

Jazykové korpusy (lingvistika, filologie, výuka jazyků)

1

KLÁRA OSOLSOBĚ
OSOLSOBE@PHIL.MUNI.CZ
G 32 9.10-10.50

O čem budeme mluvit

2

- Krátký historický exkurz
- Definice korpusu v moderním slova smyslu
- Dva metodologické přístupy k vytěžování korpusu
- Dva pohledy na korpus (lingvista a informatik)
- Filologie a korpusy
- Výuka jazyků a korpusy

Krátký historický exkurz

3

- Myšlenka korpusu
- Korpusová lingvistika – empirická disciplína
- Data a introspekce
- Technický pokrok

Konkordance – KWIC (Key Word In Context)

4

Výskytů: 213 | i.p.m. 1,76 (vztaženo k celému "omezeni/syn2015") | ARF 88,32 | Výsledek je promíchán

1 / 6

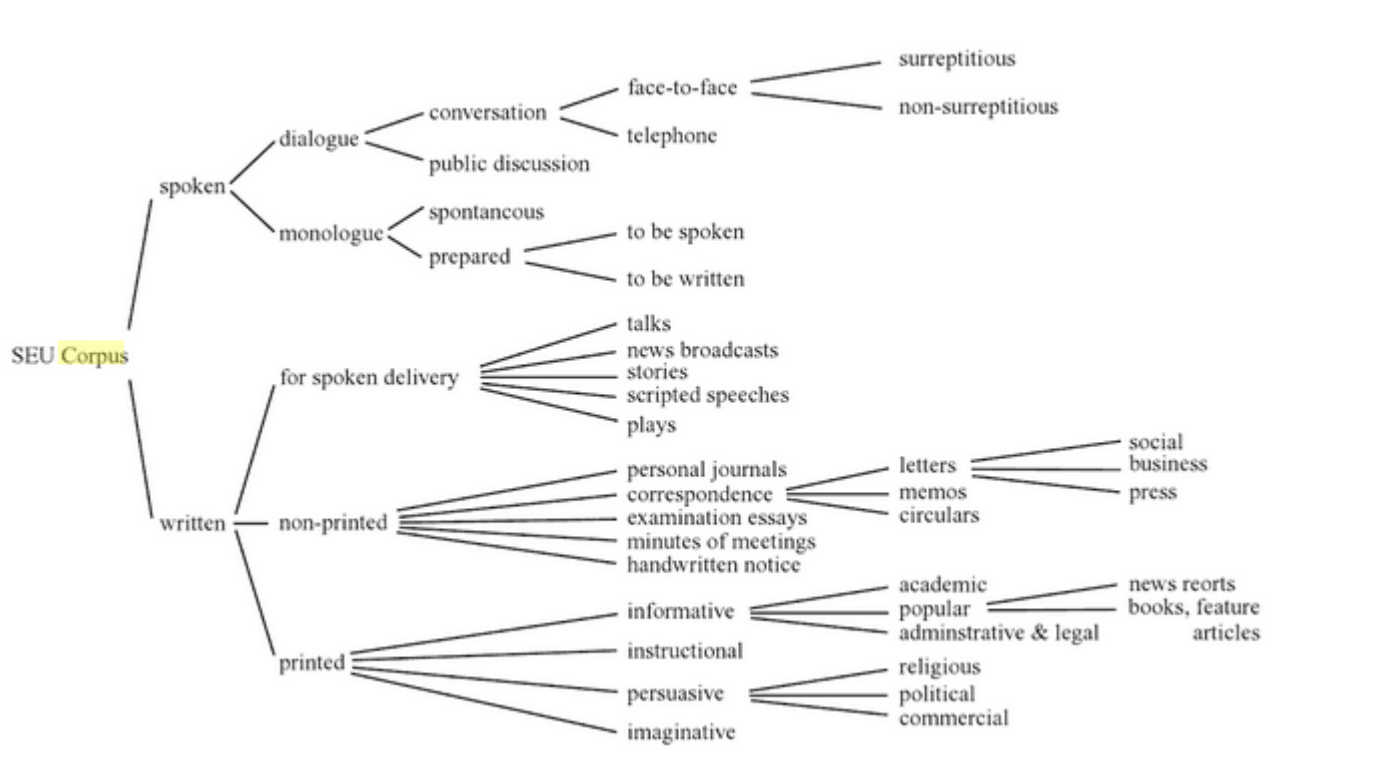
Výběr řádků: základní | Atributy:

