

Koncovky *-a* a *-u* genitivu singuláru maskulin neživotných s ohledem na korpusový přístup ve výuce češtiny jako cizího jazyka

Adrian Jan Zasina

ABSTRACT:

Endings *-a* and *-u* of genitive singular inanimate masculine nouns: a corpus approach to teaching Czech as a foreign language. The aim of this study is the issue of competitive endings *-a* and *-u* of genitive singular inanimate masculine nouns in the context of foreign language teaching. The emphasis is on a systematic description of genitive endings *-a* and *-u* using corpus methods. In the first step, we analysed the part of the learner corpus including texts of Slavic language speaking students. The results have shown that students quite often confuse both of these endings: the correct ending *-u* frequently substituted by *-a*. Next, we examined the competition of genitive endings within the corpus of contemporary Czech using the Morfio tool that identifies relevant word pairs for further analysis. The identified pairs were divided into three categories: a) nouns with the same etymological origin and meaning, b) nouns with the same etymological origin but different meaning, and c) nouns with inconclusive competition of genitive endings. A systemised list of pairs, along with the proportional and absolute frequency given, is a source of information on the use of appropriate endings, with respect to frequency. The information is crucial for students of Czech as a foreign language in order to choose an appropriate variant with the ending that is closest to the current usage. Based on the analysed material, we proposed three model corpus exercises: two direct exercises for determining a more frequent variant, and one indirect exercise — taking into account semantic differences in the usage of particular endings.

KLÍČOVÁ SLOVA / KEYWORDS

čeština, genitiv, korpusová lingvistika, korpusový přístup ve výuce, morfologie, výuka cizího jazyka, žákovský korpus

corpus linguistics, Czech, DDL, foreign language teaching, Genitive case, learner corpus, morphology

1. ÚVOD

Kolísání genitivních koncovek *-a* a *-u* v singuláru maskulin neživotných (dále jen Gsg Mi) je zapříčiněno dosud nezavršeným přechodem z kmenového deklinačního systému na rodový (Vážný, 1964, s. 9), přičemž se tento proces netýká pouze češtiny, ale je přítomen i v ostatních slovanských jazycích (srov. Bígl, 2013; Janda, 1996, s. 121–145; Zasina, 2017; Žigo, 2012).¹ V českém prostředí se této tematice věnuje řada studií (např. Bermel, 2010; Rusínová, 1992; Šimandl, 2003) a gramatik (Cvrček et al., 2010;

1 Variantnost koncovek *-a* a *-u* může být rovněž zapříčiněna fakultativní životností (viz Bermel, 2011, s. 517), která spočívá v nabývání gramatických vlastností názvů živých bytostí bytostmi neživými (Шарабурак, 2012, s. 362).



Karlík, Nekula, & Rusínová, 1995). Poslední dobou vznikají i korpusové studie zaměřené na morfosyntaktické variace mj. genitivu (Bermel & Knittl, 2012a, 2012b; Bermel, Knittl, & Russell, 2014, 2018). Tyto práce primárně analyzují korpusová data kvalitativním způsobem a následně prostřednictvím dotazníků zkoumají přijatelnost dubletních tvarů rodilými mluvčími. Experimenty vždy probíhaly na dvou rovinách: dotazovaní jednak sami doplňovali genitivní koncovku do modelových vět, jednak dvěma dubletám v kontextu přiřazovali hodnotu přijatelnosti. Výsledky průzkumu ukázaly souvislost proporcionální frekvence tvarů v jazykovém korpusu s přijatelností hodnocenou rodilými mluvčími.

I přes značnou pozornost věnovanou genitivním koncovkám *-a* a *-u* nebylo dosud toto téma zcela vyčerpáno. V této studii se na danou problematiku zaměřujeme nově z hlediska výuky češtiny jako cizího jazyka (ČCJ). Klademe důraz na systematický popis zkoumaného jevu korpusovými metodami s cílem nabídnout žákům² přehled o variantních koncovkách Gsg Mi³ tvrdého vzoru a zároveň identifikovat ty varianty, jež jsou v českém jazyce nejběžnější, a tudíž jejich použití cizincem nebude vnímáno v komunikaci s rodilým mluvčím jako rušivé nebo neobvyklé. Na základě získaných výsledků následně uvádíme příkladová korpusová cvičení, která uvádějí žáky do problematiky konkurence koncovek Gsg Mi a umožňují tuto látku procvičovat.

Problematice záměny koncovek *-a* a *-u* je třeba věnovat zvýšenou pozornost zvl. u žáků se slovanským mateřským jazykem, v jejich případě totiž bývá nejčastějším důvodem chyb interference z mateřského jazyka. Existují substantiva, která mají ve slovanských jazycích stejnou koncovku, např. čes. *kousek chleba*, pol. *kawałek chleba*, rus. *кусок хлеба*, srb. *комад хлеба*, u řady slov se však genitivní koncovky napříč slovanskými jazyky neshodují, srov. čes. *slovníku* vs. pol. *słownika*, rus. *словаря*, srb. *речника*. Proto je důležité slovanské žáky na tuto skutečnost upozornit.

V následující kapitole se zaměříme na popis dat a metodologii výzkumu: nejdříve shrneme chyby v žákovském korpusu a následně identifikujeme množinu konkurenčních substantiv v korpusu synchronní češtiny. Ve třetí kapitole získaný materiál roztřídíme do tří kategorií. Ve čtvrté kapitole pak uvádíme přímá a nepřímá korpusová cvičení a na závěr shrneme nejpodstatnější zjištění.

2. POSTUP A ZKOUMANÝ MATERIÁL

Analýza žákovského korpusu CzeSL-SGT (Šebesta et al., 2014), jež se soustředila na subkorpus obsahující pouze texty žáků se slovanským mateřským jazykem (dále jen slovanský subkorpus), odhalila problém cizinců s Gsg Mi, kdy dochází v rámci tvrdého vzoru ke konkurenci koncovek *-a* a *-u*. K identifikaci chyb v prvním kroku byla použita automatická chybová anotace⁴ (Petkevič et al., 2012), která vyhodnocovala

² Termín *žák* zde používáme ve významu *učící se* (ang. learner), srov. Gráf, 2019, s. 16–17.

³ Opomíjíme zde jména s výhradní koncovkou *-a*, jejich výčet prezentuje např. Bermel (2011, s. 513).

⁴ Seznam identifikovaných chyb viz <http://utkl.ff.cuni.cz/~rosen/public/SeznamAutoChybROR1.html>



problémy se záměnou jednoho znaku za druhý (SingCh). V druhém kroku v rámci typu chyb se záměnou znaku byly ručně identifikovány obtíže s Gsg Mi. Otázka konkurence koncovek v genitivu se ukázala nejčastější obtíží u 21 různých typů chybných slovních tvarů, činících dohromady 189 tokenů. Tabulka 1 prezentuje podrobnější data spolu s absolutní frekvencí a počtem žáků, kteří se dané chyby dopustili.

	Chybný tvar	Frekvence	Počet studentů
1.	mosta	23	21
2.	kluba	21	17
3.	životu	21	18
4.	pátka	13	13
5.	obchoda	11	11
6.	byta	9	7
7.	stola	9	9
8.	lesu	9	7
9.	světu	9	9
10.	hrada	8	8
11.	páteka	8	8
12.	pateka	8	8
13.	Vyšehrada	6	6
14.	názora	6	6
15.	bara	5	5
16.	výleta	4	4
17.	hotela	5	5
18.	roka	5	4
19.	důma	5	4
20.	stůla	2	2
21.	diploma	2	2
Celkově		189	174

TABULKA 1: Problematické slovní tvary substantiv ve slovanském subkorpusu CzeSL-SGT.⁵

Ukazuje se, že častěji dochází k výměně *-u* za *-a* – celkově v 18 typech chybných tvarů (např. **pátka*, **hrada*, **názora*) —, ve zbývajících 3 případech studenti vyměnili *-a* za *-u* (**životu*, **lesu*, **světu*).

