

SLAN

(Slovotvorný analyzátor pro výuku češtiny pro cizince)

projekt aplikovaného výzkumu č. TL03000293 (podpořen TAČR)

Ivana Kolářová, Anastasija Sokolova, (Adriana Válková)

Brno
29. 3. 2023

MUNI
PED

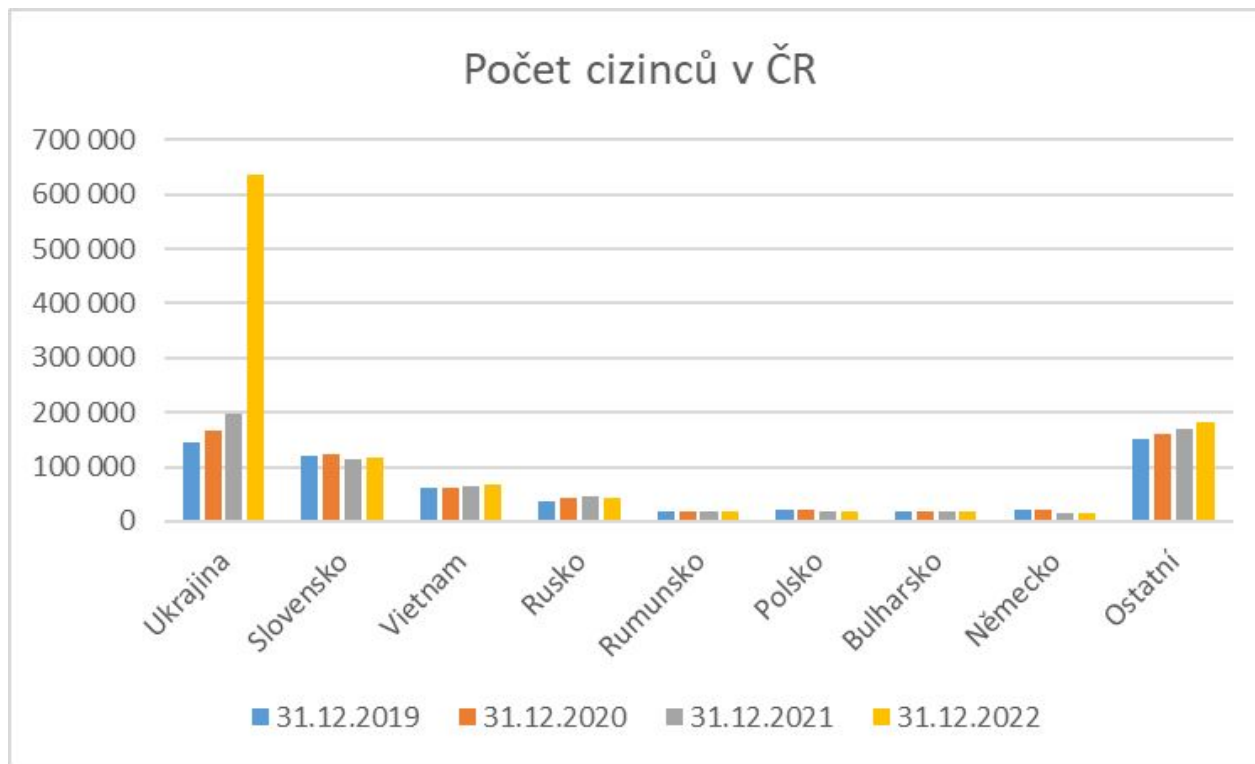
T A
Č R

Tento projekt je spolufinancován se státní podporou Technologické agentury ČR v rámci Programu ÉTA.

www.tacr.cz
Výzkum užitečný pro společnost.



Motivace: extralingvistické faktory



Zdroj dat: https://www.czso.cz/csu/cizinci/4-ciz_pocet_cizincu#cr.

Graf: vlastní zpracování.

