

Základy využití korpusu v praxi **cjbb75_LS_2022**

Kontext – **příkazový řádek** – vyhledávání podle atributů (**word, lemma, tag, lc**), **cql, regulární výrazy, KWIC, konkordanční seznam a frekvenční seznam** jako zdroj poznání o jazykových vlastnostech vyhledaných jednotek (způsob psaní/pravopis: přejatá slova /*mailovat/mejlovat/majlovat*/, aplikace pravidel – varianty /ismus/izmus/, varianty pravopisu zdobnělin vlastních jmen – Terinka/Terynka).

https://wiki.korpus.cz/doku.php/manualy:kontext:novy_dotaz

Typ dotazu

Typy dotazů a jejich využití

Typ dotazu	Nač je vhodný	Jak funguje	Co umí	Příklady
Základní dotaz	pro první seznámení s korpusem	Vyhledá vložený výraz jako slovní tvar bez ohledu na velikost písmen: jde-li zároveň o základní slovníkový tvar (lemma), vyhledají se také všechny jeho tvary.	bez regulárních výrazů (RE), case-insensitive (nerozlišuje malá a velká písmena)	černý kočka > černá kočka, černou kočku, černých koček... černá kočka > černá kočka
Lemma	pro analýzu celého paradigmatu/lexému	Vyhledá všechny tvary přiřazené k danému lemmatu .	RE (možnost použít regulární výrazy), case-sensitive (rozdílí malá a velká písmena), možnost upřesnit slovní druh	černý > černý, černému, černá, černé, černými... kočka > kočka, kočku, koček, kočkám...
Fráze	pro vícelslovnou kombinaci v daném tvaru	Vyhledá přesně zadanou frázi.	RE, case-sensitive	černý pes > černý pes český pes > český pes černého psa > černého psa
Slovní tvar	pro analýzu jednoho konkrétního tvaru	Vyhledá přesně zadaný tvar.	RE, case-in/sensitive (možnost volby Shoda velikosti)	jakkoli > jakkoli jakkoliv > jakkoliv jakkoli.* > jakkoli, jakkoliv, Jakkoli, JAKKOLIV...
Podřetězec	pro vyhledání řetězce znaků kdekoliv ve slově	Vyhledá po sobě následující znaky v rámci jednoho slova.	RE, case-sensitive	pra > praděda, praxe, doprava, lepra... křá > pookřát, křáp, Jiskřákovi...
CQL	pro vyhledání všeho, co lze pomocí korpusového manažeru vyhledat	CQL je Corpus Query Language , korpusový dotazovací jazyk (do něhož si rozhraní KonText samo interně převádí všechny předchozí typy dotazů).	RE, case-sensitive , CQL syntax	[lemma="kočka"] > kočka, kočku, koček, kočkám... [word="černá"] > černá [lemma="číst"] [tag="N.*"] > číst levity, četli článek, nečtete noviny...

Cvičení

Umíme česky?


Sloveso péci je odvozovacím základem řady příbuzných slov. Při tvoření slov dochází v češtině k hláskovým alternacím. Lexikální morf mívá řadu alomorfů. Doplňte následující tabulku o všechny možné doklady příbuzných slov. Pokud si nebudete moci vzpomenout, zda je některá kombinace přípustná, vezměte si na pomoc korpus.

	P-K	P-C	P-Č
é	pék	péc-t	pěč
í	pík	píc	píč
e	pek	pec	peč

Jak česky píšeme?

mail

? mailovat, meilovat, majlovat, mejlovat ?

Korpus: 

Typ dotazu:

Lemma:

Celkem: 69 (2 str.)

