

***JAK LZE Z JAZYKOVÉHO KORPUSU
VYTVOŘIT SLOVNÍK SLANGU:
PŘÍKLAD KORPUSU FRANKOFONNÍCH
RAPOVÝCH PÍSNÍ***

Alena Polická

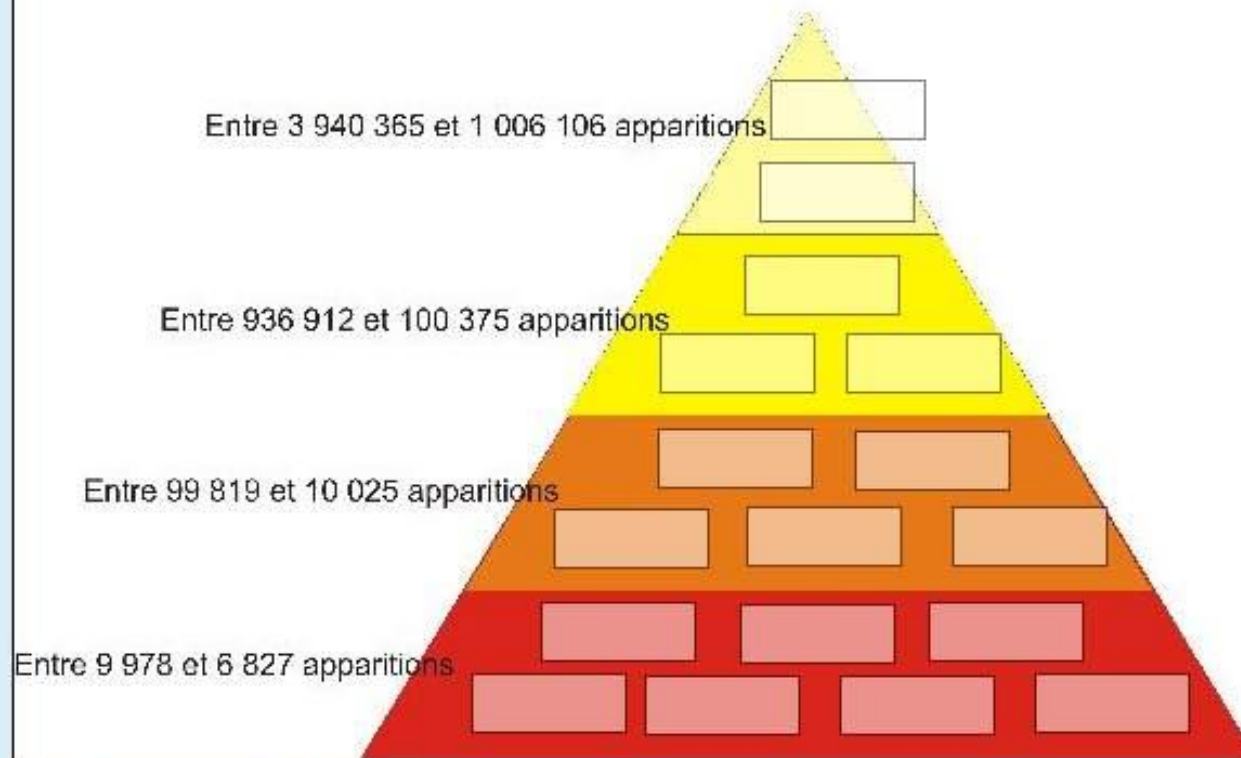
Ústav románských jazyků a literatur

Filozofická fakulta MU



LEXIKÁLNÍ PYRAMIDA

Pyramide lexicométrique de la langue française

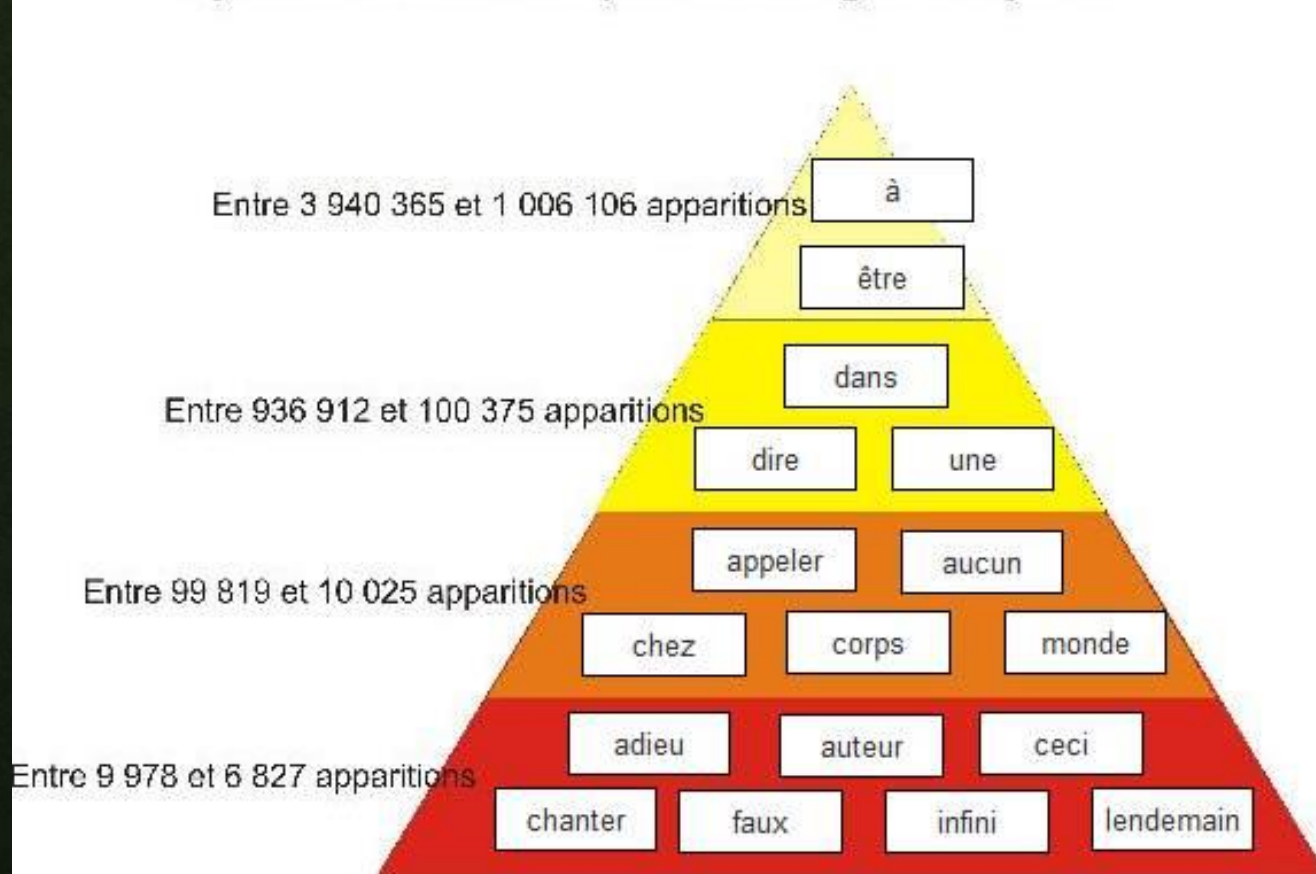


être	à	dans	dire	une	monde	chez	appeler
corps	lendemain	faux	auteur	chanter	adieu	ceci	aucun
infini							

Frekvence slov fr. jazyka (1789-1965) podle 70 milionového korpusu *Trésor de la langue Française* (TLF)

SYNSÉMANTIKA VS AUTOSÉMANTIKA

Pyramide lexicométrique de la langue française



12 najfrekventovanějších slov v korpusu TLF:
de, la, être, et, que, le, à, avoir, les, il, ne, je

62 vysokofrekvenčních slov (žlutá zóna):
un, se, des, en, qui, une, dans, ce, du, pas, elle, pour, me, vous, plus, au, on, sur, par, nous, mais, son, ce, lui, faire, comme, tout, avec, dire, sa, si, y, cette, pouvoir, ses, même, bien, ils, mon, ces, ou, sans, autre, où, aller, homme, moi, voir, aux, tu, grand, leur, deux, savoir, ma, vouloir, encore, tous, jour, là, petit, dont

➤ pokles podílu synsémantik ve prospěch autosémantik

FREKVENCE SLOV V MENŠÍCH KORPUSECH

Příklad frekvenčních slovníků ve 2 manažerech:
Sketch Engine vs TXM pro 2 letošní verze RapCoru

721,7 tis. tokenů (březen 2019) a
745,6 tis. tokenů (září 2019)

