



Užití spojkových obsahových vět v češtině nerodilých mluvčích

Andrea Hudoušková

ABSTRAKT:

Na základě dat z žákovského korpusu češtiny CzeSL-SGT a jejich srovnání s korpusy rodilých mluvčích (SYN2020, SKRIPT2012) jsou zachycena specifika užití obsahových vět se spojkami *že, aby, ať, jestli, zda, zdali* a *-li* v jazyce nerodilých mluvčích, s přihlédnutím k jejich jazykovým skupinám (indoevropské, neindoevropské a slovanské jazyky). Analýza se zaměřuje na typické kolokace, jejich nadužívání či podužívání v žákovském jazyce a popis jazykových jevů, jež se odlišují od úzu rodilých mluvčích. Data ukazují, že nerodilí mluvčí mají tendenci nadužívat významově obecné a jim dobře známé jazykové prostředky na úkor sémanticky specifičtějších lexikálních jednotek. Rozdíly mezi žákovským jazykem a územ rodilých mluvčích jsou zřejmé u kolokací, různých slovesných vazeb a napříč jazykovými skupinami. Popis žákovského jazyka má význam pro výuku češtiny jako L2, neboť umožňuje lépe didakticky zpracovat zkoumané jazykové jevy, podporovat jejich aktivní užívání a odstraňování typických chyb, a to i s ohledem na konkrétní jazykovou skupinu.

KLÍČOVÁ SLOVA:

čeština jako L2, žákovský korpus CzeSL-SGT, Český národní korpus, mezijazyk, jazykové skupiny, obsahová věta, spojka, kolokace, frekvence, nadužívání

ABSTRACT:

Based on data from the Czech learner corpus CzeSL-SGT and a comparison with native speaker corpora (SYN2020, SKRIPT2012), the paper accounts for the specific use of content clauses with the conjunctions *že, aby, ať, jestli, zda, zdali* and *-li* in the language of non-native speakers, taking into consideration their language group (non-Indo-European, Indo-European and Slavonic languages). The article analyzes typical collocations, their over- and underuse by learners; it also describes language phenomena that differ from native usage. The data show that non-native speakers tend to overuse semantically broad lexical items, which they are familiar with, at the expense of more specific ones. The differences between learner and native usage are apparent in collocations, different types of verb complements and across language groups. The description of learners' languages is helpful in teaching L2 Czech, since it facilitates a better didactic approach to specific language phenomena, promoting their active usage and the correction of typical errors with respect to particular language groups.

KEYWORDS:

L2 Czech, Learner corpus CzeSL-SGT, Czech national corpus, Interlanguage, Language groups, Content clause, Conjunction, Collocation, Frequency, Overuse

0. ÚVOD

Cílem článku je na základě dat z žákovského korpusu češtiny CzeSL popsat způsob užití obsahových vět se spojkami *že, aby, ať, jestli, zda, zdali, -li* u nerodilých mluvčích češtiny s ohledem na jazykovou skupinu, do níž patří jejich mateřský jazyk,

tj. jazyky indoevropské (s výjimkou slovanských), jazyky neindoevropské a jazyky slovanské.

V úvodu bude stručně představen žákovský korpus CzeSL, způsoby zkoumání žákovského mezijazyka a možnosti jeho srovnání s jazykem rodilých mluvčích na základě jiných korpusů v rámci Českého národního korpusu (ČNK). Dále se pozornost zaměří na jednotlivé typy spojkových obsahových vět a zkoumání toho, jakým způsobem se jejich užití v jazyce nerodilých mluvčích liší od úzu v češtině jako mateřském jazyce. V závěru bude nastíněno, co nám analýza žákovského jazyka obecně umožňuje zjistit a jak mohou být tyto poznatky využity při výuce češtiny jako cizího jazyka.

1. ŽÁKOVSKÝ KORPUS CZESL-SGT

Akviziční korpus CzeSL-SGT je jedním z řady elektronických korpusů v rámci ČNK (Českého národního korpusu), který zachycuje češtinu nerodilých mluvčích v různých fázích jejího osvojování. Celkem zahrnuje 1,15 milionu slov. Korpus je vybaven automatickou anotací, která zahrnuje lemmatizaci, morfologické tagování a anotaci chyb. K dispozici jsou rozsáhlá metadata, která zachycují jak charakteristiky samotných textů (téma, slohový postup, limity času a rozsahu textu, možnost užití pomůcek, zkouška, atd.), tak i jejich autorů (věk, pohlaví, mateřský jazyk a jeho jazyková skupina, znalost dalších jazyků, jazyková úroveň podle SERR, délka a způsob osvojování češtiny, atd.). Zatímco určení jazykové úrovně v korpusu je do značné míry nespolehlivé, a nelze ho proto důsledně zohledňovat při studiu žákovského mezijazyka, velkou výhodou jsou metadata týkající se mateřských jazyků nerodilých mluvčích a příslušných jazykových skupin, jichž bude využito v tomto výzkumu. Jedná se o skupinu jazyků indoevropských s výjimkou slovanských (CzeSL-IE), skupinu jazyků neindoevropských (CzeSL-nIE) a skupinu jazyků slovanských (CzeSL-S). Výzkumná data ukazují, že v některých rysech se způsob užití obsahových vět mezi různými jazykovými skupinami liší.

2. ŽÁKOVSKÝ MEZIJAZYK

Termín *mezijazyk* je českým ekvivalentem anglického pojmu *interlanguage* (Selinker, 1972; Corder, 1981), který vystihuje vývojové fáze cílového jazyka, jimiž cizojazyčný mluvčí prochází v průběhu osvojování druhého jazyka (L2) (v českém kontextu viz Rosen — Hana — Vidová Hladká — Jelínek — Škodová — Štindlová, 2020; Romaševská, 2012; Šebesta — Škodová, 2012; Šebesta, 2010). Žákovský mezijazyk lze zkoumat na základě frekvence určitých jazykových jevů, tj. jejich nadužívání či podužívání ve srovnání s jazykem rodilých mluvčích. Mezi další možnosti jeho poznávání patří mj. studium kolokací a také analýza jazykových chyb. Všechny tyto přístupy budou zastoupeny při zkoumání obsahových vět v následujících oddílech.





2.1 APROXIMACE MEZIJAZYKA

Zkoumání mezijazyka v žákovském korpusu je vždy pouhou aproximací, což vyplývá z několika faktorů. Prvním z nich je automatická anotace korpusu CzeSL-SGT, která vzhledem k tomu, že byla provedena na mezijazyce nerodilých mluvčích, obsahuje řadu nepřesností a chyb. Proto se ne vždy podaří vyhledat všechny relevantní doklady a výsledný obraz žákovského jazyka může být zkreslen.