	Filtr	lemma	Frekvence
1.	p/ n	mail	18 839
2.	p/ n	mailový	3 773
3.	p/ n	mailovat	1 357
4.	p/ n	mailbox	138
5.	p/ n	mailing	129
6.	p/ n	mailování	117
7.	p/ n	mailer	102
8.	p/ n	mailto	41
9.	p/ n	mailingový	18
10.	p/ n	maillot	13
11.	p/ n	mailserveru	11
12.	p/ n	mailerový	11
13.	p/ n	mail-artu	9
14.	p/ n	mailovací	8
15.	p/ n	mailově	6

Složitější formulace dotazu

[lemma="m(ai|ei|aj|ej)l.*"]

https://wiki.korpus.cz/doku.php/manualy:kontext:frekvencni_distribuce

Menu: Frekvence

Pod označením *Frekvence* se v menu rozhraní KonText nachází funkce pro vytváření **frekvenčních distribucí**. Pomocí ní je možné získat přehled **typů** (např. různých slov) ve výsledku hledání spolu s jejich frekvencí. Vyhledáme-li např. všechna substantiva v gen. pl., můžeme pomocí této funkce zjistit, jaké **tvary** se v tomto pádě a čísle vyskytují a jak často. Stejně tak je ale možné frekvenční distribuci použít na zjišťování frekvencí jednotek předcházejících či následujících, počítání **lemmat** v **konkordanci**, nebo pro zjištění distribuce hledaného jevu napříč různými texty či jejich skupinami (podle **genre**, **txttype** apod.).

Frekvenční distribuce umožňují jednak vlastní (obecné) nastavení a jednak **rychlé volby** (ty jsou přístupné v druhé úrovni menu):

1. **Lemmata** - vyhodnotí dotaz (KWIC) a vypíše všechna různá lemmata (atribut **lemma**) spolu s jejich frekvencí¹⁾
2. **Slovní tvary** - vyhodnotí dotaz (KWIC) a vypíše všechny různé tvary (atribut **word**) spolu s jejich frekvencí
3. **Dokumenty** - vyhodnotí celou **konkordanci** a vypíše názvy textů (**strukturní atribut** **nazev**), v nichž se hledaný jev vyskytuje, spolu s frekvencí tohoto jevu v jednotlivých textech
4. **Typy textů** - vyhodnotí celou **konkordanci** a vypíše přehled strukturních atributů²⁾, které se vztahují k typu textu (**strukturní atributy** **txttype_group**, **txttype**, **med**, **srclang**), spolu s jejich frekvencí (význam jednotlivých zkratk je k dispozici v **seznamu zkratk** a kódů)

Podobnou funkcionalitu, která se ovšem většinou vztahuje na celý korpus (nikoli pouze na jednu konkrétní konkordanci), poskytují také funkce **Nový dotaz** → **Seznam slov**.

Obsah
- Menu: Frekvence
- Vlastní nastavení frekvenční distribuce
- Frekvenční distribuce podle pozičních atributů
- Frekvenční distribuce podle metainformací
- Frekvenční seznam (výpis)

Celkem: 154 (4 str.)

	Filtr	lemma	Frekvence
1.	p/n	mail	18 839
2.	p/n	mailový	3 773
3.	p/n	mailovat	1 357
4.	p/n	majlant	785
5.	p/n	mej1	249
6.	p/n	mejlit	161
7.	p/n	mailbox	138
8.	p/n	mailing	129
9.	p/n	mailování	117
10.	p/n	mailer	102
11.	p/n	mejlka	59
12.	p/n	majland	58
13.	p/n	mailto	41
14.	p/n	mejlovat	22

Můžeme sledovat frekvence pravopisných variant. Pracujeme s korpusem psaných textů, které navíc prošly jazykovou kontrolou. Úzus, který není jednotný, se může stát zdrojem pro úvahy o kodifikaci, která má odrážet úzus.

Úkol:

Jak se v jazykových korpusech dodržuje norma?