Pokud prověříme chybné slovní tvary identifikované v korpusu CzeSL-SGT na základě jazykové úrovně, zjistíme, že nejvíce chyb se podle očekávání dopustili žáci na základních úrovních A1 a A2 (tabulka 2), chybování se ale nevyhnuli ani žáci úrovně B1, B2 a ojedinele ani C1. Chybovost genitivních koncovek *-a* a *-u* je vůči zvyšující se jazykové úrovni nepřímě úměrná (na rozdíl od případu zakončení *-ání* a *-aní*, kde

5 V případě chybných tvarů *světu*, *roka*, *stůla*, *diploma* byly frekvence a počet studentů upraveny, poněvadž se v rámci těchto typů vyskytly i jiné chyby (např. v rodě, tvaru plurálu), které byly vyloučeny na základě ruční analýzy.



nejčastěji chybují úrovně A1 a B2 (Zasina, 2019), a také na rozdíl od obtíží s přičes-tím minulým, kde nejčastěji chybují úrovně A1, A2 a B1 (Zasina, 2018, s. 289)). Z toho vyplývá, že během vyučovacího procesu je tato konkrétní obtíž nivelována s růstem jazykové kompetence žáka. Proto je třeba se v případech žáků na základních úrovních zaměřit na jasný výklad dané látky, zatímco v případech těch pokročilých bude spíše potřeba tento jazykový jev opakovat.

Úroveň	Frekvence	ipm ^[1]
A1	76	311,87
A2	68	272,73
B1	30	215,89
n/a	6	181,99
B2	8	92,3
C1	1	57,16

TABULKA 2: Úroveň studentů podle SERR, u nichž se vyskytla chyba v Gsg Mi.

[1] Vztaheno k jednotlivým úrovním.

V dalším kroku jsme se zaměřili na prozkoumání konkurence genitivních koncovek *-a* a *-u* v českém jazyce na základě dat synchronního korpusu SYN2015 (Křen et al., 2015, 2016). K získání relevantních slov jsme použili nástroj Morfio (Cvrček & Vondříčka, 2013a; Cvrček & Vondříčka, 2013b), který odhaduje rozsah a produktivitu slovotvorných modelů v češtině.⁶ Aplikace postihuje slovotvorný vztah⁷ jednak na základě formální shody/podobnosti v určitých částech slova (tzv. báze), jednak na základě formálních odlišností v částech specifických (tzv. formanty). Morfio tak dokáže vyhledat dvojice (popř. trojice, čtveřice) shodující se v bázi a lišící se pouze specifikovanými formanty. V naší analýze posloužil k nalezení morfologických variant.

V Morfiu jsme hledali dvojice slov se stejnouází a odlišnými koncovkami *-a* a *-u*. Specifikovali jsme rovněž dotaz vlastním tagem NNIS2.*⁸, abychom získali pouze tvary Gsg Mi. Frekvence byla pragmaticky stanovena na minimálně 4 výskyty pro obě varianty; vyhledávali jsme jednotlivé tvary a ty následně vyhodnocovali. Tímto jsme získali 123 dvojic, jež bylo nutné ještě manuálně protřídit. Nejčastěji jsme naráželi na problém životného substantiva, které v Gsg tvrdého vzoru má výhradní koncovku *-a*, chybně označovaného jako maskulina neživotného. Vždy se to týkalo jednoho lexému v páru (např. *Austina* 'jméno' × *Austina* 'město', *favorita* 'oblíbenec' × *favoritu*⁸ 'značka auta', *Hudsona* 'jméno' × *Hudsona* 'řeka'). Tyto výsledky byly z analýzy vy-

⁶ Nově i v polštině, srov. Zasina, 2017, a Zasina & Škrabal, 2018.

⁷ Vztah sémantický je těžce uchopitelný u korpusů, u nichž nemáme k dispozici sémantickou anotaci. Proto dále získaný materiál třídíme ručně, abychom homonymní tvary analyzovali jako samostatné instance.

⁸ V celé práci soustavně používáme malé počáteční písmeno pro značky aut, které se v mnoha případech taktéž vyskytují v originálních textech. Toto rozhodnutí je motivováno faktem, že použitá propria fungují v textu jako apelativa nahrazující lexém *auto*.

loučeny. K dalším vyloučeným případům patří lexémy formálně označené jako Gsg, které jsou však funkčně Asg (*Gsg jointa* : *Gsg jointu* ‘droga’), a také lexémy, jež nemají stejný etymologický původ (*turka* ‘turecká káva’ × *Turku* ‘toponymum’).



3. ANALÝZA

Výsledný materiál jsme setřídili do tří skupin po vzoru obdobného výzkumu na polštině (Zasina, 2017). Ve většině sporných případů, kde bylo určení kategorií nejasné, rozhodoval o zařazení do skupiny kontext. Navíc jsme si jasně stanovili charakteristiku jednotlivých kategorií takto:

1. dvojice substantivních tvarů se stejným etymologickým původem i významem; koncovky *-a* a *-u* se liší frekvencí a/nebo je jedna preferována v určitých kolokačních spojeních;
2. dvojice substantivních tvarů se stejným etymologickým původem, avšak rozdílným významem — tj. substantiv polysémních; koncovky *-a* a *-u* diferencují význam;
3. dvojice substantivních tvarů, u kterých je konkurence genitivních koncovek zpochybnitelná z důvodu vysoké chybovosti anotace nebo malého zastoupení jedné z variant v korpusu.

Následující tabulky 3, 5, 6 prezentují výše popsané tři skupiny substantiv Gsg Mi, s uvedeným významem, absolutní frekvencí výskytu v korpusu a proporcionální frekvencí.⁹ Jednotlivé lexémy jsou seřazeny podle abecedy.

První kategorie obsahuje celkem 102 dvojic, u nichž se koncovky *-a* a *-u* liší frekvencí. Nicméně absolutní frekvence zcela nevypovídá o převládající tendenci použití jedné z variant, proto uvádíme rovněž procentový podíl konkurujících tvarů (neboli proporcionální frekvenci — Bermel et al., 2014), který můžeme dále využít pro určení frekvenčních pásem (viz dále tabulka 4). Ta vypovídají o preferované dubletě, která by z hlediska didaktického měla být cizinci nabízena a upevňována během vyučovacího procesu. Pro češtinu škálu pro klasifikaci variant podrobně popsali Cvrček a Kodýtek (2013) a v této podobě ji zde používáme. Tato klasifikace je také respektována v MSČ1 (Cvrček et al., 2010).¹⁰ Skládá se celkově ze sedmi pásem: *nikdy nebo skoro nikdy* (v rozmezí 0–1 %), *zřídka* (1–10 %), *někdy* (10–35 %), *stejně* (35–65 %), *často* (65–90 %), *zpravidla* (90–99 %), *vždycky nebo skoro vždycky* (99–100 %). Tabulka 4 seskupuje lexémy první kategorie (řazené podle frekvence) na základě frekvenčních pásem:

Podíváme-li se na lexémy uvedené v tabulce 4 s ohledem na význam, zjistíme, že tu figuruje několik tematických skupin jako např. toponyma, značky aut či geometrické útvary. Na základě frekvenčních pásem lze jednoznačně konstatovat, že

⁹ Proportionální frekvence, tj. procentový podíl konkurujících tvarů (viz Bermel et al., 2014).