LISTE DE MOTS [DEV] RapCor 1211

lemme (7 501 éléments | 640 097 fréquence totale)

Lemme	↓	Fréquence ?	Lemme	↓	Fréquence ?
1 le		50 328 ...	11 du		9 500 ...
2 de		26 475 ...	12 mon		8 615 ...
3 être		21 133 ...	13 on		8 322 ...
4 je		19 961 ...	14 pas		8 079 ...
5 que		15 003 ...	15 en		6 852 ...
6 un		13 657 ...	16 dans		6 406 ...
7 ce		13 331 ...	17 tu		6 328 ...
8 avoir		12 179 ...	18 qui		6 181 ...
9 et		9 837 ...	19 se		6 118 ...
10 à		9 811 ...	20 pour		6 082 ...

Seuils: Fmin: 1 Fmax: 9999999

lemme	Fréquence T=745596
le	52041
,	33264
de	27452
être	21814
je	20524
que	15604
un	14043
ce	13794
avoir	12622
.	10737
à	10247
et	10056
du	9846
mon	8932

Specifičnosti orálních korpusů => zápornka *ne* – až 23. pozice

JAK ZJISTIT FREKVENCI SLOV V KORPUSU

- Př. ČNK - použité frekvenční míry:
- i.p.m. (jednotka relativní frekvence slova vztažená na 1 mil. slov),
- ARF (průměrná redukováná frekvence – částečně eliminuje problém idiolektu)

ČNK – nová (od září 2019) a uživatelsky přívětivá aplikace: *Slovo v kostce* = *Word at a Glance* (WaG)

The screenshot displays the 'Word at a Glance' (WaG) web application. The search results for the word 'slang' are shown. The interface includes a search bar with the word 'slang' entered and a 'Search' button. Below the search bar, it indicates 'Showing results for: slang (noun)'. The main content area is divided into three panels:

- Basic characteristics:** Shows the entered form 'slang', the lemma 'slang', the part of speech 'noun', and the frequency band.
- Word forms:** Shows the word form 'slangem'.
- Frequency according to a text type:** Shows a bar chart with two categories: 'Beletrie' and 'Oborová literatura'.

▼ Frequency information ?

Basic characteristics ?

entered form:

slang

lemma:

slang

part of speech:

noun

frequency band:



occurrences per million words:

0.99

words with similar frequency:

startér, pobožnost, infocentrum, vibrační, pravomocně, nebetýčný, rozměňňovat, třískat, hejtmanství, hard, scestí, naznačující, vyluxovat, mrholení, exoticky, šalotka, posuv, ohavnost, lakomý, sendvičový, zkracování, namontovaný, piercing, maratónský

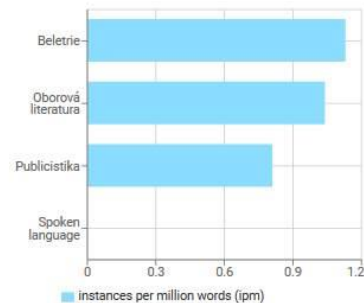
Source: syn2015

Word forms ?

slang slangu slangem

Source: syn2015

Frequency according to a text type ?



Source: syn2015 + oral_v1, more detailed information: typy textů v SYN2015 (KonText)®, výsledky v ORAL verze 1 (KonText)®

▼ Written language ?



Source: syn2015, more detailed information: Kolokace v SYN2015 (KonText)®

1 Členové ruskojazyčných kriminálních webů psali komentáře z lokálním slangem, kterému by porozuměli možná nějaký lingvista, avšak jen ti nejlepší z nejlepších by ho mohli napodobit dostatečně přesvědčivě, aby nezbudili podezření.

2 Členové ruskojazyčných kriminálních webů psali komentáře v lokálním slangem, kterému by porozuměli možná nějaký lingvista, avšak jen ti nejlepší z nejlepších by ho mohli napodobit dostatečně přesvědčivě, aby nezbudili podezření.

3 Mluví to slangem automaticky.

4 Připomínkou Odyssea Jamese Joyce a Ignota Christophera Marlowa povýšil přizemní slangem intelektuální úroveň.

5 Vzhledem k tomu, kde O Žije, by aspoň trochu španělsky rozumět měla, ale kromě nějakého toho slangem několikrát frází, aby si dokázala objednat tacos, neumí nic.

