p / n	jet	4347	
16	p / n	vyjít	4041	
17	p / n	ukázat	3974	
18	p / n	vzít	3697	
19	p / n	týkat	3191	
20	p / n	otevřít	2979	

Co opomíjejí první dva dotazy?

	Filtr	lemma	Freq	
1	p / n	začít	12136	
2	p / n	stát	10283	
3	p / n	řící	10108	
4	p / n	dostat	8330	
5	p / n	zůstat	5735	
6	p / n	nabídnout	2922	
7	p / n	rozhodnout	2684	
8	p / n	vzniknout	2457	
9	p / n	zvednout	2272	
10	p / n	dopadnout	2218	
11	p / n	zvládnout	2193	
12	p / n	napadnout	2156	
13	p / n	přestat	2026	
14	p / n	táhnout	1970	
15	p / n	nastat	1847	
16	p / n	proběhnout	1814	
17	p / n	padnout	1543	
18	p / n	plynout	1540	
19	p / n	vládnout	1486	
20	p / n	vytáhnout	1452	

21	p / n	dosáhnout	1331	
22	p / n	všimnout	1328	
23	p / n	poskytnout	1279	
24	p / n	spadnout	1076	
25	p / n	zapomenout	998	
26	p / n	vstát	993	
27	p / n	dotknout	965	
28	p / n	sednout	837	
29	p / n	přikývnout	824	
30	p / n	připomenout	796	
31	p / n	pohlédnout	795	
32	p / n	stihnout	770	
33	p / n	hnát	736	
34	p / n	vzpomenout	732	
35	p / n	stáhnout	681	
36	p / n	zasáhnout	673	
37	p / n	natáhnout	654	
38	p / n	tisknout	652	
39	p / n	klesnout	608	
40	p / n	sáhnout	559	

Podívejme se ještě jednou na tabulku a všimněme si společných a odlišných vlastností sloves podle tříd/vzorů

třída		vzor
-e	nese	nést
	peče	péci/péct
	bere	brát, vybrat
	maže	mazat
	umře	umřít
-ne	tiskne	tisknout
	mine	minout
	začne	začít
-je	kryje	krýt, pít
	kupuje	kupovat
-í	prosí	prosit, ctít
	trpí	trpět
	sází	sázet
-á	dělá	dělat, tkát

Čeho si všímáme?

- Je třeba vyjít ze znalosti gramatiky. Co je rozhodující pro zařazení slovesa ke třídě?
- Jaké jsou společné vlastnosti kmene minulého/infinitivního sloves, patřících ke I., II., III. a IV. třídě?
- Definujeme formální vlastnosti a převedeme je do podoby dotazu v jazyce cql

Triviální vyhledávání sloves podle třídy a vzoru

- Jaké jsou společné vlastnosti tvaru infinitivu sloves I. třídy vzorů *nést* a *péct*?
- Jak formulujeme dotaz na lemma (infinitiv) na základě uvedených vlastností?

Jak na vzory *nést/péct*?












- lemma končí buď na *-ci*, nebo na souhlásku (? libovolnou), za níž následuje koncovka infinitivu *-t*.
- Do podmínky zadáme i to, že chceme vyhledat pouze slovesa. Uvedená zakončení lemmatu mohou mít i další slovní druhy:
- ... bez ohledu na to, do jaké velikosti kdy **dorostu** ... × když jsem skončil v **dorostu** ...
- [lemma=".*([bcdfghjklmnpqrstvzščžřďťň]t|ci)" & tag="V.*"]

Výsledek - konkordance

Výskytů: **983 208** | i.p.m.: **8 070,54** (vztaženo k celému korpusu) | ARF: **585 056,61** | Výsledek je promíchán















