

PLIN033_2

MORFIO

Aplikace k odhadování rozsahu a produktivity slovotvorných modelů v češtině na základě korpusových dat

Co dnes chceme?

- představit aplikaci *MORFIO*
- ukázat na konkrétním příkladu postup práce při extrakci podkladů pro lingvistickou analýzu slovotvorných formací automaticky získaných prostřednictvím aplikace *MORFIO*
- zadat úkoly na příště

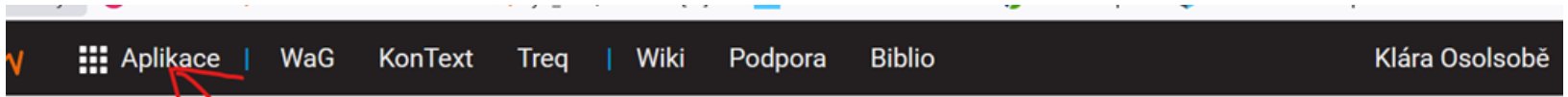
Co je to *MORFIO* ?

- Aplikace *Morfio* slouží k odhadování rozsahu a produktivity slovotvorných modelů v češtině na základě korpusových dat.
- slovotvorný vztah - vytvářen 1) formální shodou/podobností v určitých částech slova, tzv. báze (např. *dřev-* je část společná pro slova *dřevo* i *dřevěný*) a 2) formálními odlišnostmi v částech specifických, tzv. formantech (morfy *-o* a *-ěný* v předchozím příkladu).
- Cílem aplikace je najít všechny dvojice, resp. trojice nebo čtveřice, jednotek v korpusu, které se shodují v bázi a liší se pouze specifikovanými formanty.
- Výstupem aplikace *Morfio* není a nemůže být lingvisticky bezchybný výstup, spíš se jedná o pomůcku, která množství dat dokáže pro lingvistické účely předzpracovat.

MORFIO

- <http://morfio.korpus.cz/>
- <http://ucnk.ff.cuni.cz/bonito/znacky.php>
- http://wiki.korpus.cz/doku.php/pojmy:morfologicka_analyza
- <http://wiki.korpus.cz/doku.php/pojmy:tag>

Jak vypadá přístup



Morfio

Jazyk: čeština ▾

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

vzor 1: vše ▾ .*

vzor 2: vše ▾ .*

Přidat vzor

Korpus: ▾ Frekvence vyšší než: ▾ Hledat: ▾ Vyhodnotit: ▾

A = a

Hledat

Volba korpusu

- SYN2010
- SYN2005
- SYN2015
- Araneum CS Minus

Další volby

- Frekvence vyšší než (nastavíme-li 10, pak sice redukuje přegenerování, můžeme ovšem také podgenerovat, neboť produktivita je schopnost být modelem pro nová tvoření)
- Hledat / Vyhodnotit – záleží na zadání Morfologické specifikace
- Alternace – opět je třeba zohlednit míru pře-/pod-generování a také sladit zadání alternace se zadáním dotazu

[https://wiki.korpus.cz/doku.php/cnk:u](https://wiki.korpus.cz/doku.php/cnk:uvod) vod

Přehled dostupných korpusů

| Korpusy psaného jazyka (synchronní) | | | | | |
|-------------------------------------|-----------------------|-------------|---------------------|-------------------------------|-------------------------------------------------------------------------------------------------|
| korpus | velikost (počet slov) | lemmatizace | morfologické značky | rok zveřejnění ⁽¹⁾ | charakteristika korpusu |
| Obecné korpusy | | | | | |
| SYN (verze 8) | 4,5 mld. | ✓ | ✓ | 2010-2019 | verzovaný korpus, spojující synchronní psané korpusy řady SYN a další, dosud nezveřejněné texty |
| SYN2020 | 100 mil. | ✓ | ✓ | 2020 | referenční reprezentativní korpus, převažují texty z let 2015–2019 |
| SYN2015 | 100 mil. | ✓ | ✓ | 2015 | referenční reprezentativní korpus, převažují texty z let 2010–2014, s novou klasifikací textů |
| SYN2013PUB | 935 mil. | ✓ | ✓ | 2013 | referenční korpus publicistických textů z let 2005–2009 |
| SYN2010 | 100 mil. | ✓ | ✓ | 2010 | referenční reprezentativní korpus, převažují texty z let 2005–2009 |
| SYN2009PUB | 700 mil. | ✓ | ✓ | 2010 | referenční korpus publicistických textů z let 1995–2007 |
| SYN2006PUB | 300 mil. | ✓ | ✓ | 2006 | referenční korpus publicistických textů z let 1989–2004 |
| SYN2005 | 100 mil. | ✓ | ✓ | 2005 | referenční reprezentativní korpus, převažují texty z let 2000–2004 |
| SYN2000 | 100 mil. | ✓ | ✓ | 2000 | referenční reprezentativní korpus, převažují texty z let 1990–1999 |

Zadání dotazu

- regulární výrazy
- morfologické značky (viz výše)
- další možnosti zobecnění (rozlišení samohlásek, souhlásek, ... Viz **Nápověda**)
- volby různých typů alternací

Co chceme?

- Kolik a která substantiva typu *učitel* mají ženský protějšek (přechýlování/moce) tvořený příponou *-ka*
- Jaké další dvojice tohoto typu známe?
- Jak je můžeme popsat?
- Jaké mají formální vlastnosti?

-tel / -telka

- Substantivum maskulinum životné
- Lemma končí na *tel*
- Substantivum femininum
- Lemma končí na *telka*

