

CJBB75

15. 4. Je možné získat z korpusů informace o slovtvorbě?

1. Hledání slov dle počátečního/koncového řetězce

Mějme např. za úkol vyhledat slova tvořená příponou *-tel*. Má-li jít o slova tvořená touto příponou, pak je můžeme hledat dle kombinace lemmatu s koncovým řetězcem *tel*.

[lemma="*.tel"]

[https://kontext.korpus.cz/first?shuffle=1&reload=&corpname=omezeni%2Fsyn2010&usesubcorp=&queryselector=cqlrow&iquery=&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=\[lemma%3D%22.*tel%22\]&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&sca_opus.nazev=](https://kontext.korpus.cz/first?shuffle=1&reload=&corpname=omezeni%2Fsyn2010&usesubcorp=&queryselector=cqlrow&iquery=&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=[lemma%3D%22.*tel%22]&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&sca_opus.nazev=)

Celkem: 933 (19 str.)

	lemma	Frekvence	Freq [%]	
1.	p/n ředitel	24 415	8.3	
2.	p/n přítel	23 968	8.2	
3.	p/n obyvatel	19 459	6.7	
4.	p/n majitel	15 626	5.3	
5.	p/n postel	15 031	5.1	
6.	p/n kostel	13 339	4.6	
7.	p/n hotel	12 802	4.4	

Vidíme, že nechceme např. lemma *postel* (slovo je utvořeno ze slovesa *stlát*) a řetězec *tel* není sufix, ale sč. kořene.

Dotaz přegenerovává.

Co lze dále využít?

Jaké vlastnosti mají jména tvořená příponou *-tel*?

Podíváme-li se do konkordančního seznamu, vidíme, že jsou to názvy osob jako (*věřitel*, *objevitel*, *učitel*).

Lze tohoto pozorování (empirický výzkum) využít?

Ano, můžeme omezit dotaz a vyhledat pouze maskulina životná (využít tagu).

P/filtr

[lemma="*.tel" & tag="NNM.*"]

https://kontext.korpus.cz/filter?within=0&q=~YVQLpvoE&ctxattrs=word&pagesize=40&ref_s=doc%2C%3Dopus.nazev&viewmode=kwic&attrs=word%2Clemma%2Ctag&corpname=o

[mezeni%2Fsyn2010&structs=s&attr_allpos=kw&pnfilter=p&filfl=f&filfpos=0&filtpos=0&inclkwic=True&queryselector=cqlrow&iquery=&lemma=&lpos=&phrase=&word=&wpos=&char=&cql=\[lemma%3D%22.*tel%22+%26+tag%3D%22NNM.*%22\]&default_attr=word&ca_opus.nazev=](#)

Celkem: 477 (10 str.)

	lemma	Frekvence	Freq [%]	
1.	p/n ředitel	24 415	10.4	
2.	p/n přítel	23 968	10.3	
3.	p/n obyvatel	19 459	8.3	
4.	p/n majitel	15 626	6.7	
5.	p/n učitel	10 943	4.7	
6.	p/n podnikatel	9 283	4.0	
7.	p/n uživatel	8 373	3.6	
8.	p/n spisovatel	7 544	3.2	
9.	p/n nepřítel	7 377	3.2	
10.	p/n představitel	6 912	3.0	
11.	p/n velitel	6 557	2.8	
12.	p/n zaměstnavatel	5 810	2.5	
13.	p/n pachatel	5 612	2.4	
14.	p/n zastupitel	5 086	2.2	
15.	p/n dodavatel	4 822	2.1	
16.	p/n pořadatel	3 701	1.6	
17.	p/n spotřebitel	2 839	1.2	
18.	p/n provozovatel	2 763	1.2	
19.	p/n zakladatel	2 718	1.2	
20.	p/n skladatel	2 411	1.0	

Jak jsou tato podstatná jména utvořena?

Co je základové slovo?

Jaký mají význam?

