

# 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

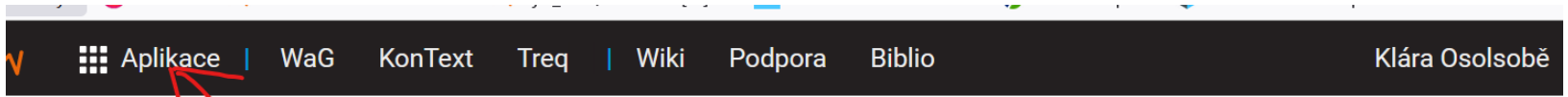
# Co je to *MORFIO* ?

- 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:morfologicka_analyza)
- <http://wiki.korpus.cz/doku.php/pojmy:tag>

# Jak vypadá přístup



## Morfio

The Morfio search interface is shown with several red arrows pointing to specific elements:

- Jazyk:** čeština (dropdown)
- společný** (dropdown) and **odlišný** (dropdown) (both with red 'X' icons)
- Morf. specifikace:** (dropdown) with a red arrow pointing to it
- vzor 1:** .+ (dropdown)
- vzor 2:** .+ (dropdown)
- Přidat vzor** (text)
- Korpus:** SYN2015 (dropdown) with a red arrow pointing to it
- Frekvence vyšší než:** 10 (input field) with a red arrow pointing to it
- Hledat:** lemmata (dropdown) with a red arrow pointing to it
- Vyhodnotit:** lemmata (dropdown) with a red arrow pointing to it
- A = a** (checkbox)
- Alternace** (button)
- Hledat** (button)
- Nové zadání** (button)
- Nápověda** (button)

# 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>

## vod

### Přehled dostupných korpusů

Korpusy psaného jazyka (synchronní)					
korpus	velikost (počet slov)	lemmatizace	morfologické značky	rok zveřejnění <sup>1)</sup>	charakteristika korpusu
<b>Obecné korpusy</b>					
<a href="#">SYN (verze 8)</a>	4,5 mld.	✓	✓	2010-2019	<a href="#">verzovaný</a> korpus, spojující synchronní psané korpusy řady SYN a další, dosud nezveřejněné texty
<a href="#">SYN2020</a>	100 mil.	✓	✓	2020	<a href="#">referenční</a> reprezentativní korpus, převažují texty z let 2015–2019
<a href="#">SYN2015</a>	100 mil.	✓	✓	2015	<a href="#">referenční</a> reprezentativní korpus, převažují texty z let 2010–2014, s novou <a href="#">klasifikací textů</a>
<a href="#">SYN2013PUB</a>	935 mil.	✓	✓	2013	<a href="#">referenční</a> korpus publicistických textů z let 2005–2009
<a href="#">SYN2010</a>	100 mil.	✓	✓	2010	<a href="#">referenční</a> reprezentativní korpus, převažují texty z let 2005–2009
<a href="#">SYN2009PUB</a>	700 mil.	✓	✓	2010	<a href="#">referenční</a> korpus publicistických textů z let 1995–2007
<a href="#">SYN2006PUB</a>	300 mil.	✓	✓	2006	<a href="#">referenční</a> korpus publicistických textů z let 1989–2004
<a href="#">SYN2005</a>	100 mil.	✓	✓	2005	<a href="#">referenční</a> reprezentativní korpus, převažují texty z let 2000–2004
<a href="#">SYN2000</a>	100 mil.	✓	✓	2000	<a href="#">referenční</a> 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: .+ ▾      tel ▾      vlastní tag > ▾      NNM.\*

