

# SLAN

## (Slovotvorný analyzátor pro výuku češtiny pro cizince)

projekt aplikovaného výzkumu č. TL03000293 (podpořen TAČR)

Ivana Kolářová, Anastasija Sokolova

*Brno*  
15. 06. 2023

**MUNI  
PED**

**T A  
Č R**

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

[www.tacr.cz](http://www.tacr.cz)  
Výzkum užitečný pro společnost.

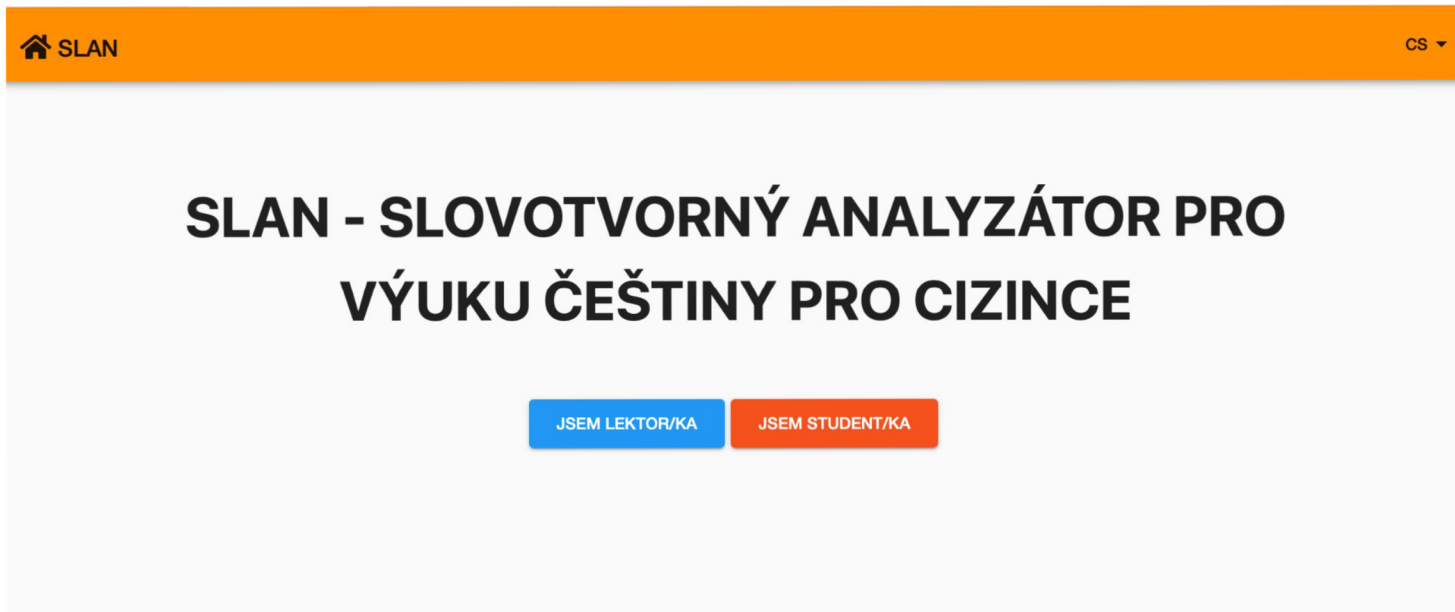


# Co je SLAN?

1. Učebnicový korpus UčebKo (9 učebnic, úrovně A2–B2)

2. Online nástroj  
(software)

<https://slan.ped.muni.cz>



SLAN - SLOVOTVORNÝ ANALYZÁTOR PRO  
VÝUKU ČEŠTINY PRO CIZINCE

JSEM LEKTOR/KA JSEM STUDENT/KA

MUNI

T A  
Č R  
Tento projekt je spolufinancován se státní  
podporou Technologické agentury ČR v rámci  
Programu ETA.  
www.tacr.cz  
Výzkum užitečný pro společnost.

eve  
technologies

## Databáze SLAN zahrnuje:

1. SEZNAMY SLOV podle sémantických skupin, sufixů/prefixů, úrovní A2, B1, B2 – slova odvozená POUZE OD JEDNOHO SLOVA ZÁKLADOVÉHO, tj. tvořená nejjednodušším způsobem.

2. CVIČENÍ

3. PŘIŘAZENÍ SLOV k tématu (např. rodina, vzdělání, dovolená...) – bude doplněno později

## **Databáze SLAN nezahrnuje:**

- slova složená (samoplátce)
- slova odvozená cirkumfixy
- slova vzniklá univerbizací

### **Důvod:**

slovotvorné způsoby náročnější na pochopení cizincem,  
vzniká-li jedno slovo minimálně ze dvou slov základových



# 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

Tento korpus je možno užit pouze při vědeckém výzkumu, jehož účelem není dosažení přímého nebo nepřímého hospodářského nebo obchodního prospěchu, a jen v rozsahu odpovídajícímu sledovanému účelu. Vždy je však nutno uvést jméno autora korpusu, jeho název a pramen.

Tento korpus je vytvořen z následujících děl pro úroveň A2: 1) Český krok za krokem 1 (od 13. kapitoly; Lída Holá, 2016), 2) Český, prosím II (Jitka Cvejnová, 2012), 3) Čeština pro cizince A1 a A2 (od 5. kapitoly; Marie Boccou Kestřánková et al., 2017),

pro úroveň B1: 1) Český krok za krokem 2 (Lída Holá, 2014), 2) Český, prosím III (Jitka Cvejnová, 2016), 3) Čeština pro cizince B1 (Marie Boccou Kestřánková et al., 2016)

a pro úroveň B2 z děl: 1) Čeština pro azylanty a cizince (Ana Adamovičová et al., 2006), 2) CZech it UP! B2 (Darina Hradilová a Kopečková, Michaela, 2020), 3) Čeština pro cizince B2 (Marie Boccou Kestřánková et al., 2020).