Motivace: teoretická východiska

Slovotvorba v procesu osvojování (cizího) jazyka

- Argus, R. and V. Kazakovskaya (2018) Acquisition of noun derivation in Estonian and Russian L1. *Eesti Rakenduslingvistika Ühingu aastaraamat / Estonian Papers in Applied Linguistics*, 14, 23–39.
- Bauer, L. and P. Nation (1993) Word families. *International Journal of Lexicography*, 6 (4), 253–279.
- Clark, E. V., and R. A. Berman (1984) Structure and Use in the Acquisition of Word Formation. *Language*, 60 (3), 542–590.
- Hacken, ten P., A. Abel and J. Knapp (2006) Word Formation in an Electronic Learners' Dictionary: ELDIT. *International Journal of Lexicography*, 19 (3), 243–256.
- Madsen, R. S. (2019) Word-formation preferences of non-natives. *SKASE Journal of Theoretical Linguistics*, 16 (1), 162–176.
- Mattes, V., W. U. Dressler, K. Korecky-Kröll and S. Sommer-Lolei (2021) *The Acquisition of Derivational Morphology*. Amsterdam: John Benjamins Publishing Company.
- Nádeníček, P. (2017) Krátké zamyšlení nad postavením slovtvorby ve výuce češtiny jako cizího jazyka. In: Nekula M. and K. Šichová (Ed.) *Variety češtiny a čeština jako cizí jazyk*, 171–183. Praha: Akropolis.
- Olshtain, E. (1987) The Acquisition of New Word Formation Processes in Second Language Acquisition. *Studies in Second Language Acquisition*, 9 (2), 221–231.
- Tschichold, C. and P. ten Hacken (2015) Word-formation in second language acquisition. In: Müller P. O., Ohnheiser I., Olsen S. and F. Rainer (Ed.) *Volume 3. Word-formation: an international handbook of the language of Europe*, 2137–2154. Berlin: Mouton de Gruyter.

Slovotvorba ČJ, RJ a dalších slovanských jazyků

- Bozděchová, I. (2016) Czech. In: Müller P. O., Ohnheiser I., Olsen S. and F. Rainer (Ed.) *Volume 4. Word-formation: an international handbook of the language of Europe*, 2872–2891. Berlin: Mouton de Gruyter.
- Janda, L. (2010) Russian word-formation in contrast with Czech and Norwegian. *Oslo Studies in Language (OSLa)*, 2 (2), 243–259.
- Körtvélyessy, L. (2016) Word-formation in Slavic languages. *Poznan Studies in Contemporary Linguistics*, 52 (3), 455–501.
- Uluhanov, I. S. (2016) Russian. In: Müller P. O., Ohnheiser I., Olsen S. and F. Rainer (Ed.) *Volume 4. Word-formation: an international handbook of the language of Europe*, 2953–2978. Berlin: Mouton de Gruyter.

Východisko dat: učebnicový korpus UčebKo

Proč korpus?

Korpus

- rozsáhlý soubor elektronicky uložených a statisticky zpracovaných textů
- obecné x specializované
- synchronní x diachronní
- psané x mluvené
- anotované x neanotované
- ...

Učebnicový korpus

- korpus tvořený z učebnic (češtiny pro cizince)
- zachycuje slovní zásobu cizinců na jednotlivých jazykových úrovních

Východisko dat: učebnicový korpus UčebKo

UčebKo

- tři subkorporusy: UčebKo-A2, UčebKo-B1 a UčebKo-B2
- 9 učebnic češtiny pro cizince (3 pro každou úroveň)
- výběr učebnic: 1. dospělí, 2. komunikativní charakter, 3. aktuálnost (ne starší 15 let), 4. různí autoři, 5. užívání ve výuce
- přístupný přes SketchEngine pouze autorizovaným uživatelům pro vědecké účely
- výhoda tvorby korpusu SketchEngine – anotovaný korpus