<input type="checkbox"/>	Brno Business	se ho 4379 lidí . Vyplynulo z něj , že	Češi jsou	se svými šéfy poměrně spokojeni . Šéfů , kteří se
<input type="checkbox"/>	Reflex	jsou hlavní autority , pak je to svědectvím , že	Češi jsou	národ-nenárod , jenž se vždycky sjednotí na nějakém kýči ,
<input type="checkbox"/>	Maxim	bezpečnost . Žádný slovenský záchranář si nezapomene rýpnout , že	Češi jsou	blázní . Devět z deseti mrtvých v Tatrách jsou Češi
<input type="checkbox"/>	Týden	, který máme dobýt obchvatem shora , horskými hřebenovkami .	Češi jsou	na Tematině ostatně dávnými hosty , hrad pokořil už Přemysl
<input type="checkbox"/>	Sport	. Česká vzpoura v Letošní NHL je v polovině a	Češi jsou	vidět ! Po několika hubenějších sezonách je tu konečně vydařený
<input type="checkbox"/>	Reflex	vnímáno hlavně jako neustálý konflikt dvou zneprátelených národů , kde	Češi jsou	vždy ti malí , ale spravedliví , utlačovaní , ale
<input type="checkbox"/>	Mladá fronta Dnes	by to ryzí zázrak . Matematicky to spočítat nelze .	Češi jsou	odkázáni na pomoc shůry . Kromě toho , že sami
<input type="checkbox"/>	Lidé a země	předpisu rakouské armády z roku 1749 . " Proslulí nadšenci	Češi jsou	v Evropě obecně známi jako jedni z nejlepších „ vojáků
<input type="checkbox"/>	Reflex	, že mají zhasínat , pokud jdou z místnosti .	Češi jsou	národem s maloměstským myšlením a jsou závistiví . Pozorujete to
<input type="checkbox"/>	Reflex	koncem dubna byl projednán návrh zákona o její regulaci .	Češi jsou	totiž vůči prostituci značně liberální . A nejen vůči ní
<input type="checkbox"/>	Mladá fronta Dnes	Alsasané ve Francii skoro dodnes . Řekla bych , že	Češi jsou	dnes s vyrovnáváním se s touto částí naší minulosti dál
<input type="checkbox"/>	Týdeník Květy	v hrubém domácím produktu na hlavu je přímo propastný :	Češi jsou	35x bohatší ! Za zmínku určitě také stojí , že
<input type="checkbox"/>	Parlamentní listy	Češi leniví a s tím klesá jejich ochota recyklovat »	Češi jsou	ochotni odnášet vyřazené světelné zdroje maximálně do vzdálenosti 7
<input type="checkbox"/>	Mladá fronta Dnes	tomu není žádný důvod , nic se nemění , ale	Češi jsou	velmi finančně negramotní . Nepřehnal se to v Česku s
<input type="checkbox"/>	Reflex	naprosto zřejmé , že dává přednost startování auta klikou ?	Češi jsou	někdy trochu záhadní . Taxikář se mě znovu zeptal na
<input type="checkbox"/>	Rodinný dům	několik hollywoodských komedií o Santa Clausovi , testující , zda	Češi jsou	již převychování , takže lze očekávat sérii sebevražd několika program
<input type="checkbox"/>	Respekt	, že oslepl právě po vodce této společnosti . „	Češi jsou	nejen světovými přeborníky v pití piva , ale podle Světové
<input type="checkbox"/>	Metro	na horách , v Paříži nebo překvapivě doma . „	Češi jsou	tedy spokojení nebo bez fantazie , „ vyplývá z průzkumu
<input type="checkbox"/>	Parlamentní listy	Dánsko , Švédsko , Německo , Velká Británie) ,	Češi jsou	se 17 % polepšených kuřáků v dolní polovině tabulky .
<input type="checkbox"/>	Lidové noviny	poklesu spotřebitelské poptávky . Není důvodem i to , že	Češi jsou	příliš opatrní a skeptičtí ohledně stavu země a jejího dalšího