# 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

CONCORDANCE UčebKo

lemma čeština 212 (585.61 per million)

Details sentence

1	<input type="checkbox"/>	① doc#0	<s> Ahoj, pro naši jazykovou školu bych chtěl objednat 25 kusů učebnice <b>češtiny</b> Česky krok za krokem. </s>	<input type="checkbox"/>
2	<input type="checkbox"/>	① doc#0	<s> A co <b>čeština</b> ? </s>	<input type="checkbox"/>
3	<input type="checkbox"/>	① doc#0	<s> <b>Čeština</b> je těžká, ale zajímavá. </s>	<input type="checkbox"/>
4	<input type="checkbox"/>	① doc#0	<s> Tom a Marina jsou ve škole na lekci <b>češtiny</b> . </s>	<input type="checkbox"/>
5	<input type="checkbox"/>	① doc#0	<s> Existují dialekty i v jiných jazycích než v <b>češtině</b> ? </s>	<input type="checkbox"/>
6	<input type="checkbox"/>	① doc#0	<s> Mluví spolu obecnou <b>češtinou</b> , která je typická pro Prahu a střední Čechy. </s>	<input type="checkbox"/>
7	<input type="checkbox"/>	① doc#0	<s> Jde mi <b>čeština</b> . </s>	<input type="checkbox"/>
8	<input type="checkbox"/>	① doc#0	<s> Martin Brown má každé úterý v 10 hodin dopoledne lekci <b>češtiny</b> . </s>	<input type="checkbox"/>
9	<input type="checkbox"/>	① doc#0	<s> A co <b>čeština</b> ? </s>	<input type="checkbox"/>
10	<input type="checkbox"/>	① doc#0	<s> <b>Čeština</b> je těžká, ale zajímavá. </s>	<input type="checkbox"/>
11	<input type="checkbox"/>	① doc#0	<s> Hele Evo, dneska odpoledne je ta přednáška o té obecné <b>češtině</b> , že jo? </s>	<input type="checkbox"/>
12	<input type="checkbox"/>	① doc#1	<s> Navštěvuji přípravný kurz <b>češtiny</b> pro cizince. </s>	<input type="checkbox"/>
13	<input type="checkbox"/>	① doc#1	<s> Ano, navštěvuje kurz <b>češtiny</b> . </s>	<input type="checkbox"/>
14	<input type="checkbox"/>	① doc#1	<s> Který den v týdnu se koná víkendový kurz <b>češtiny</b> ? </s>	<input type="checkbox"/>
15	<input type="checkbox"/>	① doc#1	<s> Kolik stojí ten kurz <b>češtiny</b> ? </s>	<input type="checkbox"/>
16	<input type="checkbox"/>	① doc#1	<s> Přihlašuji se závazně na víkendový kurz <b>češtiny</b> v termínu od 20. března do 5. června. </s>	<input type="checkbox"/>
17	<input type="checkbox"/>	① doc#1	<s> Je <b>čeština</b> těžká? </s>	<input type="checkbox"/>

## **Teoretická východiska pro třídění slov do sémantických kategorií a pro slovotvornou analýzu**

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

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		

<b>sémantická kategorie</b>	<b>počet prefixů</b>	<b>počet lemmat</b>	<b>nejzastoupenější prefixy (počet lemmat)</b>	<b>příklady (jazyková úroveň)</b>
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>
<b>CELKEM:</b>		26		

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

# Polysémie v korpusu Učebko

## výsledek děje – prostředek činnosti

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

# Polysémie v korpusu Učebko

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>



# Odborná literatura (vybrané zdroje)

## 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.

# Publikace o projektu

- Sokolova, A. and A. Válková. Slovtvorný analyzátor pro výuku češtiny pro cizince (SLAN) – didaktická pomůcka s korpusově podloženými daty. In: L. Suchomelová. *Sborník Asociace učitelů češtiny jako cizího jazyka 2021*. 1. vyd. Praha: Akropolis, 2021. s. 41-48. ISBN 978-80-7470-419-2.
- Válková, A. and I. Kolářová. CZEDD a SLAN — online nástroje zaměřené na práci se slovtvorbou češtiny jako druhého/cizího jazyka. *Nová čeština doma a ve světě*. Praha: Filozofická fakulta Univerzity Karlovy, 2021, roč. 2021, č. 2, s. 78-92. ISSN 1805-367X.
- Válková, A. Building Czech Textbook Corpora (UcebKo) for Word-formation Research of Czech as a Second Language. *Jazykovedný časopis*. Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied, 2021, roč. 72, č. 2, s. 631-640. ISSN 0021-5597. doi:10.2478/jazcas-2021-0057.
- Sokolova, A., Kolářová, I. and A. Válková. Word-formation Analysis Software Tool for Teaching Czech for Foreigners (SLAN). An interim report. *Linguistica Pragensia* (in press).

# Praktická část :)

## SLAN - SLOVOTVORNÝ ANALYZÁTOR PRO VÝUKU ČEŠTINY PRO CIZINCE

JSEM LEKTOR/KA

JSEM STUDENT/KA

MUNI

T A  
Č R

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

[www.taocr.cz](http://www.taocr.cz)  
Výzkum užitečný pro společnost.