Korpus UčebKo (Válková 2021)

	UčebKo-A2	UčebKo-B1	UčebKo-B2
number of words	91,561	122,604	89,697
number of sentences	15,099	16,993	9,739
average sentence	6 words / sentence	7 words / sentence	9 words / sentence

noun (701 items | 6,228 total frequency)



Lemma	Absolute Frequency ?	Lemma	Absolute Frequency ?	Lemma	Absolute Frequency ?	Lemma	Absolute Frequency ?	Lemma	Absolute Frequency ?
1 republika	335 ...	11 řeka	82 ...	21 linka	56 ...	31 dívka	38 ...	41 prohlídka	31 ...
2 kolika	203 ...	12 procházka	78 ...	22 válka	52 ...	32 podmínka	37 ...	42 chřipka	30 ...
3 babička	193 ...	13 památka	70 ...	23 holka	52 ...	33 pohádka	36 ...	43 Monika	30 ...
4 zkouška	182 ...	14 učitelka	69 ...	24 Šárka	50 ...	34 sekretářka	36 ...	44 vstupenka	30 ...
5 maminka	172 ...	15 zastávka	67 ...	25 láska	50 ...	35 doktorka	35 ...	45 Lenka	30 ...
6 kamarádka	151 ...	16 banka	65 ...	26 taška	49 ...	36 peněženka	35 ...	46 Amerika	29 ...
7 ruka	138 ...	17 polévka	63 ...	27 nabídka	47 ...	37 křižovatka	34 ...	47 gramatika	28 ...
8 manželka	121 ...	18 schůzka	58 ...	28 herečka	47 ...	38 studentka	34 ...	48 motorka	28 ...
9 otázka	98 ...	19 značka	57 ...	29 písnička	42 ...	39 přednáška	32 ...	49 známka	27 ...
10 matka	97 ...	20 kočka	56 ...	30 sbírka	38 ...	40 stránka	32 ...	50 omáčka	25 ...

Rows per page: 50 1-50 of 701 < < 1 / 15 > >

Databáze SLAN zahrnuje:

1. SEZNAMY SLOV podle sémantických skupin, sufixů/prefixů, úrovní A2, B1, B2

Zahrnují slova odvozená

POUZE OD JEDNOHO SLOVA ZÁKLADOVÉHO, tj. slova tvořená nejjednodušším způsobem.

Teoretická východiska

Dokulil, M.: *Tvoření slov v češtině I. Teorie odvozování slov*. Praha: Academia, 1962.

Dokulil, M. a kol.: *Tvoření slov v češtině II. Odvozování podstatných jmen*. Praha: Academia, 1967.

Štícha F. a kol. *Velká akademická gramatika spisovné češtiny. I. Morfologie. Část 1, Část 2*. Praha: Academia, 2018

Databáze SLAN zahrnuje:

2. ÚKOLY (CVIČENÍ) K PROCVIČOVÁNÍ

- hledání slov s určitým významem a určitým sufixem v textu
- poznání významu utvořeného slova

Rozlišení podle úrovně A2, B1, B2

Databáze SLAN nezahrnuje:

- slova složená

příklad z UčebKa: *studujou tam **samoplátci** z celého světa (A2)*

- slova odvozená cirkumfixy

příklad z UčebKa: *šli k **náměstí** (A2)*

- slova vzniklá univerbizací

příklad z UčebKa: *bydlí v bytě v **paneláku** (A2)*

Důvody:

- časové možnosti projektu
- slootovorné způsoby náročnější na pochopení cizincem, vzniká-li jedno slovo minimálně ze dvou slov základových