psjc Příruční slovník jazyka českého

vroucí adj. **vařící** **horký** **vřelý** Vzala čajovou konvičku, hodila tam bezový květ a nalila na to vroucí vody. **Vrch.** Vroucí vodu napouštěla na odvar čaje ve sklenicích. **Mach.** Vařil na lihovém kahanu lék. Odbíhaje od vroucího plecháčku k oknu své pracovny, pečlivě pozoroval noční oblohu. **John.** Jakási jiná, vroucí, tmavá krev byla nalita v žilách těchto lidí. **R. Svob.** Z nedovřených rtů sálal jí vroucí dech. **Čap. Ch. D zř. vařící se, kolotající, vířící.** Vlny bičované a rozryté tisícerými křižujícími se jizvami a vroucí tisícerými víry. **Pašek. D citově opravdový; hluboký, vřelý.** Mluvil tlumeně, vroucím hlasem. **Jir.** U příležitosti pátého výročí osvobození Československa posílám Vám své vroucí přátelské pozdravy. **R. právo.** Dívka zvedla kvapně skloněnou hlavu, vroucí pohled utkvěl na okamžik na mladém, statečném muži. **Jir.** Bůh vroucí modlitby její neoslyšel. **Arb.** Přitiskla vroucí políbení na jeho ruku. **Schulz.** Zasloužíš si takto někdy ještě matčin vroucí dík. **Zey.** Má vroucí duše žije jen tobě, přírodo. **Vrch.** Ty, od jakživa tak vroucí ctitel všeho venkovského, jsi nevěren svým zásadám? **Just.** Byl vroucím přívržencem josefinského osvícenství. **J. Vlč.** Je tu mnoho hluboce procítěného a opravdu vroucího. **Vrch.**

SEU (SURVEY OF ENGLISH USAGE)



BROWN CORPUS

W. Nelson Francis - Henry Kucera

7

- 1964
- 1. elektronicky zpracovaný korpus
- 1 milion slovních tvarů
- britská a americká angličtina
- pečlivý výběr textů
- vzorky

Definice korpusu v moderním slova smyslu

8

- Elektronické uložení
- Elektronická přístupnost
- Definovaný obsah (ČEHO) a rozsah (KOLIK)
- Standardní anotace – metada a interpretace jazykových jednotek
- Rychlost, spolehlivost a opakovatelnost vyhledávání a kvantifikace nalezeného

Dva metodologické přístupy k vytěžování korpusu

9

- Corpus based / korpusem ověřovaný, na korpusu založený výzkum
- Pravidlo/výjimka – otevřený/uzavřený seznam, frekvence
- Corpus driven / korpusem inspirovaný výzkum, korpusem řízený výzkum
- Výzkum kolokací /lexical bundles

Maskulina mají v češtině v gen. pl. koncovku –ů (pánů, hradů, mužů, strojů, předsedů, soudců).
Z tohoto pravidla existují výjimky. Které? Kolik?

[lc!=".*ů" & tag="NN[MI]P2.*"]

10

za nímž sedí u svého náčiní rybáři . Ostatním vyjma	rozhodčích/rozhodčí/NNMP2-----A-----	je tam po dobu soutěže zakázán vstup . " Letos
ský Beroun - Policisté obvinili osmadvacetiletou chovatelku	koní/kůň/NNMP2-----A-----	z Moravského Berouna na Bruntálsku z týrání zvířat . Podle
v Mosambiku za posledních 50 let postihly zatím asi milion	lidí/člověk/NNMP2-----A-----	. Oficiálně je registrováno kolem 200 obětí , skutečný počet
a Pospěcha spokojen . " Potřebovali bychom však takových	lidí/člověk/NNMP2-----A-----	víc , " poznamenal . Radní chtějí ve Zlíně zrušit
je prý marné . Tištěné ploše dominují názory starostů a	radních/radní/NNMP2-----A-----	, odlišným sdělením je věnováno jen 2,6 % této plochy
, jako 71 : hráče v pořadí . V dresu	Flyers/flyers/NNIP2-----A-----	odehrál 64 zápasů s bilancí 11 gólů a 26 přihrávek
Regio Taxis Bohemia . Ten by zahrnoval asi dvacet tisíc	lidí/člověk/NNMP2-----A-----	, to znamená pětaticet obcí od Benátek nad Jizerou až
místě tragédie policejní komisař Emmanuel Adebayo . Z pěti	lidí/člověk/NNMP2-----A-----	, kteří byli nejbliž místu výbuchu , zbyly jen zčernalé
oho ze zkušených ruských letců , podle něhož smrt desítek	lidí/člověk/NNMP2-----A-----	zavinila lidská hamižnost . Pilot Vjačeslav Achremenko potv
. " Přijeli lidé z tréninkových skupin z Plzně ,	Klatov/klatovy/NNIP2-----A-----	, Přeštic , Rokycan a Třemošné , " uvedl šéf
něj u nás na každých 100 osob připadá téměř 7	nemocných/nemocný/NNMP2-----A-----	a délka nemocenské je mnohdy delší než 30 dní .
Tu pořádá Občanské sdružení Omega - Jméno neokrajní	lidí/člověk/NNMP2-----A-----	něže a stále obedi budeš . Ve středu převzeme první

Celkem: 5056 (102 str.)

