Frekvenční slovníky

* Frekvence je jednou z nejpopulárnějších charakteristik slov také kvůli jejímu jednoduchému použití.
* Frekvence je závislá na 1. **délce slova** – čím kratší slovo, tím větší frekvence; 2. **slovním druhu** – funkce slova závisí na úloze slova v textu (funkce ve větě – auto/synsémantická= nižší frekvence); 3. **sémantice** (slova monosémní/ polysémní); 4. **centrum/periferie** (centrum = neutrálnost, periferie = příznakovost 🡪 místo v systému jazyka)
* Frekvenční slovníky se řadí do skupiny nevýkladových slovníků a pro český jazyk jich existuje několik:
* JELÍNEK, Jaroslav, Josef V. BEČKA a Marie TĚŠITELOVÁ: *Frekvence slov, slovních druhů a tvarů v českém jazyce*. 1 vyd. Praha: Státní pedagogické nakladatelství, 1961. 585 s.
* Práce byly započaty r. 1940 a výzkum byl ukončen v roce 1953. Slovník však vyšel až po několika odkladech v roce 1961, a to zásluhou prof. Bohuslava Havránka.
* Již v roce vydání nebyl slovník zcela aktuální - vzhledem k převratu v r. 1948 a tím i změně životního stylu, se slova jako soudruh, kombajn, brigáda… atd. používala v daleko větší míře než na počátku výzkumu.
* Slovník čítá 1 623 527 hesel pocházejících ze 75 pramenů. Nejprve je uveřejněno 10 tisíc vůbec nejfrekventovanějších slov a v druhé části jsou pak všechna hesla řazena dle abecedy.
* Hesla pocházejí z literárních děl a tiskovin rozmanitého charakteru, které vyšly po roce 1930 (to proto, aby korpus co nejvíce odpovídal současné podobě jazyka). Tato díla byla posléze rozřazena do osmi tematicky, stylisticky i slohově odlišných kategorií jako: beletrie, poezie, literatura pro mládež, drama, odborná literatura, žurnalistika, vědecká literatura a mluvené projevy uveřejněné tiskem.
* Slovník zachycuje nejen frekvenci tvarů, ale i dvojtvarů. Za samostatná slova jsou vnímány předložky, členy sousloví a spojky. Za jedno slovo považují autoři také slovesné tvary složené, slovesa pojící se se zvratným zájmenem, příslovečné spřežky či zájmena (bůhvíkdo, kdekdo sotvajaký) a složeniny (sebevětší).
* Pořadí slov dle frekvence bylo uspořádáno tak, že začíná slovem s nejvyšší frekvencí a končí s frekvencí jedna. Pořadí slova je dáno číslem na levé straně – napravo pak najdeme 3 čísla (první: absolutní frekvence v materiálu, druhé: počet skupin výskytu a třetí: výskyt v pramenech).

 Obr. 1 (*Frekvence slov* str. 101)



Obr. 2 (*Frekvence slov* str. 267)

* ČERMÁK, František a kol. : *Frekvenční slovník češtiny*. Vyd. 1. Praha: Nakladatelství Lidové noviny, 2004. 595 s. ISBN 80-7106-676-1.
* Základem a referenčním zdrojem slovníku je korpus *SYN 2000*, který byl vytvořen Ústavem Českého národního korpusu na FF UK v Praze a který díky rozsahu (100 miliónů slov) představuje nejrozsáhlejší zdroj informací o psané češtině.
* Hesla ve slovníku mají výrazné procento znaků mluvnosti a pojetím vycházejí z reálné, autentické a mnohdy nekodifikované jazykové formy. Slovník je velmi ovlivněn dobou a vlivem médií.
* Hlavní část slovníku, která je řazena abecedně, obsahuje 50 000 nejběžnějších apelativ s údaji o frekvenci. Dále je uvedeno 20 000 nejběžnějších apelativ řazených dle frekvence, 2 000 nejčastějších proprií českých i cizích v závěru je 1 000 nejčastějších zkratek. V dodatcích jsou také uvedeny nejpoužívanější interpunkční znaménka nebo písmena.
* *Frekvenční slovník češtiny* je tedy řazen jak abecedně, tak i podle absolutní frekvence. Podobná hesla jsou navíc sloučená do jednoho hesla. Např. nezapomenutelný + zapomenutelný.
* Terminologie: **Rank ARF** znamená pořadí hesla podle průměrné redukované frekvence (zohledňuje počet pramenů, v nichž se heslo nachází); **ARF** je hodnota průměrné redukované frekvence; **Rank FRQ** je označení pro pořadí hesla podle absolutní frekvence a **FRQ** je počet výskytů daného slova v korpusu.
* **10 nejfrekventovanějších apelativ podle *FSČ*** (řazeno dle FRQ):A (2792011), V/ VE (2487420), SE (2298438), BÝT/ BEJT (2042223), NA (1643934), TEN (1268161), S/ SE (879319), ŽE (865069) a Z/ ZE (844271)
* Nejčastějším propriem je Praha (Brno až na 8. místě); nejpoužívanějšími zkratkami jsou ČR, tzv., USA, aj. a Kč. Nejfrekventovanějším

interpunkčním znaménkem je čárka a nejčastějším grafémem je o. Obr. 3 (*FSČ* str. 197)

