

Univerzita Karlova
Filozofická fakulta
Ústav Českého národního korpusu
Matematická lingvistika

TEZE DISERTAČNÍ PRÁCE

Adrian Jan Zasina

Korpusový přístup ve výuce češtiny jako cizího jazyka
Corpus approach to teaching Czech as a foreign language

vedoucí práce: Mgr. Lucie Lukešová, Ph.D.

Praha 2019

1 Úvod

Rapidní vývoj informačních technologií a možností, které s sebou přinášejí, ovlivnil v posledních desetiletích takřka všechny obory lidské činnosti. V lingvistice a příbuzných disciplínách se jejich vliv odrazil především v dostupnosti velkého množství dat, zejména v podobě jazykových korpusů, a nástrojů na jejich analýzu. Současné trendy v přístupu k jazyku, které propaguje například korpusová lingvistika, zahrnují kromě autentického jazykového materiálu i důraz na význam, reálné užití a kombinovatelnost jazykových jednotek. Tento přístup sdílí mnoho společných rysů s novodobými trendy ve výuce jazyků, není proto divu, že v posledních letech dochází k nárůstu zájmu o začlenění korpusových metod a zdrojů do jazykové výuky, ať už přímo v hodinách či nepřímo v podobě jazykových učebnic a podobných materiálů.

Využití jazykových korpusů v běžné jazykové výuce je v souladu s nejnovějšími poznatky o vzdělávání, především pak s tzv. *data-driven learning* (DDL), výukou založenou na datech, která studentům umožňuje samostatně objevovat nové poznatky a učit se zábavnou, ale efektivní formou (Bernardini, 2004; Frankenberg-Garcia, 2012; Vališová, 2011, 2012). Mezi prvními projevy zájmu o využití jazykových korpusů při akvizici cizího jazyka na konci minulého století angličtí lingvisté, např. Johns (1991), který se věnoval výuce angličtiny. V českém kontextu se otázka využití korpusových metod a dat ve výuce objevuje ve větší míře až na začátku 21. století, a to především v souvislosti s výukou cizích jazyků (hlavně angličtiny) nebo češtiny pro cizince (Lukšija, 2012; Osolsobě, 2010; Vališová, 2011, 2012). Jednou z prvních publikací, která se snažila popsat možnosti přímého využití korpusového materiálu ve výuce angličtiny, je článek Thomase (2006), propagující možnosti rozhraní Sketch Engine.

Přestože lze ve výuce češtiny jako cizího jazyka zaznamenat rostoucí zájem o využití korpusů a podobných zdrojů, metodologické otázky týkající se konkrétních aplikací doposud zůstávaly stranou zájmu badatelů. Dosavadní práce představují spíše úvod do problematiky a obecné seznámení s možnostmi jazykových korpusů, ale zpravidla se systematicky nezabývají výhodami ani úskalími korpusového přístupu ve výuce a jeho konkrétními aplikacemi. Dosud také nevznikla práce, jež by se pokoušela o validaci vlivu korpusových cvičení na efektivnost výuky češtiny. Cílem této práce je proto jednak vyplnit mezeru v popisu metodologie korpusového přístupu ve výuce češtiny jako cizího jazyka, jednak formulovat konkrétní přístup při sestavování korpusových cvičení a experimentálně ověřit vliv korpusového přístupu na výuku češtiny jako cizího jazyka.

1.1 Téma práce

Tématem této práce je aplikace korpusového přístupu k jazyku ve výuce češtiny jako cizího jazyka (ČCJ). Náš výzkum si klade za cíl systematicky zmapovat možnosti začlenění korpusů do běžné jazykové výuky, a to s ohledem na konkrétní jazykové potřeby cizinců na základě jejich typických chyb. Práce využívá jak korpusy žakovské s chybovou anotací (pro identifikaci obtížných jazykových jevů), tak i reprezentativní korpusy současné češtiny, které poslouží jako zdroj dat pro konkrétní cvičení. Cílem práce je nabídnout nejen ucelený metodologický rámec, ale i praktickou ukázkou práce s korpusovými daty, která bude využitelná nejen pro odborníka vzdělaného v korpusové lingvistice, ale především pro

běžného učitele. Hlavním předmětem zájmu je navrhnout takový postup k řešení vybraných jazykových problémů, jenž bude z hlediska studenta-cizince i jeho učitele přehledný a srozumitelný a zároveň bude vycházet z autentického materiálu a nejnovějších poznatků o jazyce, opírajících se o výzkumy v oblasti korpusové lingvistiky. Součástí práce budou i návrhy konkrétních cvičení vč. uvedené úrovně obtížnosti podle Společného evropského referenčního rámce (SERR) a klíče k výsledkům.

1.2 Struktura práce

Práce se skládá z šesti základních kapitol. Po souhrnném úvodu následuje kapitola 2, jež stručně uvádí čtenáře do problematiky využití korpusů ve výuce jazyků a nastiňuje původ metody *data-driven learning*. V této části jsou rovněž krátce představeny průkopnické práce, které se tomuto tématu věnovaly, načež následuje popis stávající situace v České republice. Pozornost je také věnována možnostem a mezím korpusového přístupu ve výuce, které mají přímou souvislost s výukou jazyka.

Hlavní těžiště práce spočívá ve výzkumné části, jež je rozdělena do čtyř kapitol. Kapitola 3 popisuje použitá korpusová data a vysvětluje principy metodologie celého výzkumu, vycházejícího výhradně z korpusových dat. Nabízí ucelený rámec pro postup při analýze dat, která nejdříve identifikuje problematický jazykový jev v žakovském korpusu a následně ho na základě autentických dat z korpusu klasifikuje, což umožňuje generalizaci pravidel o fungování daného jazykového jevu.

Kapitola 4 se pokouší o zmapování nejproblematictějších oblastí ve výuce češtiny jako cizího jazyka na základě žakovského korpusu CzeSL-SGT. Analýza se zaměřuje především na autosémantika a pronomina, která jsou zkoumána v rámci pěti nejčastějších typů chyb identifikovaných pomocí automatické chybové anotace. Výsledky analýzy jsou východiskem pro další zkoumání v následující kapitole.

Systematický popis korpusového přístupu ve výuce lze najít v kapitole 5, jež je rozčleněna do tří případových studií. Tato část práce na konkrétních příkladech prezentuje postup při řešení jazykových obtíží pomocí korpusu. První studie (podkapitola 5.1) se zaměřuje na problém s kvantitou v deverbativech zakončených na *-ání* a *-aní*. Druhá (podkapitola 5.2), rovněž beroucí v potaz problematickou kvantitu, se soustředí na případy přičestí minulého. Třetí studie (podkapitola 5.3) se zabývá konkurencí genitivních koncovek singuláru *-a* a *-u* u neživotných maskulín.

Kapitola 6 se zaměřuje na didaktický experiment, jehož cílem bylo otestovat korpusový přístup ve vysokoškolské výuce s ohledem na jeho uplatnění při jazykové výuce a přínos v jazykovém vývoji studentů. V této části je detailně popsán průběh výuky. Nechybí informace o recepci korpusových cvičení ze strany studentů a dále výsledky srovnání korpusové výuky s výukou tradiční.

V závěrečné kapitole 7 jsou shrnuty výsledky celé práce – důraz je kladen na ucelený a systematický přístup ke zkoumání korpusových dat s ohledem na výuku, jenž se u nás teprve rozvíjí.

Práce je doplněna kompletním seznamem použité literatury a seznamy použitých tabulek, obrázků, grafů a korpusových cvičení. Dále se v přílohách nacházejí dodatečné materiály s analyzovanými daty a použitými výukovými materiály.

2 Využití korpusů ve výuce

Tato část se věnuje obecnému přehledu možností, jak zapojit jazykové korpusy do běžné jazykové výuky. Dále nastiňuje původ metody *data-driven learning* (česky doslova: daty řízená výuka) a zmiňuje některé průkopnické práce v této oblasti v českém prostředí.

2.1 Obecný přehled

Zájem o využití jazykových korpusů ve výuce jazyků přichází se začátkem 90. let 20. století, kdy se použití korpusů začalo rozšiřovat i na jiné disciplíny mimo okruh korpusové lingvistiky (Granath, 2009, p. 47). Jako jeden z prvních zapojuje korpusy do výuky Johns (1991), jenž systematicky popisuje metodu tzv. *data-driven learning* (DDL), využívanou ve výuce angličtiny v univerzitních jazykových kurzech (viz 2.2). Rostoucí zájem badatelů vedl k tomu, že se v roce 1994 konala první vědecká konference *Teaching and Language Corpora Conference*, zaměřená na využití korpusů ve výuce (Römer, 2009, p. 84); ta se pravidelně koná dodnes. I přesto je korpusový přístup uplatňován spíše na akademické půdě korpusovými lingvisty a jen těžce se promítá do běžné výuky na základních a středních školách (Frankenberg-Garcia, 2012; Römer, 2009), kde učitelé nemají veliké zkušenosti s prací s korpusem. Proto v poslední době vznikají publikace věnující se výuce s využitím jazykových korpusů (Bennett, 2010; Breyer, 2011; Reppen, 2010; Thomas, 2016), díky nimž mají nejen učitelé, ale i studenti možnost seznámit se s touto metodou. Zároveň se objevují další pomůcky v podobě deskriptivních gramatik (Conrad & Biber, 2009; Cvrček et al., 2010), zohledňujících poznatky korpusové lingvistiky, a také frekvenčních slovníků (Čermák & Křen, 2011; Davies & Gardner, 2013), jež pomáhají v doplňování a zpřesňování nejběžnější slovní zásoby v daném jazyce.

Korpusová lingvistika má bezesporu velký potenciál obohatit pedagogickou praxi. Učitelé mají možnost doplnit jazykovou výuku o autentické užití jazyka a také najít odpovědi na lingvistické otázky, jež se vymykají intuici anebo nejsou přiměřeně pojednány v mluvnících a slovnících (Frankenberg-Garcia, 2012: 38). Výchozím bodem korpusového zkoumání bývá zpravidla tzv. konkordance¹. Konkordance umožňuje vyhledávání určité textové jednotky (slovo, část slova, fráze, posloupnost znaků, kolokace) spolu s jejím pravým a levým kontextem, přičemž každý jednotlivý příklad se nachází na jednom (konkordančním) řádku (McEnery & Hardie, 2012: 35–37). Takové zobrazení usnadňuje pozorování korpusových dat, která jsou poté analyzována studenty. Díky konkordanci lze zkoumat jazyk na několika rovinách, což studentům dává možnost např. cvičit morfologii (Osolsobě 2014) či gramatické pády (Vališová 2011, 2012) nebo rozlišovat valenci slov majících obdobný význam, ale jinou kolokabilitu, jako např. v případě anglických slov *convince* a *persuade* (Johns, 1991). Dále pak konkordance umožňuje hledání překladových ekvivalentů pomocí paralelních korpusů (Frankenberg-Garcia, 2012: 44–46; Barlow, 2000) nebo méně typicky pomocí dvou jednojazyčných korpusů (Lewandowska, 2014) a také zkoumání kolokací, jež jsou typické pro rodilé mluvčí (Römer, 2009).

¹ pojmy:konkordance. (2016, Jun 7). V *Příručka ČNK*. Vytaženo 12:28, March 29, 2019, od <http://wiki.korpus.cz/doku.php?id=pojmy:konkordance&rev=1465291976>

K jazykové výuce mohou být využity různé typy korpusů. Široce jsou používány korpusy žákovské, jež jsou sestavovány z textů vytvořených nerodilými mluvčími a slouží k analýze chyb (Gilquin, Granger, & Paquot, 2007; Petkevič et al., 2012). Dále lze těžit z korpusů současného jazyka, které ukazují přirozený jazyk rodilých mluvčích, a také korpusů mluvených (též mluvených žákovských korpusů, srov. Gráf, 2017). Ke zkoumání problémů s negativní interferencí vznikající vlivem překladu mohou posloužit korpusy paralelní.