¹⁰ Podrobnější diskusi na téma dalších možných přístupů ve zkoumání frekvenčních pásem vedou Bermel et al. (2014, s. 218–221).



Číslo	Koncovka						Význam
	-a			-u			
	Lexém	Frekvence	Proportionální frekvence	Lexém	Frekvence	Proportionální frekvence	
1.	Androida	13	5,58 %	Androidu	220	94,42 %	operační systém
2.	antona	25	58,14 %	antonu	18	41,86 %	policejní vozidlo
3.	bavoráka	7	28,00 %	bavoráku	18	72,00 %	značka auta
4.	betléma	38	40,00 %	betlému	57	60,00 %	jesličky
5. ^[1]	Blanika	7	22,58 %	Blaníku	24	77,42 %	toponymum
6. ^[2]	Branika	19	70,37 %	Braníku	8	29,63 %	toponymum
7.	budíka	6	5,61 %	budíku	101	94,39 %	zařízení k buzení
8.	Cimburka	10	71,43 %	Cimburku	4	28,57 %	toponymum
9.	Čeljabinska	4	22,22 %	Čeljabinsku	14	77,78 %	toponymum
10.	čtvrtka	401	83,72 %	čtvrtku	78	16,28 %	den v týdnu
11.	debla	5	33,33 %	deblu	10	66,67 %	dvouhra
12.	dneška	1 028	97,35 %	dnešku	28	2,65 %	dnešní den
13.	dvorka	12	8,63 %	dvorku	127	91,37 %	malý dvůr
14.	fernet	8	24,24 %	fernetu	25	75,76 %	alkohol
15.	fiata	21	11,67 %	fiatu	159	88,33 %	značka auta
16.	forda	45	15,90 %	fordu	238	84,10 %	značka auta
17.	Frymburka	9	56,25 %	Frymburku	7	43,75 %	toponymum
18.	galavečera	59	93,65 %	galavečeru	4	6,35 %	slavnostní společenské setkání
19.	Gdaňska	30	41,10 %	Gdaňsku	43	58,90 %	toponymum
20.	golfa	10	3,83 %	golfu	251	96,17 %	značka auta
21.	góla	4	0,91 %	gólu	437	99,09 %	branka
22.	hřiba	5	16,13 %	hřibu	26	83,87 %	houba
23.	Chabarovska	4	14,81 %	Chabarovsku	23	85,19 %	toponymum
24.	chevroleta	6	12,24 %	chevroletu	43	87,76 %	značka auta
25.	chlívka	9	14,75 %	chlívku	52	85,25 %	chlév
26.	jaguara	6	16,22 %	jaguaru	31	83,78 %	značka auta
27.	Javorníka	5	20,83 %	Javorníku	19	79,17 %	toponymum
28.	jazyka	3 716	96,62 %	jazyku	130	3,38 %	mluva; orgán
29.	Jeseníka	5	3,03 %	Jeseníku	160	96,97 %	toponymum
30.	kalicha	54	43,9 %	kalichu	69	56,10 %	nádoba
31.	kláštera	1 597	97,68 %	klášteru	38	2,32 %	církevní budova
32.	klína	831	91,22 %	klínu	80	8,78 %	vnější část těla
33.	komína	476	72,12 %	komínu	184	27,88 %	zařízení odvádějící kouř
34.	kopečka	27	35,53 %	kopečku	49	64,47 %	malý kopec
35.	kostelíčka	14	70,00 %	kostelíčku	6	30,00 %	malý kostel
36.	kostelíka	99	77,34 %	kostelíku	29	22,66 %	malý kostel

Číslo	Koncovka						Význam
	-a			-u			
	Lexém	Frekvence	Proportionální frekvence	Lexém	Frekvence	Proportionální frekvence	
37.	kouska	50	16,84 %	kousku	247	83,16 %	část celku
38.	koutka	13	4,14 %	koutku	301	95,86 %	prostor mezi dvěma stěnami
39.	kouta	792	75,29 %	koutu	260	24,71 %	prostor mezi dvěma stěnami
40.	Kurska	19	67,86 %	Kursku	9	32,14 %	toponymum
41.	Kynšperka	14	60,87 %	Kynšperku	9	39,13 %	toponymum
42.	lesíka	103	77,44 %	lesíku	30	22,56 %	malý les
43.	lincolna	18	48,65 %	lincolnu	19	51,35 %	značka auta
44.	Lipníka	16	32,65 %	Lipníku	33	67,35 %	toponymum
45.	Mělníka	121	93,08 %	Mělníku	9	6,92 %	toponymum
46.	mercedesa	11	3,97 %	mercedesu	266	96,03 %	značka auta
47.	Minska	9	11,39 %	Minsku	70	88,61 %	toponymum
48.	mlýna	448	94,32 %	mlýnu	27	5,68 %	budova na výrobu mouky
49.	mnohoúhelníka	6	33,33 %	mnohoúhelníku	12	66,67 %	geometrický útvar
50.	národa	2 308	96,93 %	národu	73	3,07 %	společenství lidí
51.	Nazareta	23	56,10 %	Nazaretu	18	43,90 %	toponymum
52.	Norimberka	41	35,96 %	Norimberku	73	64,04 %	toponymum
53.	Nymburka	436	91,98 %	Nymburku	38	8,02 %	toponymum
54.	obdélníka	43	25,29 %	obdélníku	127	74,71 %	geometrický útvar
55.	oběda	818	98,79 %	obědu	10	1,21 %	polední jídlo
56.	octa	485	97,98 %	octu	10	2,02 %	potravinářská kapalina
57.	opela	9	7,89 %	opelu	105	92,11 %	značka auta
58.	Orlíka	5	26,32 %	Orlíku	14	73,68 %	toponymum
59.	Oseka	30	63,83 %	Oseku	17	36,17 %	toponymum
60.	pětiúhelníka	4	23,53 %	pětiúhelníku	13	76,47 %	geometrický útvar
61.	peugeota	8	6,78 %	peugeotu	110	93,22 %	značka auta
62.	popela	583	96,2 %	popelu	23	3,80 %	zbytek po spálení něčeho
63.	potoka	1 320	97,78 %	potoku	30	2,22 %	vodní tok
64.	Proseka	8	57,14 %	Proseku	6	42,86 %	toponymum
65.	Příbora	17	65,38 %	Příboru ^[3]	9	34,62 %	toponymum
66.	renaulta	13	9,22 %	renaultu	128	90,78 %	značka auta
67.	roka	841	1,21 %	roku	68 745	98,79 %	365 dní