Za druhé je podoba mezijazyka ovlivňována množstvím aspektů, jež při jeho analýze není možné zcela zohlednit. Patří sem zejména jazyková úroveň mluvčích s ohledem na nespolehlivost těchto metadat v CzeSL, jejich mateřština, doba a způsob učení se češtině a znalost dalších jazyků. Významnou roli hraje také zadání, na jehož základě texty vznikaly. Je zřejmé, že užití jazykových prostředků v neformálním dopise kamarádovi se bude značně lišit od prostředků využitých např. v argumentativní eseji. K žádnému z těchto faktorů při zkoumání užití obsahových vět nebylo přihlíženo.

Za třetí hraje důležitou roli metoda vyhledávání kolokací a určování frekvence jazykových jevů. Experimentálně bylo zjištěno, že v případě kolokací predikátů s obsahovými spojkami dává ve většině případů nejlepší výsledky kontext čtyř pozic. Jedná se ovšem o arbitrární rozhodnutí, které rovněž může ovlivnit výsledky vyhledávání v korpusu.

V některých případech je omezujícím faktorem malý počet dokladů v korpusu CzeSL. Výzkum se ve všech případech zaměřil na prvních 50 kolokací, takže některé, i když pravděpodobně málo významné jevy, byly opomenuty.

2.2 SROVNATELNOST ŽÁKOVSKÉHO MEZIJAZYKA S JAZYKEM RODILÝCH MLUVČÍCH

Otázkou je také to, s čím má být žákovský mezijazyk zjištěný na základě dat z korpusu CzeSL srovnáván, aby bylo možné vyhodnotit jeho odlišnosti od jazyka rodilých mluvčích. V rámci ČNK se jako vhodné korpusy jeví SYN2020 a SKRIPT2012, i když oba z nich mají svá omezení.

Korpus SYN2020 je reprezentativní synchronní korpus současné psané češtiny, který obsahuje 100 milionů slov. Problematické z hlediska srovnatelnosti může být jeho žánrové složení (beletrie, oborová literatura, žurnalistika), které se značně liší od textů v korpusu CzeSL.

SKRIPT2012 je akviziční korpus češtiny rodilých mluvčích. Jedná se o korpus školních písemných prací českých žáků, který se svým charakterem více blíží korpusu CzeSL. Nevýhodou může být jeho menší rozsah (590 tisíc slov), a z toho plynoucí nízká frekvence či absence některých jazykových jevů.

Při srovnávání žákovského mezijazyka s češtinou rodilých mluvčích proto bude přihlíženo k oběma zmíněným korpusům.

3. OBSAHOVÉ VĚTY V ČEŠTINĚ NERODILÝCH MLUVČÍCH

Vedlejší věty obsahové jsou obligatorním sémantickým doplněním predikátu věty nadřazené, případně jeho nominalizace. Tento predikát má „buď význam v širokém smyslu slova mentální — pak má inkorporovaná věta roli ‘informace’ —, nebo jde o výrazy hodnotící nebo modální.“ (Daneš — Grepl — Hlavsa, 1987, s. 503). Obsahové věty tedy vyjadřují obsah toho, co je ve větě řídicí vyjádřeno predikátem či dějovým substantivem, případně obsah toho, co se ve větě řídicí hodnotí predikativním adjektivem či adverbium (Karlík, 2017; Panevová, 1971; Bauer, 1960; Gebauer — Ertl, 1926). Může se přitom jednat o predikáty slovesné, sponové či neslovesné.

Vedlejší obsahové věty spojkové jsou k řídicí větě připojovány spojkami *že, aby, ať, jestli, zda, zdali* a *-li*. Spojka *že* je bezpříznaková a prezentuje propozici jako fakt. V případě ostatních spojek je pravdivost či reálnost propozice otevřená, spojky *aby* a *ať* přitom implikují modální význam žádoucnosti (Grepl — Karlík, 1998, s. 42n).

Zkoumání užití obsahových vět v češtině nerodilých mluvčích se tedy zaměřuje na kolokace spojek, které připojují obsahové věty, s predikáty a predikativními substantivy, adjektivy a adverbii, jež budou porovnávány s kolokacemi typickými pro češtinu rodilých mluvčích.

3.1 OBSAHOVÉ VĚTY SE SPOJKOU ŽE

Obsahové věty se spojkou *že* se typicky spojují s predikáty vyjadřujícími řeč, myšlení a vnímání, hodnocení či jistotní modalitu.

Jak je patrné z tabulky 1, nejčastější kolokace se spojkou *že* se ve zkoumaných korpusech na předních místech do značné míry překrývají.

CzeSL-SGT	SKRIPT2012	SYN2020
1. myslit	2. myslet	3. myslet
3. doufat	3. říci	4. vědět
4. říci	5. vědět	6. být
5. vědět	7. doufat	7. říci
6. myslet	8. být	10. říkat
11. říkat	13. říkat	15. zdát
13. být	15. zjistit	18. mít
14. rád	19. rád	19. zjistit
17. znamenat	20. věřit	20. tvrdit
21. rozhodnout	21. uvědomit	21. znamenat

TABULKA 1: Seznam deseti nejčastějších kolokací (lemmat) se spojkou *že*¹

1 Kolokace byly vyhledávány v kontextu čtyř pozic doleva. Číslo před lemmaty označuje rank. Ze seznamu byla vypuštěna lemmata, která skutečnými kolokacemi s obsahovou větou nejsou, tj. interpunkční znaménka, zájmena, předložky, spojky, částice, modální slovesa, některá příslovce, substantiva a adjektiva.



Při pohledu na frekvence jednotlivých kolokací ovšem vystávají rozdíly. Týká se to zejména nadužívání výrazů *myslit / myslet, doufat, rád* ve spojení se spojkou *že* v žákovském korpusu CzeSL ve srovnání s korpusem rodilých mluvčích, naopak výrazného podužívání kolokace *zjistit, že* v žákovském mezijazyce. Mezi prvními padesáti kolokacemi v korpusu CzeSL se objevila lemmata *rozumět, líbit, slyšet, pamatovat*, která se v korpusech SKRIPT a SYN mezi prvními padesát kolokací nedostala. Srovnání relativní míry výskytu jednotlivých kolokací v uvedených korpusech zachycuje tabulka 2.

nadužívání	<i>myslit / myslet, doufat, rád</i>
srovnatelné se SKRIPT a SYN	<i>vědět, říkat</i>
srovnatelné se SKRIPT	<i>řící, rozhodnout</i>
srovnatelné se SYN	<i>znamenat, jistý</i>
pouze CzeSL (prvních 50 kolokací)	<i>rozumět, dobře, slyšet, pamatovat</i>
chybí v CzeSL (prvních 50 kolokací)	<i>zjistit, uvědomit, pocit, jasný, domnívat</i>

TABULKA 2: Relativní míra výskytu kolokací s obsahovou spojkou *že* v CzeSL ve srovnání se SYN a SKRIPT

Z hlediska zkoumání žákovského mezijazyka mají největší význam kolokace, jejichž frekvence se výraznějším způsobem liší od frekvencí v korpusech rodilých mluvčích.