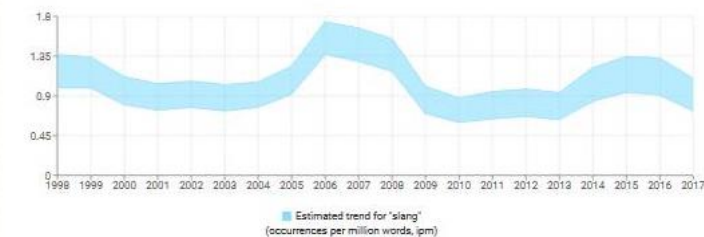
6 Členové ruskojazyčných kriminálních webů psali komentáře v lokálním slangem, kterému by porozuměli možná nějaký lingvista, avšak jen ti nejlepší z nejlepších by ho mohli napodobit dostatečně přesvědčivě, aby nezbudili podezření.

7 Členové ruskojazyčných kriminálních webů psali komentáře v lokálním slangem, kterému by porozuměli možná nějaký lingvista, avšak jen ti nejlepší z nejlepších by ho mohli napodobit dostatečně přesvědčivě, aby nezbudili podezření.

8 Mluví to slangem automaticky.

Source: syn2015

Frequency of the word during time ?



Estimated trend for 'slang' (occurrences per million words, ipm)

Př.
Slang at a glance

Slovo v kostce

Vyhledat slovo | [Hledat ve dvou jazycích / přeložit](#)

Čeština Français slang

Zobrazují výsledky pro: slang (podstatné jméno)

▼ Překlad na základě paralelního korpusu | ?

Doložené překlady

jargon argot

Zdroj: InterCorp, podrobnější informace: Trec[®]

Překlad v různých typech textu

	Publicistika	Beletrie	Parlamentní debaty	Evropská legislativa	Filmové titulky
argot					
jargon					

Zdroj: InterCorp

Ukázky překladů

1 Nějaký bílý slang?

C'est quoi, de l'argot de plouc surliste?

Jednací kolo z Dauhá klopýtalo již na začátku a celý rok se nacházelo ve slepé uličce, a to kvůli zásadnímu problému, kterým je konkrétně to, že systém pro všechny země dosáhl svých tolerovatelných limitů - pro rozvinuté, rozvíjející se nebo nejméně rozvinuté země, jak se v mezinárodním **slangu** říká těm zemím, jež jsou těžce postiženy chudobou, a které jsou nuceny začlenit se do mimořádně soutěživých světových trhů, jež je pohlcují.

1 Si les négociations du cycle de Doha piétinent depuis le début et sont au point mort depuis un an, c'est qu'il y a un problème profond: le système a atteint ses limites supportables pour tous, pays développés, émergents, ou moins avancés, comme le **jargon** international appelle ces États qui plongent dans la misère, forcés de s'intégrer dans un marché mondial ultra-concurrentiel qui les lamine.

Zdroj: intercorp_v11_cs

frekvenční pásmo:



kolikrát v milionu slov:

0,99

slova s podobně častým výskytem:

startér, pobožnost, infocentrum, vibrační, pravomocně, nebetyčný, rozměňovat, třískat, hejtmanství, hard, scestí, naznačující, vyluxovat, mrhnutí, exoticky, šalotka, posuv, ohavnost, lakomý, sendvičový, zkracování, namontovaný, piercing, maratónský

▼ Překlad na základě paralelního korpusu | ?

Doložené překlady  

jargon **argot**

Abs. frekvence 21
Rel. frekvence 80,8
Nalezené varianty [argot](#)

Zdroj: [InterCorp](#), podrobnější informace: [Treq](#)

Treq

DATABÁZE
PŘEKLADOVÝCH
EKVIVALENTŮ

Výchozí jazyk

Čeština

Cílový jazyk

Francouzština

Omezit na ?