Domácké podoby rodných jmen (*Klárinka, Lukí/Luky*)

Užíváním domáckých podob vlastních jmen vyjadřujeme určitý citový vztah k nositeli jména, proto se při jejich tvoření uplatňují hlavně různé citově zabarvené jazykové prostředky, např. zdobnělny. V běžném hovoru také často slycháme oslovení jako *Šáři, Štěpi, Tomi, Viki* apod. Jsou to tvary domáckých jmen vzniklé původně jako zkrácená podoba zdobnělnin utvořených příponou *-inka, -ička* nebo *-ik, -iček*: *Šárinka – Šáři, Stázička/Stázin(K)a – Stázi, Tomík/Tomíček – Tomi*. Potřeba tvořit citově zabarvené podoby jmen je však natolik silná, že tvary na *-i* pronikly a nadále pronikají také ke jménům, která zdobnělny s *-i/-* v příponě obvykle netvoří, např. *Věra – (Věruška, Věrunka) – Věři, Lukáš – (Lukášek) – Lukí*.

Domácké podoby jmen se uplatňují téměř výhradně v soukromých, důvěrných projevech, a to především v mluvené podobě. V písmu se objevují jen v omezené míře – v soukromé komunikaci, v neformálních textech či v beletrii, kde signalizují vztahy mezi jednotlivými postavami. Při řešení jejich pravopisných podob lze vycházet ze způsobu tvoření nebo z hláskové stavby tvaru jména a dojit tak k různým podobám.

Vlivem analogie s typy, které mají *-i* ve zdobňující příponě (*Stázička/Stázinka – Stázi, Tomík/Tomíček – Tomi*), obvykle píšeme měkké *i* u všech domáckých podob zakončených na *-i*: *Lukí, Viki, Šáři, Věři* apod. Právě živě a frekventovaně zdobnělny s *-i/-* v příponě jsou pro psaní *Lukí, Šáři* atd. silnou oporou, stejně jako přípony zvětčující až hanlivé: *Stáziina, Šářiina*. Psaní s *-i* tedy považujeme za základní, a to bez ohledu na způsob tvoření a v některých případech i navzdory tomu, že koncovému *-i* předchází pravopisně tvrdá souhláska.

Předchází-li koncovému *-i* tvrdá souhláska, nemůžeme označit za chybné ani podoby s *-y*: *Věři/Věry, Lukí/Luky* atd. Předchází-li koncovému *-i* souhláska *d, t, n*, dává se přednost psaní *-i*, nebo *-y* podle výslovnosti předchozí souhlásky: *Hanička – Hani [-ň-], Vlasta – Vlasti [-t-], ale Rudolf – Rudy, Antonín – Tony*. Méně obvyklé (ale přípustné) je psát též *Rudí [rudy]*.

Jak shromáždíme relevantní data pro corpusově založený výzkum?

1. Spolehneme se na dodržení normy a vyhledáme lemmata, která splňují dvě podmínky, a sice začínají velkým písmenem a končí na řetězec inka před nímž předchází souhláska, po nichž se v češtině obvykle nepíše *i*. `[lemma="[:upper:]].*[rhk]inka"`

- Zkontrolujeme, zda se norma vždy dodržuje. Vyhledáme lemmata, která splňují dvě podmínky, a sice začínají velkým písmenem a končí na řetězec ynka před nímž předchází souhláska, po nichž se v češtině obvykle nepíše *i*. [lemma="[:upper:].*[rhk]ynka"]
- Pozorujeme odchylky od normy. Vyhledáme lemmata, která splňují dvě podmínky, a sice začínají velkým písmenem a končí na řetězec inka nebo ynka před nímž předchází souhláska, po nichž se v češtině obvykle nepíše *i*. [lemma="[:upper:].*[rhk][iy]nka"]

Obrázek k dotazu 1

Korpus: syn v9 | Dotaz: [[:upper:].*[rhk]ynka (17 062 výskytů) ▶ Promíchat: ✓

Frekvenční seznam

1 / 5 ▶

Minimální frekvence: 1

Celkem: 239 položek (5 stránek)