2.2 DDL

Jak již bylo řečeno, průkopníkem v oblasti metody *data-driven learning* (daty řízená výuka) je Johns, jenž ve své studii zdůrazňuje, že „*úkolem studenta je »objevovat« cizí jazyk a úkolem učitele je zaručit kontext, ve kterém student bude schopen rozvinout strategii pro toto objevování*“² (Johns, 1991, p. 1). Autor (ibid., pp. 1–4) tím míní, že se u studenta vyvine schopnost „učit se, jak se učit“ (*learn how to learn*), a také zdůrazňuje, že student se v tomto případě stává zároveň vědcem, kterého korpusová data vedou k závěrům a popisu gramatických pravidel. Na rozdíl od běžných metod ve výuce je to přístup induktivní, který se nejdříve zakládá na korpusových datech, díky nimž student kategorizuje daný jev a generalizuje gramatická pravidla. Učitel přesně neví, jaké jevy a pravidla student objeví, proto je učitelovou rolí především být koordinátorem studentova výzkumu. Učitel tak již není autoritou, ale spíše partnerem. Neznamená to však, že tato pozice ubírá na učitelově významu, naopak zapojuje učitele do společné práce se studentem, kdy oba čelí problémům a snaží se přijít na vhodné řešení. Student díky této svobodě zvyšuje svou samostatnost a učitel pouze dohlíží na studentův postup, aby ho v případě obtíží nasměroval. Tento přístup umožňuje studentovi přijít na něco dosud neobjeveného, anebo přistupovat ke známým objevům z jiného úhlu pohledu, což příznivě působí na pochopení gramatických pravidel nerodilým mluvčím. Práce je koncipována tak, že učitel zadává studentovi problematickou úlohu, při jejímž řešení student sám objeví zákonitosti, aniž by mu byly předem vysvětleny učitelem.³ DDL si tak zakládá na gramatických pravidlech objevených studentem, která jsou založena na autentickém použití jazyka. Díky tomu má student možnost prohloubit své jazykové znalosti příjemným způsobem a osamostatnit se ve výuce, což pozitivně ovlivňuje jeho výsledky (Bernardini, 2004; Lewandowska, 2014).

DDL samozřejmě není přístupem absolutním v tom smyslu, že by měl nahradit veškerou výuku. Tato metoda by měla být spíše doplňkem běžné jazykové výuky (jednotlivá korpusová cvičení) anebo s ní být paralelní (kurzy zaměřené na použití korpusů ve výuce vedle běžných jazykových kurzů). Takový přístup má šanci se postupně dostat do povědomí jak učitelů, tak i studentů a napomoci při řešení složitých jevů, které nejsou dostatečně vysvětleny stávajícími gramatikami nebo pro studenta nejsou (z hlediska jeho mateřského jazyka) srozumitelné. Problematické jevy lze vysledovat v žákovských korpusech a takto získané poznatky pak využít v praktických cvičeních.

DDL se tradičně dělí na dvě odvětví podle metody práce (Leech, 1997), která je buď přímá (tzv. *hard DDL* nebo *hands-on*), kdy studenti pracují přímo s korpusovým vyhledávačem, anebo nepřímá (*soft DDL* nebo *hands-off*), kdy studenti řeší úkoly připravené na základě

² „[T]he task of the learner is to »discover« the foreign language, and (...) the task of the language teacher is to provide a context in which the learner can develop strategies for discovery.“ (překlad do češtiny A. J. Z.)

³ Např. učitel nevysvětlí studentovi, jak se tvoří minulý čas, nýbrž student sám bude na základě pozorování korpusových dat schopen popsat pravidla, a tak si je i lépe vštípit.

korpusových dat, např. ve formě vytištěných konkordancí. Pod pojmem korpusy ve výuce si lze představit tři různé způsoby využití (Geist & Hahn, 2012, p. 124): a) výuka o korpusech (v rámci lingvistických kurzů na univerzitě); b) využití korpusů ve výuce jazyků; c) výuka o využití korpusů. Nicméně, jak ukázaly výzkumy (Römer, 2009, p. 87; Frankenberg-Garcia, 2012, p. 37), nejsou ani přes snahu korpusových lingvistů korpusy mezi učiteli středních škol populární a v akademickém prostředí jsou propagovány hlavně korpusovými specialisty. Frankenberg-Garciová (2012) zdůrazňuje nutnost zapojení korpusů do každodenní výuky cizích jazyků pomocí dřívější přípravy učitelů k práci s korpusem už v rámci jejich studia, a také, jak poukazuje Römerová (2009), je v tomto ohledu nepostradatelná pomoc korpusových lingvistů, kteří by měli vyjít vstříc očekáváním učitelů a vytvářet nové nástroje nebo sestavovat ukázková cvičení pro studenty.

2.3 Korpusový přístup ve výuce v českém prostředí

Zájem o využití korpusů ve výuce se v České republice začíná projevovat až v prvním desetiletí 21. století nejprve ve výuce angličtiny. Na Masarykově univerzitě akademický učitel Thomas (2006) jako jeden z prvních zapojuje nástroj Sketch Engine (Kilgarriff, Rychlý, Smrž, & Tugwell, 2004) do práce se studenty anglické filologie. Postupně se korpusový přístup ve výuce rozšiřuje i na češtinu: vznikají první práce opírající se o frekvenci jevů v korpusu. Např. Vališová (2009) ve své diplomové práci opírající se o frekvenční data z korpusu systematicky popisuje českou konjugaci z pohledu cizince. Lukšijová (2010) se ve své bakalářské práci snaží o systematický popis předložek *do/na* s místním směrovým významem, následně se v práci diplomové (Lukšija, 2012) zabývá českou deklinací a nabízí korpusová cvičení pro začátečníky. Osolsobě (2010) nabízí studentům letního kurzu češtiny v Brně možnosti využití korpusových nástrojů a dále zapojuje jazykové korpusy i do kurzu morfologie pro české bohemisty (Osolsobě, 2014). Schöneová (Schöne, 2015) zkoumá frekvenční distribuci substantivních pádových tvarů a vybízí s ohledem na didaktické účely zahrnovat ve slovníkových částech učebnic nejen tvary nominativu plurálu, ale i dalších frekventovaných pádů s uvedením nejběžnějších kolokací. Vališová (2016) systematizuje základní typologii korpusových cvičení. Zájem o využití korpusových metod ve výuce se v průběhu let zvyšuje (Hudousková, 2014; Kočařová, 2013; Vaculová, 2013; Vališová & Osolsobě, 2012), což mj. dokládají práce s ukázkovými cvičeními pro studenty a návodem, jak korpusové nástroje používat a vytěžovat data (Konečná & Zasina, 2014; Vališová, 2011, 2012, 2016; Zasina, 2018a). Nicméně stále tu z velké části chybí systematický popis, jenž by se soustředil na zpřesnění (z hlediska cizince) toho, jaký charakter mají vznikající jazykové problémy, do jaké roviny jazyka patří (morfologické, pravopisné, flektivní atd.), jak je lze korpusově uchopit, jak pracovat s autentickými daty v korpusu a – v neposlední řadě – jak sestavit konkrétní cvičení pro procvičování dané látky. Ačkoli některé práce určitá korpusová cvičení zahrnovaly, důraz byl spíše kladen na to, aby učitel pracující se studenty byl zároveň korpusovým specialistou (Vališová, 2011, p. 319). Pokud však přístup k sestavování cvičení bude jednotný a metodologicky podložený, budou moci z korpusových materiálů těžit i učitelé bez korpusových zkušeností nebo s jen minimálními zkušenostmi s prací s korpusem. Cílem by mělo být šířit užitečnost korpusů mezi všemi učiteli, nikoli se omezovat jen na úzký okruh těch, kteří už s korpusy pracují.

3 Data a metodologie

Celá práce vychází výhradně z korpusových dat a jejich analýz, přičemž autentická data využívá dvojnásobně: jednak jako východisko pro analýzu jazyka nerodilých mluvčích (korpus CzeSL-SGT), jednak jako zdroj dat o podobě současné češtiny rodilých mluvčích (korpus SYN2015). Ze srovnání obou zdrojů a na základě použití metod korpusové lingvistiky (především frekvenčních seznamů a kontextu) pak vyplývají nejen specifické požadavky na výuku nerodilých mluvčích, ale i vlastnosti českého jazyka v podobě, která v běžných učebnicích nebývá zachycena. V této části věnujeme pozornost podrobnému popisu použitých korpusových dat a metodologii k jejich vytěžování.

3.1 Data

V první fázi tohoto výzkumu využíváme žákovský korpus CzeSL-SGT, tj. *Czech as Second Language with Spelling, Grammar and Tags* (Šebesta, 2010; Šebesta, et al., 2014), který obsahuje přepisy písemných prací nerodilých mluvčích češtiny s automaticky provedenou anotací (o evaluaci chybové anotace viz Štindlová, 2013). Písemné práce byly vytvořeny v letech 2009 až 2013. Korpus navazuje na část korpusu CzeSL-plain⁴, jež se skládá z esejí cizinců, a je doplněn o další texty sebrané v roce 2013. Celková velikost korpusu činí 1 147 477 pozic (tj. slov vč. interpunkce).

Kromě standardní morfologické anotace obsahuje tento korpus i chybovou anotaci zohledňující nejběžnější chyby nerodilých mluvčích.⁵ Každému slovnímu tvaru jsou přiřazeny slovní druh, morfologická značka a základní slovníkový tvar. Některé slovní tvary byly opraveny a texty prošly opětovnou automatickou morfologickou analýzou. Druh jazykové chyby byl stanoven na základě porovnání původního a opraveného slovního tvaru automaticky, proto je potřeba počítat s určitou mírou nepřesnosti.

V naší práci jsme využili pouze část korpusu zahrnující studenty s mateřským slovanským jazykem. Zvolený subkorpus je dále podrobně popsán v části 4.3.1 *Typologie chyb a obecná statistika*, která je východiskem pro identifikaci jazykových jevů, problematických pro studenty ČCJ. Identifikace těchto nedostatků u studentů-cizinců posloužila dále jako základ pro provedení tří případových studií popsáných v kapitole 5.

V druhé fázi výzkumu je hlavním zdrojem dat reprezentativní vyvážený korpus současné psané češtiny SYN2015 (Cvrček, Čermáková, & Křen, 2016; Křen et al., 2015; Křen et al., 2016), obsahující 120 748 715 pozic. Tento korpus je posledním přírůstkem do řady korpusů SYN, vydávaných v pětiletých intervalech. Každý korpus z této řady odráží pětileté období předcházející jeho zveřejnění (především v publicistické části): SYN2015 tedy zahrnuje zejména texty z let 2010–2014, čímž se snaží zachycovat co nejvíce aktuální jazyk.

Korpus obsahuje také standardní anotaci, je lemmatizovaný, morfologicky a syntakticky označovaný. Pokud jde o klasifikaci textů, dělí se na tři hlavní skupiny: beletrii, oborovou

⁴ cnk:czesl-plain. (2018, Aug 7). V *Příručka ČNK*. Vytaženo 15:42, March 4, 2019, od <http://wiki.korpus.cz/doku.php?id=cnk:czesl-plain&rev=1533638229>.

⁵<http://utkl.ff.cuni.cz/~rosen/public/SeznamAutoChybR0R1.html>

literaturu (odborná a populárně naučná literatura) a publicistiku. Dále je členěn do specifických kategorií.⁶ Klasifikace vychází z externích, vnětých kritérií.⁷

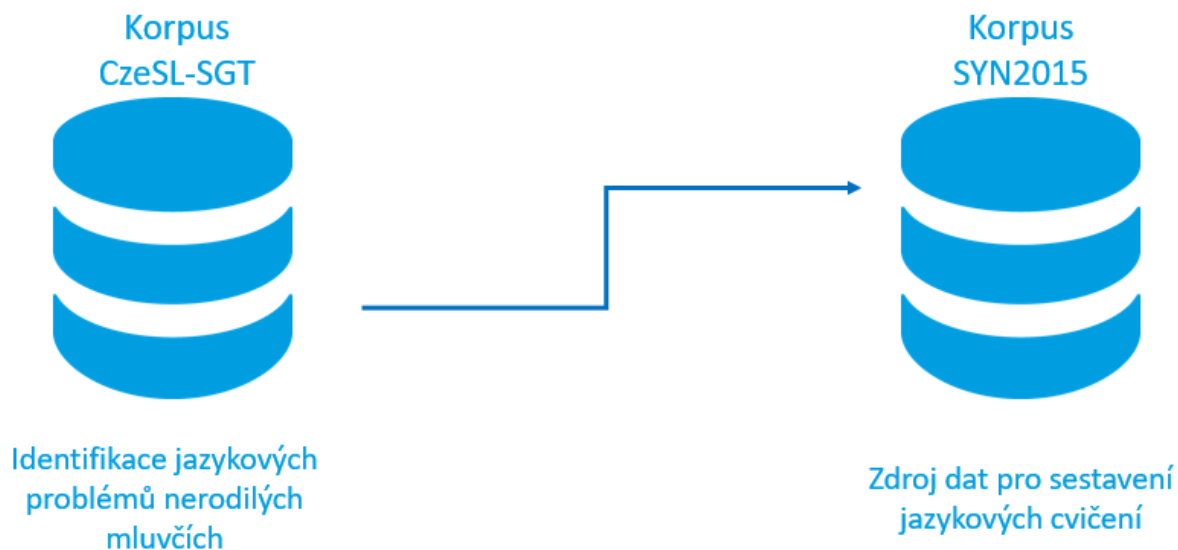
V našem výzkumu využíváme celý korpus SYN2015 – ten nám jednak slouží k prozkoumání jazyka rodilých mluvčích a dále k vyhodnocení pravidelnosti jazykových jevů v češtině, mimoto je také zdrojem pro sestavení korpusových cvičení.