Je-li základové slovo sloveso, které označuje činnost, jíž se zabývá osoba pojmenovaná odvozeným podstatným jménem, pak se tažme:

Patří sem slovo *přítel/nepřítel*?

Od kterého slovesa je odvozeno?

Existuje nějaké pravidlo, jak se od sloves jména tvoří?

Co je součástí přípony/ od kterého slovesného kmene se tvoří takováto podstatná jména?

Co provází odvozování takových substantiv?











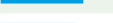



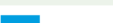

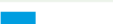

Mohou na *-tel* končit i slova, která nesplňují podmínky, které jsme formulovali výše a která přesahují vlastnosti, jež jsme na základě pozorování dat učinili?

Mohou na *tel* končit substantiva náhodně?

Jsou suffixem *-tel* tvořeny jenom jména osob?

[lc=".*tel"]

Celkem: 841 (17 str.)

	lc	lemma	pos	Frekvence	
1.	p/n	ředitel	N	14 876	
2.	p/n	obyvatel	N	9 348	
3.	p/n	majitel	N	5 592	
4.	p/n	přítel	N	4 652	
5.	p/n	učitel	N	3 463	
6.	p/n	přítel	N	3 355	
7.	p/n	kostel	N	3 198	
8.	p/n	velitel	N	3 153	
9.	p/n	postel	N	3 136	
10.	p/n	hotel	N	3 049	
11.	p/n	spisovatel	N	2 566	
12.	p/n	podnikatel	N	2 357	
13.	p/n	pachatel	N	2 019	
14.	p/n	zaměstnavatel	N	1 915	
15.	p/n	nepřítel	N	1 494	
16.	p/n	uživatel	N	1 219	
17.	p/n	pytel	N	1 095	
18.	p/n	Intel	N	1 063	

Existují nějaké formální (odhlédnuto od významu) vlastnosti jmen osob tvořených ze sloves suffixem *-tel*, jimiž jsme na základě pozorování korpusových dat schopni zobecnit, proč např. *kostel/postel* nepatří mezi příslušná substantiva?

Mohou na *tel* končit tvary slov, která nejsou substantiva?

[lc=".*tel"]

Celkem: 3 (1 str.)

	pos	Frekvence	
1.	p/n N	94 802	
2.	p/n X	941	
3.	p/n V	19	

Je to, že přestože *-l* je jednou výraznou tvarovou koncovkou sloves v češtině, tak na *tel* žádné česlé l-ové přičestí nekončí, náhoda?

<+ společný odlišný +> Morf. specifikace:

vzor 1:

vzor 2:

Další vzor

Korpus: Frekvence vyšší než: Hledají se: Vyhodnocují se:

Velikost písmen: ignorovat

zachovat záznam procesu

	vzor 1 ▲▼ (fq ▲▼)	vzor 2 ▲▼ (fq ▲▼)
1	ředitel (24415)	ředitelka (4025)
2	přítel (23968)	přítelka (2)
3	obyvatel (19459)	obyvatelka (406)
4	majitel (15626)	majitelka (1321)
5	kostel (13339)	kostelka (40)
6	učitel (10943)	učitelka (3293)
7	podnikatel (9283)	podnikatelka (334)
8	uživatel (8373)	uživatelka (22)
9	spisovatel (7544)	spisovatelka (837)
10	představitel (6912)	představitelka (384)
11	velitel (6557)	velitelka (129)
12	zaměstnavatel (5810)	zaměstnavatelka (54)
13	pachatel (5612)	pachatelka (54)
14	zastupitel (5086)	zastupitelka (189)
15	dodavatel (4822)	dodavatelka (5)
16	pořadatel (3701)	pořadatelka (59)
17	spotřebitel (2839)	spotřebitelka (17)
18	činitel (2802)	činitelka (7)
19	provozovatel (2763)	provozovatelka (55)
20	zakladatel (2719)	zakladatelka (174)

Která dvojice do seznamu nepatří?

Co je to *kostelka*?