vzor 2: .+ ▾      telka ▾      vlastní tag > ▾      NNF.\*

Přidat vzor

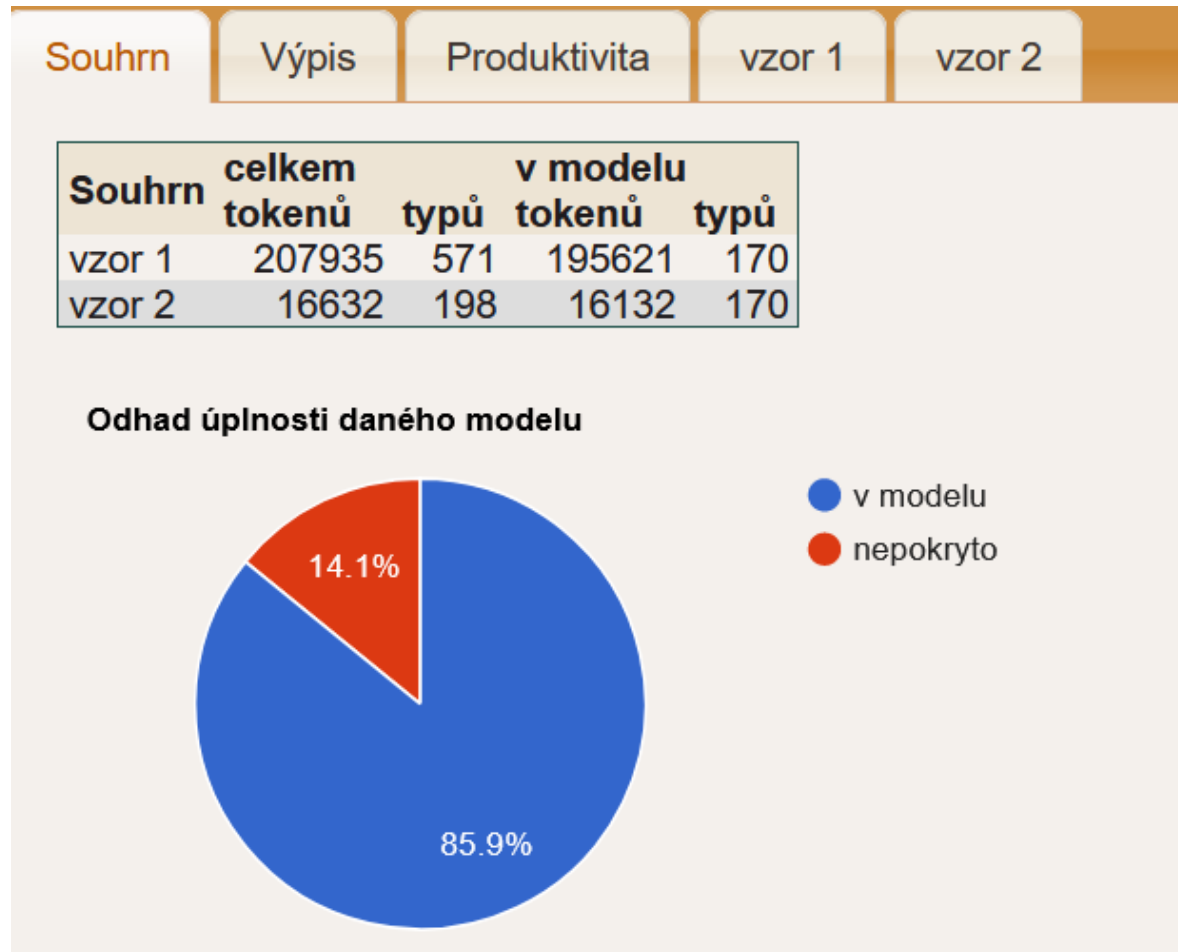
Korpus: SYN2015 ▾      Frekvence vyšší než: 0      Hledat: lemmata ▾      Vyhodnotit: lemmata ▾