Zde popsaná data jsou dále zmiňována i v praktické části, kde jsou případně doplňována v souladu s účelem jednotlivých studií, popřípadě jsou uváděna pro připomenutí a zachování logické návaznosti.

3.2 Metodologie

Předkládaná disertace nabízí nový metodologický přístup ve výzkumu problematických oblastí češtiny u nerodilých mluvčích a v jejich výuce. Využívá nejprve žakovský korpus k identifikaci jazykových problémů, pro něž se následně hledají konkrétní pravidla použití v korpusu synchronní češtiny, a na základě takové klasifikace jazykového jevu je sestaven návrh postupu ve výuce v podobě korpusových cvičení. Tato metoda jednoduše využívá informace z jednoho zdroje dat, které jsou následně uplatněny při vyhledávání v druhém zdroji, kde slouží k sestavení pravidla o fungování hledaného jazykového jevu. V této souvislosti můžeme tedy mluvit o metodě *od korpusu ke korpusu*, kterou znázorňuje níže uvedený obrázek.

Obrázek 1. Grafické znázornění přístupu od korpusu ke korpusu



Tento postup nejenže umožňuje empirický výzkum, ale zároveň objektivně přistupuje ke zkoumaným jazykovým jevům. Díky tomu výzkum provedený v rámci této disertace odhaluje prominentní obtíže cizinců s češtinou, které jsou v jazyce nerodilých mluvčích nejčastější, a zároveň nabízí jejich možné řešení.

⁶ cnk:klasifikace_textu_syn2015. (2016, Jan 25). V *Příručka ČNK*. Vytlačeno 09:04, June 8, 2019, od http://wiki.korpus.cz/doku.php?id=cnk:klasifikace_textu_syn2015&rev=1453733752.

⁷ V brzké době se rovněž chystá vnitrotextová klasifikace vycházející z multidimenzionální analýzy češtiny (viz. Cvrček et al., 2018a).

Metodologie celého výzkumu tak sestává ze tří fází:

1. analýzy chyb v žákovském korpusu, jež je východiskem pro zkoumání konkrétních jazykových problémů (kapitola 4),
2. návrhu korpusového přístupu ve třech případových studiích, který zohledňuje zvolené problematické oblasti zjištěné v první fázi (kapitola 5),
3. didaktického experimentu, v němž se testuje korpusový přístup ve vysokoškolské výuce (kapitola 6).

1. Při analýze žákovského korpusu vycházíme z automatické chybové anotace gramatických a pravopisných jevů, které dále zkoumáme s ohledem na slovní druh. Na základě frekvence obdobných jevů je pak možné jednotlivé problematické oblasti kategorizovat a podrobně popsat.

2. Na základě analýzy žákovského korpusu jsme zvolili tři nejproblematictější oblasti, jimž jsou věnovány tři samostatné případové studie. Ty zpracovávají jevy na rovině gramatické i pravopisné a jejich cílem je systematicky popsat různé aspekty korpusového přístupu ve výuce ČCJ a nabídnout konkrétní metodologický rámec postupu při zpracovávání korpusového materiálu. Důraz je kladen jak na analýzu dat, tak na jejich interpretaci.

3. Práce dále zahrnuje experimentální výzkum provedený na dvou skupinách vysokoškolských studentů češtiny pro cizince. S jednou skupinou byla látka probírána pomocí korpusových metod, zatímco druhá skupina probírala stejnou látku tradičním způsobem. Pro získání co nejlepších výsledků a minimalizaci vedlejších faktorů majících vliv na osvojování cizího jazyka byly obě skupiny homogenní, co se týče jejich rodného jazyka L1 (polština) a věku (19–24 let). Cílem experimentu bylo prověřit, zda korpusový výzkum zlepšuje efektivnost výuky a zda navržený postup pro vysvětlení nějakého problematického jevu – tj. a) identifikace problému, b) řešení problému pomocí korpusu, c) interpretace výsledků – je skutečně v praxi funkční. Díky dlouhodobé přímé práci se studenty (během jednoho semestru) bylo možné pozorovat klady a zápory navrhované metody ve výuce a na základě nových poznatků pak mohl být postup vyhodnocen.

K probraným otázkám uvádíme i cvičení, která mohou využívat jak studenti, tak i učitelé. Cvičení vycházejí jednak z korpusových analýz provedených v kapitole 5, jednak jsou odpovědí na přímé potřeby studentů během výuky (cvičení prezentována v kapitole 6).

4 Analýza chyb v žákovském korpusu CZESL-SGT

Cílem této kapitoly je zmapování nejproblematictějších oblastí ve výuce ČCJ na základě žákovského korpusu CZESL-SGT. Tato analýza poslouží jako zdroj dat pro další identifikaci konkrétních problémů cizinců s češtinou, kterými se budeme zabývat v kapitole 5. Při zkoumání žákovského korpusu budou zohledněny nejen nejčastější typy chyb, ale i jejich rozložení v rámci slovních druhů. Pozornost bude věnována především autosémantikům (substantivům, adjektivům, verbům a adverbiím) a také zčásti synsémantikům zde zastoupeným pronominy.⁸

S ohledem na rozdílné postoje vůči termínu „jazyková chyba“ v jazykové výuce je v následující podkapitole vysvětleno pojetí chyby, jež budeme dále v této studii používat. Podkapitola 4.2 prezentuje popis dat a metodologický postup při jejich analýze. Podkapitola 4.3 podrobně analyzuje nejčastější typy chyb s ohledem na slovní druh a snaží se identifikovat oblasti, v nichž mají studenti-cizinci největší obtíže. V závěru jsou zdůrazněny nejčastější problémy a uvedena další doporučení pro zkoumání problematických oblastí.

4.1 Pojetí chyby v jazykové výuce

Než se dostaneme k analýze nejčastějších chyb z hlediska studenta-cizince (viz dále podkapitola 4.3), je třeba věnovat pozornost samotnému pojetí chyby v jazykové výuce.

Každý proces učení je neoddelitelně spojen s metodou pokus–omyl, při níž se člověk učí z neúspěšných pokusů a postupně si osvojuje funkční, přijatelné řešení, které se nejvíce osvědčilo. Stejný princip můžeme vztáhnout i na proces osvojování jazyka. Běžně se v něm setkáváme s termínem *jazyková chyba*, což je poněkud kontroverzní koncept, vycházející ze vztahu k určitému ideálu, obvykle k nějaké standardní podobě jazyka. Jazyková chyba je brána jako odchylka od normy, a je tedy úzce spjata s tím, co za normu v dané situaci považujeme. V oblasti jazykové výuky (zvláště pak cizinců) je nutné pracovat s takovým konceptem normy, aby si studenti byli schopni osvojit spolehlivý a pokud možno neutrální komunikační prostředek, jenž bude pokrývat co největší množství komunikačních situací. Za takovýto neutrální prostředek se v češtině tradičně považuje spisovný jazyk, tedy podoba jazyka využívaná především ve formálním písemném a ústním projevu. Dá se totiž očekávat, že použití jiné než standardní podoby jazyka bude mít ve většině oficiálních komunikačních situací pro mluvčího závažnější následky než použití standardního jazyka v neformálním písemném či ústním projevu.

Z tohoto důvodu jsme se rozhodli s tímto konceptem pracovat i v této práci a využít definici, která popisuje jazykovou chybu jako „určitý nežádoucí odklon od jazykové normy (v jistých případech rovněž od řečové normy), [jež] bývá obvykle jedním z prvků, které s vysokou mírou spolehlivosti odlišují (pochopitelně, že při užívání jednoho a téhož jazykového kódu) mluvčího rodilého od mluvčího jinojazyčného“ (Hrdlička, 2012, pp. 89–90). Za jazykovou normu pro češtinu je zde tedy považována spisovná čeština.

⁸ V souladu s Čermákem (2010, p. 172) chápeme pronomina jako synsémantika.

4.2 Postup a zkoumaný materiál

Jako zdroj empirických dat k naší analýze chyb posloužil žakovský korpus CzeSL-SGT (Šebesta et al., 2014), obsahující texty nerodilých mluvčích češtiny s automaticky provedenou anotací (viz podkapitola 3.1). Korpus má velikost 1 147 477 pozic. Kromě standardní morfologické anotace obsahuje i chybovou anotaci zohledňující nejběžnější chyby nerodilých mluvčích.⁹

Korpus na rozdíl od běžných korpusů obsahuje dvojí sadu atributů, takže kromě standardních atributů *word*, *lemma* a *tag*, které se vztahují ke konkrétní původní lexikální jednotce v textech nerodilých mluvčích, má ještě atributy *word1*, *lemma1* a *tag1*, vztahující se k cílové hypotéze¹⁰, pokud původní tvar *word* byl vyhodnocen jako chybný. Dalšími atributy jsou *gs* (údaj o pravopisné nebo gramatické chybě) a *err* (typ chyby).¹¹

V naší studii budeme rovněž operovat s některými dynamickými atributy, jež jsou odvozeny od značek *tag* a *tag1*. Díky nim můžeme specifikovat hledané morfologické kategorie, což usnadňuje analýzu podle chybného (*k*), nebo správného slovního druhu (*kl*) či chybného (*c*) a správného pádu (*cl*).

Korpus zahrnuje i údaje o textech a jejich autorech, jako např. pohlaví, věk, mateřský jazyk, informace o studiu češtiny atd., z nichž nejpodstatnější budou prezentovány v podkapitole 4.3.1.

Následující sekce se zabývá podrobně analýzou chyb v žakovském korpusu, který posloužil jako východisko pro identifikaci jazykových jevů, jež jsou pro studenty ČCJ v rámci jednotlivých typů chyb problematické. Důraz bude kladen zejména na slovansky mluvící cizince a jejich charakteristické chyby. Cílem této analýzy je odhalit problematické oblasti a zákonitosti v chybování studentů, které budeme poté dále zkoumat v rámci tří případových studií v kapitole 5.

4.3 Analýza žakovského korpusu

Cílem analýzy chyb v žakovském korpusu je monitorovat, s jakými slovy a typy chyb mají studenti největší problémy a na jaké jazykové úrovni. Pozornost je věnována pěti nejčastějším typům chyb u slovansky mluvících studentů. V následujících podkapitolách se nejprve zaměříme na obecné statistiky analyzovaného subkorpusu a poté popíšeme každý typ chyby zvlášť.

⁹ <http://utkl.ff.cuni.cz/~rosen/public/SeznamAutoChybR0R1.html>

¹⁰ Cílovou hypotézou se zde myslí správný tvar, který se student pokusil – v případě chyby neúspěšně – zkonstruovat. Tento pojem souvisí se změnou náhledu na akvizici cizího jazyka od konce 60. let. „Nabývání cizího jazyka se tehdy začíná chápat jako kreativní proces konstruování systému, kterým žáci testují své hypotézy o cílovém jazyce. Chyby v žakovském jazyce nejsou vnímány jako povrchové deviace, tj. nepřínosně negativně, nýbrž jako důkazy žakovy snahy rekonstruovat pravidly řízený jazykový systém a jako možnost sledovat proces zpracovávání jazykových dat jinojazyčnými mluvčími“ (Štindlová 2013, p. 25).

¹¹ <http://utkl.ff.cuni.cz/~rosen/public/2014-czesl-sgt-cs.pdf>

4.3.1 Typologie chyb a obecná statistika

Jak již bylo naznačeno, analýze byla podrobena pouze ta část korpusu CzeSL-SGT obsahující texty studentů, jejichž mateřským jazykem byl jeden z jedenácti slovanských jazyků¹² (viz tabulka 1); tuto část zde označujeme jako slovanský subkorpus. Slovanský subkorpus zahrnuje 5 625 textů napsaných 1 181 různými studenty, čili v průměru jeden student vyprodukoval 4,76 textu, nicméně chyby byly identifikovány v 5 611 textech. Celková velikost subkorpusu činí 769 126 pozic, z čehož 99 825 pozic bylo automatickou chybovou analýzou označeno za chybné, tj. 12,98 % z celkové velikosti slovanského subkorpusu. Naše pozornost však bude věnována pouze pěti nejčastějším typům chyb (viz tabulka 2), jejichž celková frekvence činí 74 222 a pokrývá 9,65 % z celkové velikosti slovanského subkorpusu, což činí 74,35 % z celkového počtu chyb v subkorpusu. Tato analýza pokrývá většinu prominentních obtíží cizinců s českým jazykem.