Vyplnění formuláře

Jazyk: čeština ▾

<+ ✖ společný ▾ ✖ odlišný ▾ +> Morf. specifikace:

vzor 1: ▾ ▾ ▾

vzor 2: ▾ ▾ ▾

Přidat vzor

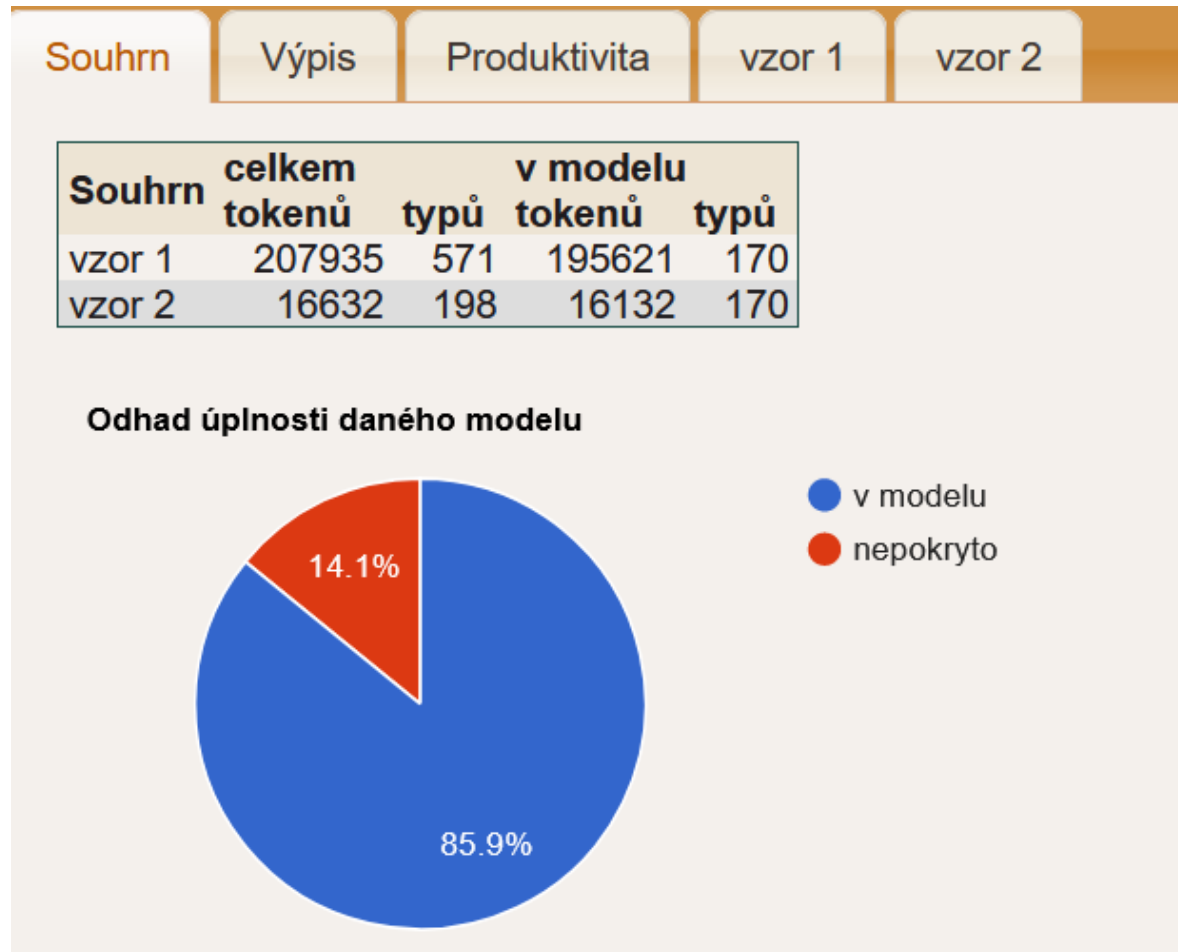
Korpus: ▾ Frekvence vyšší než: Hledat: ▾ Vyhodnotit: ▾

A = a

Odkaz na toto zadání: <http://morfio.korpus.cz/kPj5YJXM>

SOUHRN

V záložce **souhrn** jsou uvedeny počty typů s nadlimitní frekvencí a součet jejich výskytů. Jedna sada údajů (sloupec "celkem") se vždy týká vzoru samotného (chápaného izolovaně), druhá sada (sloupec "v modelu") pak odkazuje k těm jednotkám příslušejícím ke vzoru, které zároveň patří do analyzovaného slovtvorného modelu, tj. slova, která mají k sobě odvozeninu identifikovanou v rámci druhého vzoru.



Výpis

V tabulce jsou uvedeny všechny doklady ze všech vzorů, které vstupují do zadaného modelu. Červená část slov označuje společnou bázi (ta se může lišit pouze v případě aplikace alternací). V závorkách uvedený údaj představuje celkovou frekvenci jednotky ve zvoleném korpusu. Tabulku je možné přetřídit podle libovolného sloupce a to jak abecedně, tak frekvenčně pomocí šipek v záhlaví tabulky. Každé slovo zároveň funguje jako odkaz směřující k ukázce konkordancí ve zvoleném korpusu.

| | Souhrn | Výpis | Produktivita | vzor 1 | vzr |
|----|--------------------------|---------------------------|--------------|--------|-----|
| | vzor 1 ▲▼ (fq ▲▼) | vzor 2 ▲▼ (fq ▲▼) | | | |
| 1 | badatel (1662) | badatelka (68) | | | |
| 2 | bořitel (18) | bořitelka (1) | | | |
| 3 | buditel (69) | buditelka (1) | | | |
| 4 | budovatel (112) | budovatelka (4) | | | |
| 5 | cestovatel (1255) | cestovatelka (97) | | | |
| 6 | ctitel (379) | ctitelka (49) | | | |
| 7 | cvičitel (210) | cvičitelka (96) | | | |
| 8 | čekatel (144) | čekatelka (12) | | | |
| 9 | činitel (1569) | činitelka (6) | | | |
| 10 | dobyvatel (356) | dobyvatelka (7) | | | |
| 11 | dodavatel (5024) | dodavatelka (10) | | | |
| 12 | dohlížitel (92) | dohlížitelka (2) | | | |
| 13 | dopisovatel (158) | dopisovatelka (23) | | | |
| 14 | doručitel (51) | doručitelka (1) | | | |
| 15 | doručovatel (114) | doručovatelka (39) | | | |
| 16 | dovolatel (18) | dovolatelka (8) | | | |
| 17 | držitel (1446) | držitelka (215) | | | |
| 18 | exředitel (38) | exředitelka (4) | | | |
| 19 | hlasatel (526) | hlasatelka (119) | | | |
| 20 | hodnotitel (221) | hodnotitelka (7) | | | |
| 21 | hostitel (1315) | hostitelka (354) | | | |
| 22 | hubitel (11) | hubitelka (2) | | | |
| 23 | hybatel (178) | hybatelka (6) | | | |
| 24 | chovatel (1790) | chovatelka (107) | | | |
| 25 | jednatel (1067) | jednatelka (103) | | | |