	Filtr	word	Frekvence
1.	p/ n	lidí	2 515 537
2.	p/ n	peněz	647 968
3.	p/ n	dní	465 084
4.	p/ n	obyvatel	243 036
5.	p/ n	přátel	125 885
6.	p/ n	koní	102 894
7.	p/ n	hostí	92 521
8.	p/ n	rozhodčích	78 016

	Filtr	word	Frekvence
1.	p/ n	milionů	1 400 417
2.	p/ n	metrů	806 199
3.	p/ n	bodů	725 571
4.	p/ n	hráčů	588 016
5.	p/ n	kilometrů	504 610
6.	p/ n	dolarů	491 535
7.	p/ n	měsíců	475 891
8.	p/ n	milionů	474 446
9.	p/ n	mužů	453 972

Jaké je mínění o Češích?

Kolokace na pozici 1-3 vpravo od KWIC <Češi jsou> seřazené podle míry **MI-score**

11

	Filtr		Frekvence	MI	T-score
1.	p/n	pověstní	3	16.851	1.732
2.	p/n	remcalové	4	13.793	2.000
3.	p/n	rovnostář	6	13.506	2.449
4.	p/n	dobírkový	4	13.414	2.000
5.	p/n	neotevření	3	12.812	1.732
6.	p/n	smějící	42	12.415	6.480
7.	p/n	neznabůh	3	12.385	1.732
8.	p/n	zaprděný	3	12.267	1.732
9.	p/n	šetřivý	11	11.980	3.316
10.	p/n	koumák	18	11.942	4.242
11.	p/n	ateistický	35	11.917	5.915
12.	p/n	pivařský	6	11.913	2.449
13.	p/n	vyčůraný	4	11.859	1.999
14.	p/n	spořivý	30	11.780	5.476
15.	p/n	považování	3	11.474	1.731
16.	p/n	rasista	62	11.451	7.871
17.	p/n	závistivý	26	11.401	5.097

Dva pohledy na korpus (lingvista a informatik)

12

- Nástroje NLP a korpusy
- Konverzní programy, vertikál, tokenizér
- Korpusové manažery
- Automatické analyzátory
- Lingvistické interpretace v korpusech

Vyhledání slovního tvaru *jít*

13

" Můj partner na rakovinu zemřel , protože se bál	jít	k lékaři . Z vlastní zkušenosti tedy vím , jak
,“ uvedl Malášek . Nevyloučil možnost , že může	jít	o vojáka nebo policistu . K druhé loupeži došlo jen
pádu . ^ * S výjimkou případů , kdy má	jít	k veterináři a vy se ji marně snažíte napěchovat do
mrazu . (Lidovci už ztrácejí i Moravu . Zkusí	jít	do měst a zezelenat Jihomoravští lidovci získali v komunálních volbách
patriotům se změna názvu zřejmě líbit nebude , nemusí však	jít	o řešení trvalé . Vše nyní i v budoucnosti závisí
. Tyto potíže se léčí laserem . Po operaci může	jít	pacient hned domů Prof. MUDr. Pavel Kuchynka , CSc. ,
v žádném případě . Je pravda , že jsme mohli	jít	na Švédsko a Kanadu . A nebudu lhát , že
v jedenáctimilionové zemi . ATÉNY „ Nevíme , kam máme	jít	, přijeli jsme lodí z Lesbosu , moje žena je
více Dominiků , než bývá obvyklé . A nemuselo hned	jít	o zapálené hokejové fanoušky . Stačí , že se jméno
unie přijala směrnici o nedovolené podpoře podnikání , která neumožňuje	jít	s nabídkou pod určitou minimální částku , " vysvětlil mluvčí
řadových odborářů , s jakým argumentačním mandátem k němu mají	jít	. Doslova v odborářském šoku jsem po té , co
trasy k nejzajímavějším místům parku . Můžete použít lodě ,	jít	pěšky nebo jet džípem a bahnem se dobrodit až k

Vyhledání lemmatu *jít* (KWIC+lemma+tag)