Tabulka 1. Mateřské jazyky studentů ve slovanském subkorpusu

	Jazyk	Počet tokenů	Počet studentů
1.	ru – ruština	684 177	1 036
2.	uk – ukrajinština	49 436	76
3.	pl – polština	19 304	37
4.	bg – bulharština	5 541	10
5.	be – běloruština	3 624	6
6.	mk – makedonština	2 534	4
7.	sk – slovenština	1 504	4
8.	sr – srbština	1 397	3
9.	hr – chorvatština	819	1
10.	sl – slovinština	470	2
11.	sh – srbochorvatština ¹³	320	2
	Celkem	769 126	1 181

Tabulka 2. Pět nejčastějších automaticky identifikovaných chyb v subkorpusu

	Typ chyby	Popis chyby	Počet studentů	Počet tokenů
1.	Quant0	chyba v diakritice – chybí čárka nad vokálem	1 163	37 574
2.	SingCh	chyba vzniklá záměnou jednoho znaku za druhý	1 146	13 898
3.	Quant1	chyba v diakritice – čárka nad vokálem navíc	1 075	10 029
4.	Caron0	chyba v diakritice – chybí háček	1 045	7 565
5.	Cap1	chybně použité velké písmeno	937	5 156
	Celkem			74 222

¹² Případy, kde prvním jazykem byla čeština (převážně texty Čechů žijících mimo ČR), byly vyloučeny z analýzy.

¹³ Dnes již nepoužívaný pojem srbochorvatština vychází z údajů o metainformacích přiřazených k jednotlivým textům.

Nejpočetnější chyby byly identifikovány u autosémantik (substantiv, verb a adjektiv), proto dále v rámci jednotlivých typů chyb budou slovní druhy analyzovány podrobněji. Tabulka 3 prezentuje počet automaticky identifikovaných chyb u slovních druhů v rámci jednotlivých typů chyb.

Tabulka 3. Slovní druhy s největší chybovostí vůči nejčastějším typům chyb

Typ chyby Slovní druh	Quant0	SingCh	Quant1	Caron0	Cap1	Ostatní typy chyb
Substantivum	35,69 %	39,98 %	34,95 %	36,60 %	35,86 %	37,59 %
Verbum	30,73 %	21,57 %	34,34 %	25,14 %	10,45 %	22,68 %
Adjektivum	17,84 %	13,51 %	8,69 %	9,68 %	27,44 %	17,61 %
Pronominum	7,93 %	11,41 %	8,40 %	9,08 %	9,48 %	6,12 %
Adverbium	6,19 %	7,51 %	9,75 %	15,35 %	4,97 %	6,16 %
Ostatní	1,62 %	6,02 %	3,87 %	4,15 %	11,79 %	9,84 %
Všechny	100 %	100 %	100 %	100 %	100 %	100 %

Za zmínku stojí vliv jazykové úrovně studentů podle SERR na jednotlivé typy chyb. Jelikož počet textů pro jednotlivé úrovně je odlišný, pro srovnání jsme použili relativizovanou míru ipm (*instances per milion*), která měří výskyt jednotlivého jevu v hypotetickém milionovém korpusu. Tabulka 4¹⁴ potvrzuje očekávání, že největší problémy mívají studenti na nižších úrovních znalosti jazyka.

Tabulka 4. Úroveň studentů podle SERR, u kterých se vyskytla chyba

Úroveň	Chybné tokeny		
	Počet tokenů	ipm	Počet studentů
A1	34 712	142 440,35	397
A2	32 064	128 598,59	366
B1	17 135	123 308,87	223
B2	10 646	122 829,49	119
n/a¹⁵	3 842	116 533,71	46
C1	1 426	81 509	30

Z hlediska pohlaví byla většina tokenů vyprodukována ženami (525 767 vs. 239 598), ale podíváme-li se na obě pohlaví z hlediska jazykových chyb, na základě relativizované míry ipm zjistíme, že častěji se chyb dopouštěli muži.

¹⁴ V korpusu CzeSL-SGT jsou úrovně A1 a A2 rozdělené do dvou podskupin. Abychom ponechali tradiční způsob dělení úrovně jazykových dovedností podle SERR, spojili jsme obě podskupiny v rámci jedné úrovně v tabulce 4.

¹⁵ V případě chybějící informace v metadatech je položka označena jako „n/a“.

Tabulka 5. Počet chyb podle pohlaví v slovanském subkorpusu

Pohlaví	Chybné tokeny		
	Počet tokenů	ipm	Počet studentů
žena	62 930	119691,8	748
muž	36 268	151370,21	425
n/a	627	166710,98	8

Dalším hlediskem je věková kategorie studentů.¹⁶ Texty subkorpusu se ve většině skládají z textů studentů ve věku 16 let a starších. Pokud vezmeme v potaz relativizovanou míru ipm, chyb se častěji dopouštějí žáci mladší 16 let.

Tabulka 6. Počet chyb podle věkové kategorie v slovanském subkorpusu

Věková kategorie	Chybné tokeny		
	Počet tokenů	ipm	Počet studentů
6–11	68	373 626,37	2
12–15	446	149 213,78	10
≥16	98951	129 784,91	1 165
n/a	360	101 925,25	4

Zajímavým údajem je počet chyb s ohledem na učebnici, kterou používali studenti ke studiu češtiny. Výsledky ukazují, že se nejčastěji chybné tokeny vyskytují u studentů, kteří studovali z učebnic *Basic Czech*, *New Czech Step by Step*, *Easy Czech Elementary* a *Communicative Czech*, patrně z toho důvodu, že jde o materiály pro začátečníky, kteří se v počáteční fázi dopouštějí chyb výrazně častěji.

Tabulka 7. Počet chyb ve slovanském subkorpusu podle učebnice používané při studiu češtiny

Učebnice	Chybné tokeny		
	Počet tokenů	ipm	Počet studentů
BC – Basic Czech	451	166 175,39	5
NCSS – New Czech Step by Step	14 627	138 992,36	176
ECE – Easy Czech Elementary	39	133 105,8	2
CC – Communicative Czech	57 688	128 822,49	618
CE – Čeština pro ekonomy	5 276	128 002,33	80
jiné	25 923	125 740,31	324
n/a	7 674	117 308,96	125

¹⁶ Korpus CzeSL-SGT obsahuje i konkrétní údaj o věku jednotlivých studentů.

CMC – Chcete mluvit česky?	10 712	113 451,74	61
CMS – Chcete ještě lépe mluvit česky?	1 320	107 658,43	6
CpC – Čeština pro cizince	187	81 339,71	2

4.3.2 Pět nejčastějších chyb podle slovního druhu

Tato podkapitola podrobně analyzuje pět nejčastějších automaticky identifikovaných chyb ve slovanském subkorpusu (viz tabulka 2), které pokrývají 74,35 % všech chyb v tomto subkorpusu. Jedná se o chyby v diakritice, kde chybí čárka nad vokálem (Quant0), chyby vzniklé záměnou jednoho znaku za druhý (SingCh), chyby v diakritice, kde je čárka nad vokálem navíc (Quant1), chyby v diakritice, kde chybí háček (Caron0), a také chybné použití velkého písmena (Cap1). Součástí analýzy jsou i seznamy nejproblematictějších slov. V hlavním textu jsou prezentovány seznamy slov z frekvenční špičky.

4.3.3 Shrnutí nejčastějších problémů

Podrobili jsme analýze pět nejčastějších typů chyb v korpusu CzeSL-SGT: Quant0, SingCh, Quant1, Caron0 a Cap1. Tabulka 8 prezentuje jejich procentuální podíl na celkovém počtu chyb ve zvoleném slovanském subkorpusu.

Tabulka 8. Procentuální podíl zkoumaných typů chyb na celkovém počtu chyb

	Quant0	SingCh	Quant1	Caron0	Cap1	Zbytek chyb	Celkem
Počet tokenů	37 574	13 898	10 029	7 565	5 156	25 603	99 825
Procento	37,64 %	13,92 %	10,05 %	7,58 %	5,17 %	25,65 %	100 %

Výběr prvních pěti typů chyb umožnil prozkoumat nejběžnější problémy studentů-cizinců. Naše analýza pokrývá většinu chybných tvarů, jež se vyskytly ve slovanském subkorpusu; byly zde analyzovány téměř tři čtvrtiny všech chyb ve zkoumaném subkorpusu. V analyzovaných typech chyb se jednalo o chyby gramatické a pravopisné, které jsme vždy zkoumali s ohledem na slovní druh.

V některých případech analýza ukázala společné tendence napříč typy chyb. Nejvíce společných problémů se pochopitelně projevilo u chyb týkajících se délkou nad vokálem, tj. Quant0 a Quant1, obdobné problémy však nastávaly i v případě dalších chyb SingCh a Caron1. Společnou obtíž pro tyto čtyři typy chyb byly kořenové hlásky a koncovky. Typ Cap1 zastupoval odlišný problém, u nějž nebylo vždy možné určit příčinu. Tabulka 9 srovnává nejčastější specifikace chyby v rámci typů chyb napříč slovními druhy.

Tabulka 9. Nejčastější chyby v rámci typů chyb napříč slovními druhy

Typ chyby Slovní druh	Quant0	SingCh	Quant1	Caron0	Cap1
N	kořenová hláska	koncovka	kořenová hláska	kořenová hláska	velké písmeno
V	kořenová hláska	kořenová hláska	vokál před -t	kořenová hláska	velké písmeno

A	koncovka	koncovka	kořenová hláska	kořenová hláska	velké písmeno
P	koncovka	zájmeno <i>mně/mě/mi</i>	koncovka	předpona <i>ně-</i>	velké písmeno
D	kořenová hláska	zakončení slova	zakončení slova	zakončení slova	velké písmeno

Z tabulky jsou jasně patrné společné tendence různých typů chyb. Největší problém představují kořenové hlásky, dále koncovky a u adverbí zakončení. Naším cílem bylo v těchto chybách najít pravidelnosti, které lze jednoduše popsat, generalizovat na jejich základě pravidla a následně podle nich vytvořit korpusová cvičení pro studenty. Tyto poznatky tak využijeme v dalších kapitolách této práce a zjištěné problémy podrobně rozebereme s ohledem na výuku ČCJ pro slovansky mluvící studenty.

4.4 Shrnutí analýzy chyb

Cílem této kapitoly bylo zmapování nejproblematictějších oblastí ve výuce ČCJ na základě žákovského korpusu CzeSL-SGT. Díky analýze chyb jsme identifikovali oblasti, v nichž nastávají obtíže s českým jazykem u studentů-cizinců a lze je řešit za pomoci korpusových nástrojů. K těm nejzajímavějším, kterým se budeme věnovat v další kapitole v konkrétních případových studiích, patří:

- deverbativní zakončení na *-ání* vs. *-aní*,
- tvary přičestí minulého,
- konkurence genitivních koncovek singuláru *-a* a *-u*.

Navíc analýza chyb v žákovském korpusu odhalila další chyby, které lze jednoduše řešit prostými korpusovými cvičeními, jež nabízíme v didaktickém experimentu v kapitole 6:

- časování verb v přítomném čase s ohledem na kvantitu,
- koncovky v gramatických pádech,
- kořenové délky autosémantik a zájmen.

Provedená analýza chyb tak slouží jako zdroj dat pro další identifikaci konkrétních problémů cizinců s češtinou, kterými se zabývá následující kapitola. V ní se snažíme nabídnout na konkrétních příkladech ucelený rámec, jak přistupovat ke korpusovým datům ve výuce cizích jazyků.

5 Korpusový přístup ve výuce: tři případové studie

Tato kapitola nabízí ucelený rámec korpusového přístupu ve výuce, který se skládá ze tří hlavních etap: identifikace jazykového problému, jeho kategorizace a generalizace pravidel. V neposlední řadě jsou zde nabízena konkrétní korpusová cvičení určená k procvičování dané problematice látky. Ve třech případových studiích se snažíme přesně popsat, jak ke korpusovým datům přistupovat a jak z nich vytěžit maximum. Námi prezentovaný postup má také za cíl napomoci jak učitelům, tak i studentům, kteří by měli zájem korpusové metody využívat a chybí jim přesný postup a konkrétní příklady.

V této části jsou zevrubně rozebírány tři jazykové jevy, jež byly identifikovány jako problematické v předchozí kapitole: deverbativní zakončení na *-ání* vs. *-aní* (5.1), tvary přičestí minulého s ohledem na kvantitu hlásky před příponou *-l* (5.2) a konečně konkurence genitivních koncovek v singuláru *-a* a *-u* (5.3).