Výpis lze

- seřadit podle frekvence 1. nebo 2. členu dvojice

| | vzor 1 ▲▼ (fq ▲▼) | vzor 2 ▲▼ (fq ▲▼) |
|----|----------------------|----------------------|
| 1 | ředitel (24415) | ředitelka (4025) |
| 2 | obyvatel (19459) | obyvatelka (406) |
| 3 | majitel (15626) | majitelka (1321) |
| 4 | učitel (10943) | učitelka (3293) |
| 5 | podnikatel (9283) | podnikatelka (334) |
| 6 | uživatel (8373) | uživatelka (22) |
| 7 | spisovatel (7544) | spisovatelka (837) |
| 8 | představitel (6912) | představitelka (384) |
| 9 | velitel (6557) | velitelka (129) |
| 10 | zaměstnavatel (5810) | zaměstnavatelka (54) |
| 11 | pachatel (5612) | pachatelka (54) |
| 12 | zastupitel (5086) | zastupitelka (189) |
| 13 | pořadatel (3701) | pořadatelka (59) |
| 14 | spotřebitel (2839) | spotřebitelka (17) |
| 15 | provozovatel (2763) | provozovatelka (55) |
| 16 | zakladatel (2718) | zakladatelka (174) |
| 17 | skladatel (2411) | skladatelka (69) |
| 18 | vyšetřovatel (2394) | vyšetřovatelka (95) |
| 19 | žadatel (2171) | žadatelka (23) |
| 20 | pozorovatel (2129) | pozorovatelka (21) |

Výpis lze

- seřadit abecedně podle 1. nebo 2. členu dvojice

| | vzor 1 ▾ (fq ▲▼) | vzor 2 ▲▼ (fq ▲▼) |
|----|-------------------|--------------------|
| 1 | badatel (1421) | badatelka (33) |
| 2 | cestovatel (1662) | cestovatelka (68) |
| 3 | ctitel (473) | ctitelka (60) |
| 4 | cvičitel (159) | cvičitelka (52) |
| 5 | čekatel (306) | čekatelka (12) |
| 6 | dopisovatel (188) | dopisovatelka (13) |
| 7 | doručovatel (136) | doručovatelka (50) |
| 8 | držitel (1729) | držitelka (177) |
| 9 | hlasatel (489) | hlasatelka (114) |
| 10 | hostitel (1586) | hostitelka (352) |
| 11 | chovatel (1107) | chovatelka (53) |
| 12 | jednatel (1294) | jednatelka (121) |
| 13 | krotitel (110) | krotitelka (59) |
| 14 | léčitel (400) | léčitelka (75) |
| 15 | majitel (15626) | majitelka (1321) |
| 16 | nakladatel (796) | nakladatelka (22) |
| 17 | navrhovatel (352) | navrhovatelka (40) |
| 18 | ničitel (89) | ničitelka (14) |
| 19 | nositel (1936) | nositelka (178) |
| 20 | obdivovatel (628) | obdivovatelka (89) |

Kliknutím na lemma bylo možné získat konkordanci v příslušném korpusu tato funkce je „mrtvá“

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

Fuks, Ladislav: Vévodkyně a kuchařka

Strachey, Lytton: Alžběta a Essex

Y: Austrálie

Gieselmann, David: Pan Kolpert

Mácha, Karel: Integrální antropolo...

Vacková, Jarmila: Odpovědi obrazů - Mi...

Gurevič, Aron Jakovle...: Nebe, peklo, svět - ...

Gurevič, Aron Jakovle...: Nebe, peklo, svět - ...

Brunschvicg, Léon: Evropský duch

Bauman, Zygmunt: Úvahy o postmoderní době

Potůček, Martin (ed.): Manuál prognostickýc...

Keller, Jan: Teorie modernizace

Y: Ikarie, č. 4/2005

Studnička, Miloslav: Masožravé rostliny

Benda, Ivo A. (et al.): VI. rozhovory s pouč...

Y: Technik, č. 7/2005

Koukolík, František -...: Šimpanz a vesmír

Štoll, Ivan: Podivuhodné přírodní úkazy

Wolf, Josef: Kdo je kdo není v hn...

Y: S tebou mě baví život, č. 2/2008

Kalina, Pavel - Koťát...: Praha 1310-1437

Y: Právo, 3. 9. 2005

Y: Lidové noviny, 17. 7. 2006

Y: Mladá fronta DNES, 25. 4. 2007

Y: Sedmička, č. 18/2009

společníci - kapitán navíc jistý **badatel** , cestovatel a znalec Afriky
úzkostném napětí a některé moderní **badatele** vedla k domněnce , že
Průkopníci a badatelé Mnoho evropských **badatelů** , včetně Edwarda Eyra a
takovou naprosto banální odpověď : **badatel** v oblasti chaosu - tedy
snad zdát tato formulace konkrétnímu **badateli** v některé přírodovědní disciplíně příliš
Malerei (1925) německý **badatel** M. J. Friedländer s tím
s tvou služebnicí ? " **Badatelé** spatřují v rituálu " snižování
Evropě 13 . století odhalují **badatelé** na několik stovek , vedlo
: " Cítím se být **badatelem** . Žízním po úplném poznání
městské kultury ; pro mnoho **badatelů** se stal zevloun hlavním symbolem
souvislosti , které nejsou zřejmé **badatelům** v individuálních výzkumných oblastech .
sponzorů ztrácejí univerzitní učitelé a **badatelé** status nezávislých odborníků . Část
Cristea , občan Ukrajiny a **badatel** . Měl jednu bodnou ránu
darlingtonii a pavoukem zmiňují různí **badatelé** , včetně historických zpráv zmíněné
Uvědomte si proto , pozemští **badatelé** , že to , co
let dvacátého století se **badatelům** nepodařilo dostat se při snižování
řikají stochastickými - vlivy . **Badatelé** zkoumající dlouověkost drosofil se samozřejmě
1597) , holandský polární **badatel** , když v roce 1596
se vyvíjela od původního kroužku **Badatelů** Bible v sedmdesátých letech 19
F. X. Šalda . Literární **badatel** Radko Pytlík však shrnul spisovatelovo
jižní pilíř unikál pozornosti novodobých **badatelů** . Je to patrně dáno
z pozdní doby bronzové našel **badatel** spolupracující s mladoboleslavským muzeem v
je nezískal vůbec . Cambridgeští **badatelé** řadu let zkoumali třináct skupin
poškozené spisy neměly být přístupné **badatelům** . " Můžeme ho ukázat
 , spoustu zajímavých informací však **badatelé** objev i v hradeckém Okresním