* TĚŠITELOVÁ, Marie et al. *Frekvenční slovník češtiny věcného stylu*. Praha: Ústav pro jazyk český ČSAV, 1983. 329 s.
* Tento slovník je jedním ze závěrů zkoumání v oblasti psané a mluvené češtiny 70. - počátku 80. let. Shrnuje výsledky kvantitativní analýzy lexika, složek publicistiky, věcného stylu, odborné češtiny a administrativy. Podrobně se všem těmto odvětvím věnují slovníky, které byly vydávány jen jako interní tisky Ústavu pro jazyk český ČSAV (viz níže).
* Rozsah korpusu je 540 000 slov. Z toho 180 000 hesel pochází z publicistiky, 300 000 z odborné češtiny a 60 000 z administrativy). Z toho 75% hesel pochází z psaných a 25% z mluvených projevů.
* *Frekvenční slovník češtiny věcného stylu* lemmanizuje (tzn., že pod základním heslem se ukrývají všechny jeho pádové formy a tvary) a vychází ze spisovné češtiny*.*
* Hesla jsou uspořádána dle absolutní frekvence (sestupně) a uvádí jen hesla do frekvence 5.
* **10 nejfrekventovanějších slov v *FSČVS***: A 19017, V/VE 12345, BÝTI/NE 12345, NA 8937, TEN 8839, KTERÝ V 5341, ZE 4552, Z/ E 4529, TENTO/TENHLE/TEN 4413, S/E 4410
* ČERMÁK, František a kol. *Frekvenční slovník mluvené češtiny*. Vyd. 1. Praha: Karolinum, 2007. 510 s. ISBN 978-80-246-1425-0
* *FSMS* představuje autentickou mluvenou češtinu - jak lidé skutečně mluví v 21. století. Hesla ve slovníku jsou bez kodifikačních zásahů. Obr. 4 (*FSMČ* str. 196)
* Vychází z *Pražského mluveného korpusu* (PMK), získaného na základě magnetofonových nahrávek v terénu (v letech 1988 - 1992 a 1994 - 1996) v podobě rozhovorů s 278 mluvčími, které měly být cca 2300 slov, tedy 20 min. (Obsahuje 2 typy rozhovorů: jeden volný, druhý podle předem připraveného dotazníku).
* Slovník má rozsah asi 780 000 slov a je rozdělený do dvou částí: Abecední slovník mluvené češtiny (obsahuje všechna stylová lemmata a jejich tvary, která se v korpusu vyskytla více než jednou) a Abecední slovník málo frekventovaných slov (slova s frekvencí 1= lemmata).
* **10 nejfrekventovanějších slov dle *FSMČ***: TEN40271, BEJT 20255, A 19586, ŽE 13542, JÁ 8743, MÍT 7871, NO 7784, V 7716,NA 7067 aJAKO 6717
* Dále jsou uvedena nejčastější autosémantika (bejt), synsémantika (ten), substantiva (dítě), adjektiva (dobrej), verba (bejt) a adverbia (tam).
* Kromě výše zmíněných slovníků máme i speciální FS:
* TĚŠITELOVÁ, Marie a kol. *Frekvenční slovník současné české publicistiky*. Praha: Ústav pro jazyk český ČSAV, 1980. 189 s.
* TĚŠITELOVÁ, Marie, ed. a kol. *Frekvenční slovník současné české administrativy*. Praha: Ústav pro jazyk český Československé ak. věd, 1980. 86 s.
* TĚŠITELOVÁ, Marie, ed. aj. *Frekvenční slovník současné odborné češtiny*. Praha: Ústav pro jazyk český ČSAV, 1982. 229 s.
* Prvních 10 slov v našich frekvenčních slovnících velmi dobře ukazuje nejčastější slova v češtině vůbec. Jsou to:
* spojky A, ŽE
* předložky, tzv. primární, jednoznačně předložka V, NA; s jistým kolísáním v pořadí předložky S a Z
* sloveso BÝT (ve významu spony)
* zájmena (ukazovací) jednoznačně zájmeno TEN, dále vztažné nebo tázací zájmeno KTERÝ
* Nejčastější slova v češtině jsou tedy slova formální, mluvnická a jsou většinou velmi krátká. Jde tu o projev ekonomizace jazyka.
* Při frekvenci slov rozlišujeme tři skupiny druhů slov:
* jmennou: podstatná jména, přídavná jména a předložky
* slovesnou: slovesa, zájmena, příslovce, spojky
* neutrální: číslovky, citoslovce, částice
* Použitá literatura:

JELÍNEK, Jaroslav, Josef V. BEČKA a Marie TĚŠITELOVÁ: *Frekvence slov, slovních druhů a tvarů v českém jazyce*. 1 vyd. Praha: Státní pedagogické nakladatelství, 1961. 585 s.

ČERMÁK, František a kol. : *Frekvenční slovník češtiny*. Vyd. 1. Praha: Nakladatelství Lidové noviny, 2004. 595 s. ISBN 80-7106-676-1

TĚŠITELOVÁ, Marie et al. *Frekvenční slovník češtiny věcného stylu*. Praha: Ústav pro jazyk český ČSAV, 1983. 329 s.

ČERMÁK, František a kol. *Frekvenční slovník mluvené češtiny*. Vyd. 1. Praha: Karolinum, 2007. 510 s. ISBN 978-80-246-1425-0