<https://slan.ped.muni.cz/#/o-projektu>

chci úroveň:
 A2 B1 B2

chci najít:
 podstatné jméno přídavné jméno

má význam:
místa

začíná na:

končí na:
iště

chci to vysvětlit v:
 češtině ruštině angličtině

Vyhledat

bydlet → **bydliště**

hrát → **hřiště**

letět → **letiště**

(nastoupit) → **nástupiště**

parkovat → **parkoviště**

schod → **schodiště**

sportovat → **sportoviště**

stání → **stanoviště**

hledět → **hledišťe**

možnosti vyhledávaných slov **podle zvolených vlastností**

výsledek vyhledávání (příklad)

SLAN - SLOVOTVORNÝ ANALÝZÁTOR PRO VÝUKU ČEŠTINY PRO CIZINCE

chci úroveň:

 A2 B1 B2

chci najít:

 podstatné jméno přídavné jméno

má význam:

místa

začíná na:

končí na:

iště

chci to vysvětlit v:

 češtině ruštině angličtině

Vyhledat

Chci najít slova

Chci si procvičovat

bydlet → **bydliště**hrát → **hřiště**letět → **letišťe**(nastoupit) → **nástupiště**parkovat → **parkoviště**schod → **schodiště**sportovat → **sportoviště**stání → **stanoviště**hledět → **hledišťe**• **rod:** střední (to)• **příklady:***domácí/ fotbalové/ golfové
dětské hřiště (A2), být
blízko hřiště, chodit s
kamarády na hřiště*• **pozor:** mění se r → ř

CZEDD

M U N I

T A
Č R

Tento projekt je spolufinancován se státní
podporou Technologické agentury ČR v rámci
Programu ÉTA.

www.taocr.cz

Výzkum užitečný pro společnost.



SLAN - SLOVOTVORNÝ ANALÝZÁTOR PRO VÝUKU ČEŠTINY PRO CIZINCE

chci úroveň:
 A2 B1 B2

chci najít:
 podstatné jméno přídavné jméno

má význam:
místa

začíná na:

končí na:
iště

chci to vysvětlit v:
 češtině ruštině angličtině

Vyhledat

bydlet → **bydliště**hrát → **hřiště**letět → **letišťe**(nastoupit) → **nástupiště**parkovat → **parkoviště**schod → **schodiště**sportovat → **sportoviště**stání → **stanoviště**hledět → **hledišťe**

- **род:** средний (то)
- **примеры:**
*домáci/fotbalové/golfové
dětské hřiště (A2), být
blízko hřiště, chodit s
kamarády na hřiště*
- **внимание:** происходит
изменение г → ř

Chci najít slova

Chci si procvičovat

CZEDD

MUNI

T A
Č R

Tento projekt je spolufinancován se státní podporou Technologické agentury ČR v rámci Programu ĚTA.

www.tacr.cz
Výzkum užitečný pro společnost.



Co najdeme u jednotlivých derivátů?

základové slovo:

učit → učitel

v češtině, ruštině, angličtině:

- gramatické informace o rodě substantiv
- příklady kolokací z korpusu Učebko nebo ze SYN2020
- informace o hláskových změnách při odvozování, pokud se vyskytují
- informace o variantách sufixu, pokud se vyskytují

Seznamy derivátů podle sémantických skupin a sufixů - podstatná jména

Sufixy psané versálou se ukázaly jako nejvíce produktivní

SEZNAMY SLOV - podstatná jména se sufixy

sémantická skupina

sufixy

názvy osob

0	a	áč	ák	AN
ANT/ ENT	AŘ/ ÁŘ	át	CE	Č
ček	čí	ec	eň	ér/ iér
ÍK/ NÍK	ik	ina	íř	ISTA
ka	l	och	oun	TEL