Lze ovšem zadat dotaz do korpusu [tag="N.F.*" & lemma="hodnotitelka"]

Korpus: [syn2015](#) | Dotaz: [N.F.*](#), [hodnotitelka](#) (7 výskytů) ▶ Promíchat: ✓

Výskytů: 7 | i.p.m.: 0,06 (vztaženo k celému korpusu) | ARF: 3,64 | Výsledek je promíchán 1 / 1

Výběr řádků: základní ▾

| | | | | |
|--------------------------|--|---------------------------------------------------------------------|----------------------|------------------------------------------------------------------------|
| <input type="checkbox"/> | | co vybírat . klára Donathová , recenzentka restaurací Autorka je | hodnotitelkou | Scuk . cz Müller Thurgau 2011 kabinet VÍNO / PIVO |
| <input type="checkbox"/> | | vymyslel i další příšernosti , nad kterými by musela centralizovaná | hodnotitelka | mé práce ržát blahem . A i kdyby neržála a |
| <input type="checkbox"/> | | na mase . To platilo i tehdy , když sama | hodnotitelka | vegetariánka nebyla . Podle Havlíčka páchnou muži , kteří jedí |
| <input type="checkbox"/> | | opatření a aktivit na podporu rodin v obcích . „ | Hodnotitelkám | se velmi líbilo kunovické Komunitní centrum , „ poodhaluje podrobnosti |
| <input type="checkbox"/> | | anketám . Svoje místo na slunci už získaly i ženské | hodnotitelky | , které vyhláší Světové auto roku podle žen . A |
| <input type="checkbox"/> | | . A myslím , že v dobrém slova smyslu jsme | hodnotitelky | „ dorazili “ prezentací folklorních souborů , sportovních klubů a |
| <input type="checkbox"/> | | je – podle vlastního názoru i dle hodnocení několika nezávislých | hodnotitelek | – nejskvělejší kluk ve Wolverhamptonu . Má kompletní antologii Byrds |

Vzor 1

- Výsledky analýzy jednotlivých vzorů jako samostatných dotazů jsou prezentovány ve formě tabulky jednotek (slovních tvarů nebo lemmat) spolu s jejich frekvencemi ve zvoleném korpusu. Tabulku je možné doplnit i o jednotky, které v modelu nebyly brány v potaz, protože jejich frekvence byla nižší než uživatelem stanovený limit. Údaje zvýrazněné barevným pozadím se účastní slovotvorného modelu (tj. existuje k nim v druhém vzoru protějšek se stejnou bází, lišící se pouze formanty).

tel (vzor 1)

- Žlutě jsou ty členy, k nimž byl nalezen vzor 2

skryt formy s podlimitní frekvencí

skupiny +/-

| abc/cba ▲▼ | fq ▲▼ |
|-------------|-------|
| badatel | 1421 |
| bortel | 33 |
| bořitel | 15 |
| buditel | 76 |
| budovatel | 159 |
| cestovatel | 1662 |
| ctitel | 473 |
| cvičitel | 159 |
| čekatel | 306 |
| činitel | 2277 |
| čítatel | 32 |
| datel | 135 |
| dělitel | 13 |
| dobyvatel | 441 |
| dodavatel | 4822 |
| dohlížitel | 30 |
| dohlížitel | 61 |
| dopisovatel | 188 |
| doručitel | 40 |
| doručovatel | 136 |
| dovolatel | 14 |
| dražitel | 73 |
| držitel | 1729 |
| expřitel | 31 |
| exředitel | 28 |
| gerstel | 11 |

Co lze vyčíst ze Vzoru 1?

- Na výsledcích vidíme nedostatky malého korpusu. Spočívají v tom, že méně obvyklá přechýlená jména v korpusu menšího rozsahu nejsou doložena. V důsledku toho je výzkum produktivity prováděný na jednotlivých menších korpusech nutně zkreslený.
- v SYNv11 jsou doložena přechýlená jména: *buditelka* - https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=lemma_lc&attrs=tag&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.title&q=~rGmeAsycq0OS&cutoff=
exředitelka
https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=lemma_lc&attrs=tag&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.title&q=~o8giMAAwIKwu&cutoff=0 a
další

telka (vzor2)