Tabulka 3 zachycuje relativní frekvence kolokací s lemmaty, jež jsou v CzeSL ve srovnání s korpusem SYN a SKRIPT nadužívána.

Kolokace ²	CzeSL	SKRIPT	SYN
<i>myslit/myslet + že</i>	2 312,9	911,57	378,69
<i>myslit/myslet + na AK</i>	128,98	57,58	60,68
<i>doufat + že</i>	668,42	361,24	100,88
<i>rád + že</i>	223,1	129,82	51,65

TABULKA 3: Relativní frekvence (i.p.m.)³ kolokací s lemmaty *myslit / myslet, doufat, říci / říkat a rád* v korpusech CzeSL, SKRIPT a SYN

V tabulce 3 je vidět, že lemma *myslit / myslet* má ve srovnání s korpusem SKRIPT 2,5krát vyšší relativní frekvenci, ve srovnání s korpusem SYN více než šestinásobnou. Nadužívání tohoto lemmatu se projevuje i ve spojení s předložkovou vazbou, i když již ne tak výrazně. Podobný poměr nacházíme i u lemmatu *doufat*. Lemma *rád* je v CzeSL hypoteticky téměř dvakrát častější než v korpuse SKRIPT a čtyřikrát častější než v korpuse SYN.

2 Kolokace *myslit / myslet, že* a *doufat, že* byly vyhledávány v kontextu čtyř pozic, ostatní v kontextu dvou pozic, tj. [lemma="myslit|myslet"] [] {0,3} [word="že"]; [lemma="rád"] [] {0,1} [word="že"].

3 Relativní frekvence i.p.m. (items per milion) vyjadřuje průměrný počet výskytů dané jednotky přepočtený na hypotetický korpus o velikosti jednoho milionu slov. Tím je zajištěna srovnatelnost frekvencí v různě velkých korpusech.



Tabulka 4 zachycuje relativní frekvenci kolokací spojky *že* s lemmaty *zjistit*, *domnívat*, *uvědomit*, *pocit* a *jasný*, které mezi prvními padesáti kolokacemi v CzeSL-SGT na rozdíl od korpusů SKRIPT a SYN chybí. V tomto případě je relativní frekvence kolokací s těmito lemmaty v CzeSL několikanásobně nižší než v korpusech rodilých mluvčích. Důvodem je pravděpodobně fakt, že lemma *myslet* / *myslit*, které je v korpusu CzeSL nadužíváno, pravděpodobně zastupuje sémanticky významově specifitější lexikální jednotky jako *domnívat se* a *uvědomit si*. K možnému vysvětlení nízké relativní frekvence slovesa *zjistit* v CzeSL viz 3.1.2.

Kolokace ⁴	CzeSL	SKRIPT	SYN
<i>zjistit</i> + <i>že</i>	22,66	183,44	112,31
<i>zjistit</i> + AK	17,43	52,21	39,33
<i>uvědomit</i> + <i>že</i>	7,84	81,45	139,7
<i>uvědomit</i> + AK	1,74	42,33	18,05
<i>domnívat</i> + <i>že</i>	20,04	59,27	55,04
<i>pocit</i> + <i>že</i>	13,94	101,6	88,11
<i>jasný</i> + <i>že</i>	6,1	91,72	59,59

TABULKA 4: Relativní frekvence (i. p. m.) kolokací s lemmaty *zjistit*, *uvědomit*, *domnívat*, *pocit*, *jasný* v korpusech CzeSL, SKRIPT a SYN

3.1.1 OBSAHOVÉ VĚTY SE SPOJKOU ŽE V SUBKORPUSECH CZESL-SGT

Zajímavá zjištění o rozdílech mezi nerodilými mluvčími češtiny přináší zkoumání subkorpusů CzeSL na základě tří jazykových skupin, tj. CzeSL-IE (indoevropské jazyky s výjimkou slovanských), CzeSL-nIE (neindoevropské jazyky) a CzeSL-S (slovanské jazyky).

Pro subkorpus CzeSL-IE je v porovnání s ostatními subkorpusemi příznačná vyšší frekvence lemmatu *znamenat*, naopak menší výskyt lemmatu *rozhodnout*.

V rámci subkorpusu CzeSL-nIE je ve srovnání s ostatními subkorpusemi zajímavá vyšší frekvence sloves smyslového vnímání. Lemma *vidět* je v kolokaci se spojkou *že* častější než u jiných jazykových skupin, lemma *slyšet* je mezi prvními padesáti kolokacemi doloženo pouze v tomto subkorpuse. Mezi další lemmata, která se objevila v seznamu kolokací pouze v CzeSL-nIE, patří lemma *omlouvat* a dále substantiva *problém*, *štěstí* a *pravda*.

Co se týče subkorpusu slovanských jazyků, jsou v seznamu kolokací zastoupena lemmata *možná* a *jistý*, která mezi kolokacemi v ostatních subkorpusech doložena nejsou, a lemma *pochopit*, které se v rámci prvních padesáti kolokací neobjevilo ani v ostatních subkorpusech, ani v korpusech SKRIPT a SYN.

⁴ Kolokace *zjistit* / *uvědomit* / *domnívat* ..., *že* byly vyhledávány v kontextu čtyř pozic, ostatní v kontextu dvou pozic.



3.1.2 ZVLÁŠTNOSTI ŽÁKOVSKÉHO MEZIJAZYKA U OBSAHOVÝCH VĚT SE SPOJKOU ŽE

Zvláštnosti žákovského mezijazyka se týkají za prvé kolokací, které nejsou doloženy mezi prvními padesáti v korpusech rodilých mluvčích, za druhé způsobu užití některých lemmat ve spojení s obsahovými větami.

Do první skupiny patří kolokace s lemmaty *líbit*, *dobře*, *slyšet*, *rozumět* a *pamatovat*. Z pohledu do tabulky 5 vyplývá, že relativní frekvence kolokací spojky *že* s těmito lemmaty je výrazně vyšší než v korpusech rodilých mluvčích.