Kolekce: 6

slang

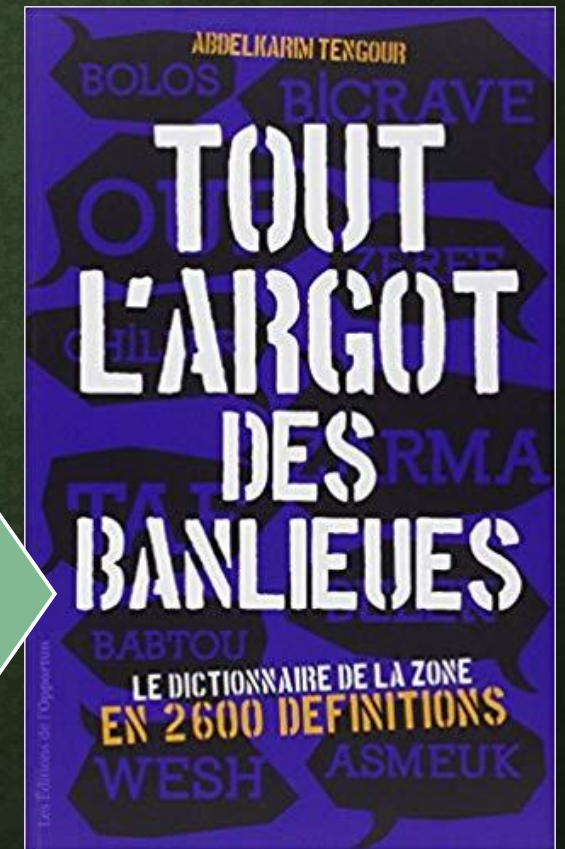
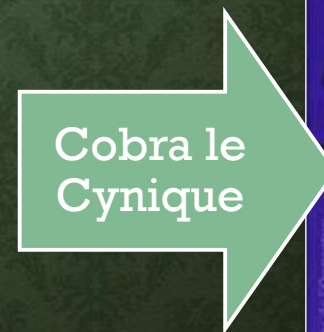
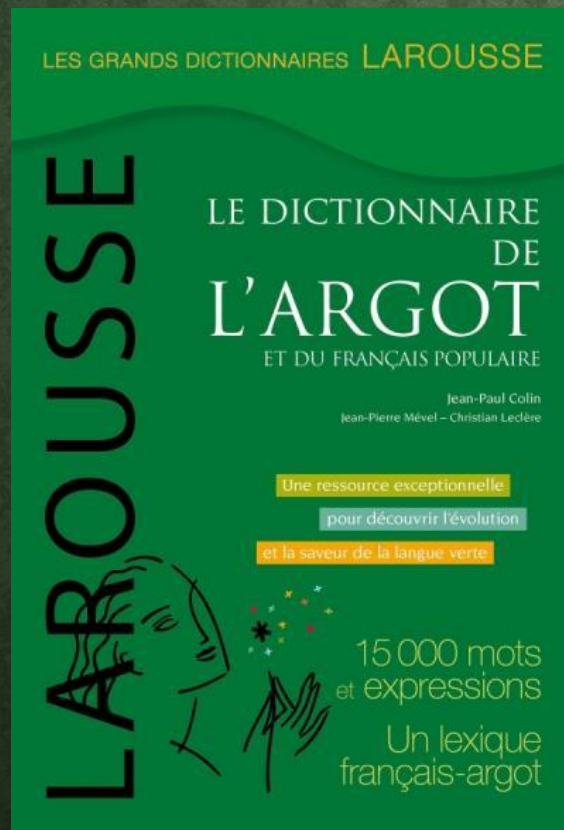
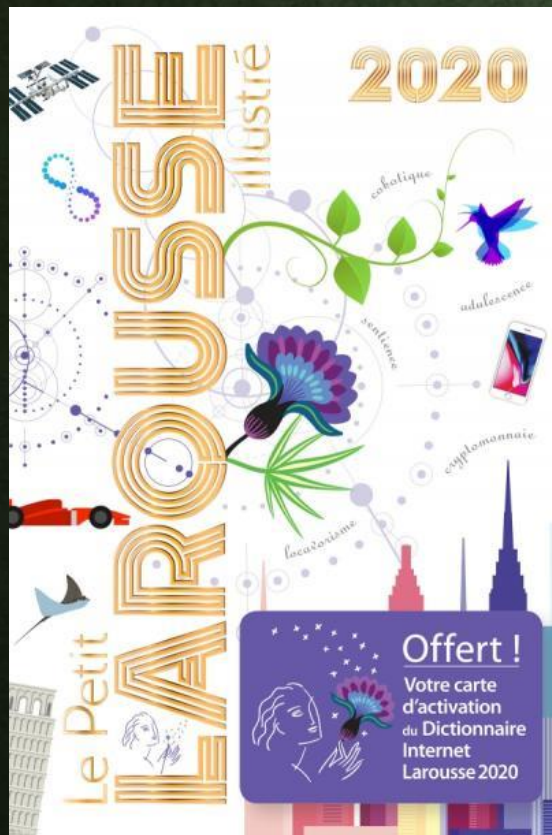
Hledej

 Lemmata ? Viceslovné ? Regulární ? A = a ?