5.1 Deverbativní zakončení na *-ání* vs. *-aní*

Jak bylo zjištěno na základě automatické analýzy chyb v žákovském korpusu CzeSL-SGT, psaní dlouhého vokálu je nejčastější chybou v textech nerodilých mluvčích češtiny. Tento problém se může projevat jak v kořeni slova, tak v koncovekách.

Tato studie se zabývala problematikou dlouhého vokálu u verbálních substantiv končících na *-aní*, anebo *-ání*. Na základě dat z korpusu SYN2015 jsme identifikovali množinu verbálních substantiv s oběma zakončeními. Ověřili jsme, že volba mezi zakončením *-ání*, nebo *-aní* nastává pouze v případě dvouslabičných verbálních substantiv, jak bylo popsáno v dřívějších pracích (Gebauer, 1958, pp. 107–108; Zdeňková, 2018). Víceslabičná verbální substantiva mají vždy zakončení *-ání*. Správnou variantu dvouslabičných lexémů lze z hlediska jazykové výuky určit na základě infinitivu, u kterého buď dochází ke krácení kořenové samohlásky u prefigovaných infinitivů (zakončení *-aní*), nebo k ponechání délky (zakončení *-ání*). Dále jsme nabídli konkrétní *hands-off* cvičení, abychom demonstrovali konkrétní korpusové řešení tohoto problematického jevu.

Tato práce přistupuje novým způsobem k tematice verbálních substantiv na *-ání* a *-aní* z hlediska jazykové výuky. Pomocí korpusových dat jasně klasifikuje lexémy a určuje pravidla pro použití vhodné varianty. Taktéž rozporuje tvrzení, že verbální substantiva na *-ání* a *-aní* mají nepravidelnou distribuci obou variant (Karlík et al., 1995, p. 149). Lze konstatovat, že i přes určité výjimky je soudobý stav nejfrekventovanějších verbálních substantiv na *-ání* a *-aní* stabilní a naprosto pravidelný. Můžeme však samozřejmě očekávat, že v následujících dekádách se zde prezentované kolísající varianty mohou ustálit ve prospěch jedné z nich.

5.2 Tvary přičestí minulého

Současný výzkum vychází z dřívější studie (Zasina, 2014), v níž však nebyl brán zřetel na jazykovou výuku a data ze žákovského korpusu. Proto jsme tento výzkum zopakovali a tentokrát použili nová data, abychom mohli vytvořit relevantní podklady pro studenty ČCJ.

Nová analýza byla publikována v konferenčním sborníku (Zasina, 2018b) a zde ji předkládáme v rozšířené podobě.

Tato případová studie se obdobně jako předchozí práce v podkapitole 5.1 zabývala otázkou kvantity. V této části jsme se soustředili na problém s kvantitou samohlásky před příponou *-l* v přičestí minulém, která již sice byla dříve diskutována (viz Osolsobě 2011), ale pouze z úhlu pohledu rodilého mluvčího. Proto bylo nutné prozkoumat tuto oblast s ohledem na studenty-cizince. Provedený výzkum je tak prvním pokusem o systematický popis skupiny přičestí činných zachovávajících si dlouhou samohlásku před příponou *-l* s ohledem na výuku ČCJ.

Za pomoci korpusové analýzy jsme identifikovali analyzovaná slovesa a pokusili jsme se o jejich klasifikaci. Nabídlí jsme rozdělení přičestí činných do příbuzných skupin podle slovo tvorné báze, což napomáhá ve studiu/výuce ČCJ. Přijaté rozdělení na derivační skupiny usnadňuje zapamatování si více než stovky jednotlivých tvarů, sdružených pod dvacet kategorií. Na základě výzkumu byl dále sestaven seznam nejfrekventovanějších sloves, jež jsou podstatná při osvojování tohoto jazykového jevu nerodilým mluvčím. Seznam může být dále rozšiřován podle jazykové úrovně studentů.

V neposlední řadě byla nabídnuta čtyři korpusová cvičení jako příklad aktivit, které nevyžadují přímou práci s počítačem. Jsou ukázkou aplikace korpusového přístupu ve výuce a postupně uvádějí studenty do problematiky kvantity přičestí minulých.

5.3 Konkurence genitivních koncovek singuláru *-a* a *-u*

Konkurence koncovek *-a* a *-u* v genitivu singuláru maskulin neživotných (dále jen Gsg Mi) je další problematickou oblastí pro studenty-cizince, na kterou poukázala analýza chyb v žákovském korpusu. V rámci typu chyb SingCh se jednalo o nejpočetnější jazykový jev, v němž studenti chybují.

Tematické konkurenčních koncovek Gsg Mi bylo již věnováno mnoho pozornosti, dosud však nebyl brán zřetel na popis tohoto jevu ve výuce ČCJ pomocí korpusových metod. Za 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* je vůči zvětšující se jazykové úrovni nepřímě úměrná a častěji dochází k nahrazování správné koncovky *-u* koncovkou *-a*. Dále byla na základě dat ze synchronního korpusu češtiny variantní substantiva seříděna do tří kategorií: substantiva mající stejný etymologický původ i význam, substantiva mající stejný etymologický původ, avšak rozdílný význam a substantiva 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 z hlediska četnosti výskytu vhodnější pro používání nerodilými mluvčími. Nadto je zdrojem pro sestavení vhodných korpusových cvičení, jejichž tři příklady zde prezentujeme. Dvě přímá cvičení se zabývají konkurenčními koncovkami a student díky nim identifikuje častější variantu, zatímco nepřímé cvičení upozorňuje studenta na možné sémantické rozdíly v použití té které koncovky.

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. Tato procedura systematizuje práci s korpusem pro vyučovací účely a zároveň může být pokynem a inspirací pro pedagogy, jak s korpusem pracovat.

Korpusová data v rámci substantiv se stejným etymologickým původem a významem 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*. Substantiva se zakončením na *-in*, *-en*, *-ol* a *-id* mají tendenci vyskytovat se s koncovkou *-u* u názvů chemických látek. Dále preferují koncovku *-u* zpravidla názvy geometrických útvarů a značek aut.

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

6 Didaktický experiment

V českém prostředí nebyla aplikace korpusového přístupu v přímé akademické výuce nikdy dříve podrobena analýze z hlediska funkčnosti a vlivu na studenty. Pro účel tohoto experimentu tak byl vytvořen na Jagellonské univerzitě speciální vysokoškolský kurz pro bohemisty zaměřený na gramaticko-lexikální cvičení. Aby byla možná evaluace korpusového přístupu ve výuce, byl experiment proveden na dvou skupinách, z čehož jedna byla skupinou kontrolní, vedenou tradičním způsobem výuky, kdežto druhá skupina využívala většinou korpusové materiály. Náplň kurzu vyplývala z dřívější identifikace obtížných oblastí v akvizici češtiny (viz kapitola 4), z plánů výuky studentů a také z učitelových poznatků. Takto připravený předmět zahrnoval následující oblasti: vybrané pády substantiv, adjektiv a zájmen, problémy s kvantitou u časování sloves, dále kvantitu v přičestí minulém, deverbativa, rozlišování zájmen *mně/mě* a *můj/svůj*, tvoření adverbíí, vokalizace předložek, stylové rozdíly mezi psanou a mluvenou češtinou, přirovnání popisující vlastnosti, ustálená a neustálená slovní spojení a frazémy.

V této části jsme podrobně popsali průběh celého semestrálního kurzu s ohledem na využití korpusových metod tak, aby tento výzkum mohl být replikovatelný. Každý korpusový seminář byl krátce popsán a ke každému byl přiložen detailní plán hodiny a handout pro studenty. Následně jsme předložili analýzu výsledků výuky, založenou pouze na výsledcích cvičení, která byla během semestru vyučována korpusově, a zkoumali jsme zlepšení korpusové skupiny oproti tradiční. Zde jsme porovnávali výsledky vstupního a závěrečného testu. Ukázalo se, že skupina s experimentální výukou vynikala v těchto oblastech: gramatický rod, vokalizace předložek, genitiv singuláru a plurálu, měkká a tvrdá adjektiva, osobní zájmena. Tyto poznatky potvrdila další analýza pěti vybraných cvičení, jež byla stejná ve vstupním, druhém průběžném a závěrečném testu. Vývojová tendence ve všech pěti cvičeních u experimentální skupiny byla kontinuální a postupně stoupala, zatímco v případě tradiční skupiny došlo k propadu u jednoho cvičení a u ostatních byla tendence vzrůstu oproti korpusové skupině výrazně mírnější. V poslední fázi jsme podrobili analýze všechny čtyři testy. Ukázalo se, že rozdíly mezi skupinami nejsou výrazné, i když korpusová skupina dosáhla mírně lepších výsledků, avšak byla zastoupena menším počtem studentů, což nám znemožňuje větší generalizaci. Srovnání mediánu všech testů však naznačilo, že v dlouhodobé perspektivě má korpusová skupina tendenci k výraznějším zlepšením. Je třeba také zdůraznit, že jsme se ve svém experimentu zaměřovali na několik různých jazykových jevů, což rozhodně mělo vliv na to, že rozdíly mezi oběma skupinami nejsou tak výrazné. Samozřejmě bychom mohli klást větší zřetel na jeden konkrétní jev a zabývat se jím do detailu, to by však nebylo vhodné pro komplexní jazykovou výuku. Ve vyučování je totiž podstatné zajistit studentům co největší pestrost materiálu a rozvíjet podle možností veškeré jazykové dovednosti.

Cílem této studie bylo ověřit vliv korpusového přístupu ve výuce na efektivitu jazykové výuky a zda navrhovaná procedura, od identifikace jazykového problému přes jeho kategorizaci na základě korpusových dat až ke generalizaci výsledků, je funkční v praxi. Experiment potvrdil, že korpusová výuka stejně jako výuka tradiční dosahuje obdobných výsledků, ačkoli v několika oblastech má korpusový přístup šanci umožnit studentům dosažení lepších výsledků v dlouhodobé perspektivě. Navíc jsme potvrdili, že zde

prezentovaná jasná a systematizovaná procedura přispívá k upevnění probírané látky, což potvrdily i odpovědi ve studentském hodnocení. Korpusová skupina, i přes svou neochotu k práci s počítačem, dosáhla pozitivních výsledků.

Experiment však mohl být proveden pouze na malém počtu studentů (zvláště v korpusové skupině) a jeho výpovědní hodnota je tedy omezená, čehož jsme si vědomi. Domníváme se nicméně, že i tak lze konstatovat, že zavedení korpusové výuky má široký potenciál do budoucna, poněvadž, jak jsme prokázali, zvláště v určitých oblastech přispívá k výraznému zlepšení. Navíc je zpestřením nabídky jazykových kurzů a má jistě šanci na své uplatnění v moderní výuce.

Budeme rádi, když se náš experiment stane inspirací pro další badatele a učitele a poslouží k postupnému pronikání korpusových metod do výuky jazyků.

7 Závěr

Tato disertační práce je zacílena na korpusový přístup ve výuce češtiny jako cizího jazyka a snaží se systematicky zmapovat možnosti začlenění korpusů do běžné jazykové výuky, a to s ohledem na konkrétní jazykové potřeby cizinců na základě jejich typických chyb. I přes rostoucí zájem v této oblasti pořád zůstávaly stranou metodologické otázky týkající se konkrétních aplikací korpusů ve výuce. Tato práce se proto zaměřuje jak na formulování konkrétního přístupu při sestavování korpusových cvičení, tak na ověření funkčnosti nové metody v akademické výuce studentů bohemistiky. Základ práce tvoří podrobná analýza nejčastějších problematických oblastí studentů češtiny s jiným slovanským rodným jazykem a tři případové studie, jež zevrubně zkoumají nejčastější problematické jevy. Navíc zjištění plynoucí z analýzy jsou podpořena experimentem ověřujícím možnosti začlenění korpusové výuky v akademickém prostředí a zkoumajícím vliv nové metody na jazykový vývoj studentů.

V teoretické části seznamujeme čtenáře s obecným přehledem korpusového přístupu ve výuce a jeho vývojem v českém prostředí. Věnujeme se rovněž výhodám vyplývajícím z práce s korpusem, k nimž patří autentičnost textů, přirozený kontext hledaných lexikálních jednotek, způsob zobrazení výsledků ve formě konkordance, uspořádání textů podle žánrů umožňující zkoumání stylových rozdílů, možnost sledování syntagmatických vztahů (kolokabilita), rychlý přístup k datům či zpestření jazykové výuky. Nenecháváme však stranou ani omezení korpusového přístupu, k nimž patří nestandardní prvky v textech roditelých mluvčích, automatická anotace textů a její limity, nedostatečné pokrytí periferních jazykových jevů v menších korpusech, složitost korpusových nástrojů nebo rezervovanost učitelů a studentů vůči novým technologiím.