Číslo	Koncovka						Význam
	-a			-u			
	Lexém	Frekvence	Proporcionální frekvence	Lexém	Frekvence	Proporcionální frekvence	
68.	Rumburka	17	26,56 %	Rumburku	47	73,44 %	toponymum
69.	rybníčka	16	14,29 %	rybníčku	96	85,71 %	malý rybník
70.	rybníka	928	87,05 %	rybníku	138	12,95 %	umělá vodní nádrž
71.	saaba	14	17,72 %	saabu	65	82,28 %	značka auta
72.	Saturna	5	6,1 %	Saturnu	77	93,90 %	planeta
73.	saturna	5	35,71 %	saturnu	9	64,29 %	značka auta
74.	silvestra	18	78,26 %	silvestru	5	21,74 %	poslední den v roce
75.	Smolenska	7	30,43 %	Smolensku	16	69,57 %	toponymum
76.	sna	226	17,44 %	snu	1 070	82,56 %	iluze během spánku
77.	Sokola	499	97,08 %	Sokolu	15	2,92 %	tělovýchovná jednota
78.	Suchdola	36	76,60 %	Suchdolu	11	23,40 %	toponymum
79.	sýra	712	88,12 %	sýru	96	11,88 %	jídlo
80.	šajna	13	68,42 %	šajnu	6	31,58 %	ponětí, zdání
81.	šestiúhelníka	7	26,92 %	šestiúhelníku	19	73,08 %	geometrický útvar
82.	Šternberka	365	94,32 %	Šternberku	22	5,68 %	toponymum
83.	Štramberka	19	61,29 %	Štramberku	12	38,71 %	toponymum
84.	Šumperka	119	76,28 %	Šumperku	37	23,72 %	toponymum
85.	Tábora	339	99,12 %	Táboru	3	0,88 %	toponymum
86.	tábora	2 366	99,75 %	táboru	6	0,25 %	ubytovací zařízení
87.	taxíka ^[4]	5	1,31 %	taxíku	378	98,69 %	dopravní prostředek
88.	trabanta	23	62,16 %	trabantu	14	37,84 %	značka auta
89.	trojúhelníka	75	23,36 %	trojúhelníku	246	76,64 %	geometrický útvar
90.	týla	163	68,31 %	týlu	28	31,69 %	část hlavy
91. ^[5]	Újezda	33	45,21 %	Újezdu	40	54,79 %	toponymum
92.	Úštěka	5	38,46 %	Úštěku	8	61,54 %	toponymum
93.	velína	12	35,29 %	velínu	22	64,71 %	ovládací stanoviště
94.	Vimperka	23	69,70 %	Vimperku	10	30,30 %	toponymum
95.	vnějška	4	2,76 %	vnějšku	141	97,24 %	vnější prostor
96.	všehomíra	71	92,21 %	všehomíru	6	7,79 %	vesmír
97.	walkmana	31	68,89 %	walkmanu	14	31,11 %	přehrávač kazet



Číslo	Koncovka						Význam
	-a			-u			
	Lexém	Frekvence	Proporcionální frekvence	Lexém	Frekvence	Proporcionální frekvence	
98.	Zábřeha	11	15,28 %	Zábřehu	61	84,72 %	toponymum
99.	záchoda	53	14,10 %	záchodu	323	85,90 %	toaleta
100.	zákona	10 607	99,48 %	zákonu	55	0,52 %	ustálené pravidlo
101.	Zbiroha	16	64,00 %	Zbirohu	9	36,00 %	toponymum
102.	zítřka	433	98,19 %	zítřku	8	1,81 %	zítřejší den

TABULKA 3: Dvojice substantivních tvarů mající stejný etymologický původ a význam (kategorie 1).

- [1] Frekvence je zde ručně ověřena, upravena a významově desambiguována.
 [2] Frekvence je zde ručně ověřena, upravena a významově desambiguována.
 [3] Frekvence je zde ručně ověřena a upravena.
 [4] Frekvence je zde ručně ověřena a upravena.
 [5] Frekvence je zde ručně ověřena a upravena.

geometrické útvary mají zpravidla koncovku *-u* (srov. *mnohoúhelníku*, *šestiúhelníku*, *obdélníku*, *pětiúhelníku*, *trojúhelníku*). Rovněž značky aut mají větší tendenci vyskytovat se s koncovkou *-u*, nicméně existují případy, kdy obě koncovky mají víceméně stejné postavení (*trabant*, *lincoln*, *saturn*).¹¹ U toponym však není situace jednoznačná: vyskytují se jak lexémy s genitivní koncovkou *-a* (*Tábor*, *Šternberk*, *Mělník*, *Nymburk*), tak s *-u* (*Jeseník*), kolísání mezi jednou a druhou koncovkou je však vysoké. MSC₁ (Cvrček et al., 2010, s. 165) udává, že:

[k]oncovku *-a* mají v Gsg. vždy nebo zpravidla názvy měst a obcí zakončené na *-ov* (*Prostějov*, *Mnichov* atd.) a řada dalších: *Zlín*, *Londýn*, *Temelín*, *Berlín*, *Řím* atd., je i *Egypt* — *Egypta*. U názvů českých a moravských míst záleží i na místním úzu a jeho rozšíření: jen do *Zlína*, většinou z *Mělníka*, zřídka i z *Mělníku*, jen do *Českého Dubu*, ale místně do *Duba*.

Internetová jazyková příručka¹² popisuje situaci obdobně: u vlastních jmen místních zakončených na *-ov* nebo *-ín* se používá koncovka *-a* (např. *Benešova*, *Lvova*, *Kojetína*, *Londýna*). Navíc komentuje, že v případě „nevymezitelného množství“ substantiv dochází k uplatnění obou koncovek (např. *Fulneku/Fulneka*, *Mělníka/Mělníku*, *Lipníka/Lipníku*, *Příbora/Příboru*, *Zábřehu/Zábřeha*). Dále tvrdí, že uplatnění obou koncovek je možné u některých jmen na *-ík* (např. u geometrických termínů: *-úhelník*, *-běžník*, *-dělník*; a také dalších jmen: *bochník*, *budík*, *holubník*). Korpusová data však ukazují, že

¹¹ Může to také souviset s fakultativní životností, která je u standardních názvů značek aut (např. *peugeot*, *mercedes*) řídkým jevem, častá je naopak u substandardních tvarů (např. *ba-voorák*; viz Bermel, 2011, s. 520).

¹² <http://prirucka.ujc.cas.cz/?id=221&dotaz=koncovka#nadpis5> (cit. 5. 3. 2019)



Frekvenční pásma	-a	-u
100-99 %	tábor, zákon, Tábor	gól
99-90 %	oběd, zítřek, ocet, potok, klášter, dnešek, Sokol, národ, jazyk, popel, mlýn, Šternberk, galavečer, Mělník, všehomír, Nymburk, klín	rok, taxík, vnějšek, Jeseník, golf, mercedes, koutek, Android, budík, Saturn, peugeot, opel, dvorek, renault
90-65 %	sýr, rybník, čtvrtek, silvestr, lesík, kostelík, Suchdol, Šumperk, kout, komín, Cimburk, Braník, kostelíček, Vimperk, walkman, šajn, týl, Kursk, Příbor	Minsk, fiat, chevrolet, záchod, rybníček, chlívek, Chabrovsk, Zábřeh, ford, hřib, jaguar, kousek, sen, saab, Javorník, Čeljabinsk, Blaník, trojúhelník, pětiúhelník, fernet, obdélník, Orlík, Rumburk, šestiúhelník, bavorák, Smolensk, Lipník, debl, mnohoúhelník
65-35 %	Zbiroh, Osek, trabant, Štramberk, Kynšperk, anton, Prosek, Frymburk, Nazaret, lincoln, Újezd, kalich, Gdaňsk, betlém, Úštěk, Norimberk, saturn, kopeček, velín	velín, kopeček, saturn, Norimberk, Úštěk, betlém, Gdaňsk, kalich, Újezd, lincoln, Nazaret, Frymburk, Prosek, anton, Kynšperk, Štramberk, trabant, Osek, Zbiroh
35-10 %	mnohoúhelník, debl, Lipník, Smolensk, bavorák, šestiúhelník, Rumburk, Orlík, obdélník, fernet, pětiúhelník, trojúhelník, Blaník, Čljabinsk, Javorník, saab, sen, kousek, jaguar, hřib, ford, Zábřeh, Chabrovsk, chlívek, rybníček, záchod, chevrolet, fiat, Minsk	Příbor, Kursk, týl, šajn, walkman, Vimperk, kostelíček, Braník, Cimburk, komín, kout, Šumperk, Suchdol, kostelík, lesík, silvestr, čtvrtek, rybník, sýr
10%	renault, dvorek, opel, peugeot, Saturn, budík, Android, koutek, mercedes, golf, Jeseník, vnějšek, taxík, rok	klín, Nymburk, všehomír, Mělník, galavečer, Šternberk, mlýn, popel, jazyk, národ Sokol, dnešek, klášter, potok, ocet, zítřek, oběd
1-0 %	gól	Tábor, zákon, tábor

TABULKA 4: Frekvenční pásma lexémů první kategorie.

geometrické útvary ve skutečnosti projevují větší tendenci vyskytovat se s koncovkou *-u*.¹³ Stejná tendence převládá u ostatních lexémů zakončených na *-ík*.¹⁴ Existují však rovněž frekventovaná substantiva s preferencí koncovky *-a*, např: *rybník, Mělník, Rakovník, lesík, kostelík*.