	Filtr	lemma	Freq	
1	p / n	Barinka	5157	
2	p / n	Wawrinka	3637	
3	p / n	Márinka	1602	
4	p / n	Darinka	933	
5	p / n	Sárinka	797	
6	p / n	Horinka	728	
7	p / n	Marinka	497	
8	p / n	Laurinka	480	
9	p / n	Terinka	320	
10	p / n	Javorinka	252	










Obrázek k dotazu 2

Frekvenční seznam





1 / 2 ▶

Minimální frekvence:

Celkem: 67 položek (2 stránky)

	Filtr	lemma	Freq	
1	p / n	Kuchynka	2835	
2	p / n	Blahynka	561	
3	p / n	Machynka	185	
4	p / n	Suryнка	106	
5	p / n	Šáryнка	46	
6	p / n	Máryнка	22	
7	p / n	Teryнка	20	
8	p / n	Maryнка	15	
9	p / n	Meryнка	10	
10	p / n	Horyнка	9	

Obrázek k dotazu 3

26	p / n	Kikinka	59	
27	p / n	Šárinka	56	
28	p / n	Zrinka	49	
29	p / n	Borinka	48	
30	p / n	Šáryнка	46	
31	p / n	Irinka	41	
32	p / n	Vikinka	40	

Podívejme se na případ, kdy frekvence variant je velmi blízká a není zanedbatelně nízká, abychom ji mohli pokládat za chybu.

Šár[iy]nka

Korpus: **syn v9** | Dotaz: **Šár[iy]nka** (102 výskytů) ▶ Promíchat: ✓

Frekvenční seznam

1 / 1

Minimální frekvence: 1

Celkem: 2 položky (1 stránka)

	Filtr	lemma	Freq	
1	p / n	Šárinka	56	<div style="width: 56px; height: 10px; background-color: #007bff;"></div>
2	p / n	Šárynka	46	<div style="width: 46px; height: 10px; background-color: #007bff;"></div>







Vyskytů: 102 | l.p.m.: 0,02 (vztaženo k celému korpusu) | ARF: 39,11 | Výsledek je promíchaný 1 / 3 ▶▶▶