Kolokace ⁵	CzeSL-SGT	SKRIPT	SYN
líbit + že	35,73	4,23	8,77
dobře + že	15,69	8,47	6,05
slyšet + že	45,32	22,58	23,13
rozumět + že	61	0	3,96
pamatovat + že	46,19	19,76	14,59

TABULKA 5: Relativní frekvence (i.p.m.) kolokací s lemmaty *líbit*, *dobře*, *slyšet*, *rozumět*, *pamatovat* v korpusech CzeSL, SKRIPT a SYN

Při bližším zkoumání konkordancí v korpuse CzeSL se ukazují další vlastnosti žákovského mezijazyka. Kolokace (*být*) *dobře*, *že* je typická pro CzeSL-S, neboť 14 ze 16 jejích výskytů spadá do skupiny slovanských jazyků. Totéž platí o kolokaci *rozumět*, *že*, jež je skupinou slovanských jazyků zastoupena v 61 ze 70 výskytů.

Bližší zkoumání konkordancí s lemmatem *slyšet* odhaluje, že jeho vyšší relativní frekvence v korpuse CzeSL-SGT může být dána tím, že v některých případech toto sloveso zastupuje významově specifitější lexikální jednotku, např. sloveso *dozvědět se*, jak ukazuje tabulka 6.

nIE	píšu dopis ti . Mám se dobře v Praze .	Slyšel jsem že	budeš přijet do prahy . V sobotu . Můžeme se
S	budeš mít radost . Ach , jo . Dneska jsem	slyšela , že	tam je teplá zima - možna budeš mít možnost plavat
S	Připomeň mi kolik dnů ty ještě budeš v Praze .	Slyšel jsem , že	pojedeš na dovolenou . Kam chceš jet ? A s
S	. Ahoj , Adame ! Píšu Ti z ČR .	Slyšel jsem , že	brzy chceš jet do ČR na výlet , proto chtěl
nIE	. A proto jsem byl velmi vzrušený , když jsem	slyšel , že	kamarád z Islandu bude jet do Prahy na výlet .

TABULKA 6: Příklady konkordancí s lemmatem *slyšet* v CzeSL-SGT

Tato tendence k nahrazování významově specifického slovesa lexikální jednotkou obecnější je v CzeSL zřetelná i u slovesa *vidět*, ačkoli jeho relativní frekvence je ve srovnání s korpusem rodilých mluvčích nižší. Jak ukazují příklady v tabulce 7, sloveso *vidět* může zastupovat slovesa *všimnout si*, *zjistit* či *pochořit*.

⁵ Kolokace byly vyhledávány v kontextu čtyř pozic, pouze u kolokace *dobře*, *že* v kontextu dvou pozic: [lemma="být"] [] {0,1} [word="dobře"] [] {0,1} [word="že"] .



S	gramatiky ,Pani Karlova popovídala o kinematogra o divadle .	Viděla jsem , že	stejně jako já ráda se dívá na Divadlo Jary Cimrmana
S	slovník a trochu se na něj podíváme , dá se	vidět že	ty dva jazyky obsahují skoro padesát procent slov , která
S	dobu už jsem začal špatně vidět . Ale teď já	vidím , že	bych byl špatným optikem . Chemie a biologie vůbec mi
nIE	Prahy a budu se naučit . Ale v Praze jsem	viděla , že	čestina je těžší ja horší že jsem se myslala !
nIE	muže poslouchat tento jazyk přes rádio nebo televize atakdal a	vidím , že	média zlehčí život , třeba bez počítače v společnostech ,

TABULKA 7: Příklady konkordancí s lemmatem *vidět* v CzeSL-SGT

Nadužívané sloveso *rozumět* často vyjadřuje významy sloves *chápat / pochopit*, jak ukazují příklady v tabulce 8. Navíc, jak vyplývá ze srovnání s korpusem rodilých mluvčích, je toto sloveso v žákovském mezijazyce užíváno odlišně. Zatímco v korpuse CzeSL převládá kolokace *rozumět, že*, v korpuse SYN dominují kolokace *rozumět tomu, že*, případně *rozumět tak, že* či *rozumí se, že*.

S	ruštině , ale když jsem přišla a začala jsem studovat	rozuměla že	čeština není lehká . U první době bylo velmi těžké
IE	moje dceru . Začatek nebyl spokojený ale teď věřím že	rozuměl že	to není práce jenom pro ženy . Bude mocť žije
IE	, a popřeju něco tajemně . Kdy otevřím oči ,	rozumím , že	to všechno bylo snem , mou soukromou pohádkou
nIE	restauraci , abychom měli večere . Ale potom , jsme	rozuměli , že	tak už nepůjde , a začali jsme šetřit peníze .
S	lepší bude být jisti a nebat se . Už teď	rozumím , že	měl pravdu . Můj dědeček mám rád čtení a je

TABULKA 8: Příklady konkordancí s lemmatem *rozumět* v CzeSL-SGT

Také u lemmatu *památovat se* způsob jeho užití v CzeSL liší od korpusem rodilých mluvčích. V žákovském korpuse je z 53 konkordancí s tímto lemmatem ve spojení se spojkou *že* pouze šestnáct dokladů, kdy je slovesa ve významu „vzpomínat si“ užito náležitě v reflexivní podobě s formantem *si* nebo *se*. V ostatních případech je sloveso užito v nereflexivní podobě, která je v daném významu nenáležitá, jak ukazují příklady v tabulce 9. Naproti tomu v korpuse SKRIPT se ze 14 výskytů lemmatu *památovat* v kolokaci se spojkou *že* celkem dvanáctkrát jedná o sloveso *památovat si*; jednou je doloženo sloveso *památovat se*, jednou sloveso *památovat*. V náhodném vzorku 250 konkordancí v korpuse SYN se o nereflexivní sloveso *památovat* jedná v 58 případech. Ve většině případů se jedná o spojení *památovat (na to)*, *že* s významem „myslet na to, že“, pouze ve dvou případech jde o význam „vzpomínat si“ a v jednom případě o význam „být pamětníkem“.

IE	tatinek , moje sestra a muj bratr dojeli tam .	Pamatuju že	cesta byla strašně dlouhá , myslím že to bylo 35
S	, a co nejvažněji chodit na výlety za město .	Pamatuju , že	rád sportuješ . a proto doufám , že budeš mit
S	Její výška asi 130 metrů , ale já už přesně	nepamatuju . Já vím že	máš rád fotbal . Proto pokud budeš v Barcelóně ,
nIE	kratší nebo delší výlet , to záleží na tobě .	Pamatuju , že	řikala jsi , že v tom městě ještě jsi nebyla
nIE	jsem velkou radost , že jsem nemohla spat . Ještě	pamatuju , že	tato kniha se jmenuje " Dědeček "

TABULKA 9: Příklady konkordancí s lemmatem *památovat* v CzeSL-SGT

Poslední poznámka se týká užití slovesa *chtít*, které se v korpuse CzeSL v devatenácti případech objevuje v kolokaci se spojkou *že*, což v korpusech rodilých mluvčích doloženo není.

