

Název	Nejčastější česká slova
Štítky	ČJ, počítač, projektor, slovní zásoba, rodilý, 6.–7. roč., 8.–9. roč., SŠ, hledání, porovnávání

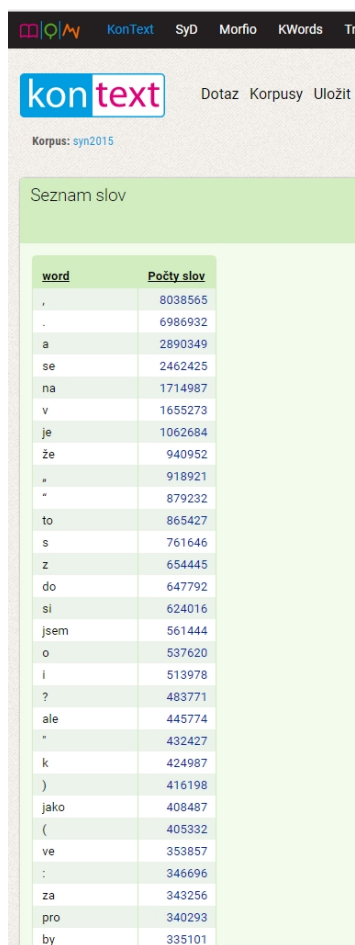
Zadání

Přemýšleli jste někdy nad tím, jaké je nejčastější české slovo? Jazykové korpusy nám v tom mohou napovědět. V rozhraní KonText (kontext.korpus.cz) lze snadno vytvořit **seznam nejčastějších českých slov** (příp. konkrétních slovních tvarů) ve formálním psaném i neformálním mluveném jazyce a porovnat, zda a případně jak se liší. Využijte funkci Seznam slov v menu Dotaz a srovnajte seznamy slov (lemma) nebo slovních tvarů (word) z **korpusu SYN2015 (psaná čeština)** a **ORTOFON (mluvená čeština)**.

Tip: Tečky v mluveném korpusu neznačí interpunkci, ale pauzy (. krátká pauza, .. delší pauza).

Klíč

Rozhraní KonText > v menu Dotaz > Seznam slov (korpus SYN2015 a ORTOFON)



word	Počty slov
.	8038565
.	6986932
a	2890349
se	2462425
na	1714987
v	1655273
je	1062684
že	940952
*	918921
"	879232
to	865427
s	761646
z	654445
do	647792
si	624016
jsem	561444
o	537620
i	513978
?	483771
ale	445774
"	432427
k	424987
)	416198
jako	408487
(405832
ve	353857
:	346696
za	343256
pro	340293
by	335101

psaný



word	Počty slov
..	85871
to	54765
.	38866
a	33880
no	31746
tak	22500
že	21158
jo	20774
je	20223
tam	16499
jako	15386
se	14872
?	14839
na	14627
hmm	14456
já	14339
+	13917
jsem	12939
ale	11677
(nadechnuti)	10517
v	9656
ty	9581
@	9548
si	9294
...	8104
ne	8022
(rušivý zvuk)	7543
co	6997
jak	6769
už	6411

mluvený

V obou seznamech se na předních příčkách objevují tečky (a další interpunkce), ovšem v psaném jazyce značí konec věty a v mluveném jazyce pauzu v řeči. Jak je obvyklé snad ve všech jazycích, ta nejčastěji používaná slova jsou velmi krátká a zpravidla neplnovýznamová, s výjimkou slovesa *být*. Na seznamu z mluveného jazyka nacházíme výplňková slova jako *no*, *jako* nebo *hmm*, kdežto na seznamu psaného formálního jazyka se vyskytuje více předložek (*z*, *do*, *s*, *ve*, *za*, *pro*, *o*). Nejčastější slovo v mluvené češtině je *to*, které má mnoho různých funkcí (*to je hezké*, *udělej to*, *myslela jsem to* . *no* . *víš ne* apod.). Mezi nejčastějšími výrazy v mluvené řeči je i *ty*, neboť spontánní mluvená řeč bývá dialogická (zahrnuje více mluvčích).

Konkrétní počty slov odrážejí velikost korpusů: psaný SYN2015 má přibližně 100 milionů slov, kdežto mluvený ORTOFON jen milion. Pokud bychom tedy chtěli porovnávat četnost nějakého výrazu v obou korpusech, museli bychom tato čísla přepočítat (buď na procenta vzhledem k celkové velikosti daného korpusu, nebo na tzv. ipm. počet výskytů na milion: $\text{ipm} = \text{počet} / \text{velikost korpusu} * \text{milion}$).

Korpusy umožňují vyhledat i nejčastější slova v rámci konkrétního slovního druhu i typu textu, např. typická adjektiva v publicistice (viz pracovní list 15 – Typická adjektiva v publicistice).