Vyber řádků: základní

<input type="checkbox"/>	Deníky Bohemia	, která po příchodu na svět vážila 2,84 kg .	Šárinka/šárinka/X@.....	má domov v Čisté u Rakovníka a bude vyrůstat s
<input type="checkbox"/>	Mladá fronta DNES	miní . Tento záměr představila ředitelka pražské Lesní mateřské školky	Šárinka/šárinka/X@.....	Linda Kubale . , Přijde mi jako vhodný nápad použít
<input type="checkbox"/>	Deníky Bohemia	, čtyřletá sestřička Romanka a teta Martina . Domov má	Šárinka/šárinka/X@.....	ve Zdicích . Jméno Daniel vybrali manželé Michaela a Jaroslav
<input type="checkbox"/>	Víkend HN	na 20 stupňů , která slouží jako zázemí lesní školky	Šárynka/šárynka/X@.....	. Děti v otepovačkách měly navrhout svoji vlastní rukavici .
<input type="checkbox"/>	Týdeník školství	jako je například vozík pro Kristinu , nevidomá Zuzanka ,	Šárinka/šárinka/X@.....	z azylového domu či studium pro Graciána Švačinu . Přehled
<input type="checkbox"/>	Strategie	, například v zahradě Domova svaté Rodiny nebo v Ekodomově	Šárynka/šárynka/X@.....	. ŽENY pomáhají podnikat ženám Karel Havlíček , předseda představenstva
<input type="checkbox"/>	Nevědčiče děti	dopis , usměje se) : Píše Mini , že	Šárinka/šárinka/X@.....	už chodí . Nikdy nepochopím , proč jí dala jméno
<input type="checkbox"/>	Víkend HN	školka Jako vzor právě vznikajícím lesním školkám by mohla posloužit	Šárynka/šárynka/X@.....	, která od roku 2009 funguje v pražském dolnošáreckém údolí
<input type="checkbox"/>	Děti a MY	žádném permanentním výletě . Pracuje jako učitelka lesní mateřské školy	Šárynka/šárynka/X@.....	. Po vzoru zejména skandinávských lesních školek se i u
<input type="checkbox"/>	Deníky Bohemia	dal k ústům trubku a spustil Vzpomínku na Zbiroh .	Šárinka/šárinka/X@.....	a Monička , dvě rozřeštěné , splašené kobylinky . Zastavím
<input type="checkbox"/>	Tina	umělé plicní ventilace , mimotělní oběhu a infuzní pumpy .	Šárinka/šárinka/X@.....	nakonec přežila . Zachránila ji skvělá péče lékařů , a
<input type="checkbox"/>	Deníky Bohemia	Kateřiny nebyly nedostatky v legislativě překážkou pro založení Dětského klubu	Šárynka/šárynka/X@.....	na rodinném pozemku v Šareckém údolí . Zázemí vytvořilo občanské
<input type="checkbox"/>	Děti a MY	školka je divoká a přátelská Říká zakladatelka Dopoledního dětského klubu	Šárynka/šárynka/X@.....	, Linda Kubale : Vzdělávací program Šárynky vychází z Rámcového
<input type="checkbox"/>	Reflex	ve mlýně zhruba rok a se Šárinkou byly kamarádky .	Šárinka/šárinka/X@.....	nic nekomentovala , jen dál mlčky malovala . SOULOŽIT A
<input type="checkbox"/>	Rodinný dům	pečuje o zdraví osázeného prostoru . V pražském ekologickém středisku	Šárynka/šárynka/X@.....	radí , Nezabýváte se cenného organického odpadu a kompostujete
<input type="checkbox"/>	Mladá fronta DNES	by tak mohla strávit alespoň část Vánoc doma . *	Šárinka/šárinka/X@.....	je na tom na jednotce intenzivní péče už lépe .
<input type="checkbox"/>	Euro	prostředky , při práci na jarním úklidu v zahradě Ekodomova	Šárynka/šárynka/X@.....	, jenž se zabývá podporou ekologie a udržitelnosti . Dobrovolníci

Všimněme si, že v případě porušení pravopisné normy jde o proprium užitá jako název instituce.

Celkem: 59 položek (2 stránky)

	Filtr	word	word	Freq	
1	p / n	.	Šárinka	11	
2	p / n	klub	Šárynka	8	
3	p / n	klubu	Šárynka	7	
4	p / n	,	Šárinka	4	
5	p / n	a	Šárinka	3	
6	p / n	školky	Šárynka	3	
7	p / n	Ekodomově	Šárynka	3	
8	p / n	Ekocentru	Šárynka	3	
9	p / n	sestřička	Šárinka	3	
10	p / n	.	Šárynka	3	
11	p / n	ekocentra	Šárynka	2	
12	p / n	malá	Šárinka	2	
13	p / n	"	Šárinka	2	
14	p / n	středisku	Šárynka	2	
15	p / n	pravnoučata	Šárinka	2	
16	p / n	posloužit	Šárynka	1	
17	p / n	vnučka	Šárinka	1	

Můžeme dále zkoumat, které z frekventovaných podob deminutiv nedodrží předepsanou pravopisnou normu:

M[áa]r[iy]nka

Frekvenční seznam

/ 1

Minimální frekvence:

[Použít](#)

Celkem: 5 položek (1 stránka)