S	, to se mi moc líbí . Kongratuluju tebe !	Chci , že	ty byla zdravá , veselá a krásná . Ale omlouvám
IE	Můj nejlepší den , myslím že ještě nebyl , protože	chtěl bych že	můj nejlepší den byl někdy v budoucnosti . Myslím že
S	Eva . Můj nejlepší kamarád je z Ruska , ale	chci , že	on bydlil v Praze , jako já . Znáš ho
nIE	, jsem jenom obyčejné . " Každá babička doufá a	chce , že	její vnoučata se vrací na návštěvu často , samozřejmě .
nIE	je ne možno . Řekl by ji , že první	chci , že	mou rutina byla vždycky zdravou a v bezpečnosti . to

TABULKA 10: Příklady kolokací lemmatu *chtít se* spojkou *že* v CzeSL-SGT



3.1.3 NEJČASTĚJŠÍ PREDIKÁTY SE SPOJKOU ŽE A JEJICH VÝSKYT V CZESL-SGT

Ze srovnání nejčastějších predikátů či predikátových výrazů se spojkou *že* na základě dat z korpusu SYN se seznamem nejčastějších kolokací s touto spojkou v korpusu CzeSL-SGT vyplývá, že nejčastější predikáty (predikátové výrazy) se objevují i v seznamu kolokací v žákovském korpusu. Týká se to lemmat *myslet, vědět, říkat, říci, zdát, znamenat, doufat, být, věřit, rozhodnout, vidět, možný*. Naopak řada z nich se mezi nejčastějšími kolokacemi v rámci CzeSL-SGT neobjevuje. Platí to zejména o predikátech, jež byly zmíněny v oddíle 3.1.2, tj. lemmat *zjistit, uvědomit, jasný, domnívat, všimnout, pochopit*, a dále lemmat *napadnout, pravda, platit, oznámit, vyplývat, cítit* apod. V závislosti na konkrétní jazykové úrovni by si tyto predikáty ve výuce češtiny jako cizího jazyka jistě zasloužily cílenou pozornost.

3.2 OBSAHOVÉ VĚTY SE SPOJKAMI *ABY, AŽ*

Obsahové věty se spojkou *aby* se kombinují zejména s predikáty vyjadřujícími příkaz, žádost či přání. Ve větách imperativního charakteru může být užito rovněž spojky *ať*, ve větách optativních je tato varianta hodnocena jako řídká a hovorová (Daneš — Grepl — Hlavsa, 1987, s. 513). Obsahové věty se spojkou *aby* se spojují také s predikáty vyjadřujícími hodnocení, jistotní a voluntativní modalitu.

Podobně jako v případě obsahových vět se spojkou *že* se i v případě spojky *aby* lemmata v nejčastějších kolokacích překrývají ve všech typech zkoumaných korpusů, jak ukazuje tabulka 11.

CzeSL-SGT	SKRIPT2012	SYN2020
1. chtít	1. přát	1. chtít
4. nutný	4. chtít	5. přát
9. snažit	9. důležitý	7. požádat
13. důležitý	11. možný	12. dovolit
18. dovolit	13. snažit	13. snažit
25. třeba	16. stačit	14. důležitý
41. stačit	20. říci	16. stačit
50. přát	25. usilovat	22. přimět
	27. říkat	23. žádat
	39. modlit	25. vyzvat

TABULKA 11: Seznam deseti nejčastějších kolokací (lemmat) se spojkou *aby*⁶

⁶ Kolokace byly vyhledávány v kontextu čtyř pozic doleva. Číslo před lemmaty označuje rank. Ze seznamu byla vypuštěna lemmata, která skutečnými kolokacemi s obsahovou větou nejsou, tj. interpunkční znaménka, zájmena, předložky, spojky, částice, modální slovesa, některá příslovce, substantiva a adjektiva. Dále byla vypuštěna slovesa, s nimiž se v spojení se spojkou *aby* jednoznačně nejedná o vedlejší větu obsahovou, nýbrž účelovou.



Z tabulky 11 je zřejmé, že pestrost kolokací s obsahovou spojkou *aby* je v žákovském jazyce nižší než u spojky *že*. Mezi prvními padesáti kolokacemi se vyskytlo pouze osm lemmat, která se spojují s obsahovou větou vedlejší se spojkou *aby*.

Na nejvyšších pozicích se mezi kolokacemi ve všech korpusech drží lemmata *chtít*, *snažit* a *důležitý*. Naopak lemmata *stačit* a *přát* mají v žákovském korpusu mnohem nižší rank a frekvenci než v korpusech rodilých mluvčích.

Ze srovnání CzeSL-SGT s korpusem rodilých mluvčích v tabulce 12 vyplývá, že v žákovském korpusu je ve spojení s obsahovou spojkou *aby* nadužíváno predikativní adjektivum *nutný*, naopak podužívána jsou lemmata *přát* a *stačit*. Na rozdíl od korpusů SKRIPT a SYN chybí v korpusu CzeSL-SGT mezi padesáti prvními kolokacemi predikativní adjektivum *možný*.

nadužívání	<i>nutný</i>
srovnatelné se SKRIPT a SYN	<i>snažit, dovolit, důležitý</i>
podužívání	<i>přát, stačit</i>
chybí v CzeSL na rozdíl od SKRIPT, SYN	<i>možný</i>
chybí v CzeSL na rozdíl od SKRIPT	<i>usilovat, modlit</i>
chybí v CzeSL na rozdíl od SYN	<i>požádat, přimět, žádat, vyzvat, přesvědčit, navrhnout, nechat, (dát) pozor, zajistit, požadovat, zabránit</i>

TABULKA 12: Relativní míra výskytu kolokací s obsahovou spojkou *aby* v CzeSL ve srovnání se SYN a SKRIPT

Srovnání relativních frekvencí lemmat, jejichž užití se v žákovském korpusu liší, opět přináší zajímavá zjištění o charakteru žákovského mezijazyka. Tabulka 13 ukazuje relativní frekvence predikativního adjektiva *nutný*, které je v žákovském jazyce v kolokaci se spojkou *aby* nadužíváno, a predikativa *třeba*, které se v korpusu CzeSL-SGT na rozdíl od korpusů rodilých mluvčích objevilo mezi prvními padesáti kolokacemi.