SEZNAMY SLOV - podstatná jména se sufixy

sémantická skupina

názvy prostředků a nástrojů

sufixy

a	ák	č	ČKA/ KA	dlo
ík/ník	na	ní/tí	tko	

názvy objektů

a	ák	ba	č	dlo	ec
ek	ice/ nice	ÍK/ NÍK	INA	ivo	KA
ko	NÍ/ TÍ	ost	ovna		

SEZNAMY SLOV - podstatná jména se sufixy

sémantická skupina

názvy látek

sufixy

dlo	INA/NINA	ivo
-----	----------	-----

názvy jednotlivostí

ek	ín	ina	inka
----	----	-----	------

názvy míst

ba	dlo	ík/ ník	INA	isko
IŠTĚ	KA	NA	ní/tí	NICE
ost/ nost	STVÍ/ CTVÍ	tko		

SEZNAMY SLOV - podstatná jména se sufixy

sémantická skupina

názvy vlastností a stavů

názvy jazyků

sufixy

a	ba	no/o
OST/ NOST	ota	STVÍ/ OVSTVÍ

a	INA
---	-----

SEZNAMY SLOV - podstatná jména se sufixy

sémantická skupina

názvy dějů

sufixy

A	BA	dlo	e/ě	EK
el/ ež	ch	í	ina	j
KA	ná/ tá	NÍ/ TÍ	o	ost/ nost
ot	ství	t	va	

názvy poplatků

É	ovné
----------	-------------

SEZNAMY SLOV - podstatná jména se sufixy

sémantická skupina

názvy číselných pojmů

sufixy

ice	ina/ tina	iny	KA
-----	--------------	-----	----

názvy hromadné

í	oví	CTVO/ STVO
---	-----	---------------

názvy zveličelé

isko

SEZNAMY SLOV - podstatná jména se sufixy

sémantická skupina

názvy zdrobnělé

sufixy

áček	átko	ečka	ečko
EK	enka	IČKA	íčka
ičko	íčko	ík	ínek
inka	KA	ko	

SEZNAMY SLOV - podstatná jména se sufixy

sémantická skupina

názvy přechýlené

sufixy

ice	KA	yně/ kyně
-----	----	--------------

názvy ostatní

čko/ ko	ík	KA	ní/tí
------------	----	----	-------

SEZNAMY SLOV - podstatná jména s prefixy (příklady)

sémantická skupina

názvy míst

prefixy

pra	před	vele	multi
-----	------	------	-------

názvy dějů

pod	před	sou
-----	------	-----

názvy vlastností a stavů

bez	nad	po	sou
-----	-----	----	-----

SEZNAMY SLOV - podstatná jména s prefixy (příklady)

sémantická skupina

názvy osob

prefixy

multi	nad	pod
pra	před	vice

Polysémie výrazů z korpusu UčebKo

výsledek děje – prostředek činnosti

	význam 1	význam 2
	výsledek děje	prostředek činnosti
zařízení	<i>Co potřebujeme k zařízení bytu?</i> (B2)	<i>mobilní zařízení, elektronická zařízení (A2, B1, B2)</i>
sprcha	<i>dát si pořádnou sprchu (B1)</i>	<i>koupelna se sprchou (A2, B1, B2)</i>

Další případy

Někdy je v korpusech Učebko slovo doloženo jen **v jednom z možných významů**

	UčebKo	syn2020
výstava	místo	děj, výsledek děje
	<i>navštívit výstavu (A2, B1, B2)</i>	<i>výstava lidového umění</i>
	místo	vlastnost, stav
pohotovost	<i>jít na pohotovost (A2, B1, B2)</i>	<i>vyhlásit stav pohotovosti, být v plné pohotovosti</i>
	místo	děj, výsledek děje
oddělení	<i>studijní oddělení, vedoucí oddělení (A2, B1, B2)</i>	<i>brzké oddělení od rodiny</i>