Východiskem práce je její metodologie, postavená na využití korpusových dat dvojným způsobem. Zprvč využívá data žákovského korpusu k identifikaci nejčastějších chyb cizinců. Zadruhé za pomoci synchronního korpusu SYN2015 s texty roditelých mluvčích identifikuje pravidelnosti současného jazyka a zároveň korpus slouží jako zdroj dat pro sestavení praktických korpusových cvičení. Metoda založená na principu „od korpusu ke korpusu“ tak zajišťuje co možná nejobektivnější přístup ke zkoumaným jevům.

Analytická část, tvořící jádro této disertace, je rozdělena do tří kapitol (kapitoly 4, 5 a 6). Čtvrtá kapitola mapuje nejproblematictější oblasti ve výuce ČCJ na základě žákovského korpusu CzeSL-SGT. V úvodu se věnujeme vymezení pojetí jazykové chyby, kterou považujeme za nestandardní tvar vůči jazykové normě, přičemž touto chybou spolehlivě odlišujeme cizojazyčného mluvčího od mluvčího roditelého. S tímto termínem pracujeme dále v analýze nejčastějších chyb. Zdrojem dat nám byl subkorpus zahrnující studenty se slovanským mateřským jazykem, v němž bylo automatickou chybovou analýzou identifikováno celkově 12,98 % chybných tokenů. K pěti nejčastějším typům chyby patří: 1) chyba v diakritice, kde chybí čárka nad vokálem (Quant0); 2) chyba vzniklá záměnou jednoho znaku za druhý (SingCh); 3) chyba v diakritice, kde čárka je nad vokálem nadbytečná (Quant1); 4) chyba v diakritice, kde chybí háček (Caron0); 5) chybně použité velké písmeno (Cap1). Tyto nejfrekventovanější chyby pokrývají celkově 9,68 % chybných tokenů. Každý z pěti typů chyb byl jednotlivě analyzován s ohledem na slovní druh. V potaz jsme brali pouze autosémantika a pronomina, kde je zkoumání chyb nejvíc relevantní. Některé typy chyb sdílely společné tendence jako např. chyby týkající se diakritiky, a některé naopak vykazovaly

úplně jiná specifika (typ Cap1). Pro substantiva byly nejproblematictější kořenové hlásky v typech Quant0, Quant1 a Caron0, zatímco pro typ SingCh koncovky. Dále pro verba v typech Quant0, SingCh a Caron0 byly problematickou oblastí kořenové hlásky a pro typ Quant1 vokál před příponou *-t*. V případě adjektiv v typech Quant0 a SingCh se ukázaly být prominentní koncovky, zatímco v typech Quant1 a Caron0 zase kořenové hlásky. U zájmen byly problematickou oblastí koncovky v typech Quant0 a Quant1, zájmeno *mně/mě/mi* v typu SingCh a prefix *ně-* v typu Caron0. V případě adverbíí byla obtížná zakončení slova v typech SingCh, Quant1 a Caron0, a kořenová hláska v typu Quant0. V případě typu chyby Cap1 se ve všech slovních druzích jednalo o problém s psaním velkých písmen, což je oblast problematická i pro rodilé mluvčí. K nejzajímavějším jevům, které jsme vybrali k další podrobné analýze v jednotlivých případových studiích, patřily deverbativní zakončení na *-ání* vs. *-aní*, tvary příčestí minulého a konkurence genitivních koncovek singuláru *-a* a *-u*.

Pátá kapitola se podrobně zaměřuje na problematické oblasti češtiny identifikované v předchozí kapitole. Je rozdělena do tří samostatných podkapitol – případových studií, kde na konkrétních příkladech ukazujeme celý postup korpusového přístupu. Tato kapitola tak nabízí ucelený rámec, jak přistupovat ke korpusovým datům ve výuce od identifikace problematického jazykového jevu v žákovském korpusu, přes jeho klasifikaci na základě korpusových dat z reprezentativního korpusu, až k finální generalizaci výsledků. Každou z těchto částí doplňují i autorská korpusová cvičení k řešení prezentovaných problémů.

Podkapitola 5.1 se zabývá problémem s kvantitou hlásky *á* u deverbativních verbálních substantiv končících na *-aní* anebo *-ání*. Na základě dat z korpusu SYN2015 jsme zjistili, že k volbě mezi zakončeními *-ání* nebo *-aní* dochází pouze v případě dvouslabičných verbálních substantiv a víceslabičná verbální substantiva mají vždy zakončení *-ání*. Díky jasné klasifikaci materiálu lze jednoduše určit správné zakončení dvouslabičného substantiva na základě jeho infinitivu. Dochází-li ke krácení kořenové samohlásky u prefigovaných infinitivů, má dvouslabičné substantivum zakončení *-aní*. Nedochozí-li ke krácení, má dvouslabičné substantivum zakončení *-ání*. Soudobý stav nejfrekventovanějších verbálních zakončení na *-ání* a *-aní* tak můžeme označit za pravidelný i přes určité výjimky. Lze očekávat, že se kolísající varianty ustálí ve prospěch jedné z nich.

V další podkapitole 5.2 jsme zkoumali slovesa zachovávající si dlouhou kořenovou samohlásku před příponou *-l* v příčestím minulém. Běžně kvantita kmenotvorné přípony infinitivu v příčestí minulém mizí, srov. *brát* : *bral*. Nicméně ve středu naší pozornosti byly případy, kdy samohláska před příponou *-l* kvantitu ponechává, např. *hrát* : *hrál*. Výzkum vycházel z dřívější analýzy (Zasina, 2014), jež tu byla provedena na nových datech a doplněna o korpusová cvičení. Navíc jsme vycházeli z výzkumu chyb v žákovském korpusu, který poukázal na důležitost tohoto problému v jazykové výuce. Díky korpusu SYN2015 jsme určili množinu všech sloves, jež si kvantitu v participiu ponechávají, a na základě analýzy gramatických konstrukcí více než stovky nalezených sloves jsme je rozdělili do 20 skupin. Klasifikace uspořádaná z derivačního hlediska umožňuje rychlejší zapamatování jednotlivých tvarů a ulehčuje jejich osvojení. Tento výzkum byl prvním pokusem o systematický popis skupiny příčestí činných ponechávajících dlouhou kořenovou samohlásku před příponou *-l* s ohledem na výuku ČJ.

Poslední případová studie, prezentovaná v podkapitole 5.3, se věnuje tématu konkurence genitivních koncovek singuláru *-a* a *-u*, kterému dosud nebyla věnována dostatečná pozornost

s ohledem na výuku. Analýza žakovského korpusu ukázala, že chybovost genitivních koncovek *-a* a *-u* je vůči zvětšující se jazykové úrovni nepřímo úměrná. Navíc si studenti častěji pletou správnou koncovku *-u* s *-a*. V další fázi jsme na základě materiálu z korpusu SYN2015 setřídili zkoumaná substantiva do tří skupin (substantiva mající stejný etymologický původ i význam, substantiva mající stejný etymologický původ, avšak rozdílný význam a substantiva s neprůkaznou konkurencí genitivních koncovek), jež systematizují popis konkurujících si dublet. Uvedení proporcionální a absolutní frekvence výskytu jednotlivých variant tak umožňuje nerodilému mluvčímu volbu vhodné dublety z hlediska četnosti výskytu. Tato studie, opírající se zároveň o hledisko kvantitativní i kvalitativní, tak přispívá k doplnění poznatků dříve provedených studií v této oblasti (srov. Bermel & Knittl, 2012a, 2012b).

Šestá kapitola je věnována experimentu provedenému v Krakově na Jagellonské univerzitě. Cílem bylo prověřit možnosti zavedení korpusových metod do jazykové výuky a zjistit v praxi, zda takový přístup přináší měřitelné výhody v jazykovém vývoji studentů. Výzkum byl proveden na skupině studentů prvního ročníku bohemistiky, kteří byli rozděleni do dvou skupin: kontrolní a experimentální. Experimentální byla vedena převážně korpusovou metodou výuky, zatímco kontrolní skupina pouze tradičním způsobem bez využití korpusových metod a materiálů. Za účelem tohoto experimentu byl připraven speciální kurz *Gramaticko-lexikální cvičení pro bohemisty*, vyučovaný v letním semestru akademického roku 2017/2018. Obě skupiny byly vedeny stejným učitelem, autorem této práce. Výuka byla sestavena na základě provedené analýzy nejčastějších problematických oblastí češtiny v žakovském korpusu a také na základě programu studia a učitelových dřívějších zkušeností s výukou. Srovnání výsledků skupin bylo provedeno na základě vstupního a závěrečného testu a rovněž dvou průběžných testů. Podrobná analýza ukázala, že rozdíly mezi skupinami nejsou nikterak dramatické a že obě dosáhly prospěšného zlepšení. Nicméně v několika oblastech (gramatický rod, genitiv singuláru a plurálu, měkká a tvrdá adjektiva, vokalizace předložek, osobní zájmena), kde byl použit korpusový přístup ve výuce, dosáhla korpusová skupina výrazně většího zlepšení. Experiment tak potvrdil, že navrhovaná procedura od identifikace jazykového problému přes jeho kategorizaci na základě korpusových dat k následné generalizaci výsledků je v praxi funkční. Navíc se ukázalo, že korpusová skupina má v dlouhodobé perspektivě tendenci k výraznějšímu a trvalejšímu zlepšení.

Náš rozsáhlý výzkum věnovaný rychle se rozvíjející oblasti korpusového přístupu ve výuce má tak dvojí hlavní přínos: zaprvé se nám podařilo identifikovat nejčastější problematické oblasti češtiny nerodilých mluvčích, zadruhé jsme se pokusili systematizovat a navrhnout ucelený přístup ke zkoumání jazykových dat s ohledem na výuku. Doufáme, že nabízený korpusový přístup ve výuce ČCJ bude mít pozitivní vliv na další vývoj této disciplíny a v neposlední řadě přispěje učitelům i studentům k hlubšímu poznání zákonitostí českého jazyka a tím i ozvláštní, usnadní a urychlí proces osvojování cizí řeči.

8 Kompletní seznam literatury

- Adamovičová, A., & Hrdlička, M. (2010). *Basic Czech III*. Praha: Karolinum.
- Adamovičová, A., Ivanovová, D., & Hrdlička, M. (2012). *Basic Czech II*. Praha: Karolinum.
- Barlow, M. (2000). Parallel texts in language teaching. In S. P. Botley, A. M. McEnery, & A. Wilson, (Eds.), *Multilingual corpora in teaching and research* (pp. 106–115). Amsterdam: Rodopi.
- Bednaříková, B. (2011). *Cvičebnice z morfologie se zaměřením na flexi přejatých slov*. Olomouc: Univerzita Palackého v Olomouci.
- Bennett, G. R. (2010). *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. Ann Arbor, MI: University of Michigan Press.
- Bermel, N. (1993). Sémantické rozdíly v tvarech českého lokálu. *Naše řeč*, 76, 192–198.
- Bermel, N. (2004). V korpuse nebo v korpusu? Co nám řekne (a neřekne) ČNK o morfologické variaci v tvarech lokálu. In Z. Hladková, & P. Karlík (Eds.), *Čeština – univerzália a specifika 5* (pp. 163–171). Praha: Nakladatelství Lidové Noviny.
- Bermel, N. (2010). Variace a frekvence variant na příkladu tvrdých neživotných maskulin. In S. Čmejrková, J. Hoffmannová, & E. Havlová (Eds.), *Užívání a prožívání jazyka* (pp. 135–140), Praha: Karolinum.
- 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. doi: 10.1515/cllt-2016-0032
- Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. McH. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 15–36). Amsterdam: John Benjamins.
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. New York, NY: Cambridge University Press.
- Bígl, R. (2013). *Vývoj lužickosrbského skloňování a stupňování*. Praha: Karolinum.
- Boccou Kestránková, M. (2016). Postavení obecné češtiny ve výuce češtiny pro cizince. *Studie z aplikované lingvistiky*, 7(2), 72–85.
- Boccou Kestránková, M., Hlínová, K., Pečený, P., & Štěpánková, D. (2013). *Čeština pro cizince – učebnice: úroveň B2*. Brno: Edika.
- Breyer, Y. (2011). *Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges*. Frankfurt am Main: Peter Lang.