- Žlutě jsou ty členy, k nimž byl nalezen vzor 1

skrýt formy s podlimitní frekvencí

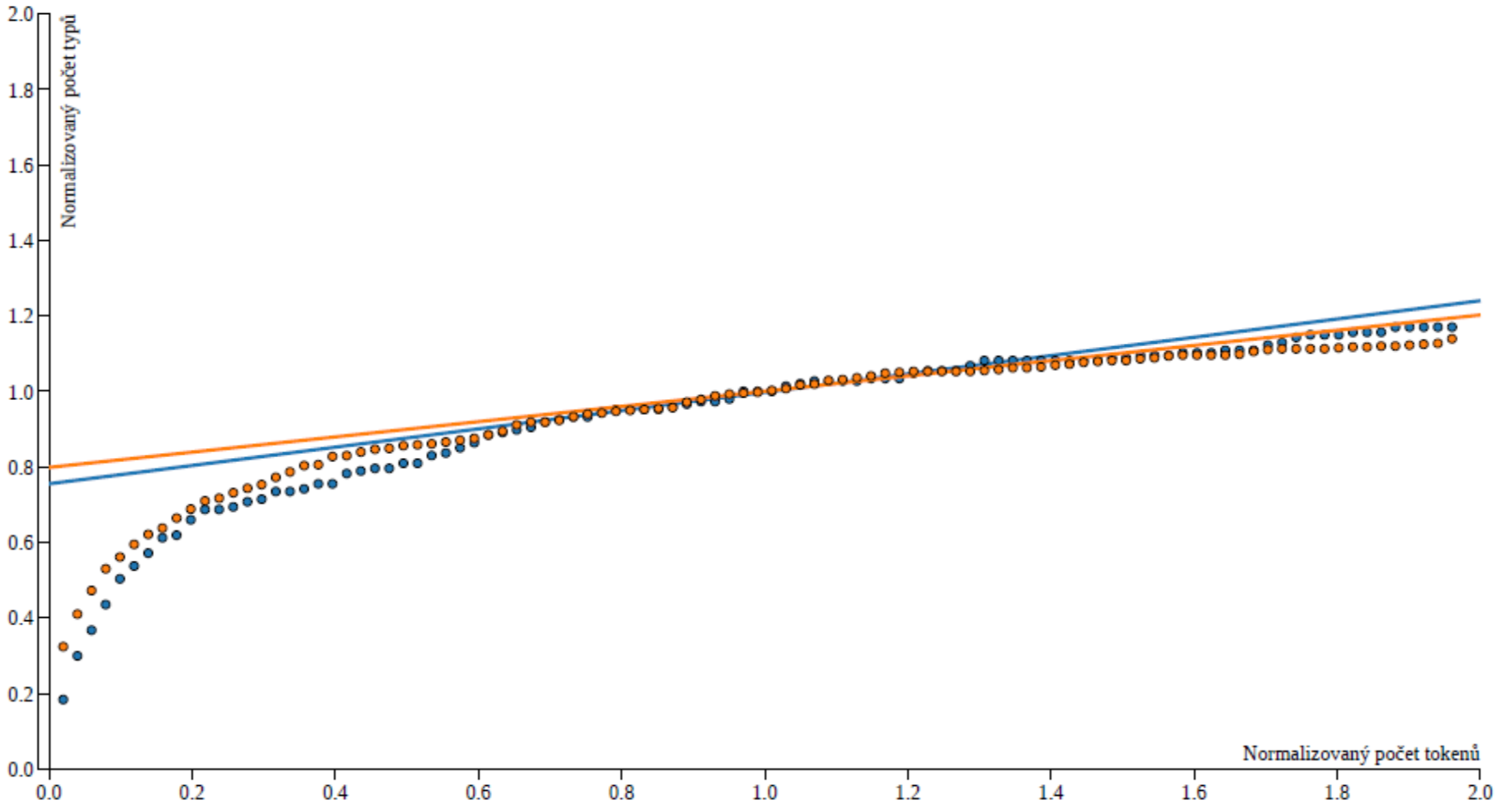
skupiny +/-

| abc/cba ▲▼ | fq ▲▼ |
|---------------|-------|
| badatelka | 33 |
| cestovatelka | 68 |
| ctitelka | 60 |
| cvičitelka | 52 |
| čekatelka | 12 |
| dopisovatelka | 13 |
| doručovatelka | 50 |
| držitelka | 177 |
| hlasatelka | 114 |
| hostitelka | 352 |
| chovatelka | 53 |
| jednatelka | 121 |
| krotitelka | 59 |
| léčitelka | 75 |
| majitelka | 1321 |
| nakladatelka | 22 |
| navrhovatelka | 40 |
| ničitelka | 14 |
| nositelka | 178 |
| obdivovatelka | 89 |
| objednatelka | 12 |
| obyvatelka | 406 |
| opatrovatelka | 16 |
| ošetřovatelka | 879 |
| oznamovatelka | 16 |
| pachatelka | 54 |
| pastelka | 236 |

Co lze vyčíst ze Vzoru 2?

- Na konci seznamu si všimněte slova *pastelka*, které splňuje formální zadání, nicméně neodpovídá mu žádné životné maskulinum. Pokud bychom nedali do dotazu *Morfia* podmínku `tag=NNM.*`, tak bychom dostali přegenerovanou dvojici *pastel/pastelka*.
- Je možné zúžit dotaz bez podmínky tagu tak, abychom přegenerovanou dvojici vyloučili?
- Je možné, aby činitelské jméno v češtině končilo na `.*stel?` (https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=lemma_l&attrs=tag&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.title&q=~0GSWOcQsYOwA&cutoff=0)

Produktivita



Měření produktivity

- Odhad produktivity obou vzorů a jejich vzájemné porovnání vychází z teoretických poznámek H. Baayena (viz [zde](#)). Morfologická produktivita se zde měří pomocí odhadu tendence přírůstku nových typů při přírůstku dokladů (tokenů) pro každý vzor samostatně. Ze srovnání pak vyplývá, který vzor je produktivní, protože počet jeho typů roste rychleji, s jeho formanty se pojí nové a nové báze, a který vzor je naopak neproduktivní a potenciálně uzavřený (i když třeba frekventovaný a rozsáhlý).

kapital.* /social.*

Morfio

<+ odlišný společný Morf. specifikace: +>

vzor 1:

vzor 2:

Další vzor

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

Velikost písmen: ignorovat

Odkaz na toto zadání: <http://morfio.korpus.cz/ct3PnOhl>

Souhrn

Výpis

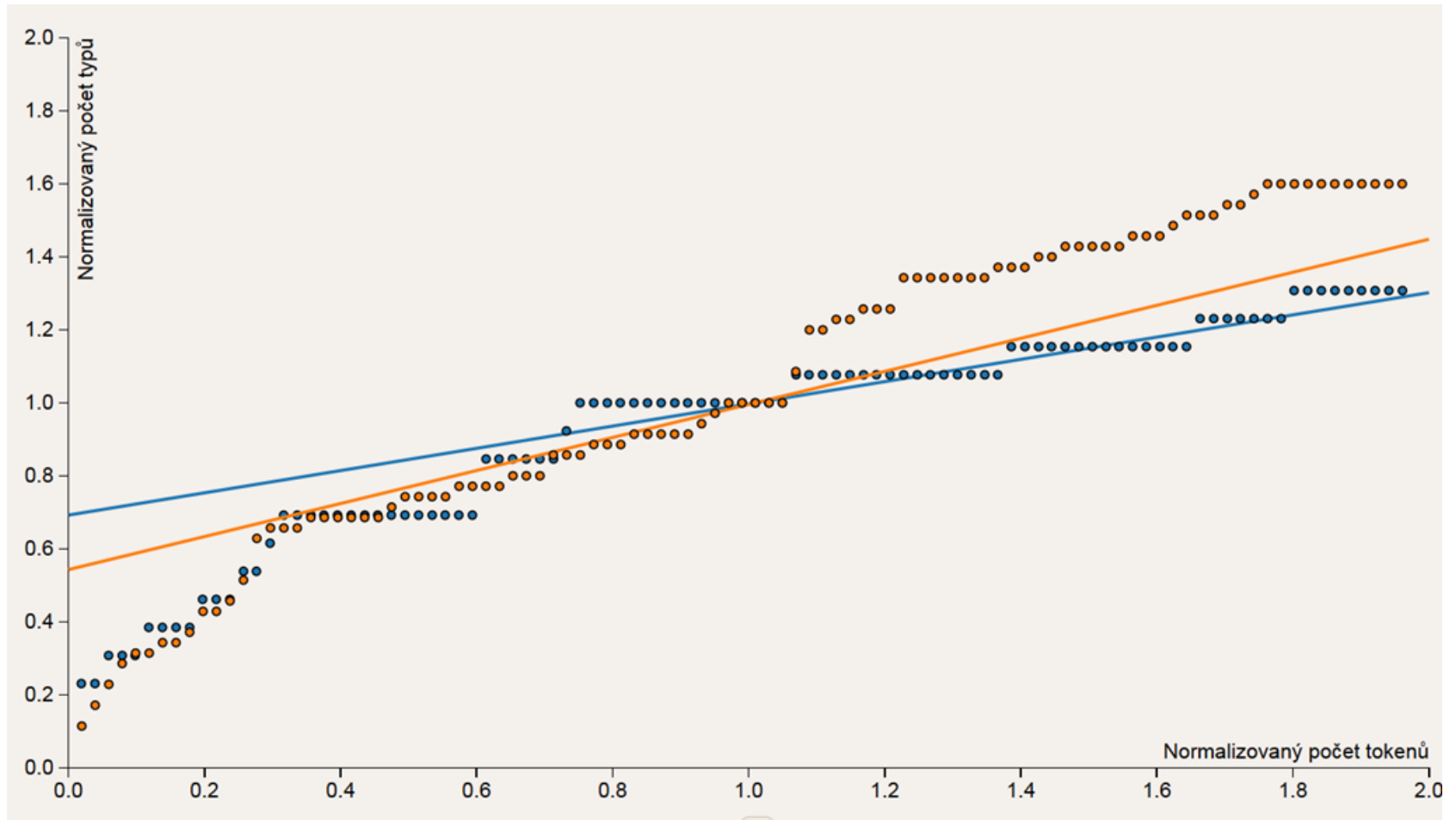
Odhad úplnosti daného modelu



- v modelu
- nepokryto

| vzor 1 ▲▼ (fq ▲▼) | vzor 2 ▲▼ (fq ▲▼) |
|------------------------|----------------------|
| 1 kapitalismus (1011) | socialismus (1497) |
| 2 kapitalista (203) | socialista (989) |
| 3 kapitalistický (538) | socialistický (2446) |
| 4 kapitalizace (68) | socializace (399) |

produktivita



Vzor 1

skupiny +/-

| abc/cba ▲▼ | fq ▲▼ |
|----------------|-------|
| kapitalismus | 1011 |
| kapitalista | 203 |
| kapitalistický | 538 |
| kapitalizace | 68 |

Vzor 2

skupiny +/-

| abc/cba ▲▼ | fq ▲▼ |
|--------------------------|-------|
| social | 190 |
| socialbakers | 31 |
| socialismus | 1497 |
| socialista | 989 |
| socialistickorealistický | 11 |
| socialisticky | 27 |
| socialistický | 2446 |
| socialistka | 20 |
| socializace | 399 |
| socializační | 74 |
| socializovaný | 29 |
| socializovat | 39 |
| socializující | 13 |

pomocí aplikace morfio vyhledejte v korpusu syn2015 dvojice
základové slovo=adjektivum, odvozené slovo= substantivum
název vlastnosti

Morfio

Jazyk: čeština

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

vzor 1: .+ [ýí] přidavná jména A.*

vzor 2: .+ ost podstatná jména N.*

Přidat vzor

Korpus: SYN2015 Frekvence vyšší než: 0 Hledat: lemmata Vyhodnotit: lemmata

A = a ▶ Alternace

Hledat Nové zadání Odkaz na toto zadání: <http://morfio.korpus.cz/ZvBgS20r> Nápověda

seznam podle frekvence adjektiv

| | vzor 1 ▲▼ (fq ▲▼) | vzor 2 ▲▼ (fq ▲▼) |
|----|-------------------|-------------------|
| 1 | velký (190639) | velkost (18) |
| 2 | nový (137415) | novost (160) |
| 3 | celý (133678) | celost (73) |
| 4 | jiný (124171) | jinost (3) |
| 5 | dobrý (116272) | dobrost (3) |
| 6 | český (102231) | českost (30) |
| 7 | malý (94847) | malost (133) |
| 8 | vysoký (78490) | vysokost (6) |
| 9 | starý (69518) | starost (9779) |
| 10 | vlastní (59617) | vlastnost (12735) |
| 11 | možný (52381) | možnost (43140) |
| 12 | mladý (48961) | mladost (72) |
| 13 | různý (48517) | různost (205) |
| 14 | jediný (46240) | jedinost (8) |
| 15 | stejný (41494) | stejnost (49) |
| 16 | důležitý (40918) | důležitost (1766) |
| 17 | známý (34311) | známost (1073) |
| 18 | evropský (31600) | evropskost (3) |
| 19 | plný (30642) | plnost (270) |
| 20 | špatný (30445) | špatnost (209) |

přidán variantní sufix -ota

Jazyk: čeština

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

vzor ~~×~~ 1: .+ [ýí] přídavná jména A.*

vzor ~~×~~ 2: .+ ost podstatná jména N.*

vzor ~~×~~ 3: .+ ota podstatná jména N.*

Přidat vzor

Korpus: SYN2015 Frekvence vyšší než: 0 Hledat: lemmata Vyhodnotit: lemmata

A = a ▶ Alternace

Hledat Nové zadání Odkaz na toto zadání: <http://morfio.korpus.cz/lhbwl2rJ> Nápověda