sémantická kategorie	počet sufixů	počet lemmat	nejzastoupenější sufixy (počet lemmat)	příklady (jazyková úroveň)
děj, výsledek děje	19	691	-ní/-tí (464)	<i>představení (A2, B1, B2) ← představen/představit</i>
názvy osob (masculine animate)	26	417	-ík/-ník (73)	<i>zákazník (A2, B1, B2) ← zakázka</i>
vlastnost, stav	6	202	-ost/-nost (173)	<i>radost (A2, B1, B2) ← rád, zkušenost (A2, B1, B2) ← zkušený</i>
zdrobněliny	15	186	-ek (77)	<i>dárek (A2, B1, B2) ← dar</i>
názvy žen	3	162	-ka* (125)	<i>kamarádka (A2, B1, B2) ← kamarád, manželka (A2, B1, B2) ← manžel</i>
místo	13	135	-na (43)	<i>kavárna (A2, B1, B2) ← káva, knihovna (A2, B1, B2) ← kniha</i>
objekty	16	115	-ka* (38)	<i>fotka (A2, B1, B2) ← fotit</i>
prostředky a nástroje	9	48	-čka/-ka* (14)	<i>motorka (A2, B1, B2) ← motor</i>
číselné výrazy	4	30	-ka* (21)	<i>desítka (A2, B1, B2) ← deset</i>
názvy jazyků	2	29	-ina* (28)	<i>angličtina (A2, B1, B2) ← anglický</i>
hromadné názvy	3	14	-ctvo/-stvo (9)	<i>obyvatelstvo (A2, B1, B2) ← obyvatel</i>
poplatky	2	10	-né (9)	<i>jízdné (A2, B1, B2) ← jízda</i>
látky	3	8	-ina* (6)	<i>tekutina (B1, B2) ← tekutý</i>
jednotliviny	4	5	-ina* (2)	<i>potravina (A2, B1, B2) ← potrava</i>
zveličení	1	1	-isko (1)	<i>psisko (B2) ← pes</i>
jiné		53		
CELKEM:		2 106		

sémantická kategorie	počet prefixů	počet lemmat	nejzastoupenější prefixy (počet lemmat)	příklady (jazyková úroveň)
děj, výsledek děje	3	3	-	<i>podnájem (A2, B1, B2) ← nájem, předprodej (A2, B1, B2) ← prodej, souhra (B1, B2) ← hra</i>
názvy osob (masculine animate)	6	12	pra- (7)	<i>prarodič (A2, B1, B2) ← rodič</i>
místo	5	5	-	<i>předsíň (A2, B1, B2) ← síň, veletrh (A2, B1, B2) ← trh, multikino (B1, B2) ← kino, mikroregion (B1, B2) ← region, prales (B2) ← les</i>
vlastnost, stav	4	6	nad- (2), sou- (2)	<i>nadváha (A2, B1, B2) ← váha, soulad (B1, B2) ← lad</i>
CELKEM:		26		

	ž. r.	stř. r.	m. r. živ.	m. r. neživ.	CELKEM:
končí na	869	634	431	172	2 106
začíná na	6	5	9	6	26
CELKEM:	875	639	440	178	2 132

chci úroveň:
 A2 B1 B2

chci najít:
 podstatné jméno

má význam:
 pojmenování žen

končí na:
 ice, ka, yně/kyně

chci to vysvětlit v:
 češtině ruštině angličtině

[Vyhledat](#)

zákazník → **zákaznice**

úředník → **úřednice**

kadeřník → **kadeřnice**

pracovník → **pracovnice**

návštěvník → **návštěv**

přítel → **přítelkyně**

kolega → **kolegyně**

průvodce → **průvodkyně**

sportovec → **sportovkyně**

zaměstnanec → **zaměstnankyně**

pěvec → **pěvkyně**

plavec → **plavkyně**

zájemce → **zájemkyně**

- rod: ženský (ta)
- příklady: nová kolegyně (A2), představit novou kolegyni (B1), seznámit se s novou kolegyní (B1)

Děkujeme za pozornost!

- Ivana Kolářová: kolarova@ped.muni.cz
- Anastasija Sokolova: sokolova@ped.muni.cz
- Adriana Válková: ad.valkova@gmail.com