Analýza korpusových dat z hlediska zakončení substantiva a obdobný přístup Divjakové (2018) pro polštinu nás může inspirovat k dalším výzkumným otázkám,

¹³ Z 664 výskytů pro dotaz [lemma="(?)i.+((úhelník)|(běžník)|(délník))" & tag="N.S2.*"] má 497 tokenů koncovku *-u*.

¹⁴ Z 32 349 výskytů pro dotaz [lemma="(?)i.+ík" & tag="N.IS2.*"] má 30 714 tokenů koncovku *-u*.



Číslo	Koncovka					
	-a			-u		
	Lexém	Frekvence	Význam	Lexém	Frekvence	Význam
1.	Betléma ^[1]	113	toponymum	betlému	57	jesličky
2. ^[2]	Blaníka	7	toponymum	Blaníku	4	opera
3.	Braníka	19	toponymum	Braníku	3	pivo; pivovar
4.	klína	831	vnější část těla	klínu ^[3]	64	kus dřeva; předmět, který ho připomíná
5. ^[4]	oriona	8	motocykl	Orionu	21	souhvězdí
6.	prostředka	17	střed	prostředku	566	předmět; opatření
7.	světa	28 571	zeměkoule, vesmír	Světu	6	rybník
8. ^[5]	týla	163	část hlavy	týlu	49	prostor za frontou
9. ^[6]	Újezda	33	toponymum	újezdu	40	vojenský újezd

TABULKA 5: Dvojice substantivních tvarů mající stejný etymologický původ, avšak rozdílný význam (kategorie 2).

^[1] Frekvence je zde ručně ověřena, upravena a významově desambiguována.

^[2] Frekvence je zde ručně ověřena, upravena a významově desambiguována.

^[3] Frekvence je zde ručně ověřena, upravena a významově desambiguována.

^[4] Frekvence je zde ručně ověřena, upravena a významově desambiguována.

^[5] Frekvence je zde ručně ověřena, upravena a významově desambiguována.

^[6] Frekvence je zde ručně ověřena, upravena a významově desambiguována.

např. zda zakončení substantiv na specifickou kombinaci grafémů mohou mít vliv na tendenci slova vyskytovat se častěji nebo výhradně s konkrétní koncovkou Gsg.

Pro neživotná maskulina v korpusu SYN2015 najdeme pro koncovky *-a* a *-u* celkově 1 451 333 tokenů, z čehož 192 177 má koncovku *-a* a 1 259 156 koncovku *-u*. Pro obě koncovky jsme vytvořili retrográdní seznamy lemmat a seřadili substantiva podle posledních dvou grafémů. V pilotní analýze jsme se soustředili pouze na široké spektrum zakončení pro genitivní koncovku *-a*.¹⁵ Rozhodnutí bylo ovlivněno vysokým počtem substantiv s koncovkou *-u* přesahujícím 13 tisíc lemmat, což je pro ruční třídnění výsledků příliš časově náročné. Identifikovali jsme celkově 73 různých zakončení a následně jsme ověřovali, zda se vyskytují také v seznamu pro koncovku *-u*. Naprostou většinu z nich najdeme v obou seznamech. Pouze pro 6 zakončení, zastoupených vždy jen jedním lemmatem, je specifická koncovka *-a*: *-év* (*chlév*), *-řt* (*buřt*), *-ýs* (*Brandýs*), *-jr* (*sejr*), *-ím* (*Řím*), *-ěb* (*chléb*), zakončení *-ow* je reprezentováno čtyřmi lexémy — *Glasgow*, *salchow*, *Seelow*, *Krakow*.

15 Analýza substantiv z hlediska kombinatoriky posledních grafémů je zajímavá, nicméně není jádrem této studie, proto zde uvádíme jen obecná zjištění pilotní analýzy.



Na základě těchto výsledků tedy nelze jednoznačně říci, že některá z koncovek má tendenci vyskytovat se s určitým zakončením. Jak však bylo naznačeno již výše, určitou preferenci koncovky můžeme sledovat na základě sémantiky. Proto mají například substantiva zakončená na *-ýn* a *-ín* tendenci vyskytovat se s koncovkou *-a*, pokud jde o toponyma (*Londýn, Rýn, Hlubětín, Zlín*), v případě zakončení *-on* se vyskytuje u toponym naopak koncovka *-u* (*Boston, Brighton*). U substantiv se zakončením na *-in, -en, -ol* a *-id* má koncovka *-u* tendenci vyskytovat se u názvů chemických látek (*pervitin, benzen, etanol, oxid*). Na podrobnější sémantickou analýzu zde vzhledem k primárnímu zaměření práce na výuku bohužel není prostor.

Druhá kategorie obsahuje celkově 9 dvojic, u nichž koncovky *-a* a *-u* rozlišují význam. Konkurence obou koncovek umožnila sémantickou diferenciaci lexémů se stejným etymologickým původem. Nejedná se tu však o konkurenci (z hlediska synchronního) jako v první kategorii. Bermel a Knittl (2012b) tuto záležitost podrobně nekomentují a zahrnují substantiva *betlém, klín, prostředek, tyl* k běžné konkurenci variant, prezentované zde v první kategorii. To by mohlo naznačovat, že považují např. varianty *prostředka, střed' × prostředku* ‚předmět‘ za dublety, ačkoli se zde jedná v podstatě o dva různé lexémy.¹⁶ Proto se domníváme, že uvádění proporcionální frekvence konkurujících si tvarů je v takovýchto případech zbytečné: z hlediska synchronního o skutečnou konkurenci nejde a oba tvary jsou rovnocenné. Naše analýza, kombinující kvantitativní i kvalitativní přístup, tak představuje doplněk k dřívějším výzkumům v této oblasti (Bermel & Knittl, 2012a, 2012b; Bermel et al., 2014, 2018).

Z hlediska jazykové výuky jsou zjištěné informace podstatné zvláště u frekventovaných apelativ *klín a prostředek*, kde si žák musí uvědomit, že zvolená koncovka Gsg diferencuje význam lexému. Třída těchto slov není obsáhlá, tudíž by jejich zapamatování nemělo být překážkou. Nicméně je důležité, aby žáci na vyšší úrovni znalosti češtiny tyto rozdíly znali.

Třetí kategorie obsahuje celkem 8 dvojic, u nichž konkurence není průkazná z důvodu vysoké chybovosti nebo malého zastoupení jedné z variant v korpusu (max. 2 výskyty). Ve většině případů byla frekvence upravena tak, aby skutečně reprezentovala pouze Gsg Mi. Tabulka 6 obsahuje jak dvojice se stejnou sémantikou, tak lexémy rozlišující význam — vzhledem k marginálnosti tohoto jevu nepovažujeme za nutné obě skupiny rozlišovat. V případě této kategorie se tedy jedná spíše o náznak potenciální konkurence v češtině, byť může jít i o nerespektování většinového jazykového úzu. Z těchto dat nemůžeme vyvodit hlubší závěry. U lexémů, u nichž užití genitivní koncovky *a-* nebo *u-* nepředstavuje změnu významu (*únor, večer*), bychom ve výuce měli za standardní varianty považovat tvary častější a jen s takovými žáky seznamovat. Značky aut (*orion, passat, volkswagen*), které obecně nepatří k jazykovému jádru, není nutné ve výuce zohledňovat.