- Brown, H. D. (2000). *Principles of language learning and teaching* (4th ed.). White Plains, NY: Longman. Retrieved from https://latestacna.files.wordpress.com/2012/04/principles_of_language_learning_and_teaching.pdf
- Conrad, S., & Biber, D. (2009). *Real Grammar: A Corpus-based Approach to English*. Upper Saddle River, NJ: Pearson Education.
- Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, 5(4), 161–170. Retrieved from <https://files.eric.ed.gov/fulltext/ED019903.pdf>
- Cvrček, V., Čermáková, A., & Křen, M. (2016). Nová koncepce synchronních korpusů psané češtiny. *Slovo a slovesnost*, 77(2), 83–101.
- Cvrček, V., & Kodýtek, V. (2013). Ke klasifikaci morfologických variant. *Slovo a slovesnost*, 74(2), 139–146.
- Cvrček, V., Kodýtek, V., Koprivová, 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., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2018a). From extra- to intratextual characteristics: Charting the space of variation in Czech through MDA. *Corpus Linguistics and Linguistic Theory*. Advance online publication. doi: 10.1515/cllt-2018-0020
- Cvrček, V., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2018b). Variabilita češtiny: multidimenzionální analýza. *Slovo a Slovesnost*, 79(4), 293–321.
- Cvrček, V., & Vondříčka, P. (2011a). SyD – Korpusový průzkum variant (Version 2.2) [Corpus tool]. Retrieved from <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* (pp. 184–195). Praha: Nakladatelství Lidové noviny.
- Cvrček, V., & Vondříčka, P. (2013a). Morfio [Corpus tool]. Retrieved from <https://morfio.korpus.cz/>
- Cvrček, V., & Vondříčka, P. (2013b). Nástroj pro slovotvornou analýzu jazykového korpusu. *Gramatika a korpus 2012*. Hradec Králové: Gaudeamus. Retrieved from http://www.ujc.cas.cz/miranda2/export/sitesavcr/ujc/veda-vyzkum/vyzkum/gramatika-a-korpus/proceedings-2012/konferencni-prispevky/CvrcekVaclav_VondrickaPavel.pdf
- Čermák, F. (2010). *Lexikon a sémantika*. Praha: Nakladatelství Lidové noviny.
- Čermák, F. (2014). *Periferie jazyka: slovník monokolokabilních slov*. Praha: Nakladatelství Lidové noviny.
- Čermák, F., & Křen, M. (2011). *A Frequency Dictionary of Czech: Core Vocabulary for Learners*. London: Routledge.
- Davies, M., & Gardner, D. (2013). *A frequency dictionary of contemporary American English: Word sketches, collocates and thematic lists*. London: Routledge.
- Divjak, D. (2018, December). *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.
- Dvořák, V. (2017). Verbální substantivum. In P. Karlík, M. Nekula, & J. Pleskalová (Eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. Retrieved from <https://www.czechency.org/slovník/VERBÁLNÍ SUBSTANTIVUM>
- Filipec, J., Daneš, F., Machač, J., & Mejstřík, V. (Eds.). (1994). *Slovník spisovné češtiny pro školu a veřejnost*. Praha: Academia.
- Frankenberg-Garcia, A. (2012). Integrating corpora with everyday language teaching. In J. Thomas, & A. Boulton, (Eds.), *Input, Process and Product: Developments in Teaching and Language Corpora* (pp. 36–53). Brno: Masaryk University Press.
- Gebauer, J. (1958). *Historická mluvnice jazyka českého: Díl III. Tvarosloví, II. Časování*. Praha: Nakladatelství Československé akademie věd.
- Geist, M., & Hahn, A. (2012). Using a corpus for written production: A classroom study In J. Thomas, & A. Boulton, (Eds.), *Input, Process and Product: Developments in Teaching and Language Corpora* (pp. 123–135). Brno: Masaryk University Press.
- Gilquin, G., Granger, S., & Paquot, M. (2007). Learner corpora: The missing link in EAP pedagogy. *Journal of English for Academic Purposes*, 6(4), 319–335. doi: 10.1016/j.jeap.2007.09.007
- Gráf, T. (2017). *LINDSEI CZ: korpus spontánní mluvené angličtiny pokročilých mluvčích*. Praha: Ústav Českého národního korpusu FF UK. Retrieved from <http://www.korpus.cz>
- Granath, S. (2009). Who benefits from learning how to use corpora? In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 47–65). Amsterdam: John Benjamins.
- Hendrich, J. (1988). *Didaktika cizího jazyka*. Praha: Státní pedagogické nakladatelství.
- Holá, L. (2006). *New Step by Step*. Praha: Akropolis.
- Holá, L. (2008). *Spisovná vs. obecná čeština ve výuce češtiny jako cizího jazyka*. Retrieved from http://www.czechstepbystep.cz/clanky/spisovna_sv_obecna.html
- Holá, L. (2016). *Česky krok za krokem 1*. Praha: Akropolis.
- Holá, L., & Bořilová, P. (2009). *Česky krok za krokem 2*. Praha: Akropolis.
- Holá, L., & Bořilová, P. (2014). *Čeština Express 3*. Praha: Akropolis.
- Hrdlička, M. (2012). Jazyková chyba a práce s ní v jazykovém vyučování. In K. Šebesta & S. Škodová (Eds.), *Čeština – cílový jazyk a korpusy* (pp. 89–108). Liberec: Technická univerzita v Liberci.
- Hudousková, A. (2014). Jmenné koncovky v češtině pro cizince – distribuce, frekvence a fonetika. první sonda. In V. Petkevič, A. Adamovičová, & V. Cvrček (Eds.), *Radost z jazyků. Sborník k 75. narozeninám prof. Františka Čermáka* (pp. 215–230). Praha: Nakladatelství Lidové noviny.
- Jalková, (1998). *Čeština pro cizince I*. České Budějovice: Jihočeská univerzita v Českých Budějovicích.
- Janda, L. A. (1996). *Back from the brink*. München: LINCOM Europa.
- Jelínek, T. (2008). Nové značkování v Českém národním korpusu. *Naše řeč*, 91(1), 13–20.

- Jelínek, T., & Petkevič V. (2011). Systém jazykového značkování současné psané češtiny. In V. Petkevič, & A. Rosen (Eds.), *Korpusová lingvistika Praha 2011, sv. 3: Gramatika a značkování korpusů* (pp. 143–170). Praha: Nakladatelství Lidové noviny.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *Classroom Concordancing: ELR Journal*, 4, 1–16.
- Junková, B. (1998). *Čeština pro cizince IV*. České Budějovice: Jihočeská univerzita v Českých Budějovicích.
- Kaniewski, T. A. (2009). *Čeština. Podstawowy kurs języka czeskiego*. Nowa Ruda: Wydawnictwo "Maria".
- Karlík, P., Nekula, M., & Rusínová, Z. (Eds.). (1995). *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.
- Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams, & S. Vessier (Eds.), *Proceedings of the Eleventh EURALEX International Congress* (pp. 105–116). Lorient: Université de Bretagne-Sud.
- Kočařová, B. (2013). *Korpus a jmenný rod ve výuce češtiny jako cizího jazyka* (Bakalářská práce, Masaryková univerzita, Brno). Retrieved from <https://is.muni.cz/th/fvi7f/>
- Kolářová, V. (2010). *Valence deverbativních substantiv v češtině: (na materiálu substantiv s dativní valencí)*. Praha: Karolinum.
- Konečná, H., & Zasina, A. J. (2014). Studium českého jazyka a internet. In E. Rusínová (Ed.), *Přednášky a besedy ze XLVII. běhu LŠSS* (pp. 104–112). Brno: Masarykova univerzita.
- Kopřivová, M., Komrsková, Z., Lukeš, D., Poukarová, P., & Škarpová, M. (2017). *ORTOFON: korpus neformální mluvené češtiny s víceúrovňovým přepisem, verze 1 z 2. 6. 2017*. Praha: Ústav Českého národního korpusu FF UK. Retrieved from <http://www.korpus.cz>
- Krejčová, E. (2013). *Gramatika českého jazyka pro cizince. Pro mírně pokročilé studenty*. České Budějovice: Jihočeská univerzita v Českých Budějovicích.
- 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. Retrieved from <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)* (pp. 2522–2528). Portorož: ELRA.
- Kupka, P. (2009). *Přehled české gramatiky*. Praha: Kupka nakladatelství.
- Ledajaksová, I., & Rozkopalová, H. (1997). *Čeština pro cizince II*. České Budějovice: Jihočeská univerzita v Českých Budějovicích.
- Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 11–23). Harlow: Addison Wesley Longman.
- Leray, M., & Tyne, H. (2016). Homophonie et maîtrise du français écrit: apport de l'apprentissage sur corpus. *Linguistik online*, 78(4). Retrieved from <https://hal.archives-ouvertes.fr/hal-01401956>

- Lewandowska, A. (2014). Using corpus-based classroom activities to enhance learner autonomy. *Konińskie Studia Językowe*, 2(3), 237–255.
- Lukšija, M. (2010). *Korpus jako zdroj dat při prezentaci předložek do/na s místním směrovým významem ve výuce češtiny pro cizince* (Bakalářská práce, Masaryková univerzita, Brno). Retrieved from http://is.muni.cz/th/217240/ff_b/
- Lukšija, M. (2012). *Korpusy a česká deklinace ve výuce češtiny jako cizího jazyka* (Diplomová práce, Masaryková univerzita, Brno). Retrieved from http://is.muni.cz/th/217240/ff_m/
- Machálek, T., & Křen, M. (2013). Query interface for diverse corpus types. In K. Gajdošová, & A. Žáková (Eds.), *Natural Language Processing, Corpus Linguistics, E-learning* (pp. 166–173). Lüdenscheid: RAM Verlag.
- Malá, Z. (2012). *Česky krok za krokem 2. Pracovní sešit 1–10*. Praha: Akropolis.
- Malá, Z. (2016). *Česky krok za krokem 2. Pracovní sešit 11–20*. Praha: Akropolis.
- Marešová, J. (2013). *Bajky*. Praha: Axióma.
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- Nekovařová, A. (2006). *Čeština pro život*. Praha: Akropolis.
- Netrvalová, S. (2015, August 24). Chaos v psaní velkých písmen skončí, vědci chystají změnu pravopisu. *Mladá fronta Dnes*. Retrieved from <https://www.idnes.cz>
- Omyła-Rudzka, M. (2018). *Komunikat z badań. Stosunek do innych narodów*. Warszawa: Centrum Badania Opinii Społecznej.
- Osolsobě, K. (2010). Jak se učit česky s korpusem. In E. Rusinová (Ed.), *Přednášky a besedy z XLIII. běhu LŠSS* (pp. 112–119). Brno: Masarykova univerzita.
- Osolsobě, K. (2011). *Morfologie českého slovesa a tvoření deverbativ jako problém strojové analýzy češtiny*. Brno: Masarykova univerzita.
- Osolsobě, K. (2014). *Česká morfologie a korpusy*. Praha: Karolinum.
- 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* (pp. 61–87). Liberec: Technická univerzita v Liberci.
- Petráčková, V., & Kraus, J. (Eds.). (1999). *Akademický slovník cizích slov*. Praha: Academia.
- Reppen, R. (2010). *Using corpora in the language classroom*. New York, NY: Cambridge University Press.
- Rešková, I., & Pintarová, M. (2009). *Communicative Czech: Elementary Czech*. Tvarožná u Brna: Ivana Rešková.
- Richards, J. C., & Schmidt, R. (Ed.) (2010). *Longman Dictionary of Language Teaching and Applied Linguistics* (4th ed.). Harlow: Pearson.
- Römer, U. (2009). Corpus research and practice. What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and Language Teaching* (pp. 83–98). Amsterdam: John Benjamins.