Příklady konkordancí pro "[lemma="(?)kotelka" & tag="N.*"]" z korpusu syn2010

Eidler, Petr: Sebranka	nezamlouvalo . " Asi myslíte	Kotelku	, ne ? " "
Eidler, Petr: Sebranka	to bylo tak primitivní ?	Kotelka	rovná se Kastelán ? "
Eidler, Petr: Sebranka	ty řečí doma . Ale	Kotelka	, ten ano . "
Eidler, Petr: Sebranka	" Po krátkém hovoru s	Kotelkou	mi Rybář sdělil , že
Eidler, Petr: Sebranka	na pozdrav . " Mistr	Kotelka	? " " Vy budete
Eidler, Petr: Sebranka	ho před něj . Pan	Kotelka	se na to podíval ,
Eidler, Petr: Sebranka	, Kotelka , Kastelán ,	Kotelka	, " opakoval jsem několikrát
Eidler, Petr: Sebranka	svět se nezbožil , pane	Kotelko	. Kdy jste ji viděl
Eidler, Petr: Sebranka	Výsledek návštěvy u amatérského fotografa	Kotelky	nakonec přinesl jen to zjištění
Y: Auto.cz, 14. 4. 2008	dlouholetým přítelem exministra obrany Miroslava	Kotelky	(za ČSSD) .
Y: Auto.cz, 14. 4. 2008	k tomu Stošek . Exministr	Kotelka	, který působí jako velvyslanec
Y: Auto.cz, 14. 4. 2008	a spolužák exministra obrany Miroslava	Kotelky	(za ČSSD) Pavel
Y: Lidové noviny, 18. 4. 2005	televize to totiž byl právě	Kotelka	, který přišel v roce
Y: Reflex, č. 15/2005	. Tehdejší ministr obrany Miroslav	Kotelka	prohlásil , že armáda bude
Y: Týden, č. 27/2005	Paroubkovi , později o generálu	Kotelkovi	a nakonec prosadil veřejnosti naprosto
Y: Právo, 4. 12. 2006	ministři Jaroslav Bašta , Miroslav	Kotelka	a František Bublan (naposled
Y: Právo, 4. 12. 2006	. Po dalších pokusech -	Kotelkovi	podle Vidíma přestal turkmenský protějšek
Y: Právo, 4. 12. 2006	řekl velvyslanec v Moskvě Miroslav	Kotelka	. Turkmenský vojenský přidělenec podle
Y: Právo, 4. 12. 2006	způsobem není , " uvedl	Kotelka	. " Turkmenská strana se
Y: Právo, 4. 12. 2006	všechna povolení , " uvedl	Kotelka	. Česká republika měla písemné
Y: Právo, 4. 12. 2006	vybrán Ašgabat , " uvedl	Kotelka	. Vztahy se středoasijskou republikou
Y: Právo, 4. 12. 2006	Saparmurat Nijazov , jsou podle	Kotelky	standardní . Referendum v Litvinově
Y: Deník Moravia, 29. 3. 2007	2004 Český ministr obrany Miroslav	Kotelka	potvrzuje existenci tajných jednání o
Y: Deník Moravia, 29. 3. 2007	2004 Český ministr obrany Miroslav	Kotelka	potvrzuje existenci tajných jednání o
Y: Právo, 17. 4. 2008	, že exministr obrany Miroslav	Kotelka	(za ČSSD) přišel

Dotaz pro chytré hlavičky:

Existuje ještě nějaká podobná dvojice?

ÚKOL na 22. 4. 2015

Ve většině českých mluvnic i dalších příruček se pracuje s dvojicí sufixů *-ský/-cký*. Vyhledejte v korpusu SYN2010 adjektiva zakončená na *cký* a na *ský* a pozorujte je. Pokuste se argumentovat proč je varianta *cký* poněkud sporná. Všimněte si, od jakých základů (slov) se adjektiva zakončená na *cký* tvoří a definujte v rámci těchto slov výrazné skupiny, které mají nějaké formální i sémantické společné vlastnosti.