▲ Frekvence ▼	▲ Procenta ▼	▲ Čeština ▼	▲ Francouzština ▼
11	84.6	slang	argot
1	7.7	slang	argotique
1	7.7	slang	Argot
13			

Starší nástroj než *WaG*: databáze překladových ekvivalentů *Treq* (od října 2015)
založený na tzv. jádru InterCorpu

CENTRUM A PERIFERIE LEXIKONU Z HLEDISKA LEXIKOGRAFICKÉHO



Francouzština	Free zdarma	Study 59 Kč	Profi 249 Kč	Platinum 499 Kč
Všeobecný slovník ▲	Kapesní	Praktický	Velký	Platinum
▶ počet hesel	92 000	92 000	107 000	178 000
▶ počet významů	144 000	144 000	169 000	271 000
▶ počet příkladů, frází a idiomů	24 000	24 000	86 000	195 000
▶ počet překladů	286 000	286 000	417 000	681 000
Namluvená výslovnost	✓	✓	✓	✓
Tabulky tvarosloví	✓	✓	✓	✓
Gramatika ⓘ	mini	✓	✓	✓
Konverzace ⓘ	mini	✓	mini	✓
Mluvník ⓘ	✗	✓	✗	✓
Užití slov ⓘ	✗	✓	✓	✓
Slovník slangu ⓘ	✗	✓	✓	✓
Ekonomický slovník	✗	✗	✓	✓
Technický slovník	✗	✗	✓	✓
			Objednat	Objednat
				Objednat

SPOLUPRÁCE S LINGEOU NA EXTRAKCI SLANGOVÝCH NEOLOGISMŮ (2012)



PAS DE BLEME ! Slovník slangu a hovorové francouzštiny



(4.5)

Celkový počet hlasů: 10

Nové vydání oblíbeného titulu. Francouzský slovník jsme značně rozšířili, přibylo na šest set hesel. Součástí je tradičně seznam zkratk a fonetického zápisu slov, která se běžně používají v e-mailech, SMS zprávách nebo na chatu. Dále zde naleznete slangová slova pocházející z arabštiny a romštiny. Novinkou je slovníček hovorové francouzštiny používané v Kanadě.

[více zde ...](#)

Vydání: 2., 2012 | **Stran:** 192 | **Formát:** 165x115 mm | **ISBN:** 978-80-87471-61-6

Vazba: lepená vazba, obal měkké lamino | **Ebook format:** epub | **Dostupnost:** skladem

kniha

cena: 289 Kč

[vložit do košíku](#)

Obsah

10 000 aktuálních slangových a hovorových výrazů

slangové a hovorové idiomy s příklady a českými překlady

slovník kanadské hovorové francouzštiny

NUTNOST KVALITNÍHO ZNAČKOVÁNÍ V RÁMCI TZV. SLANGOTVORNÝCH TÉMAT

Př. Synonyma pro lehké
drogy v korpusu RapCor
(verze březen 2019)

1	shit	36	49,88		...
2	sum	2	2,77		...
3	seum	2	2,77		...
4	herbe	2	2,77		...
5	hasch	2	2,77		...
6	HI	2	2,77		...
7	zetla	1	1,39		...
8	splif	1	1,39		...
9	popo	1	1,39		...
10	hi	1	1,39		...
11	hash	1	1,39		...
12	chichon	1	1,39		...
13	chichichichon	1	1,39		...

NUTNOST PROPRACOVANÉHO SYSTÉMU LEMMATIZACE PRO SESKUPENÍ GRAFICKÝCH VARIANT

Př. *cité* = „sídliště“

– grafická variabilita slangového
ekvivalentu *tess* {*tèce*; *tèc*; *téc*}

verlanizace:

[site] > [tesi] > [tɛs]

Word	↓	Fréquence	Fréquence/million		
1	tess	44	60,96		...
2	tèce	2	2,77		...
3	Tess	2	2,77		...
4	tés	1	1,39		...
5	téc	1	1,39		...
6	tèss	1	1,39		...
7	tèc	1	1,39		...
8	Tè-ce	1	1,39		...

JAK ZVOLIT LEMMA?