14

v Ledči . Pokud dva tři dny pořádně prší ,	jde/jít/VB-S---3P-AA---I	začít pár kilometrů výše proti proudu a zdolat Stvořidla .
to přece dělal baron Prášil ! Americký kaskadér David Smith	jde/jít/VB-S---3P-AA---I	v jeho stopách - o víkendu tenhle bláznivý kousek předvedl
takové se mezi nimi možná občas najdou - , ale	jde/jít/VB-S---3P-AA---I	o spory , jež jdou nutně ruku v ruce s
ale také ceny za téměř třicet tisíc korun . "	Jde/jít/VB-S---3P-AA---I	opravdu o nálož cen kvalitních značek , " řekl pořadatel
příplácet , když chce státní úřad . Zvlášť , když	jde/jít/VB-S---3P-AA---I	o peníze , které tak jako tak od státu předtím
v baru , bowling . Něco , co lidi přiměje	jít/jít/Vf-----A----I	tam , a ne trávit čas jinak . Vesecko nabízí
s onou šaškárnou . Upevňuje se klamně přesvědčení , že	jde/jít/VB-S---3P-AA---I	o replikaci bytostí . Že je to něco jako nesmrtelnost
rání firemního mobilního telefonu a automobilu k soukromým účelům	jdou/jít/VB-P---3P-AA---I	ruku v ruce s rostoucími ekonomickými výsledky společnosti . Zkrátka
filmy promítají . „ Máme rozpis , podle kterého měla	jít/jít/Vf-----A----I	kopie filmu od nás do Sezimova Ústí . Bohužel jsem
byl opilý . Tehdy mi řekl , že když to	nepůjde/jít/VB-S---3F-NA---I	po dobrém , půjde to po zlém , " vypověděla
na Liberec ? No , dlouho to vypadalo , že	jdeme/jít/VB-P---1P-AA---I	na Karlovy Vary , protože Litvínov v posledním kole vyrovnal
zjevně mladší a nezdravě agresivní . " Jsem ženská a	jdou/jít/VB-S---1P-AA---I	z práce . " odpověděla jsem poněkud mimo . "
v koupelně a přišla je tam Kate , když si	šla/jít/VpFS---3R-AA---I	upravit nalíčení . Pak ale Pippa začala chodit s bankéřem
neexistující exoty a kdo všechno o tom podvůdku věděl a	nešel/jít/VpMS---3R-NA---I	to oznámit na policii . Ztratit paměť je , milí
, protože se jedná o klasické inflační peníze , které	jdou/jít/VB-P---3P-AA---I	v podstatě do stavebních firem . Mnohem lepší podle mne
= hrůza ! Jaké máte zkušenosti s českými silnicemi ?	Nejde/jít/VB-S---3P-NA---I	mi dost dobře do hlavy stav českých dálnic , které
v síti STEP () . HLEdEJtE spoLEčnou cEstu Pokud	jde/jít/VB-S---3P-AA---I	o předkládání projektů na odstraňování starých ekologických zátěží , nejbližší
. Celý dospělý život jsem měla největší přání - aby	šel/jít/VpMS---3R-AA---I	bolševik od válu . A to se mi splnilo .

desambiguace

15

- Pánové, *nežeňte se.*
- Nemluv a *rožni.*
- Jan *je osel.*

Víceznačné tvary

16

- *nežeňte/(ne)hnat/V*
- *nežeňte/(ne)ženit/V*
- *se/se/P*
- *se/s/R*
- *rožni/rožnit/V*
- *rožni/rozžehnout/V*
- *rožni/rožeň/N*
- *je/být/V*
- *je/on/P*
- *osel/osel/N*
- *osel/osít/V*

hnát/ženit

17

pracujte za byt a stravu . Ale nevdávejte se a	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	jen kvůli tomu , že chcete uniknout z do
sexu	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	s učitelkou . Potkáte - li ženu , která krác
Shrnutí těchto vědeckých poznatků : Nejezděte na kole a	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	a nevdávejte) .
legendární Knoflenkou .	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	KDO S KÝM KDE
Takže se mějte a smějte (a	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	do ženění
vážný důvod , proč tam chce jet .	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	HEJNICE
Tvůrci :	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	Pot
takovýchto " žertíků " neustále a iv jiném prostředí .	<u>NEŽEŇTE/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	ZA SLEVOU , ŽÁDNÁ NENÍ
Jedn	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	z jedné činnosti do druhé .
Uvědomte si , že vaše zásoba energie je omezená .	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	7 . V
Udo Pollmer říká : " Vyhodte z bytu televizi a	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	. Svatba je počátkem tukového obalovár
ventilaci - či pootevřete e dveře . Když skončíte ,	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	sprchovat - studenou vodou , ale zůstaň
to , co vyděláte hned dál investujte .	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	do nejistých investic
3 .	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	Nezkouším
požehnaného Čechova , jenž pravil , cituji : „ '	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	, pokud se bojíte osamělosti . ' "
se mohli vrátit k zaříkadlům , která přivolávají lásku .	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	však hned do vážných známostí , nejdřív
na cizí . Sebastian Roch Chamfort	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	! Anton Pavlovič Čechov
Bojíteli se samoty ,	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	Bůh stv
Uvědomte si , že vaše zásoba energie je omezená .	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	z jedné činnosti do druhé .
7 . V	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	na sál jen proto , že kamarádi byli . Zept
Nepodléhejte trendům a	<u>nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	za detaily . Nechť je vám v myšlenkách i
vědění je chromá . Věda bez náboženství je slepá .	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	do velkých zakázek bezhlavě . Získávejte
růst zajistí .	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	do ženění . Až na svatební cestě Eddie z
5 . Budte trpěliví a nevdávejte to	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	
20.00 hodin kino v Hejnicích . Motto filmu zní :	<u>Nežeňte/hnát/Vi-P---2--N---I se/se/P7--4-----</u>	