A = a      ▶ Alternace

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

## 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		vz	
vzor 1 ▲▼ (fq ▲▼)		vzor 2 ▲▼ (fq ▲▼)							
1	<b>badatel</b> (1662)	<b>badatelka</b> (68)							
2	<b>bořitel</b> (18)	<b>bořitelka</b> (1)							
3	<b>buditel</b> (69)	<b>buditelka</b> (1)							
4	<b>budovatel</b> (112)	<b>budovatelka</b> (4)							
5	<b>cestovatel</b> (1255)	<b>cestovatelka</b> (97)							
6	<b>ctitel</b> (379)	<b>ctitelka</b> (49)							
7	<b>cvičitel</b> (210)	<b>cvičitelka</b> (96)							
8	<b>čekatel</b> (144)	<b>čekatelka</b> (12)							
9	<b>činitel</b> (1569)	<b>činitelka</b> (6)							
10	<b>dobyvatel</b> (356)	<b>dobyvatelka</b> (7)							
11	<b>dodavatel</b> (5024)	<b>dodavatelka</b> (10)							
12	<b>dohlížitel</b> (92)	<b>dohlížitelka</b> (2)							
13	<b>dopisovatel</b> (158)	<b>dopisovatelka</b> (23)							
14	<b>doručitel</b> (51)	<b>doručitelka</b> (1)							
15	<b>doručovatel</b> (114)	<b>doručovatelka</b> (39)							
16	<b>dovolatel</b> (18)	<b>dovolatelka</b> (8)							
17	<b>držitel</b> (1446)	<b>držitelka</b> (215)							
18	<b>exředitel</b> (38)	<b>exředitelka</b> (4)							
19	<b>hlasatel</b> (526)	<b>hlasatelka</b> (119)							
20	<b>hodnotitel</b> (221)	<b>hodnotitelka</b> (7)							
21	<b>hostitel</b> (1315)	<b>hostitelka</b> (354)							
22	<b>hubitel</b> (11)	<b>hubitelka</b> (2)							
23	<b>hybatel</b> (178)	<b>hybatelka</b> (6)							
24	<b>chovatel</b> (1790)	<b>chovatelka</b> (107)							
25	<b>jednatel</b> (1067)	<b>jednatelka</b> (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	<b>hodnotitelkou</b>	Scuk . cz Müller Thurgau 2011 kabinet VÍNO / PIVO
<input type="checkbox"/>		vymyslel i další příšernosti , nad kterými by musela centralizovaná	<b>hodnotitelka</b>	mé práce ržát blahem . A i kdyby neržála a
<input type="checkbox"/>		na mase . To platilo i tehdy , když sama	<b>hodnotitelka</b>	vegetariánka nebyla . Podle Havlíčka páchnou muži , kteří jedí
<input type="checkbox"/>		opatření a aktivit na podporu rodin v obcích . „	<b>Hodnotitelkám</b>	se velmi líbilo kunovické Komunitní centrum , “ poodhaluje podrobnosti
<input type="checkbox"/>		anketám . Svoje místo na slunci už získaly i ženské	<b>hodnotitelky</b>	, které vyhláší Světové auto roku podle žen . A
<input type="checkbox"/>		. A myslím , že v dobrém slova smyslu jsme	<b>hodnotitelky</b>	„ dorazili “ prezentací folklorních souborů , sportovních klubů a
<input type="checkbox"/>		je – podle vlastního názoru i dle hodnocení několika nezávislých	<b>hodnotitelek</b>	– 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 stejnouází, 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=~rGmeAsycq00S&cutoff=](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=~rGmeAsycq00S&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=~o8qiMAAwIKwu&cutoff=0](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=~o8qiMAAwIKwu&cutoff=0) a další