1 / 24 581 ▶▶▶

Výběr řádků: základní ▾

<input type="checkbox"/>		Proč by nemohli mít další práci , jež by jim	mohla	třeba na část roku přinést zaměstnání , zisk , užitek
<input type="checkbox"/>		Vánoce měla nateklou ruku . Víc rukodělné to být ani	nemohlo	. Jaká je u vás vlastně pracovní doba ? U
<input type="checkbox"/>		“ , jako by jen jen čekala , až jí	přičtete	psyché , aby si mohla nárokovat ontologickou důstojnost . V
<input type="checkbox"/>		nástupem do nemocnice absolvovat jazykové kurzy ? Ostatní evropské státy	zavedly	pro zahraniční lékaře poměrně náročné jazykové zkoušky . Tyto zkoušky
<input type="checkbox"/>		, a tak se vlastně léčíme . Zjednodušeně se dá	říct	, že každý kámen je složen z neustále kmitajících a
<input type="checkbox"/>		rovnou do Pekingu . Musí se očistit , jinak nebude	mocht	dál žít . Tady mají všichni klapky na očích ,
<input type="checkbox"/>		. Jen tak . Ale protože já jsem já ,	kladl	jsem další hloupé otázky . „ Proč já ? “
<input type="checkbox"/>		galaxií směrem k prázdnému prostoru souhvězdí Labutě se odběratelé časopisu	mohli	dočíst , že velká kupa galaxií zamířila ke kvadrantu ,
<input type="checkbox"/>		Když se jich vyzáblé tváře zeptaly , okřikli je a	přečetli	další z prohlášení , které pak přibíli na dřevěná vrata
<input type="checkbox"/>		spoluautor představení . Kdy se na Hvězdárně a planetáriu Brno	můžete	vydat blíže ke hvězdám ? Stačí se podívat na .
<input type="checkbox"/>		. Kluk , co mě té noci před osmi měsíci	odvezl	na most . Zabuším na dveře záchodů . „ No
<input type="checkbox"/>			...	

Výsledek - lemmata

Celkem: 336 položek (7 stránek)

	Filtr	lemma	Freq	
1	p / n	moci	411048	
2	p / n	říci	177516	
3	p / n	vést	44600	
4	p / n	pomoci	33684	
5	p / n	uvést	31728	
6	p / n	číst	17246	
7	p / n	přinést	17064	
8	p / n	růst	13434	
9	p / n	provést	13064	
10	p / n	nést	12811	
11	p / n	jíst	10226	
12	p / n	klást	7799	
13	p / n	dovést	7749	
14	p / n	přečíst	7357	
15	p / n	utéci	7316	

Odpovězte

- Které konsonanty mohou předcházet před *-t* u sloves typu *nést*?
- Předchází před *-t/-i* u sloves typu *péci/péct* vždy *c*?
- Je vyhledávání sloves podle těchto vzorů opravdu triviální?

Kde nastanou problémy?

- říci
- nařknout
- vyřknout
- odřici/odřeknout
- podřeknout
- [lemma=".*říc[it] | .* (řk | řek)nout"]

	Filtr	lemma	Freq	
1	p / n	říci	177516	
2	p / n	zřici	518	
3	p / n	dořici	348	
4	p / n	vyřknout	308	
5	p / n	přiřknout	255	
6	p / n	nařknout	226	
7	p / n	odřici	189	
8	p / n	zařknout	118	
9	p / n	řici být	109	
10	p / n	vyřici	100	
11	p / n	zhořknout	89	
12	p / n	střici	68	
13	p / n	prořknout	45	
14	p / n	zahořknout	41	
15	p / n	podřeknout	40	
16	p / n	přeřknout	33	
17	p / n	hořknout	26	
18	p / n	uřknout	20	
19	p / n	uřeknout	7	
20	p / n	Těžkořict	3	

Která další slovesa mají tvary podle 1. i 2. třídy?

Jazyk: čeština

<+ ~~společný~~ ~~odlišný~~ +> Morf. specifikace:

vzor 1: .+ c[it] vše .*

vzor 2: .+ knout vše .*

Přidat vzor

Korpus: SYN2015 Frekvence vyšší než: 10 Hledat: lemmata Vyhodnotit: lemmata

A = a ▶ Alternace

Hledat Nové zadání Odkaz na toto zadání: <http://morfio.korpus.cz/OLkeTeZX> Nápověda

výsledek je dosti přegenerovaný

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	obléci (1929)	obléknout (2475)
2	svléci (111)	svléknout (1628)
3	převléci (248)	převléknout (808)
4	navléci (194)	navléknout (750)
5	leci (222)	leknout (287)
6	vyvléci (197)	vyvléknout (133)
7	vysvléci (52)	vysvléknout (122)
8	stříci (30)	stříknout (99)
9	zmoci (1377)	zmoknout (91)
10	provléci (35)	provléknout (66)
11	moci (411376)	moknout (25)
12	povléci (36)	povléknout (18)

Jak na vzory *tisknout/minout*?

- lemma končí na *-nout*.
- Do podmínky zadáme i to, že chceme vyhledat pouze slovesa. Uvedená zakončení lemmatu mohou mít i další slovní druhy:
- ... že se tady lidi museli **vohnout** až na zem ... × uzavřel Emanuel **Vohnout** alias Kamil Emanuel Gott ...
- [lemma=".*nout" & tag="V.*"]











Výsledek - konkordance

Korpus: [syn2020](#) | Dotaz: [.*nout, V.*](#) (872 619 výskytů) ▶ Promíchat: ✓

Výskytů: **872 619** | i.p.m.: **7 162,78** (vztaženo k celému korpusu) | ARF: **483 865,56** | Výsledek je promíchán
