být/on

18

Není to nic moc , ale Millicent to nepozná .	Je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	. Znova - que voulez-vous ? " Prozpěvovaly kvulevu ,
Titulní strana Šípu , která o premiérovi hlásala , že	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	, či ta , kde trenér Brückner drží před začátkem
a profous považuje za svou povinnost říci mu , že	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	. Mužstvo jde na těžké cvičení do hor , ale
Nebudu dělat to , co řekne Manuel . „ Manuel	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	, Pere ! Osel je zvíře a ty přece nebudeš
městský jarmark . V roce 1616 následovala hra Ďábel	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	, v níž ani ďábel nestačí na zlo lidí ,
o kurský sedlák osít dvě míry pole krupicí . Kdo	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	, nechť se přihlásí . V říčce , do které
o rozesmálo všechny . Mě tedy nejvíc překvapilo , že	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	chytřejší než kůň . Po dotazech , které jsme měli
dy smím použít přejemnělého výrazu - jaký ten druhý	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	. Ale ne tak Evženka . Ta , když vypravuje
kde se před drátěným plotem něco pohybovalo . " To	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	! " zajásala Julie a už se k němu rozeběhla
: A co dále , Baltazare , kde hlavním hrdinou	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	nebo u S. Paradžanova : Barva granátového jablka , kde
osla a pro jistotu pod obraz napsal : " Toto	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	. " Bedlivější analýzou vyprávěných anekdot lze zjistit , že
to na tebe narařčil , ty osle ! Román Gaius	<u>je/on1/PPFP4--3----- osel/ose1/NNMS1-----A-----</u>	ukazuje , že Henry Winterfeld psal už před půlstoletím lepší
Trutnovska z obce Mladé Buky , že v ulici Tmavá	je/být/VB-S---3P-AA---I osel/ose1/NNMS1-----A-----	. „ Policisté díky místní znalosti vypátrali jeho majitele a

rožeň/rožnit/rozžehnout

19

NY . , O. LI . V . Y ,

ROŽNI/rožeň/NNIP7-----A-----

, K. I . Š , Pa . Le .

to vím jistě . Z nářečí znám třeba slovo „

rožni/rožeň/NNIP7-----A-----

“ . Ester Hozáková , 6 let , FrýdekMístek :

k paní Haleové , „ že s těmi hodinami ,

rožni/rožeň/NNIP7-----A-----

a planetárium máte doma malé muzeum . “ „ Jestli

jste naučila vy jeho ? Různé moravismy jako šufánek nebo

rožni/rožeň/NNIP7-----A-----

. Je nevěra tak zajímavá , že o ní teď

nás chodí dívat místní . Je tam obrovský krb s

rožni/rožeň/NNIP7-----A-----

, ve kterém se peče drůbež - od bažanta přes

je jen krátký úryvek z nabídky mas na grilu i

rožni/rožeň/NNIP7-----A-----

, které Farma nabízí . Mezi omáčkami , přílohami a

pro služebnictvo sestoupil do kuchyní , kde skomíral oheň pod

rožni/rožeň/NNIP7-----A-----

bez masa . Na zemi vedle stolu na porcování masa

pražsky , což rozhodně popírám ! Děti říkají fajne ,

rožni/rožeň/NNIP7-----A-----

místo rozsvíř , štramplicky místo punčocháče . Opakují slova ,

, ale také teplá kuchyně s grily , rošty a

rožni/rožeň/NNIP7-----A-----

, kde gastronomické poklady před našima očima do zlatova dozrávaly

místnosti Dechem řádky staletí s krápníky sazí z ohňů pod

rožni/rožeň/NNIP7-----A-----

prasat či skopců hovoří centrální černá kuchyně , kam by

sklad plný všelikého harampádí a jen mohutné topeniště krbu s

rožni/rožeň/NNIP7-----A-----

a naproti stojící sporák byly důkazem , že stojím v

to pyšný . V Praze pořád používám slovo žufánek nebo

rožni/rožeň/NNIP7-----A-----

. A nerozumějí mi , “ prozradil . Když Okamura

, ale také teplá kuchyně s grily , rošty a

rožni/rožeň/NNIP7-----A-----

, kde gastronomické poklady před našima očima do zlatova dozrávaly .