# telka (vzor2)

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

skrýt formy s podlimitní frekvencí

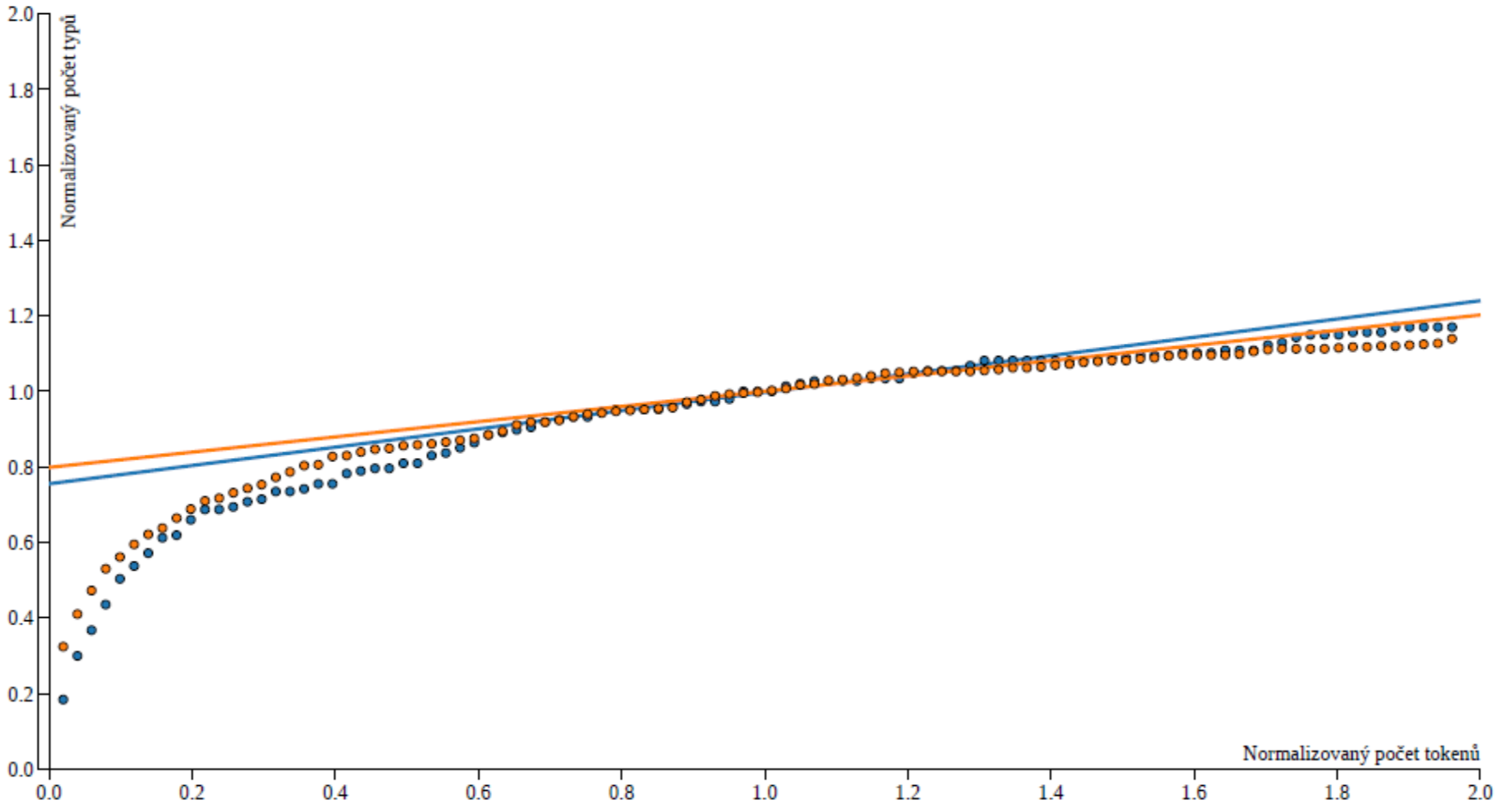
skupiny +/-

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 lc&attrs=tag&attr v mode=visible-kwic&base viewattr=word&refs=%3Ddoc.title&q=~0GSWOcQsYOwA&cutoff=0](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=lemma%20lc&attrs=tag&attr%20mode=visible-kwic&base%20viewattr=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