1 / 21 816 ▶▶▶

Výběr řádků:

<input type="checkbox"/>		se pomalu potápěla . Když vpadne Verres do koupelny ,	vybafne	hromovým hlasem : " Lucie , padni na tvář před
<input type="checkbox"/>		zvláštní je , že vědci racionálně promýšlející tyto skutečnosti neumějí	zasáhnout	v pravou chvíli ve svůj prospěch . Těšil se na
<input type="checkbox"/>		se silnými černými obroučkami . Usměje se do davu ,	mrkne	na Zou , která se rychle propletá blíž , jejich
<input type="checkbox"/>		majitele , Leopolda Mánesa , a až by před ní	smekl	klobouk , usmála by se a zastavila by se na
<input type="checkbox"/>		Tak jak nemám být ustaraná ? " Děti . Yasmin	seběhla	po schodech a v ruce držela list papíru . „
<input type="checkbox"/>		plánovali rozšířit „ dětskou republiku “ do jiných táborů ,	vyschly	bakuským psychologům finanční zdroje a museli svůj program začít omezovat
<input type="checkbox"/>		se vrátím a vyřešíme naši záležitost . " Gaston se	zvedl	a se mnou se zatočila podlaha . Musel jsem na
<input type="checkbox"/>		až tam dozadu k němu , pak ho najednou něco	popadne	, a místo aby si se Čmelákem hrál anebo se
<input type="checkbox"/>		sklidila po sobě a šla zpět do obývacího pokoje .	Odsunula	stranou stojací lampu , odsunula petlici a otevřela dveře skryté
<input type="checkbox"/>		v úvahu vozík a uspořádání budovy . Znovu jsem si	klekla	vedle křesla a pohládila Venus po tváři . Stále plakala

Výsledek - lemmata

Celkem: 1 908 položek (39 stránek)

	Filtr	lemma	Freq	
1	p / n	rozhodnout	34053	
2	p / n	vzniknout	19671	
3	p / n	zvednout	18818	
4	p / n	zapomenout	17849	
5	p / n	dosáhnout	17139	
6	p / n	napadnout	17087	
7	p / n	všimnout	16866	
8	p / n	nabídnout	13885	
9	p / n	vytáhnout	13244	
10	p / n	zvládnout	13203	
11	p / n	vzpomenout	10548	
12	p / n	odmítnout	9783	
13	p / n	padnout	9599	
14	p / n	navrhnout	9529	
15	p / n	připomenout	9241	

Odpovězte

- Jak odlišíme slovesa typu *tisknout* a *minout*?
- Vyhledali jsme i slovesa, která nepatří ani k jednomu z uvedených vzorů?
- Je vyhledávání sloves podle těchto vzorů opravdu triviální?
- Jak budeme hledat slovesa typu *začít*?

Jak na vzor *kupovat*?








Jak na vzor *prosit*?

Jak na vzory *trpět/sázet*?

- lemma končí na *-ovat*.
- [lemma=".*ovat" & tag="V.*"]
- lemma končí na *-it*.
- [lemma=".*it" & tag="V.*"]
- lemma končí na *-[eě]t*.
- [lemma=".*[eě]t" & tag="V.*"]
















Narazili jste na sloveso, které splňuje uvedené podmínky a ke vzoru *kupovat* nepatří?

Celkem: 4 428 položek (89 stránek)

	Filtr	lemma	Freq	
1	p / n	pracovat	11171	
2	p / n	pokračovat	6773	
3	p / n	sledovat	6431	
4	p / n	věnovat	5545	
5	p / n	fungovat	4102	
6	p / n	bojovat	3733	
7	p / n	potřebovat	3553	
8	p / n	chovat	3388	
9	p / n	považovat	3258	
10	p / n	reagovat	2789	
11	p / n	pozorovat	2472	
12	p / n	studovat	2390	
13	p / n	pohybovat	2240	
14	p / n	uvažovat	2218	
15	p / n	zachovat	2177	

Narazili jste na sloveso, které splňuje uvedené podmínky a ke vzoru *trpět/sázet* nepatří?