Filologie a korpusy

20

- Obecné a specializované korpusy
- Příklady z českého prostředí
- Tvorba vlastního korpusu

ÚČNK <http://ucnk.ff.cuni.cz/cs/>

21

- akademický projekt 1994
- systematicky mapuje češtinu i další jazyky
- po bezplatné registraci otevřeny všem zájemcům

Korpusy ÚČNK

https://kontext.korpus.cz/first_form?corpname=omezeni%2Fsyn2015

22



Můj seznam | **Všechny korpusy**

Zrušit výběr

řada SYN

řada ORAL

InterCorp

synchronní

diachronní

mluvený

psaný

webový

současná verze

starší verze

čeština

cizojazyčný

paralelní

srovnatelný

autorský

reprezentativní

akviziční

specializovaný

Specializovaný - příklad

23

- Korpus Jerome je **jednojazyčný srovnatelný korpus** (*monolingual comparable corpus*) speciálně sestavený pro zkoumání překladové češtiny (tedy textů přeložených do češtiny z jiných jazyků) v porovnání s češtinou nepřekladovou (původní česky psanou).

autorský

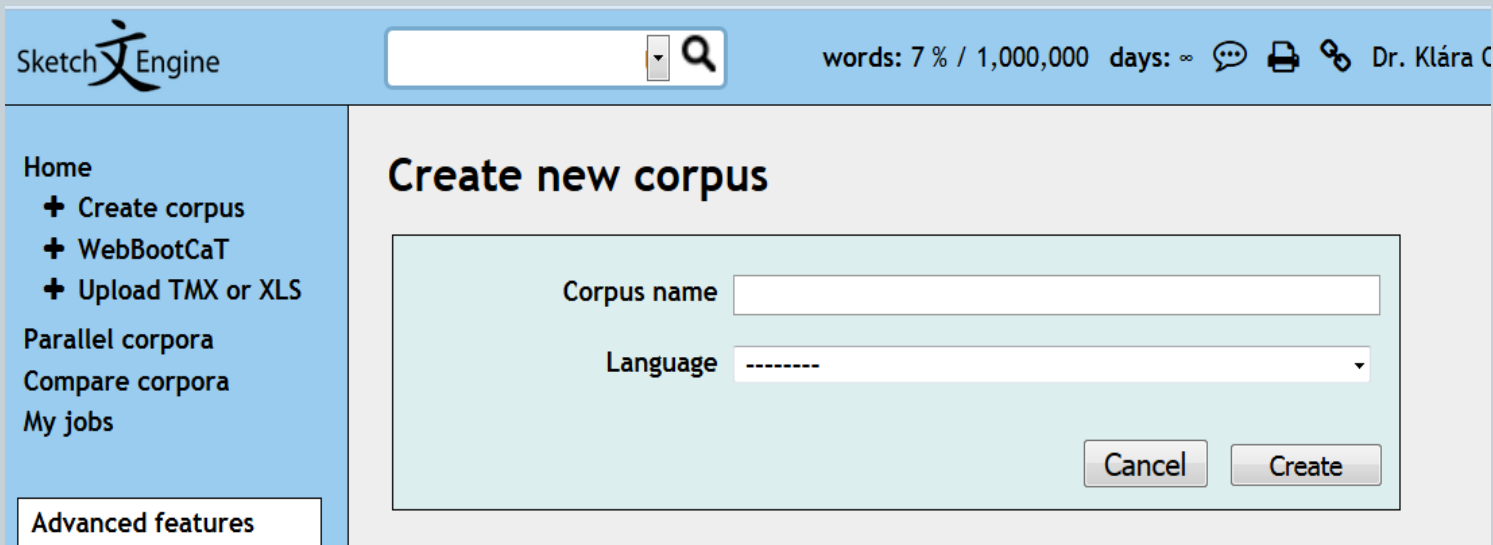
24

- Karel Čapek
- Bohumil Hrabal
- Korespondence Karla Havlíčka Borovského

Tvorba vlastního korpusu - DIY

25

- <https://nlp.fi.muni.cz/cs/JakVytvoritKorpus1>
- https://ske.fi.muni.cz/auth/create_corpus/



The screenshot shows the Sketch Engine web interface. At the top, there is a search bar and a status bar displaying "words: 7 % / 1,000,000 days: ∞" along with icons for chat, print, and share, and the name "Dr. Klára C". On the left side, there is a navigation menu with the following items: "Home", "Create corpus", "WebBootCaT", "Upload TMX or XLS", "Parallel corpora", "Compare corpora", "My jobs", and "Advanced features". The main content area is titled "Create new corpus" and contains a form with two input fields: "Corpus name" and "Language". Below the form are two buttons: "Cancel" and "Create".