¹⁶ Bermel a Knittl (2012b, s. 94–95) však komentují možný vliv syntaktických konstrukcí na volbu určité koncovky. Z jejich analýzy přijatelnosti konkurenčních koncovek rodilými mluvčími vyplývá, že v otázkách, v nichž se daný lexém bezprostředně vyskytuje po frekventované předložce, mají minoritní koncovky větší šanci získat lepší hodnocení přijatelnosti. Na druhé straně však v případě konstrukcí tvořených méně častými slovními spojeními mají vyšší přijatelnost koncovky běžné (ibid., s. 114).

Číslo	Koncovka					
	-a			-u		
	Lexém	Frekvence	Význam	Lexém	Frekvence	Význam
1.	klína	831	vnější část těla	klínu	1	vnitřní část rodidel
2.	oriona	8	motocykl	orionu	2	motocykl
3.	passata	1	značka auta	passatu	33	značka auta
4.	Saturna	5	planeta	Saturnu	2	fotbalový klub
5.	saturna	5	značka auta	Saturnu	2	fotbalový klub
6.	února	4165	druhý měsíc	únoru	2	druhý měsíc
7.	večera	4485	část dne	večeru	2	část dne
8.	volkswagena	1	značka auta	volkswagenu	256	značka auta

TABULKA 6: Dvojice substantivních tvarů s nedostatečně průkaznou konkurencí koncovek Gsg (kategorie 3).

Rozdělení substantiv Gsg Mi do tří kategorií zvýraznilo rozdílné pozadí, na němž je třeba zkoumat konkurenční substantivní koncovky *-a* a *-u*. Buď se jedná o významově stejná slova (žáci by se měli při výběru vhodné koncovky řídit četností použití), anebo koncovka diferencuje význam slova (žáci by si sémantické rozdíly měli uvědomovat). Analyzovaná slova první kategorie projevují určité tendence k volbě genitivní koncovky s ohledem na tematickou skupinu, např. geometrické útvary a značky automobilů se častěji vyskytují s koncovkou *-u*. Tato koncovka je obecně v dnešní češtině u neživotných podstatných jmen častější. Nicméně existují i frekventovaná slova s koncovkou *-a* (*života, světa, zákona*), která si nerodilí mluvčí musejí zapamatovat.

4. KORPUSOVÁ CVIČENÍ

Korpusový přístup ve výuce se zakládá na induktivní metodě analýzy dat, která vede žáky k vyvození závěrů o obecných pravidelnostech daného jazykového jevu. Korpusová cvičení jsou totiž sestavena takovým způsobem, aby naváděla žáky k určitým výsledkům. Samostatná analýza dat napomáhá si novou nebo opakovanou látku mnohem lépe zapamatovat, poněvadž není předložena učitelem, ale žák musí závěry formulovat na základě zobrazení sám (nicméně asistence učitele je pořád žádoucí). Vzhledem k tomu, že tento typ práce vyžaduje od žáků větší úsilí a zároveň více času na samostatné zpracování úkolu, není v plném rozsahu vhodný — i když je stále aplikovatelný v upravené podobě — ve výuce upřednostňující kvantitu probraného materiálu v omezeném časovém rozmezí, tj. v intenzivních kurzech češtiny, standardních jazykových kurzech nebo letních jazykových kurzech (letních školách). Žáci zde nejsou zvyklí na induktivní způsob práce a očekávají, že dané informace získají od učitele. Proto je tato metoda vhodná ve vysokoškolské výuce, kde je možné věnovat více času důkladným analýzám, vyučující nejsou omezeni rámcem vyučovacího programu a navíc se klade důraz nejen na probrání látky, ale i její důkladné upevnění a rozšíření.



V této kapitole prezentujeme cvičení, jež jsou zacílena na bohemisty slovanského původu. Na základě poznatků zjištěných korpusovou analýzou můžeme sestavit korpusová cvičení. Zde uvádíme tři příklady cvičení, dvě přímá a jedno nepřímé:¹⁷ studenti jsou nejprve uvedeni do problematiky konkurenčních koncovek *-a* a *-u* (cvičení 1) a následně procvičují získané znalosti (cvičení 2 a 3).

První korpusové cvičení je založeno na vyhledávání správného tvaru Gsg v korpusu SYN2015. Žáci ze začátku nevědí, u kterých slov dochází ke konkurenci obou koncovek, proto se v tomto úvodním cvičení soustředíme čistě na volbu správné koncovky u vybraných slov. Zvolili jsme 15 nejčastějších neživotných maskulin, která student musí najít v korpusu ve spojení s předložkou, jež se pojí s genitivem. Tím se vyhneme složitějšímu dotazu v CQL. Tento typ cvičení je vhodný již pro studenty od úrovně A2.

Cvičení 1: Pomocí rozhraní KonText v korpusu SYN2015 vyberte typ dotazu *Fráze* a pokuste se najít správnou koncovku Gsg následujících substantiv ve spojení s nejběžnější předložkou:

- a) od rok[au]
- b) do život[au]
- c) do svět[au]
- d) od čas[au]
- e) do dom[au]
- f) do systém[au]
- g) podle typ[au]
- h) podle zákon[au]
- i) do projekt[au]
- j) od stát[au]
- k) u stol[au]
- l) kolem milion[au]
- m) do program[au]
- n) do prostor[au]
- o) do klub[au]

Na základě výsledků z cvičení 1 si žák asociuje frekventovaná substantiva s preferovanou koncovkou a zároveň si uvědomí, že v Gsg je častější koncovka *-u*. Celková proporce koncovky *-a* ku *-u* v korpusu SYN2015 činí 13,24 % ku 86,76 % (v přepočtu na tokeny). Následně můžeme studenty informovat, že v některých případech dochází ke konkurenci mezi oběma koncovkami, a přejít k cvičení 2.

Cvičení 2 rovněž vyžaduje přímé zapojení studenta do hledání v korpusu: studenti musejí porovnat v přepočtu na procenta dvě varianty pro jednotlivé lexémy a identifikovat tu častější. V rozhraní KonText (Machálek & Křen, 2013) bychom museli

¹⁷ Korpusový přístup ve výuce tradičně dělíme na dvě odvětví podle metody práce (Leech, 1997), která je buď založena na práci s počítačem — přímý přístup, nebo na práci s předem upraveným korpusovým materiálem ve formě vytištěných cvičení — nepřímý přístup.



použít CQL dotaz, který však může být pro nezkušené žáky náročný. Cvičení lze však provést také pomocí nástroje SyD (Cvrček & Vondříčka, 2011a; Cvrček & Vondříčka, 2011b) bez nutné znalosti dotazování v CQL — stačí, když žák v nástroji SyD zadá dvě varianty a podívá se na výsledky v tabulce s morfologickými kategoriemi v psaném jazyce (viz tabulka 7), jež se nachází v záložce *souhrn*. Je však třeba studenty upozornit, že prezentované grafy nepodávají vždy zcela přesnou informaci o konkurenci, poněvadž ve statistice mohou být zahrnuty i tvary jiných pádů (např. lokálu, který je s genitivem často homonymní).

gram. kategorie	[1]		[2]	
celkem	100 %	14 525	100 %	1042
Podst. jm.: gen. jedn. č., rod muž. neživ.	98,61 %	(14 323)	3,26 %	(34)
Podst. jm.: dat. jedn. č., rod muž. neživ.	0,00 %	(0)	69,87 %	(728)
Podst. jm.: lok. jedn. č., rod muž. neživ.	0,00 %	(0)	24,95 %	(260)
Podst. jm.: gen. jedn. č., rod muž. živ.	1,28 %	(186)	0,00 %	(0)
Podst. jm.: dat. jedn. č., rod muž. živ.	0,00 %	(0)	0,96 %	(10)
Podst. jm.: ak. jedn. č., rod muž. živ.	0,11 %	(16)	0,00 %	(0)
Podst. jm.: lok. jedn. č., rod muž. živ.	0,00 %	(0)	0,96 %	(10)

TABULKA 7: Distribuce morfologických kategorií v psaném jazyce (podle nástroje SyD) na příkladu tvarů *zákona* vs. *zákonu*.