	Filtr	<u>lemma</u>	<u>Freq</u>	
1	p / n	Márinka	1602	<div style="width: 100%; height: 10px; background-color: #007bff;"></div>
2	p / n	Marinka	497	<div style="width: 30%; height: 10px; background-color: #007bff;"></div>
3	p / n	Marinko	25	<div style="width: 5%; height: 10px; background-color: #007bff;"></div>
4	p / n	Márynka	22	<div style="width: 5%; height: 10px; background-color: #007bff;"></div>
5	p / n	Marynka	15	<div style="width: 5%; height: 10px; background-color: #007bff;"></div>

Klár[iy]nka

Korpus: [syn v9](#) | Dotaz: [Klár\[iy\]nka](#) (151 výskytů) ▶ Promíchat: ✓

Frekvenční seznam

/ 1

Minimální frekvence:

Celkem: 2 položky (1 stránka)

	Filtr	lemma	Freq	
1	p / n	Klárinka	149	<div style="width: 100%; height: 10px; background-color: #007bff;"></div>
2	p / n	Klárýnka	2	

Korpus: [syn v9](#) | Dotaz: [Laur\[iy\]nka](#) (485 výskytů) ▶ Promíchat: ✓

Frekvenční seznam

/ 1

Minimální frekvence:

Celkem: 3 položky (1 stránka)

	Filtr	lemma	Freq	
1	p / n	Laurinka	480	<div style="width: 100%; height: 10px; background-color: #007bff;"></div>
2	p / n	laurinka	4	
3	p / n	Laurýnka	1	

Korpus: [syn v9](#) | Dotaz: [Sár\[iy\]nka](#) (799 výskytů) ▶ Promíchat: ✓

Frekvenční seznam

1 / 1

Minimální frekvence:

Celkem: 2 položky (1 stránka)

	Filtr	lemma	Freq	
1	p / n	Sárinka	797	<div style="width: 100%; height: 10px; background-color: #007bff;"></div>
2	p / n	Sárynka	2	<div style="width: 1%; height: 10px; background-color: #007bff;"></div>

Korpus: [syn v9](#) | Dotaz: [Ter\[iy\]nka](#) (410 výskytů) ▶ Promíchat: ✓

Frekvenční seznam

1 / 1

Minimální frekvence:

Celkem: 3 položky (1 stránka)

	Filtr	lemma	Freq	
1	p / n	Terinka	320	<div style="width: 100%; height: 10px; background-color: #007bff;"></div>
2	p / n	terinka	70	<div style="width: 22%; height: 10px; background-color: #007bff;"></div>
3	p / n	Terynka	20	<div style="width: 6%; height: 10px; background-color: #007bff;"></div>

Korpus: syn v9 | Dotaz: Ter[ly]nka (410 výskytů) | Promíchat: ✓ | Pozitivní filtr: Terynka (20 výskytů)

Výskytů: 20 | i.p.m.: 0 (vztaheno k celému korpusu) | ARF: 7,48 | Výsledek je promíchán

