

Srovnání nejfrekventovanějších lemmat v psané a mluvené češtině

Bc. Leona Špetová

Český jazyk – specializační studium, 3. ročník



Úvod

Náplň práce

- zkoumání frekvenční špičky (20 nejfrekventovanějších lemmat) psaného korpusu SYN2020 a mluveného korpusu ORAL

Východisko práce

- nedostatečné prozkoumání mluvené komunikace a vztahu mezi psanou a mluvenou komunikací¹

Výzkumné otázky

- Která lemmata se vyskytují ve frekvenční špičce korpusu SYN2020 a ORAL? Proč jsou tato lemmata tak frekventovaná? Proč jsou frekventovanější než jiná lemmata?
- Jaké jsou významy a funkce nejfrekventovanějších lemmat v korpusu SYN2020 a ORAL? V čem se tato lemmata shodují a v čem se liší? Užívají se v korpusu SYN2020 a ORAL shodně, nebo odlišně?
- Co lze na základě nejfrekventovanějších lemmat v korpusu SYN2020 a ORAL vyvodit ohledně psané a mluvené komunikace?

Přínos výzkumu

- prohloubení poznatků o psané a mluvené komunikaci
- materiály pro lexikografické zpracování zkoumaných lemmat
- materiály pro mezijazykové srovnání
- podklady pro výuku češtiny jako cizího jazyka

Získání dat

- vyhledání nejfrekventovanějších lemmat v korpusu SYN2020 a ORAL pomocí funkce Seznam slov webového rozhraní KonText
- získání seznamů nejfrekventovanějších lemmat v korpusu SYN2020 a ORAL
- manuální úprava obou seznamů: doplnění informace o relativní frekvenci lemmat; sjednocení odlišně lemmatizovaných jednotek (úprava lemmat být, on a co)
- stanovení 20 nejfrekventovanějších lemmat v korpusu SYN2020 a ORAL

Analýza dat



- stanoveno 20 nejfrekventovanějších lemmat v korpusu SYN2020 a ORAL
- u jednotlivých lemmat vytvořen vzorek o velikosti 100–300 konkordančních řádků z korpusu SYN2020 a ORAL
- provedena sémantická analýza lemmat na základě výkladových slovníků a dalších lingvistických prací
- srovnány výsledky sémantické analýzy lemmat v korpusu SYN2020 a ORAL

Frekvenční špičky korpusů SYN2020 a ORAL

	SYN2020		ORAL	
rank	lemma	rank v ORAL	lemma	rank v SYN2020
1	být	2	ten	6
2	a	3	být	1
3	se	5	a	2
4	v	16	no	212
5	na	14	se	3
6	ten	1	já	18
7	on	7	on	7
8	s	27	tak	26
9	že	9	že	9
10	z	31	jo	576
11	který	80	mít	12
12	mít	11	jako	19
13	do	28	tam	80
14	o	60	na	5
15	i	70	ale	16
16	ale	15	v	4
17	k	76	hmm	4104
18	já	6	vědět	57
19	jako	12	ne	90
20	moci	38	co	24

- lemmata se vyskytují ve frekvenční špičce obou korpusů
- lemmata se vyskytují pouze ve frekvenční špičce daného korpusu

Výsledky

Frekvenční špičky obou korpusů

- tvořeny převážně synsémantiky
- lemmata mají především gramatickou funkci
- lemmata se podílejí na výstavbě a organizaci textu
- lemmata mají více významů a bývají řazena k více slovním druhům

Frekvenční špička korpusu SYN2020

- více předložek; velmi frekventované jsou víceslovné předložky
- více spojek; velmi frekventované jsou víceslovné spojky

Frekvenční špička korpusu ORAL

- více zájmen a deiktických výrazů
- více diskurzních markerů²

Závěr



Psaná komunikace

- složitější syntax; vyjádření různých významových vztahů
- explicitní vyjadřování
- nejfrekventovanější lemmata odrážejí významy popsané ve slovnících
- lexikální bohatost a rozmanitost

Mluvená komunikace

- méně strukturovaná a jednodušší syntax; syntaktické celky k sobě řazeny volně
- vliv sdíleného kontextu: výrazy s vágním významem; implicitní vyjadřování; výrazy poukazující k účastníkům komunikace
- nejfrekventovanější lemmata často užitá ve významech, které nejsou popsány ve slovnících
- lexikální omezenost a repetitivnost: kumulace stejných výrazů; rutinní obraty a fráze
- vznik lineárně a v reálném čase: časté výrazy s pragmatickou funkcí

¹ srov. CARTER, R. – MCCARTHY, M. (2017): Spoken Grammar: Where Are We and Where We Are Going? Applied Linguistics 38, č.1, s. 1–20; HOFFMANNOVÁ, J. – ZEMAN, J. (2017): Výzkum syntaxe mluvené češtiny: inventarizace problémů. Slovo a slovesnost 78, č. 1., s. 45–66.

² charakteristika diskurzních markerů viz ČERMÁKOVÁ, A. – JÍLKOVÁ, L. – KOMRSKOVÁ, Z. – KOPRIVOVA, M. – POUKAROVÁ, P. (2019): Diskurzní markery. In Hoffmannová, J. – Homoláč, J. – Mrázková, K. (eds.): Syntax mluvené češtiny. Praha: Academia, s. 244–351.