Je patrné, že nadužívání kolokace *nutný, aby* v žákovském korpusu oproti korpusům SKRIPT a SYN je dáno tím, že v korpusech rodilých mluvčích je typičtější spojení lemmatu *nutný* s infinitivem a kolokace se spojkou *aby* jsou řídké. Totéž, i když v menší míře, platí též pro spojení lemmatu *třeba* se spojkou *aby* v žákovském korpusu, v důsledku čehož se tato kolokace dostala mezi prvních padesát v korpusu CzeSL-SGT, nikoli však v korpusech rodilých mluvčích, kde je mnohonásobně častější spojení tohoto predikativa s infinitivem.

Kolokace ⁷	CzeSL	SKRIPT	SYN
<i>nutný + aby</i>	8,71	0	4,82
<i>nutný + INF</i>	22,66	59,27	74,08
<i>třeba + aby</i>	1,74	0	4,35
<i>třeba + INF</i>	29,63	79,02	184,97

TABULKA 13: Relativní frekvence (i.p.m.) kolokací lemmat *nutný, třeba* se spojkou *aby* a infinitivem v korpusech CzeSL, SKRIPT a SYN

⁷ Kolokace byly vyhledávány v kontextu dvou pozic ve spojení se spojkou *aby* a v kontextu tří pozic ve spojení s infinitivem.



Tabulka 14 zachycuje lemmata, která jsou ve spojení se spojkou *aby* v žákovském korpusu oproti korpusům SKRIPT a SYN podužívána. Z údajů v tabulce vyplývá, že lemma *přát* je v korpusu CzeSL podužíváno zejména ve spojení se spojkou *aby*. Jeho poměrně vysoká relativní frekvence ve spojení s akuzativním doplněním vyplývá z vysoké četnosti kolokací typu *přát všechno nejlepší* a *přát štěstí a zdraví* v žákovském korpusu. Nízká relativní frekvence lemmatu *přát* ve spojení se spojkou *aby* či infinitivem v korpusu CzeSL-SGT může souviset také s obecně vysokou frekvencí významově blízkého lemmatu *chtít* v žákovském korpusu ve srovnání s korpusy rodilých mluvčích, jak ukazuje tabulka 15.

Lemma *snažit* je ve všech svých vazbách v žákovském mezijazyce méně frekvencované než v jazyce rodilých mluvčích.

Predikativní adjektivum *možný* je v žákovském mezijazyce podužíváno ve spojení se spojkou *aby*, a tudíž se nedostalo mezi prvních padesát nejčastějších kolokací, relativní frekvence kolokace s infinitivem se však více blíží relativní míře jeho výskytu v korpusech rodilých mluvčích.

Kolokace ⁸	CzeSL	SKRIPT	SYN
<i>přát</i> + <i>aby</i> {0,3}	2,61	32,46	24,88
<i>přát</i> + INF {0,2}	3,49	29,63	19,2
<i>přát</i> + AK {0,2}	85,4	101,6	55,7
<i>snažit</i> + <i>aby</i> {0,2}	6,1	12,7	10,86
<i>snažit</i> + INF {0,2}	115,91	529,16	314,04
<i>snažit</i> + o AK {0,3}	5,23	9,88	11,87
<i>možný</i> + <i>aby</i> {0,1}	0,87	12,7	6,85
<i>možný</i> + INF {0,2}	88,89	64,91	147,87

TABULKA 14: Relativní frekvence (i.p.m.) kolokací lemmat *přát*, *snažit*, *možný* se spojkou *aby*, infinitivem a jmenným doplněním v korpusech CzeSL, SKRIPT a SYN

Kolokace ⁹	CzeSL	SKRIPT	SYN
<i>chtít</i> + <i>aby</i> {0,2}	74,08	47,98	74,52
<i>chtít</i> + INF {0,3}	2 720,75	1 594,54	1 123,16
<i>chtít</i> + AK {0,0}	349,46	261,05	217,26

TABULKA 15: Relativní frekvence (i.p.m.) kolokací lemmatu *chtít* se spojkou *aby*, infinitivem a akuzativním doplněním v korpusech CzeSL, SKRIPT a SYN

8 Čísla ve složených závorkách udávají počet možných pozic mezi členy kolokace, tj. {0,2} označuje kontext tří pozic. Kontext vyhledávání byl přizpůsoben tak, aby výsledky v co nejmenší míře zahrnovaly nerelevantní konkordance. Přesto mohou být údaje zkresleny přítomností zejména účelových vedlejších vět se spojkou *aby* ve výsledku vyhledávání.

9 Čísla ve složených závorkách udávají počet možných pozic mezi členy kolokace. Frekvenční údaje v případě kolokací se spojkou *aby* jsou mírně zkresleny přítomností modálního užití slovesa *chtít* ve výsledcích vyhledávání.



Seznam nejčastějších kolokací se spojkou *aby* v jednotlivých subkorpusech CzeSL, zachycený v tabulce 16, odhaluje, že mluvčí indoevropských a neindoevropských jazyků využívají ve větší míře pouze sloveso *chtít*. Naopak škála výrazů užitých ve spojení se spojkou *aby* ve skupině slovanských jazyků je výrazně širší a blíží se seznamu lemmat v korpusech rodilých mluvčích.

CzeSL-IE	CzeSL-nIE	CzeSL-S
chtít	chtít důležitý	chtít nutný snažit důležitý dovolit třeba stačit přát řící dobrý dostatečný poposit

TABULKA 16: Nejčastější lemmata v kolokacích se spojkou *aby* v subkorpusech CzeSL-SGT

I u predikátů se spojkou *aby* platí, že nejčastější z nich jsou zastoupeny mezi kolokacemi v žákovském korpuse. Týká se to lemmat *přát*, *chtít*, *důležitý*, *dovolit*, *stačit*, *snažit*, *nutný*, *třeba*. Řada těchto predikátů se však v CzeSL-SGT objevuje jen okrajově či vůbec, přestože patří k typickým kolokacím s obsahovými větami se spojkou *aby*, např. *(po)žádat*, *navrhnout*, *doporučovat / doporučit*, *(po)prosit*, *možný*, *radit*, na vyšších úrovních případně i další jako *vyzvat*, *vyžadovat*, *nařídít*, *modlit*, *přikázat*, *bát* apod.