Vyběr rádků: základní

<input type="checkbox"/>	Milostenky nemilostivé	dřevěným zábradlím , které odděluje divácká místa , sedí jeho	Terynka/teryinka/X@-----	, celá v slzách , jimiž hodlá dojmout pana předsedu
<input type="checkbox"/>	Milostenky nemilostivé	frajer , vime ? "- Manželka Antonie čili	Terynka/teryinka/X@-----	jen poříkává . Celá tahle patálie , pro kterou pana
<input type="checkbox"/>	Deniky Bohemia	bude muset zkusit najít náhradu za elitního střelce soutěže Aleše	Terynka/teryinka/X@-----	, o kterého má zájem rezerva FK SIAD Most .
<input type="checkbox"/>	Deniky Bohemia	míči se tak dostal Wagner , jenž pro svého kolegu	Terynka/teryinka/X@-----	vymyslel přihrávku na míru . Terynek se tak jako první
<input type="checkbox"/>	Mladí fronta DNES	Záleží ale také na tom , kdy půjdou Eliška ,	Terynka/teryinka/X@-----	a Jindřich do světa , " dodal Ambrož . Překonal
<input type="checkbox"/>	Deniky Bohemia	, Ferda Mravenec , Menší sen , Sněhulák , Velryba	Terynka/teryinka/X@-----	a Motýl Emanuel – skutečnost ovšem předčila i nejbujnější představy
<input type="checkbox"/>	Deniky Bohemia	se znaky . Temperamentní a vhodný na zahradu . "	Terynka/teryinka/X@-----	3 roky , fenka " Nekaastrovaná fenka . Černá srst
<input type="checkbox"/>	Mladí fronta DNES	je šupa , že ho má zrovna Baník I (Terynka/teryinka/X@-----) - Tak skvělý fotbal jsem na Bazalech ještě neviděl
<input type="checkbox"/>	Deniky Bohemia	Souš spolehala . Ve třinácté minutě šli hosté po akci	Terynka/teryinka/X@-----	s Berešem do vedení a domácí museli více myslet na
<input type="checkbox"/>	Deniky Bohemia	, ale Bartůnek nastřílel pouze Soldána . Souš zahrozila hlavičkou	Terynka/teryinka/X@-----	deset minut před poločasem . Na druhé straně nedošlo na
<input type="checkbox"/>	Deniky Bohemia	SIAD Bilina . Hosty porazil 3 : 0 dvěma góly	Terynka/teryinka/X@-----	a brankou Schofe po poločasu 1 : 0 . 0
<input type="checkbox"/>	Deniky Bohemia	a 15 . minutě likvidoval Lilko střely útočného hrotu domácích	Terynka/teryinka/X@-----	. Poté hosté převzali iniciativu hry do svých rukou .
<input type="checkbox"/>	Deniky Bohemia	, nepřijemný byl zejména odchod našeho loňského nejlepšího střelce Aleše	Terynka/teryinka/X@-----	. Soupiska tak byla doplněna o dorostence , " dodává
<input type="checkbox"/>	Deniky Bohemia	výsledku . V 57 . minutě se dočkali a brankou	Terynka/teryinka/X@-----	snížili na rozdíl jediného gólu . Jirko ještě odpověděl a
<input type="checkbox"/>	Mladí fronta DNES	: Eliška , Terynka a Jindřich Dvorec - Eliška ,	Terynka/teryinka/X@-----	a Jindřich . Jmény postav z pohádky Princezna ze mlejna
<input type="checkbox"/>	Deniky Bohemia	první hrubka děčinské obrany a také první trest z kopačky	Terynka/teryinka/X@-----	- 1 : 0 . Vyrovnal mohl Zeňáček , jeho
<input type="checkbox"/>	Mladí fronta DNES	Jilek , starosta Jankova i Holašovic Tygřiči : Eliška ,	Terynka/teryinka/X@-----	a Jindřich Dvorec - Eliška , Terynka a Jindřich .
<input type="checkbox"/>	Milostenky nemilostivé	, Tak poď , paní ! " - Poslušná	Terynka/teryinka/X@-----	vyškočí , celá šťastná se zavěsí s loktem do loktu
<input type="checkbox"/>	Deniky Bohemia	Ferda Mravenec s Beruškou aneb zlaté české ručičky . VELRYBA	Terynka/teryinka/X@-----	(nahofe) a Pedro de la Vedro (vlevo
<input type="checkbox"/>	Sedmička	let Rasa : jezevčík Povaha : Kaštánek , Fialka a	Terynka/teryinka/X@-----	. Museli opustit rodinu , kde mělo dítě alergie .

Závěr: Je patrné, že analýza korpusových dat může přinést některé poznatky odporující běžným poučkám a pravidlům. Analýza deminutivního sufixu *-inka* užívaného pro tvoření podob vlastních jmen prokázala případy úzu, který je v rozporu s pravidly uvedenými v IJP. V případě frekventovanějších užití se ovšem ukazuje, že nejde vždy o chybně utvořené deminutivum. Je třeba brát v potaz např. příjmení (Terynka), kdy hrají roli mimojazykové faktory.