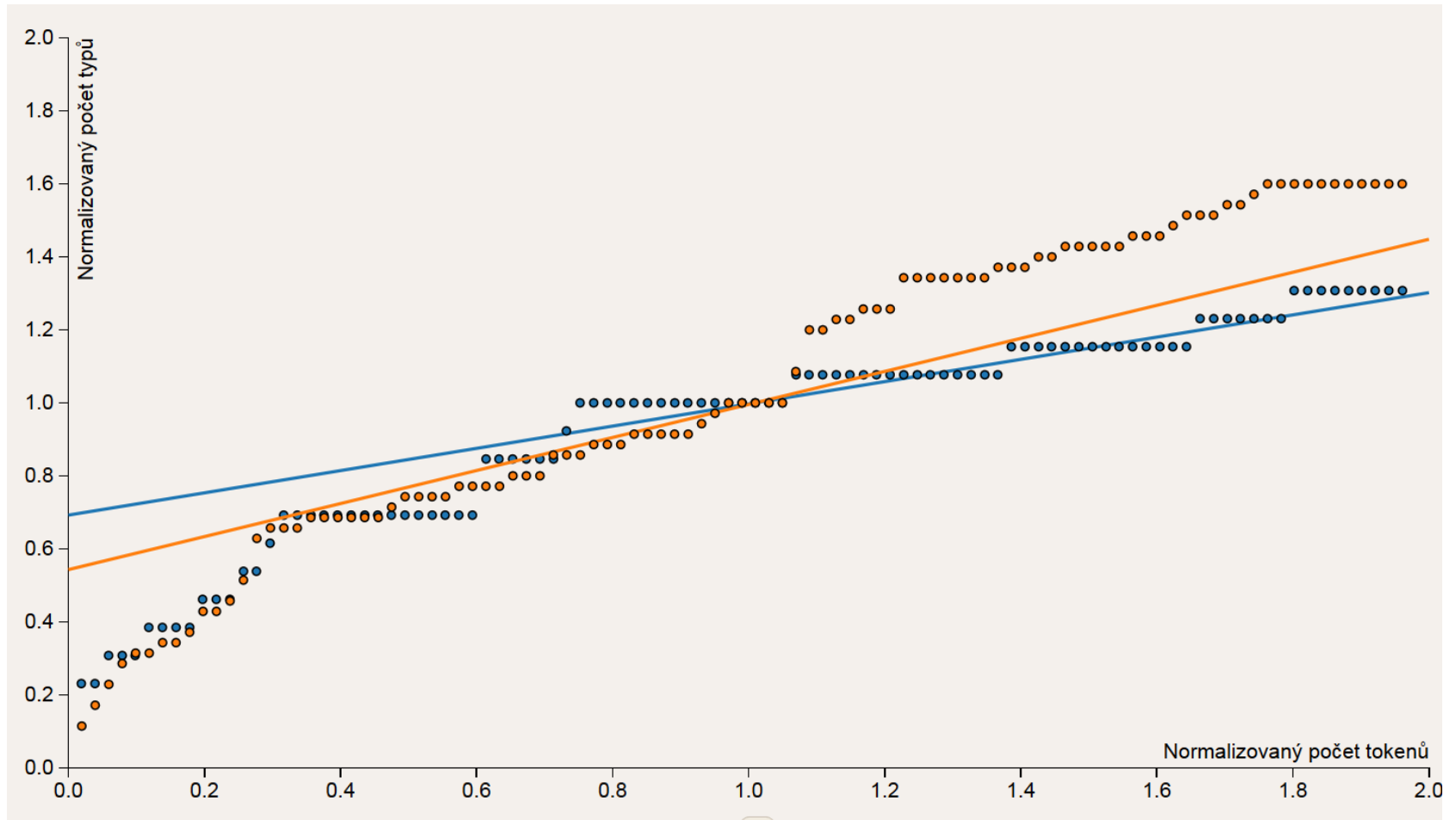
Odhad úplnosti daného modelu



- 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:

vzor 2:

Přidat vzor

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

A = a

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

# 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 <del>1</del> :	.+	[ýí]	přídavná jména	A.*
vzor <del>2</del> :	.+	ost	podstatná jména	N.*
vzor <del>3</del> :	.+	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://morfio.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](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)

# Úkol na příště

- Pomocí aplikace *MORFIO* vyhledejte v korpusu SYN2010 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**.