Stávající verze SyDu obsahuje data z korpusu SYN2010, proto frekvence výskytů neodpovídají našim datům z novějšího korpusu SYN2015 prezentovaným výše v tabulkách, nicméně ukazují stejné tendence používání. Ve cvičení 2 má žák do tabulky jednoduše doplnit informace týkající se frekvence výskytů jednotlivých variant a přepočítat čísla na procenta, aby zjistil, která varianta je častější. Určení frekvenčního pásma má pouze napomoci žákovi posoudit vhodnost dané varianty. Frekvenční pásmo určí na základě tabulky frekvenčních pásem připojené k cvičení. Díky těmto informacím bude schopen na základě dat samostatně usoudit, které koncovky by měl používat. Cvičení je doporučeno samostatným uživatelům češtiny od úrovně B1.

Cvičení 2: Porovnejte v nástroji SyD (<https://syd.korpus.cz/>) konkurenční koncovky genitivu substantiv uvedených v tabulce. Doplněte informace o frekvenci (viz tabulka morfologických kategorií), procentuálním zastoupení obou variant a frekvenčním pásmu.



Substantivum	Frekvence		Procento		Frekvenční pásmo	
	-a	-u	-a	-u	-a	-u
zákon						
jazyk						
sýr						
rybník						
trabant						
kopeček						
trojúhelník						
kousek						
taxík						
gól						

Frekvenční pásma

Procento	0-1	1-10	10-35	35-65	65-90	90-99	99-100
Pásmo	nikdy	zřídka	Někdy	stejně	často	zpravidla	vždycky

Poslední cvičení nevyžaduje přímou práci s počítačem a je založeno na informacích uvedených v tabulce 5 se substantivy, u nichž rozdílná koncovka genitivu diferencuje význam. V cvičení 3 má student za úkol doplnit do vět chybějící substantivum se správnou genitivní koncovkou s ohledem na význam věty. Věty obsažené v zadání pocházejí z korpusu současné psané češtiny SYN2015. Díky tomuto přístupu má student možnost seznámit se s nejčastějšími neživotnými maskuliny tohoto typu v jejich přirozeném kontextu. Je však třeba žáky upozornit, že se v bodě c) setkají s grafickou podobou z *prostředka*, která se běžně vyskytuje i jako adverbialní spřežka *zprostředka*. Jelikož se celkově jedná o řidší jev, je toto cvičení doporučováno pro studenty minimálně na úrovni B2.

Cvičení 3: Doplňte do níže uvedených vět správné tvary genitivu singuláru následujících lexémů:

a) betlém/Betlém

- Na samotné Vánoce by do třicetitisícového _____ mělo podle starosty Batáry dorazit až pětadvacet tisíc lidí.
- Ponocný a pekař budou součástí nového dřevěného _____, který se stane vánoční ozdobou poutního kostela.

b) klín

- Knihu si položil do _____ a rychle naplnil sklenici.
- Podobné zvuky už slyšel — po bitvě, když ranění muži zatínali zuby do stočeného hadru nebo dřevěného _____, zatímco jim rovnali kosti.

c) prostředek

- Na první pohled šílený nápad podivného dopravního _____ sdílí celá řada konstruktérů po celém světě.
- Musela jsem dostat auto z _____ ulice; stála jsem v neosvětlené zatáčce a byla tma.

Tato cvičení jsou příkladem aplikace korpusového přístupu ve vyučovacích hodinách češtiny. První dvě cvičení na sebe navazují a snaží se postupně poukazovat na problémy s koncovkami Gsg Mi, třetí cvičení je příkladem aktivity pro pokročilejší žáky. Vhodnost cvičení pro danou skupinu žáků by měl vždy určit učitel. Zde úroveň doporučujeme na základě vlastní zkušenosti s prací se Slovaný. Tato zadání jsou rovněž přiměřená jako opakování a upevňování dříve probrané problematiky Gsg Mi.

5. ZÁVĚR

Problematicke konkurenčních koncovek Gsg Mi bylo již věnováno mnoho pozornosti, dosud však nebyl brán zřetel na popis tohoto jevu ve výuce ČJ pomocí korpusových metod. Jako základ k této studii jsme použili analýzu žákovského korpusu, v němž byly problémy cizinců s Gsg Mi identifikovány. Zjistili jsme, že chybovost u genitivních koncovek *-a* a *-u* se s jazykovou úrovní snižuje. Obecně také dochází k nahrazování správné koncovky *-u* koncovkou *-a*. V další fázi byla na základě dat ze synchronního korpusu češtiny variantní substantiva roztržena do tří kategorií: dvojice substantivních tvarů mající stejný etymologický původ i význam, dvojice substantivních tvarů mající stejný etymologický původ, avšak rozdílný význam a dvojice substantivních tvarů s neprůkaznou konkurencí genitivních koncovek (danou nedostatkem dat). Toto rozdělení systematicky třídí substantiva, u nichž si buď koncovky ze synchronního hlediska vzájemně konkurují, nebo se ustálily a nesou konkrétní význam. Takové rozdělení s uvedením proporcionální a absolutní frekvence je zdrojem informací, která koncovka je frekventovaněji volena rodilými mluvčími. Nadto je zdrojem pro sestavení vhodných korpusových cvičení, jejichž tři příklady zde prezentujeme. Dvě přímá cvičení se zabývají konkurenčními koncovkami a žák díky nim identifikuje častější variantu, zatímco nepřímé cvičení upozorňuje žáka na možné sémantické rozdíly při použití různých koncovek.

Tato studie přistupuje k datům jak kvantitativně, tak i kvalitativně a je tedy komplementární k dřívějším korpusovým studiím (srov. Bermel & Knittl, 2012a, 2012b), zvláště pak v oblasti jazykové výuky. Zde prezentovaný korpusový přístup se skládá ze tří částí: identifikace problematického jevu pomocí žákovského korpusu, klasifikace tohoto jevu pomocí korpusu synchronního a konečně generalizace výsledků. Tento postup systemizuje práci s korpusem pro účely výuky a zároveň může být pokynem a inspirací pro pedagogy, jak s korpusem pracovat.

V rámci substantiv se stejným etymologickým původem a významem korpusová data rovněž potvrdila určité sémantické preference dané koncovky. Substantiva zakončená na *-ýn* a *-ín* mají tendenci vyskytovat se s koncovkou *-a*, pokud jde o toponyma, v případě zakončení *-on* u toponym naopak převažuje koncovka *-u*. Sub-



stantiva se zakončením na *-in*, *-en*, *-ol* a *-id* mají tendenci vyskytovat se s koncovkou *-u* u názvů chemických látek. Dále je koncovka *-u* preferována u názvů geometrických útvarů a značek aut.