- 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* (pp. 23–31). Brno: Masaryková univerzita v Brně.
- Schöne, K. (2015). *Zkoumání hierarchizace pádů českého substantiva v sémantických (kolokačních) třídách*. (Doktorská práce, Univerzita Karlova, Praha). Retrieved from <https://is.cuni.cz/webapps/zzp/detail/104275>
- Skoumalová, H. (2011). Porovnání úspěšnosti tagování korpusu. In V. Petkevič, & A. Rosen (Eds.), *Korpusová lingvistika Praha 2011, sv. 3: Gramatika a značkování korpusů* (pp. 199–207). Praha: Nakladatelství Lidové noviny.
- Svobodová, I. (2000). O psaní velkých písmen (Na pozadí poradenské počítačové databáze). *Naše řeč*, 83(1), 17–29.
- Svobodová, I. (Ed.) (2015). *Psaní velkých písmen v češtině*. Praha: Academia.
- Šebesta, K. (2010). Korpusy češtiny a osvojování jazyka. *Studie z aplikované lingvistiky*, 1(2), 11–33.
- 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. Retrieved from <http://www.korpus.cz>
- Šimandl, J. (2003). Od čvrtku do pátka. *Naše řeč*, 86(3), 161–164.
- Šimandl, J. (Ed.) (2016). *Slovník afixů užívaných v češtině*. Praha: Karolinum.
- Šindelářová, J. & Škodová, S. (2013). *Práce s chybou ve výuce žáků-cizinců*. Retrieved from <https://clanky.rvp.cz/clanek/c/Z/17479/prace-s-chybou-ve-vyuce-zaku-cizincu.html/>
- Štícha, F. (2009). Lokál singuláru tvrdých neživotných maskulin (ve vlaku vs. v potoce): úzus a gramatičnost. *Slovo a slovesnost*, 70(3), 193–220.
- Štindlová, B. (2013). *Žákovský korpus češtiny a evaluace jeho chybové anotace*. Praha: Univerzita Karlova v Praze, Filozofická fakulta.
- Thomas, J. (2006). Using Corpora in Language Teaching and Learning. *Teaching English with Technology, A Journal for Teachers of English*, 6(1). Retrieved from <http://www.tewtjournal.org/?wpdmact=process&did=MTkyLmhvdGxpbms>
- Thomas, J. (2016). *Discovering English with Sketch Engine: A Corpus-Based Approach to Language Exploration*. Brno: Versatile.
- Trnková, A. (2003). *Cvičebnice z české mluvnice pro cizince*. Praha: ISV nakladatelství.
- Trnková, A. (2004). *Textová cvičebnice českého jazyka pro zahraniční studenty. Díl II*. Praha: ISV nakladatelství.
- Vaculová, K. (2013). Použití korpusu ve výuce češtiny – některé interferenční chyby polských mluvčích v češtině zachycené v chybové databázi CHRUP. *Bohemistika*, XIII(1), 48–56.
- Vališová, P. (2009). *Korpus jako zdroj dat systémového popisu české konjugace při výuce češtiny jako cizího jazyka* (Diplomová práce, Masaryková univerzita, Brno). Retrieved from <https://is.muni.cz/th/v98z0/>
- Vališová, P. (2010). Korpus jako zdroj dat systémového popisu české konjugace při výuce češtiny jako cizího jazyka. *Bohemica Olomucensia*, 2, 192–201.

- Vališová, P. (2011). Výukové materiály založené na korpusu. In F. Čermák (Ed.), *Korpusová lingvistika Praha 2011. 2 Výzkum a výstavba korpusu* (pp. 313–323). Praha: Nakladatelství Lidové noviny.
- Vališová, P. (2012). Data-driven learning a výuka češtiny jako cizího jazyka. *CASALC Review*, 2, 22–39.
- Vališová, P. (2016). Korpusy ve výuce češtiny jako cizího jazyka – typy cvičení. In I. Starý Kořánová, & T. Vučka (Eds.), *Čeština jako cizí jazyk VIII* (pp. 129–141). Praha: Univerzita Karlova.
- Vališová, P., & Osolobě, K. (2012). Using data-driven method in teaching Czech as a foreign language. In J. Thomas, & A. Boulton (Eds.), *Input, Process and Product: Developments in Teaching and Language Corpora* (183–194). Brno: Masaryk University Press.
- Vážný, V. (1964). *Historická mluvnice česká. 2. Tvarosloví*. Praha: Státní pedagogické nakladatelství.
- Vlčková, Z. (2015). *Gramatika českého jazyka pro cizince. Pro začátečníky*. České Budějovice: Jihočeská univerzita v Českých Budějovicích.
- Zasina, A. J. (2014). Zachowywanie długości tematycznej samogłoski poprzedzającej wyznacznik czasu przeszłego w języku czeskim. *Bohemistyka*, XIV(1), 64–77.
- 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* (pp. 90–98). Praha: Nakladatelství Lidové noviny.
- Zasina, A. J. (2018a). Korpusy językowe w nauczaniu języków obcych – metoda, narzędzia, praktyka. In M. Jedynak (Ed.), *Specyficzne potrzeby studentów szkół wyższych a nauczanie języków obcych: Tom II: Praktyczne narzędzia* (pp. 110–124). Wrocław: Uniwersytet Wrocławski.
- Zasina, A. J. (2018b). O problémech nerodilých mluvčích s kvantitou na základě analýzy korpusových dat. In S. Škodová, & M. Hrdlička (Eds.), *Čeština jako cizí jazyk v průsečíku pohledů* (pp. 281–298). Praha: Univerzita Karlova, Filozofická fakulta.
- Zasina, A. J., & Škrabal, M. (2018, September). *Morfio.pl – the possibilities for the application of Czech corpus tools to other languages*. Paper presented at the SlaviCorp 2018. Prague. Retrieved from https://slavicorp.ff.cuni.cz/wp-content/uploads/sites/144/2018/10/Zasina_Morfio_pl.pptx
- Zdeňková, J. (2018). K vokalické kvantitě u verbálních substantiv na *-ani/-ání* ve čtyřjazyčném Nomenklátoru Daniela Adama z Veleslavína. *Naše řeč*, 101(1), 3–16.
- 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* (pp. 27–40). Bratislava: VEDA.

9 Publikační, akademická a pedagogická činnost

9.1 Publikace

- Zasina, A. J. (2014). Zachowywanie długości tematycznej samogłoski poprzedzającej wyznacznik czasu przeszłego w języku czeskim. *Bohemistyka*, XIV(1), 64–77.
- Zasina, A. J. (2016). Korespondence kolokací s lexémem *ženský/mužský* v překladu z češtiny do polštiny. *Studia Slavica*, XX (2), 123–132.
- Zasina, A. J. (2016). Překlad a korpus. Textová ekvivalence vybraných českých adjektiv. *Studia Filologiczne*, 5, 25–58.
- Zasina, A. J. (2016). Zpětný překlad kolokací s lexémem *ženský/mužský* v překladu z češtiny do polštiny. *Slované: Souznění a konflikty*, 65–74.
- 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* (pp. 90–98). Praha: Nakladatelství Lidové noviny.
- Zasina, A. J. (2018). Image of politicians and gender in Czech daily newspapers. In M. Fidler & V. Cvrček (Eds.), *Taming the Corpus: From Inflection and Lexis to Interpretation* (167–194). Chad: Springer International Publishing.
- Zasina, A. J. (2018). Językoznawstwo korpusowe. Empiryczne podejście w badaniach humanistycznych. *Dziennikarstwo i Media*, 9, 169–178.
- Zasina, A. J. (2018). Korpusy językowe w nauczaniu języków obcych – metoda, narzędzia, praktyka. In M. Jedynek (Ed.), *Specyficzne potrzeby studentów szkół wyższych a nauczanie języków obcych: Tom II: Praktyczne narzędzia* (pp. 110–124). Wrocław: Uniwersytet Wrocławski.
- 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* (pp. 281–298). Praha: Univerzita Karlova, Filozofická fakulta.

Jako spoluautor (výběr)

- Cvrček, V., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2018a). From extra- to intratextual characteristics: Charting the space of variation in Czech through MDA. *Corpus Linguistics and Linguistic Theory*. Advance online publication. doi: 10.1515/cllt-2018-0020
- Cvrček, V., Komrsková, Z., Lukeš, D., Poukarová, P., Řehořková, A., & Zasina, A. J. (2018b). Variabilita češtiny: multidimenzionální analýza. *Slovo a Slovesnost*, 79(4), 293–321.
- Konečná, H., & Zasina, A. J. (2014). Studium českého jazyka a internet. In E. Rusinová (Ed.), *Přednášky a besedy ze XLVII. běhu LŠSS* (pp. 104–112). Brno: Masarykova univerzita.

- 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. Retrieved from <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)* (pp. 2522–2528). Portorož: ELRA.
- Rosen, A., Vavřín, M., & Zasina, A. J. (2017). *Korpus InterCorp, verze 10 z 1. 12. 2017*. Praha: Ústav Českého národního korpusu FF UK. Retrieved from <http://www.korpus.cz>
- Zasina, A. J., Lukeš, D., Komrsková, Z., Poukarová, P., & Řehořková, A. (2018). *Koditex: korpus diverzifikovaných textů*. Praha: Ústav Českého národního korpusu FF UK. Retrieved from <http://www.korpus.cz>

9.2 Akademická činnost (výběr)

KONFERENCE

Podejście korpusowe w nauczaniu języka polskiego jako obcego na przykładzie alternacji *ó* : *o* w rzeczownikach. Nowości w glottodydaktyce polonistycznej. Perspektywa międzynarodowa: teoria i praktyka, 1. prosince 2018, Warszawa, Poland.

Evaluating a corpus-driven approach in L2 classroom on the example of Czech. SlaviCorp 2018, 24.–26. září 2018, Praha, Czech Republic.

Morfio.pl – the possibilities for the application of Czech corpus tools to other languages (spoluautor: M. Škrabal). SlaviCorp 2018, 24.–26. září 2018, Praha, Czech Republic.

O problémech nerodilých mluvčích s kvantitou na základě analýzy korpusových dat. IX. mezinárodní sympozium o češtině jako cizím jazyku, 14.–17. srpna 2018, Praha, Czech Republic.

Korpusový přístup jako řešení jazykových úskalí v osvojování češtiny jako cizího jazyka: případ předpon *s-* a *z-*. Słowiańskie Przyjemności 3: Wielojęzyczność, 26.–28. října 2017, Kraków, Poland.

Metoda data-driven learning w nauczaniu języka polskiego i czeskiego. Nauczanie języków słowińskich jako obcych – nowe metody i technologie w podnoszeniu jakości kształcenia, 23.–24. listopadu 2017, Wrocław, Poland.

Multi-dimensional analysis of Czech. Pilot study (spoluautoři: Cvrček V., Komrsková Z., Lukeš D., Poukarová P., Řehořková A.). Corpus Linguistics 2017, 25.–28. července 2017, Birmingham, United Kingdom.

Adjective collocations with the lexemes *muž* and *žena* in Czech journalistic texts. Young Linguists' Meeting in Poznań, 25.–27. listopadu 2016, Poznań, Poland.

Multidimenzionální analýza češtiny. Pilotní studie (spoluautoři: Cvrček V., Komrsková Z., Lukeš D., Poukarová P., Řehořková A.). Korpusová lingvistika, 17. září 2016, Praha, Czech Republic.

Korpusy językowe w nauczaniu języków obcych – metoda, narzędzia, praktyka. Konferencja Dydaktyczno-Naukowa Specyficzne Potrzeby Studentów Szkół Wyższych a Nauczanie

Języków Obcych: Kierunki Rozwoju, Nowe Wyzwania, Rekomendacje, 15.–17. září 2016, Wrocław, Poland.

Ekvivalence kolokací s lexémem *ženský/mužský* v překladu z češtiny do polštiny. Konference mladých slavistů 2015: Slované: Souznění a konflikty, 5.–6. listopadu 2015, Praha, Czech Republic.

Transfer elementów kulturowych w filmie Shrek, na podstawie angielskiej, francuskiej, czeskiej i polskiej wersji językowej (spoluautor P. Słowik). Konferencja Studencko-Doktorancka SLAVIK PEČVORK „4 STRONY ŚWIATA”, 11. června, 2015, Opole, Poland.

Textual equivalence of Czech adjectives *matný, drsný, trapný*. 2nd International Conference from the series The Cracow Young Linguists' and Translation Researchers Conference CYLAT 2014. 4.–5. dubna, 2014, Kraków, Poland.

STÁŽE

- **program Erasmus+**: Erasmus od 25. 2. 2018 do 29. 6. 2018 (Polsko, Uniwersytet Jagielloński w Krakowie).
- **letní škola korpusové lingvistiky**: Lancaster Summer Schools in Corpus Linguistics od 14. 7. 2015 do 17. 7. 2015 (Velká Británie, Lancaster University, Lancaster).

9.3 Pedagogická činnost

Český jazyk, srpen 2018, Letní škola slovanských studií, Ústav bohemistických studií FF UK.

Ćwiczenia gramatyczno-leksykalne dla bohemistów, letní semestr akademického roku 2017/2018, Filologická fakulta Jagellonské univerzity v Krakově – seminář pro studenty bohemistiky.

Gramaticko-lexikální cvičení založená na korpusu, akademický rok 2016/2017, 2017/2018 a 2018/2019, Filozofická fakulta Univerzity Karlovy – seminář pro studenty bohemistiky pro cizince a další zájemce.

Český jazyk a korpus, letní semestr akademického roku 2015/2016, Filozofická fakulta Univerzity Karlovy – seminář pro studenty bohemistiky a další zájemce.

Polský jazyk, zimní semestr akademického roku 2013/2014, Filozofická fakulta Masarykovy univerzity – jazyková výuka polštiny pro zájemce.