seznam podle substantiv na -ota

| | vzor 1 ▲▼ (fq ▲▼) | vzor 2 ▲▼ (fq ▲▼) | vzor 3 ▲▼ (fq ▲▼) |
|----|-------------------|-------------------|-------------------|
| 1 | hodný (5344) | hodnost (841) | hodnota (29676) |
| 2 | teplý (8817) | teplot (1) | teplota (15213) |
| 3 | temný (7125) | temnost (12) | temnota (2661) |
| 4 | prázdný (11068) | prázdnost (9) | prázdnota (1618) |
| 5 | dobrý (116272) | dobrost (3) | dobrota (1322) |
| 6 | nahý (4180) | nahost (5) | nahota (461) |
| 7 | slepý (3046) | slepost (4) | slepot (352) |
| 8 | nový (137415) | novost (160) | novota (331) |
| 9 | prostý (5544) | prostost (1) | prostota (262) |
| 10 | mastný (1572) | mastnost (1) | mastnota (184) |
| 11 | hluchý (722) | hluchost (3) | hluchota (140) |
| 12 | němý (1214) | němost (3) | němota (133) |
| 13 | živý (12759) | živost (124) | života (119) |
| 14 | černý (29197) | černost (3) | černota (106) |
| 15 | cizí (11286) | cizost (47) | cizota (64) |
| 16 | milý (7399) | milost (2209) | milota (45) |
| 17 | měkký (5206) | měkkost (191) | měkkota (37) |
| 18 | teskný (137) | tesknost (2) | tesknota (36) |
| 19 | slabý (9391) | slabost (2027) | slabota (25) |
| 20 | tichý (6918) | tichost (618) | tichota (18) |
| 21 | malý (94847) | malost (133) | malota (9) |
| 22 | suchý (6382) | suchost (56) | suchota (2) |
| 23 | zlý (7155) | zlost (1270) | zlota (2) |
| 24 | mladý (48961) | mladost (72) | mladota (1) |
| 25 | mokrý (3610) | mokrost (1) | mokrota (1) |

OTÁZKY

- Sledujte doklady, které se vám z nějakého důvodu zdají nepatřičné.
- Zobrazte si jejich výskyty v příslušném korpusu a snažte se najít postup, jak přeformulovat dotaz tak, aby výsledek nebyl přegenerovaný.

Jak hledat substantiva typu *soudce*

- <http://morfo.korpus.cz/cm0fitx0>
- https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=lemma&attrs=tag&attr_vmode=visible-kwic&base_viewattr=word&structs=p&refs=%3Ddoc.title&q=~kCq8Waqiiil8&cutoff=0

Výsledek

| pattern 1 ▲▼ (fq ▲▼) | pattern 2 ▲▼ (fq ▲▼) |
|--------------------------------|----------------------|
| 1 dohodit dohodnout (4848) | dohodce (5) |
| 2 dopravit (1303) | dopravce (1457) |
| 3 obhájit (1214) | obhájce (1489) |
| 4 obránit (2) | obránce (3316) |
| 5 odhadnout (2981) | odhadce (59) |
| 6 ochránit (2288) | ochránce (1353) |
| 7 opravit (3507) | opravce (6) |
| 8 popravit (898) | popravce (1) |
| 9 poradit (6766) | poradce (3150) |
| 10 přepravit (465) | přepravce (184) |
| 11 rozhodit rozhodnout (34422) | rozhodce (135) |
| 12 rozhodnout rozhodit (34422) | rozhodce (135) |
| 13 rozvozit (1) | rozvozce (5) |
| 14 soudit (4568) | soudce (5353) |
| 15 spoluvládnout (7) | spoluvládce (9) |
| 16 strážit (50) | strážce (2116) |
| 17 vinit (689) | vince (56) |
| 18 vládnout vládat (5361) | vládce (1749) |
| 19 zachránit (8534) | zachránce (365) |

Přegenerování dotazu

Concordance examples for "[lemma="(?)vince" & tag=".*"]" in the corpus syn2015

Baker, Tiffany,Sladká vůně bolehlavu

Gerritsen, Tess,Suvenýr

Gerritsen, Tess,Suvenýr

Gerritsen, Tess,Suvenýr

Gerritsen, Tess,Suvenýr

Gerritsen, Tess,Suvenýr

Gerritsen, Tess,Suvenýr

Gerritsen, Tess,Suvenýr

Herbert, Frank,Zelený mozek

Tsiolkas, Christos,Facka

Tsiolkas, Christos,Facka

Grisham, John,Calico Joe

Grisham, John,Calico Joe

Grisham, John,Calico Joe

Grisham, John,Calico Joe

Grisham, John,Calico Joe

Grisham, John,Calico Joe

Ryan, Amy Kathleen,Vesmírná pout' 2

Ryan, Amy Kathleen,Vesmírná pout' 2

Giffin, Emily,Na čem záleží

Jakubowski, Maxim (ed.),Světové krimipovídky

Jakubowski, Maxim (ed.),Světové krimipovídky

Jakubowski, Maxim (ed.),Světové krimipovídky


Dunbar, Robin,Příběh rodu Homo

Eliot, Marc,Michael Douglas

protáhlý míč se snadným půvabem **Vince** Lombardiho . Běh by byl
grilu . Bude tam i **Vince** Korsak , takže to bude
Frank , ale penzionovaný detektiv **Vince** Korsak . Pečlivě obracel š'avnaté
svoji ženu , nahradil ho **Vince** Korsak . Vládou nad grilem
Věř mi , " pokračoval **Vince** Korsak , „ když se
vždycky polda . " A **Vince** Korsak býval opravdu chytrý vyšetřovatel
Líbí se mi , když **Vince** vykládá , co zažíval při
na Korsaka . „ A **Vince** mě naučí střílet z pistole
vyštěkl . „ Tommy ! **Vince** ! Pojdme ... " Zmlkl
. Taky v něm hraje **Vince** Vaughn . Prohlédla si hercův
. Taky trochu vypadal jako **Vince** Vaughn . Pečlivě si namalovala
jednoduché . Hlasatelé Cubs , **Vince** Lloyd a Lou Boudreau ,
povede i třímetrový odpal . **Vince** Lloyd a Lou Boudreau si
v krku . V rádiu **Vince** Lloyd mluvil o nejdramatičtějších okamžiku
příbuzných shromáždily , aby poslouchaly **Vince** Lloyd a Lou Boudreaua ,
velmi smutné . " " **Vince** Lloyd a Lou Boudreau jako
? Byl jste tam . **Vince** Lloyd a Lou Boudreau opakovaně
Stráže , Harvey Markem a **Vince** Petrelli , se podívali na
v kondici jako jindy . **Vince** byl zrovna tak velký jako
Lullaby of the Leaves " **Vince** Guaraldiho . V duchu ji
nedávna taky myslel , ale **Vince** a Todd mě přesvědčili ,
akci všechno . " „ **Vince** a Todd jsou věčně vystrašení
" „ Ne , ale **Vince** a Todd ... " „
zpochybnili dva kalifornští genetici , **Vince** Sarich a Alan Wilson ,
Brubecka , Cala Tjadera a **Vince** Guaraldiho a několika živých alb