Word	↓	Fréquence	Fréquence/million		
1	Refré	8	11,08	<div style="width: 100%;"><div style="width: 100%;"></div></div>	...
2	refrès	5	6,93	<div style="width: 100%;"><div style="width: 80%;"></div></div>	...
3	refrés	3	4,16	<div style="width: 100%;"><div style="width: 60%;"></div></div>	...
4	refré	3	4,16	<div style="width: 100%;"><div style="width: 60%;"></div></div>	...
5	refrè	3	4,16	<div style="width: 100%;"><div style="width: 60%;"></div></div>	...
6	re-frès	2	2,77	<div style="width: 100%;"><div style="width: 40%;"></div></div>	...
7	reufrais	1	1,39	<div style="width: 100%;"><div style="width: 20%;"></div></div>	...
8	re-frè	1	1,39	<div style="width: 100%;"><div style="width: 20%;"></div></div>	...

Př. Neutrální informatické řešení pro *reufré* a *refré* v *Dictionnaire de la zone* (online)

reufré

[Définition](#) [Commentaires](#)

reufré *nom masculin.*
Frère.

Quand il pleut sur les bancs d'la tess, mes reufrés rouillent | Pour qu'j'vois la vie du bon côté il m'faut des lunettes infrarouges (NAKK MENDOSA, « Invincible », *Le monde est mon pays*, 2010) 📖.

Mon reufré j'espère que tu me crois ! | J'étais connu dans Paris avant d'faire du peurra ! (NEXFEL, « Ne le dis pas à ma mère feat. Alpha Wann », *En sous-marin*, 2011) 📖.

J'ai égaré mon temps reufré, j'regarde les journées défiler | Et j'reste jamais longtemps au frais comme les bouteilles de l'épicier (DEEN BURBIGO, « Un « je ne sais quoi » », *Inception*, 2012) 📖.

VAR. refré .

ÉTYM. Verlan de *frère*.

Le Dictionnaire de la Zone. Tout l'argot des banlieues. © 2000 - 2019 Cobra le Cynique.

NUTNOST KONTEXTOVÉHO DEFINOVÁNÍ HESEL

CONCORDANCE [DEV] RapCor

Obtenu

CONCORDANCE [DEV] RapCor

CQL [sens="ami"] 116 > mélanger 116 x > filterFirstHit 76 (117,47 par M) x

Fréquence [MODIFIER LES CRITÈRES](#) [RETOUR AU CONCORDANCIER](#)

Montrer la fréquence par million

	Lemma	↓	Fréquence	Fréquence/million
1	chum		21	32,46
2	pote		18	27,82
3	poteau		8	12,37
4	poto		6	9,27
5	kho		6	9,27
6	soce		4	6,18
7	frangin		4	6,18
8	pot		1	1,55

1 2005 • TAK06 • TAKTIKA
y arriver Des **chums** qui dealent la

2 2008 • MAC05 • MAC TYER
son cou Ses **potos** figé , impact

3 2008 • BES09 • BENI SNASSEN
a rendu nos **potos** pantins Certains de

4 2003 • IAM39 • FREEMAN
partie du quota **potos** A cause de

5 1999 • KOM04 • KOMA
j' inviterai mes **potes** ailleurs qu' en

6 2005 • TAK05 • TAKTIKA
la musique entre **chums** pis de rêver

7 2003 • JGD01 • MC JEAN GAB1
deutsch non mon **pote** j' suis pas

8 2012 • SPC02 • Sir Pathétik
table . Des **chums** pas fiables et

9 2005 • TAS01 • TAKTIKA ET SANS PRESSION
front avec tes **pots** mais Sans Pression

Korpusy nejsou pouze dortové: Mezinárodní den jazyků, PedF MU, 26.9.2019

ZÁVĚREM

Přínos korpusu frankofonních rapových textů (2009-2019):

- primárně didaktický (seminární a kvalifikační práce studentů)
- druhotně vědecký – reprezentativní doklad: 1) dynamického vývoje na periferii slovní zásoby
2) grafické neukotvenosti neologismů i starších slangových (tudíž primárně orálních) výrazů
3) internacionalizace slovní zásoby (nejen slov, ale i hlásek – viz *rho*, *hass*)
4) variantnosti francouzštiny na všech úrovních (diatopické, diachronní, diafazické, diastratické, diamezické)

Děkuji za pozornost.
Merci pour votre attention.

Kontakt:

podhorna@phil.muni.cz

Webové stránky korpusu RapCor: <https://is.muni.cz/do/phil/Pracoviste/URJL/rapcor/index.html>