Pravopisná norma vychází z existence deminutivní přípony *-inka*, která nemá dvě pravopisné realizace. Jestliže ovšem připustíme dvě pravopisné realizace přípony *-i*, kterou se v dnešním úzu frekventovaně tvoří domácké podoby *proprii*, pak je třeba mít na zřeteli, že tlak systému/analogie bude působit i na pravopisnou normu deminutiv na *-inka*.

Korpus: ksk dopisy | Dotaz: Šár[ly] (4 výskytů) | Promíchat: ✓

Výskytů: 4 | i.p.m.: 4,24 (vztaheno k celému korpusu) | ARF: 1,65 | Výsledek je promíchán

Vyběr rádků: základní

<input type="checkbox"/>	F2B3KF2B2000	, moc krásně . A určitě se ozví Zdraví Tě Šáři i Miša Pošlu Ti nějaké citáty : <note text="cit: " Když máš ve svém srdci nebe , pak se nebe šíří i okolo tebe . " Řekni díky , človíčku , za proud lásky , který v tichosti proudí ve světě . " Kd
<input type="checkbox"/>	F2A7ZF2B2000), ale moc se mi do toho nechce . Šáry končím , jdu na tu matiku . Měj se krásně
<input type="checkbox"/>	F2A7ZF2B2000	S pozdravem Jana 21 . září 2000 v Teplicích Ahoj Šáry ! Jako Ty nikdy nevím jak mám začínat , takové
<input type="checkbox"/>	F2B6KF2B2002	brzy ! @ Zborovice 29 . 7 . 02 Ahoj Šáry jak se máš a co pořád provádíš ? Nechbí ti

Korpus: ksk dopisy | Dotaz: Ter[ly] (3 výskytů) | Promíchat: ✓

Výskytů: 3 | i.p.m.: 3,18 (vztaheno k celému korpusu) | ARF: 1,85 | Výsledek je promíchán

Vyběr rádků: základní

<input type="checkbox"/>	F2B2KF2A1992	: Ať žijí prázdniny 9 . 6 . 1992 Nazdar Tery - i ! Předem mého dopisu tě srdečně pozdravuju a
<input type="checkbox"/>	F2B9KF2B2001	podrobná prohlížet . 14 . 9 . 2001 Hostinné Ahojda Teri , omlouvám se za zpoždění a rychle napravuju . Díky
<input type="checkbox"/>	F2B7KF2B1999	čas , až budeš mít klid . Ahoj Terry Čau Teri , nezlobím se za dopis , že jsi mi ho

Naučili jsme se:

Používat některé **regulární výrazy** v jazyce *cql*. (., [], |)

Používat **statistiky** a **zobrazení**.

Používat **Frekvence**.

Ukázali jsme, jak lze používat jazykový korpus k hledání odpovědí na otázky týkající se grafické podoby jazyka a nejen jí.

Doplňující četba:

OSOLSOBĚ, Klára a Jana MACHALOVÁ. Hypokoristika z rodných jmen v Korpusu soukromé korespondence. In Hladká, Zdeňka a kol.. *Soukromá korespondence jako lingvistický pramen*. Vyd. 1. Brno: Masarykova univerzita, 2013. s. 33-59.

Dů:

7. 3. Jaká jsou lemmata/tvary adaptovaných slov a jaké je frekvenční rozložení variantních způsobů jejich grafické realizace (slovesa utvořená od anglických základů: použít(va)t *Skype*=????, použít(va)t *google* = ???, použít(va)t *twitter* = ???). V korpusu SYNv9 vyhledejte tvary sloves tvořených od substantiva *instagram* a sledujte, ke kterým slovesným třídám patří.

https://is.muni.cz/auth/el/phil/jaro2022/CJBB75/ode/ukoly_cjbb75/1_ukol/