Podgenerování dotazu

1 / 4 ▶ (celkem: 195 položek)

Sdílet tabulku: 

| | Filtr | <u>lemma_lc [lowercase lemma]</u> | Freq ▼ | i.p.m. |
|----|-------|-----------------------------------|--------|--------|
| 1 | p / n | zástupce | 12 323 | 102,06 |
| 2 | p / n | výrobce | 12 125 | 100,42 |
| 3 | p / n | zájemce | 5 528 | 45,78 |
| 4 | p / n | soudce | 5 353 | 44,33 |
| 5 | p / n | vůdce | 4 017 | 33,27 |
| 6 | p / n | tvůrce | 3 881 | 32,14 |
| 7 | p / n | správce | 3 628 | 30,05 |
| 8 | p / n | obránce | 3 316 | 27,46 |
| 9 | p / n | poradce | 3 150 | 26,09 |
| 10 | p / n | prodejce | 2 603 | 21,56 |
| 11 | p / n | předchůdce | 2 445 | 20,25 |
| 12 | p / n | strážce | 2 116 | 17,52 |
| 13 | p / n | průvodce | 2 005 | 16,61 |
| 14 | p / n | důchodce | 1 931 | 15,99 |
| 15 | p / n | nástupce | 1 864 | 15,44 |
| 16 | p / n | vládce | 1 749 | 14,49 |

Pozorování

Dotaz na deverbativa na *ce* (N.M.*, .*ce) dává mnohem více výsledků.

Při pohledu na ně lze vyvodit, proč je výsledek získaný aplikací *Morfio* podgenerovaný.

V české tradici se rozlišují jména činitelská a konatelská. Liší se pouze slovnědruhovou příslušností základového slova. U tzv. root-derivací (do derivace nevstupuje slovesný kořen+kmenotvorná přípona, ale pouhý kořen) je někdy nesnadné odlišit, kdy je základovým slovem sloveso a kdy deverbativum vzniklé konverzí.

Pozorujeme

zástupce ← zastoupit/zastupovat x zástup

výrobce ← vyrobit/vyrábět x výroba

zájemce ← zajmout/zajímat x zájem

vůdce ← vodit/vést x *vod

tvůrce ← tvořit/tvář[ie]t x tvor

správce ← spravit/spravovat x správa

obránce ← obránit x obrana

prodejce ← prodát/prodávat x prodej

předchůdce ← předejít/*předchodit x *předchod

průvodce ← provádět/provést x průvod

důchodce ← docházet x důchod

nástupce ← nastoupit/nastupovat x nástup

https://linguisticapragensia.ff.cuni.cz/wp-content/uploads/sites/12/2022/11/Magda_Sevcikova_173-197.pdf

Vyzkoušejte opakování dotazu s alternacemi

<http://morfio.korpus.cz/KKs0LTk1>

<http://morfio.korpus.cz/7fgquwNR>

| pattern 1 ▲▼ (fq ▲▼) | pattern 2 ▲▼ (fq ▲▼) |
|-------------------------------------------|----------------------|
| 1 zastoupit (1723) | zástupce (12323) |
| 2 vyrábět vyrobit (13981) | výrobce (12125) |
| 3 zajímat (16137) | zájemce (5528) |
| 4 soudit soužit (4710) | soudce (5353) |
| 5 vodit vozit (2920) | vůdce (4017) |
| 6 tvořit (21495) | tvůrce (3881) |
| 7 spravit (856) | správce (3628) |
| 8 obránit (2) | obránce (3316) |
| 9 poradit pořádat pražit ... (17543) | poradce (3150) |
| 10 strážit stražit střádat ... (535) | strážce (2116) |
| 11 provdat provádět (10421) | průvodce (2005) |
| 12 dochodit (11) | důchodce (1931) |
| 13 nastoupat nastoupit nastoupnout (8126) | nástupce (1864) |
| 14 vložit vládat vlašit ... (11096) | vládce (1749) |
| 15 odpárat (45) | odpůrce (1628) |
| 16 přejímat přijímat (5745) | příjemce (1587) |
| 17 oblékat (1014) | oblékač (1480) |

Alternace

Přegenerování
Podgenerování

Zobrazení substantiv bez odpovídajících sloves

| abc/cba ▲▼ | fq ▲▼ |
|------------|-------|
| zástupce | 12323 |
| výrobce | 12125 |
| zájemce | 5528 |
| soudce | 5353 |
| vůdce | 4017 |
| tvůrce | 3881 |
| správce | 3628 |
| obránce | 3316 |
| poradce | 3150 |
| prodejce | 2603 |
| předchůdce | 2445 |
| strážce | 2116 |
| průvodce | 2005 |
| důchodce | 1931 |
| nástupce | 1864 |
| vládce | 1749 |
| odpůrce | 1628 |
| příjemce | 1587 |
| obhájce | 1489 |
| zastánce | 1483 |
| dopravce | 1457 |
| žalobce | 1452 |

Která slova fundují substantiva na *ce* ?

Slovesa

Konvertovaná jména

? *zastánce*

Jak na substantiva tvořená suxem -ec

<http://morfio.korpus.cz/2CXjd9wa>

| | |
|-------------|-------|
| otec | 38344 |
| zaměstnanec | 13937 |
| chlapec | 13439 |
| herec | 9740 |
| jedinec | 8105 |
| vědec | 8064 |
| němec | 7866 |
| umělec | 6789 |
| poslanec | 6037 |
| cizinec | 5049 |
| sportovec | 3530 |
| milenec | 3024 |
| jednotlivec | 3018 |
| střelec | 2968 |
| příznivec | 2848 |
| jezdec | 2390 |
| samec | 2362 |
| stařec | 2333 |
| lovec | 2297 |

Úkol na příště

- Pomocí aplikace *MORFIO* vyhledejte v korpusu SYN2020 kandidáty na deriváty **substantiv tvořených ze sloves** sufixem **–č**, přičemž chceme pouze názvy **živých bytostí (hrát/hráč)**, takže např. dvojice jako *vařit-vaříč* nás zajímá pouze v případě, že jde u vzoru dvě o označení člověka vyrábějícího pokoutně drogy a nikoli o neškodnou část běžného kuchyňského vybavení.
- **Popište problémy**, na které jste při práci narazili a připravte si **dotazy k technickým problémům**.