Spojka *ať* je v žákovském korpuse velmi řídká, byly zjištěny pouze tři případy jejího spontánního užití, které jsou zachyceny v tabulce 17.

nIE	v realitních kancelářích minulý týden . Poprosili jsme ho ,	ať	nám najde nějakých bytu 3+1 , cena od 15,000 -
	nějaké problémy vypravěla mi řešení a říkala mi , že	ať	jsem pevná a zažím je dobře . Miluju ji a
nIE	jeden týden . Moje kamarádka bydlí tam a říkala ,	ať	navštěvu u ní . A proto jsem přemyslela stále .

TABULKA 17: Konkordance se spojkou *ať* v korpuse CzeSL-SGT

3.3 OBSAHOVÉ VĚTY SE SPOJKAMI *JESTLI*, *ZDA*, *ZDALI*, *-LI*

Obsahové věty se spojkami *jestli*, *zda(li)* a *-li* se typicky spojují s predikáty vyjadřujícími zjišťování či nerozhodnutost, případně jistotní modalitu.

Jak ukazuje tabulka 18, tři nejčastější lemmata v kolokaci s obsahovou spojkou *jestli* se překrývají ve všech zkoumaných korpusech.



CzeSL-SGT	SKRIPT2012	SYN2020
1. zeptat	1. vědět	1. zeptat
2. vědět	2. ptát	2. vědět
3. ptát	3. zeptat	3. ptát
4. prosit	4. otázka	4. jistý
6. otázka	5. zajímat	5. přemýšlet
7. záležet	6. zkontrolovat	6. zjistit
9. vadit	7. přemýšlet	7. napadnout
10. poprosit	8. obava	8. otázka
12. říkat	9. jistý	9. zajímat
17. důležitý	10. uvažovat	10. uvažovat

TABULKA 18: Seznam deseti nejčastějších kolokací (lemmat) se spojkou *jestli*¹⁰

Jak je patrné z tabulek 19–21, významnější frekvenci v kolokacích se spojkou *jestli* má pouze lemma *zeptat*, jehož frekvence v korpusu CzeSL-SGT je srovnatelná s korpusy SKRIPT a SYN, a lemma *vědět*. Relativní frekvence ostatních lemmat doložených jako kolokace v žakovském korpusu je velmi nízká. V korpusu CzeSL-SGT jsou nadužívány kolokace lemmat *napsat* a *důležitý*, které se v korpusech rodilých mluvčích jeví jako velmi řídké.

srovnatelné se SKRIPT a SYN	<i>zeptat</i>
podužívání	<i>vědět, ptát, jistý, přemýšlet</i>
chybí ve SKRIPT, SYN	<i>napsat, důležitý</i>

TABULKA 19: Relativní míra výskytu kolokací s obsahovou spojkou *jestli* v CzeSL ve srovnání se SYN a SKRIPT

Kolokace ¹¹	CzeSL-SGT	SKRIPT	SYN
<i>zeptat</i>	13,94	16,93	17,62
<i>vědět</i>	27,89	91,72	36,18
<i>ptát</i>	3,49	18,34	11,39
<i>jistý</i>	2,61	7,06	9,73
<i>přemýšlet</i>	0	8,47	7,26

TABULKA 20: Relativní frekvence (i.p.m.) kolokací lemmat *zeptat, vědět, ptát, jistý, přemýšlet* se spojkou *jestli* v korpusech CzeSL, SKRIPT a SYN

¹⁰ Kolokace byly vyhledávány v kontextu čtyř pozic doleva. Číslo před lemmaty označuje rank. Ze seznamu byla vypuštěna lemmata, která skutečnými kolokacemi s obsahovou větou nejsou, tj. interpunkční znaménka, zájmena, předložky, spojky, částice, modální slovesa, některá příslovce, substantiva a adjektiva.

¹¹ Všechny kolokace byly vyhledávány v kontextu čtyř pozic.



Kolokace	CzeSL-SGT	SKRIPT	SYN
napsat {0,3}	2,61	1,41	0,13
důležitý {0,1}	1,74	0	0,37

TABULKA 21: Relativní frekvence (i.p.m.) kolokací lemmat *napsat*, *důležitý* se spojkou *jestli* v korpusch CzeSL, SKRIPT a SYN

Z nejčastějších predikátů se spojkou *jestli* se nejfrekventovanější z nich objevují i mezi prvními padesáti kolokacemi zjištěnými v korpusu CzeSL-SGT, tj. lemmata *ptát*, *přemýšlet*, *zeptat*, *vědět*, *otázka*, *(po)prosit*, *záležet*, *říkat*, *jistý*. Jiné, podobně časté predikáty však mezi nimi zastoupeny nejsou, jako např. lemmata *zajímat*, *uvažovat*, *uvidět*, *váhat*, *(z)kontrolovat*, *poznat*, *zjistit*, *požádat*, *rozmyslet*, *podívat*, *rozhodovat*, *zkoušet*, *jasný*, *zavolat* atd. Tyto predikáty by si ve spojení s obsahovou spojkou *jestli* na odpovídajících úrovních pokročilosti jistě zasloužily více pozornosti.

Spojka *-li* připojující obsahovou větu je v korpusu CzeSL doložena 26krát, z toho 23krát v subkorpusu slovanských jazyků. Příklady konkordancí jsou zobrazeny v tabulce 22.

S	rodné zemi , jaké zážitky budeš mít a bude -	li	čeština zařazena do tvého osobního života , budeš - li
S	i nevládní chyby . Ale otázkou je , budem -	li	kdykoliv absolvovat , a když budem , co je za
S	, není zklamaný a bude zaležet na něm chce -	li	něco dosáhnout v životě . Myslím , to by bylo
S	zjišťovat své výsledky . Na nich záleží - budu -	li	žít v České republice a studovat tady , nebo -
nIE	Máte zajímavou nabídku , ale jste nenapsal , máte -	li	tam garaž nebo ne ? To je důležité pro mě

TABULKA 22: Příklady konkordancí se spojkou *-li* v korpusu CzeSL-SGT

Co se týče spojky *zda*, je v korpusu CzeSL-SGT omezena na 12 výskytů, a jedná se tedy v žákovském mezijazyce o obsahovou spojkou periferní. Varianta *zdali* je doložena v jediném případě. Příklady některých konkordancí přináší tabulka 23.

S	smutně se usmála . Často si přemyslím nad tím ,	zda	vůbec chci být stará . Jestli chci dožít do osmdesáti
S	, avšak se v tuto chvíli tak připadám . Zjistit	zda	se rozum obejde bez citu a cit bez něj ,
IE	domací vylohy . Kromě finanční zaležitosti , je otázka ,	zda	mít děti nebo ne ? Někdy , děti zpevní manželství
nIE	na velikosti své hodnoty , ale záleží na tom ,	zda	je přenášena a jak dobře se šíří . " Mnoho
S	mi nešlo (jako malému dítěti) o to ,	zda	mít či nemít v lásce tuto zemi . Já jakožto

TABULKA 23: Příklady konkordancí se spojkou *zda* v korpusu CzeSL-SGT

4. ZÁVĚR

Analýza dat z žákovského korpusu CzeSL-SGT umožňuje odhalit specifika žákovského jazyka jak v obecných rysech, tak v konkrétních případech.