Hlavním cílem bylo prozkoumat variantní tvary Gsg Mi a nabídnout jednotný korpusový přístup ve výuce této problematice látky. V této oblasti je nicméně zapotřebí podrobnějšího výzkumu sémantických rozdílů konkurenčních genitivních koncovek, který by měl být proveden v samostatné studii. Nové poznatky týkající se významových rozdílů by dozajista přispěly i ke zkvalitnění výuky ČČJ.

LITERATURA

- Bermel, N. (2010). Variace a frekvence variant na příkladu tvrdých neživotných maskulin. In S. Čmejrková, J. Hoffmannová, & E. Havlová (Eds.), *Užívání a prožívání jazyka* (s. 135–140), Praha: Karolinum.
- Bermel, N. (2011). Variantnost ve flexi tvrdých neživotných maskulin. In F. Štícha (Ed.), *Kapitoly z české gramatiky* (s. 509–522). Praha: Academia.
- Bermel, N., & Knittl, L. (2012a). Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory*, 8, 241–275.
- Bermel, N., & Knittl, L. (2012b). Morphosyntactic variation and syntactic constructions in Czech nominal declension: corpus frequency and native-speaker judgments. *Russian linguistics*, 36(1), 91–119.
- Bermel, N., Knittl, L., & Russell, J. (2014). Absolutní a proporcionální frekvence v ČNK ve světle výzkumu morfosyntaktické variace v češtině. *Naše řeč*, 97(4–5), 216–227.
- Bermel, N., Knittl, L., & Russell, J. (2018). Frequency data from corpora partially explain native-speaker ratings and choices in overabundant paradigm cells. *Corpus Linguistics and Linguistic Theory*, 14(2), 197–231.
- Bigl, R. (2013). *Vývoj lužickosrbského skloňování a stupňování*. Praha: Karolinum.
- Cvrček, V., & Kodýtek, V. (2013). Ke klasifikaci morfologických variant. *Slovo a slovesnost*, 74(2), 139–146.
- Cvrček, V., Kodýtek, V., Kopřivová, M., Kovářiková, D., Sgall, P., Šulc, M., ... Waclawičová, M. (2010). *Mluvnice současné češtiny 1. Jak se píše a jak se mluví*. Praha: Karolinum.
- Cvrček, V., & Vondříčka, P. (2011a). SyD — Korpusový průzkum variant (Version 2.2) [Corpus tool]. Dostupné z <http://syd.korpus.cz>
- Cvrček, V., & Vondříčka, P. (2011b). Výzkum variability v korpusech češtiny. In F. Čermák (Ed.), *Korpusová lingvistika Praha 2011. 2. Výzkum a výstavba korpusů* (s. 184–195). Praha: Nakladatelství Lidové noviny.
- Cvrček, V., & Vondříčka, P. (2013a). Morfio [Corpus tool]. Dostupné z <https://morfio.korpus.cz/>
- Cvrček, V., & Vondříčka, P. (2013b). Nástroj pro slovotvornou analýzu jazykového korpusu. *Gramatika a korpus 2012*. Hradec Králové: Gaudeamus. Dostupné z http://www.ujc.cas.cz/miranda2/export/sitesavcr/ujc/veda-vyzkum/vyzkum/gramatika-a-korpus/proceedings-2012/konferencni-prispevky/CvrcekVaclav_VondrickaPavel.pdf
- Divjak, D. (2018). *How machines that learn language can help language instructors teach language: a case study on the genitive singular of masculine inanimate nouns in Polish*. Paper presented at the conference Nowości w glottodydaktyce polonistycznej, Warszawa, Poland.
- Gráf, T. (2019). Teorie osvojování jazyka a korpusový výzkum. In K. Šormová & K. Šebesta (Eds.), *Korpusy v jazykovém vyučování* (s. 16–33). Praha: Filozofická Fakulta Univerzity Karlovy.
- Janda, L. A. (1996). *Back From The Brink*. München: LINCOM Europa.

- Karlík, P., Nekula, M., & Rusínová, Z. (Eds.). (1995). *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.
- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., ... Zasina, A. (2015). *SYN2015: reprezentativní korpus psané češtiny*. Praha: Ústav Českého národního korpusu FF UK. Dostupné z <http://www.korpus.cz>
- Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., ... Zasina, A. J. (2016). SYN2015: Representative Corpus of Contemporary Written Czech. In N. Calzolari et al. (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (s. 2522–2528). Portorož: ELRA.
- Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora* (s. 11–23). Harlow: Addison Wesley Longman.
- Machálek, T., & Křen, M. (2013). Query interface for diverse corpus types. In K. Gajdošová & A. Žáková (Eds.), *Natural Language Processing, Corpus Linguistics, E-learning* (s. 166–173). Lüdenscheid: RAM Verlag.
- Petkevič, V., Rosen, A., Štindlová, B., Jelínek, T., Hnátková, M., & Jäger, P. (2012). Anotace chybových textů v českém žákovském korpusu. In K. Šebesta & S. Škodová (Eds.), *Čeština — cílový jazyk a korpusy* (s. 61–87). Liberec: Technická univerzita v Liberci.
- Rusínová, Z. (1992). Některé aspekty distribuce alomorfů: (genitiv a lokál sg. maskulin). In V. Podborský (Ed.), *Sborník prací Filozofické fakulty brněnské univerzity. A, Řada jazykovědná XLI* (s. 23–31). Brno: Masaryková univerzita v Brně.
- Šebesta, K., Bedřichová, Z., Šormová, K., Štindlová, B., Hrdlička, M., Hrdličková, M., ... Rosen, A. (2014). *CzeSL-SGT: korpus češtiny nerodilých mluvčích s automaticky provedenou anotací, verze 2 z 28. 7. 2014*. Praha: Ústav Českého národního korpusu FF UK. Dostupné z <http://www.korpus.cz>
- Šimandl, J. (2003). Od čvrtku do pátku. *Naše řeč*, 86(3), 161–164.
- Vážný, V. (1964). *Historická mluvnice česká. 2. Tvarosloví*. Praha: Státní pedagogické nakladatelství.
- Zasina, A. J. (2017). Konkurence koncovek -a a -u v genitivu singuláru neživotných maskulin v polštině. In M. Stluka & M. Škrabal (Eds.), *Lifka a czban — Sborník příspěvků k 70. narozeninám prof. Karla Kučery* (s. 90–98). Praha: Nakladatelství Lidové noviny.
- Zasina, A. J. (2018). O problémech nerodilých mluvčích s kvantitou na základě analýzy korpusových dat. In S. Škodová, & M. Hrdlička (Eds.), *Čeština jako cizí jazyk v průsečíku pohledů* (s. 281–298). Praha: Univerzita Karlova, Filozofická fakulta
- Zasina, A. J. (2019). *Dlouhá, nebo krátká? Verbální substantiva na -ání a -aní ve výuce cizinců*. Manuscript in preparation.
- Zasina, A. J., & Škrabal, M. (2018). *Morfio.pl — the possibilities for the application of Czech corpus tools to other languages*. Paper presented at the SlavicCorp 2018. Prague. Dostupné z: https://slaviccorp.ff.cuni.cz/wp-content/uploads/sites/144/2018/10/Zasina_Morfio_pl.pptx
- Žigo, P. (2012). Variantnosť v retrospektíve. In K. Buzássyová, B. Chocholová, & N. Janočková (Eds.), *Slovo v Slovníku. Aspekty lexikálnej sémantiky — gramatika — štylistika (pragmatika). Na počesť Alexandry Jarošovej* (s. 27–40). Bratislava: VEDA.
- Шарабуряк, А. А. (2012). Rozszerzenie klasy gramatycznej rzeczowników żywotnych we współczesnym języku polskim. *Наукові записки. Серія: Філологічна*, 26, 362–364.