Výuka jazyků a korpusy

26

- Metoda DDL (Tim Johnes)
- Žákovské korpusy (Learner Corpora)
- Učebnicové korpusy

Žákovské korpusy

27

- Texty mluvčích L2
- Výzkum interlanguage
- Zpětná vazba na základě analýzy chyb
- Sledování růstu jazykových kompetencí

Učebnicový korpus Učko

28

1	Adamovičová, A. – Ivanovová, D. (2007): <i>Basic Czech I, II</i> . Praha: Karolinum.
2	Bischofová, J. et al. (2007): <i>Čeština pro cizince a azylanty B1</i> . Brno: SOZE.
3	Bořilová, P. – Holá, L. (2011): <i>Česky krok za krokem 2</i> . Praha: Akropolis.
4	Cvejnová, J. (2011): <i>Česky, prosím</i> . Praha.
5	Čechová, E. – Remediosová, H. (2005): <i>Chcete mluvit česky? 1, 2</i> Liberec: Harry Putz.
6	Froulíková, L. (2008): <i>Adam a Eva v Českém ráji</i> . Praha: Academia.
7	Hádková, M. (2005): <i>Čeština pro cizince a azylanty A1, A2</i> . Brno: SOZE.
8	Holá, L. (2006): <i>New Czech Step by Step</i> . Praha: Akropolis.
9	Holá, L. (2010): <i>Čeština Express 1, 2</i> . Praha: Akropolis.
10	Kestráňková, M. et al. (2010): <i>Čeština pro cizince B1</i> . Brno: Cpress.
11	Matula, O. (2007): <i>Český den. Kurz českého jazyka pro azylanty navazující na Manuál pro učitele českého jazyka pro cizince bez znalosti latinky</i> . Praha: Člověk v tísni o.p.s., Projekt Varianty.
12	Nekovářová, A. (2006): <i>Čeština pro život - 15 moderních konverzačních témat</i> . Praha: Akropolis.
13	Parolková, O. (2004): <i>Czech for foreigner s 1, 2</i> . Praha: Bohemika.
14	Pintarová, M. – Režková, I. (1995): <i>Communicative Czech. Elementary Czech. Intermediate Czech</i> . Praha: Univerzita Karlova.
15	Rigerová, K. (2000): <i>Czech for Everyone</i> . Praha: K. Rigerová.
16	Štindl, O. (2008): <i>Easy Czech. Elementary</i> . Praha: Akronym.
17	Štindlová, B. (2008): <i>Česky v Česku 1, 2</i> . Praha: Ústav jazykové a odborné přípravy UK: Akropolis.

chodit do/na (porovnání obecného a učebnicového korpusu)

29

Celkem: 17391 (348 str.)

	Filtr	word	lc [lowercase word]	Frekvence
1.	p/ n	do	školy	21 199
2.	p/ n	do	práce	15 196
3.	p/ n	na	procházky	4 640
4.	p/ n	do	kina	3 661
5.	p/ n	do	kostela	2 940
6.	p/ n	do	posilovny	2 785
7.	p/ n	do	školky	2 693
8.	p/ n	na	fotbal	2 153
9.	p/ n	do	divadla	2 104
10.	p/ n	do	lesa	2 063
11.	p/ n	do	zaměstnání	1 808
12.	p/ n	na	tréninky	1 629
13.	p/ n	do	hospody	1 540
14.	p/ n	na	pivo	1 534
15.	p/ n	na	houby	1 380

	lemma	word	Frekvence
P N	chodit na	procházky	5
P N	chodit do	kina	4
P N	chodit do	školy	3
P N	chodit do	divadla	3
P N	chodit na	oběd	2
P N	chodit na	dlouhé	2
P N	chodit na	balet	2
P N	chodit do	kostela	2
P N	chodit na	různé	1
P N	chodit na	ryby	1
P N	chodit na	lekce	1
P N	chodit na	koncerty	1
P N	chodit na	houby	1
P N	chodit na	hodinu	1
P N	chodit na	dlouhou	1
P N	chodit na	brigádu	1
P N	chodit do	přírody	1
P N	chodit do	práce	1
P N	chodit do	lesa	1
P N	chodit do	kin	1
P N	chodit do	hospodv	1

Závěr

30

- Historie a současnost – technický pokrok a metodologické přístupy
- Rychlost- spolehlivost – opakovatelnost experimentu
- Zdroje nespolehlivosti
- Co je k dispozici a co si mohu sám udělat
- Na co nezbyl čas

Děkuji vám za pozornost

31