Celkem: 1 460 položek (30 stránek)








	Filtr	lemma	Freq	
1	p / n	muset	196280	
2	p / n	vědět	151504	
3	p / n	vidět	106933	
4	p / n	myslet	81430	
5	p / n	slyšet	33812	
6	p / n	sedět	31883	
7	p / n	držet	27486	
8	p / n	umět	26493	
9	p / n	odpovědět	25620	
10	p / n	jet	25591	
11	p / n	nabízet	24767	
12	p / n	ležet	21905	
13	p / n	chybět	20107	
14	p / n	smět	17725	
15	p / n	uvidět	17677	

1	p / n	příspět
2	p / n	chvět
3	p / n	prospěť
4	p / n	uspět
5	p / n	přimět
6	p / n	dospět
7	p / n	zet
8	p / n	spět
9	p / n	zachvět
10	p / n	pět

slovesa na -ít:

[lemma=".*ít" & tag="V.*"]
















Celkem: 371 položek (8 stránek)

	Filtr	lemma	Freq	
1	p / n	mít	745135	
2	p / n	chtít	192375	
3	p / n	jít	163640	
4	p / n	začít	103080	
5	p / n	přijít	79762	
6	p / n	najít	64261	
7	p / n	vzít	44812	
8	p / n	žít	41122	
9	p / n	dojít	39279	
10	p / n	otevřít	29644	
11	p / n	vyjít	23950	
12	p / n	odejít	20290	
13	p / n	použít	18832	
14	p / n	využít	17980	
15	p / n	projít	15895	

slovesa na –řít:

[lemma=".*řít" & tag="V.*"]
















Celkem: 81 položek (2 stránky)

	Filtr	lemma	Freq	
1	p / n	otevřít	29644	
2	p / n	zemřít	14572	
3	p / n	zavřít	12346	
4	p / n	uzavřít	8302	
5	p / n	umřít	5265	
6	p / n	opřít	5165	
7	p / n	sevřít	3168	
8	p / n	otřít	1867	
9	p / n	upřít	1561	
10	p / n	potřít	1346	
11	p / n	popřít	1265	
12	p / n	utřít	1023	
13	p / n	dřít	912	
14	p / n	zapřít	873	
15	p / n	třít	861	

slovesa na -át:
















[lemma=".*át" & tag="V.*"]

Celkem: 160 položek (4 stránky)

	Filtr	lemma	Freq	
1	p / n	stát	148241	
2	p / n	dát	110864	
3	p / n	hrát	46160	
4	p / n	znát	39951	
5	p / n	zdát	29643	
6	p / n	psát	26156	
7	p / n	brát	23008	
8	p / n	bát	19284	
9	p / n	ptát	17837	
10	p / n	přát	14684	
11	p / n	spát	14345	
12	p / n	vyhrát	12762	
13	p / n	usmát	12427	
14	p / n	smát	10751	
15	p / n	vstát	9743	

slovesa na -ýt a -out (nikoli nout):
[lemma=".*(ý|ou)t" & lemma!=".*nout" & tag="V.*"]

Celkem: 711 položek (15 stránek)

	Filtr	lemma	Freq	
1	p / n	být	4003105	
2	p / n	aby být	254698	
3	p / n	kdyby být	55659	
4	p / n	přijmout	14295	
5	p / n	selbýt	12332	
6	p / n	zaujmout	7109	
7	p / n	zbýt	6060	
8	p / n	obejmout	4073	
9	p / n	přibýt	3244	
10	p / n	ujmout	3008	
11	p / n	zakrýt	2765	
12	p / n	pokrýt	2369	
13	p / n	ty být	2113	
14	p / n	pojmout	2065	
15	p / n	skrýt	2028	

K čemu je takové cvičení dobré?

- Formální znalosti týkající se vlastností, které mají slovesa patřící k jedné třídě/vzoru nám mohou pomoci v případě, že pátráme po formalizovatelných vlastnostech deverbativ.
- Deverbativa se v češtině tvoří od tří základů: kořene, kmene, tvaru.
- V případě derivátů od kmene a tvaru sdílí deriváty od jedné třídy/vzoru stejné formální vlastnosti (kmenotvornou příponu kmene, kmenoutvornou příponu + tvarovou koncovku tvaru).
- Pokud zahrneme formální vlastnosti do podmínky, můžeme lépe popsat formální vlastnosti derivátu (slovesného tvaru).

Otázky

- Může l-ové přídělníky v češtině končit na řetězec **ol**?
- Častými příjmeními jsou tvary homonymní s l-ovými přídělníky (**Drbal, Zlámal, Nezval, Přecechtěl, Vyskočil, ...**). Mohla by k nim patřit i příjmení **Obrtel, Šarel, Herel**?
- Od sloves se tvoří sufixem **č** jména osob (činitelská jména) i jména prostředků činnosti (pojmenované základovým slovesem). Která z následujících substantiv mohou být utvořena popsáním způsobem: **kalač, Kaloč, zteč, kolotoč, Poreč**?
- Které ze sloves nepatří do II. třídy slovesné: *klene, upne, dožene, zatne, vzpomene*?

Úkol na 28. 4. 2021

- Popište, jak lze vyhledat v korpusu substantiva tvořená ze sloves typ *kutil*, tedy jak apelativa, tak propria tvořená substantivizací I-ových příčestí.