Obecným rysem žákovského mezijazyka je užívání „bezpečných“ jazykových prostředků, tj. takových, které jsou mluvčímu dobře známé a významově dostatečně široké. Ve spojení s vedlejšími větami obsahovými to platí například o nadužívání sloves *myslet* a *chtít*, a to na úkor sémanticky specifitějších lexikálních jednotek jako *domnívat se*, *uvědomit si*, resp. *přát si*. U kolokací s obsahovými spojkami *že* a *aby* jsou



doložena lemmata, která jsou v porovnání s jazykem roditelých mluvčích nadužívána, neboť zastupují významově specifičtější synonymické výrazy.

Tatáž tendence platí i pro samotné spojky. U spojky *že* je v žákovském mezijazyce největší pestrost lemmat a kolokace dosahují nejvyšších frekvencí. Pestrost i četnost kolokací postupně klesá u spojek *aby* a *jestli*. Spojky *-li*, *zda*, *zdali* a *ať* jsou v žákovském mezijazyce prostředkem periferním.

Na základě dat z korpusu CzeSL-SGT lze také zkoumat preference v realizaci valenčních doplnění. Nejzřetelnější rozdíly se projevily u lemmatu *přát*, které je oproti úzu roditelých mluvčích v žákovském mezijazyce mnohonásobně častější ve spojení s akuzativním komplementem než ve spojení s obsahovou vedlejší větou či infinitivem. Disproporce ve způsobu užití byly sledovány také u lemmat *nutný* a *třeba*.

Na jednotlivých příkladech lze zjišťovat konkrétní specifika žákovského mezijazyka, a to jednak v závislosti na mateřském jazyce mluvčích, či nezávisle na něm. Například pro skupinu slovanských jazyků je typická kolokace (*být*) *dobře, že* či užití lemmatu *rozumět* ve významu „pochopit, chápat“. Nezávisle na rodném jazyce se objevovaly například problémy s užitím sloves *pamatovat* (významově nenáležitě nereflexivní užití) a *chtít* (kolokace *chtít, že*).

Tato zjištění mohou vést k zamyšlení nad příčinami nadužívání či podužívání jazykových prostředků, jež byly v textu v některých případech naznačeny. Dalším krokem může být analýza učebních materiálů a míry pozornosti, která je v nich daným jazykovým prostředkům věnována. Vyšší míru zaměření na prostředky, které v žákovském mezijazyce nejsou dostatečně zastoupeny, může dále podpořit mapování častých kolokací frekventovaných predikátů na základě autentického jazyka roditelých mluvčích a jejich důkladné didaktické zpracování s ohledem na SERR. Cílené procvičování těchto prostředků ve výuce by se pak mělo zaměřit jak na rozšiřování škály aktivně užívaných jazykových prostředků, tak na odstraňování typických chyb či odlišností v jejich užívání.

BIBLIOGRAFIE:

- Bauer, J. (1960): *Vývoj českého souvětí*. Brno: FF BU.
- Corder, S. P. (1981): *Error Analysis and Interlanguage*. Oxford: OUP.
- Daneš, F. — Grepl, M. — Hlavsa, Z. (1987): *Mluvnice češtiny. 3. Skladba*. Praha: Academia.
- Gebauer, J. — Ertl, V. (1926): *Gebauerova Mluvnice česká*. Praha: Česká grafická Unie.
- Grepl, M. — Karlík, P. (1998): *Skladba češtiny*. Olomouc: Votobia.
- Karlík, P. (2017). Obsahová věta, in: Karlík, P. — Nekula, M. — Pleskalová, J. (eds.), *CzechEncy — Nový encyklopedický slovník češtiny*. [online] [cit. 2021-09-17]. Dostupné z [www: <https://www.czechency.org/slovník/OBSAHOVÁ_VĚTA>](https://www.czechency.org/slovník/OBSAHOVÁ_VĚTA).
- Panevová, J. (1971): Vedlejší věty obsahové, *Slovo a slovesnost* 32. 4, s. 289–300.
- Romaševská, K. (2012): Teorie mezijazyka a její význam pro výuku druhého jazyka, *Nová čeština doma a ve světě* 1, s. 62–74.
- Rosen, A. a kol. (2020): *Compiling and annotating a learner corpus for a morphologically rich language*. Praha: Karolinum.
- Selinker, L. (1972): *Interlanguage, International Review of Applied Linguistics in Language Teaching* 10, s. 209–231.
- Šebesta, K. (2010): Korpusy češtiny a osvojování jazyka, *Studie z aplikované lingvistiky* 2, s. 11–33.
- Škodová, S. — Šebesta, K. (2012): *Čeština — cílový jazyk a korpusy*. Liberec: TUL.

KORPUSY

Křen, M. — Cvrček, V. — Henyš, J. —
 Hnátková, M. — Jelínek, T. — Kocek, J. —
 Kovářiková, D. — Křivan, J. — Milička, J. —
 Petkevič, V. — Procházka, P. —
 Skoumalová, H. — Šindlerová, J. —
 Škrabal, M.: SYN2020: *reprezentativní korpus
 psané češtiny*. Ústav Českého národního
 korpusu FF UK, Praha 2020.

Šebesta, K. — Bedřichová, Z. — Šormová, K. —
 Štindlová, B. — Hrdlička, M. —
 Hrdličková, T. — Hana, J. — Petkevič, V. —
 Jelínek, T. — Škodová, S. — Poláčková, M. —
 Janeš, P. — Lundáková, K. —
 Skoumalová, H. — Sládek, Š. —

Pierscieniak, P. — Toufarová, D. —
 Richter, M. — Straka, M. — Rosen, A.:
*CzeSL-SGT: korpus češtiny nerodilých mluvčích
 s automaticky provedenou anotací, verze 2
 z 28. 7. 2014*. Ústav Českého národního
 korpusu FF UK, Praha 2014.

Šebesta, K. — Goláňová, H. — Jelínek, T. —
 Jelínková, B. — Křen, M. — Letafková, J. —
 Procházka, P. — Skoumalová, H.: *SKRIPT2012:
 akviziční korpus psané češtiny — přepisy
 písemných prací žáků základních a středních
 škol v ČR*. Ústav Českého národního korpusu
 FF UK, Praha 2013.

