Filozofická fakulta Univerzity Karlovy v Praze

BAKALÁŘSKÁ PRÁCE

2018 **Nikola Piálková**

Filozofická fakulta Univerzity Karlovy v Praze

Ústav českého jazyka a teorie komunikace

BAKALÁŘSKÁ PRÁCE

Nikola Piálková

Čeština čínských mluvčích v korpusu CzeSL-SGT

Czech of Chinese Native Speakers in CzeSL-SGT Corpus

Praha 2018

Vedoucí práce: prof. PhDr. Karel Šebesta, CSc.

Ráda bych poděkovala vedoucímu své práce, prof. Karlu Šebestovi, za věnovaný čas a cenné rady. Taktéž bych chtěla projevit vděčnost doc. Lukáši Zádrapovi za nápomocné konzultace týkající se čínské lingvistiky.

Prohlašuji, že bakalářskou práci jsem vypracovala samostatně, řádně jsem citovala všechny použité prameny a že tato práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

ABSTRAKT

Práce se zabývá specifickou částí Českého národního korpusu. Konkrétně se jedná o data žákovského korpusu CzeSL-SGT, jež byla získána od studentů češtiny, jejichž prvním jazykem je čínština. V první kapitole je tato část uvedena kontextem akvizičního korpusu CzeSL, do nějž je začleněna. Druhá, hlavní, kapitola nahlíží korpus prostřednictvím metadat, jež jsou součástí každého studentského textu. Poslední část je zaměřena na výběr chyb, které jsou typické pro rodilé mluvčí analytického jazyka během nabývání jazyka syntetického.

Klíčová slova: čeština jako druhý jazyk, žákovské korpusy, CzeSL-SGT

ABSTRACT

The present thesis deals with a part of the Czech National Corpus: to be specific, it deals with a collection of data from CzeSL-SGT, a student corpus, that have been obtained from students of the Czech language whose L1 is Chinese. This section is introduced in the first chapter of the thesis with regard to the context of the acquisition corpus CzeSL. The second – main – chapter examines the corpus through metadata that can be found in every student text. The final section describes a selection of typical errors made by native speakers of an analytic language during a synthetic language acquisition.

Key words: Czech as a second language, learner corpora, CzeSL-SGT

Obsah

ÚVOD 8

SLOVNÍČEK POJMŮ 9

1. ŽÁKOVSKÝ KORPUS CzeSL-SGT 11

1.1. CzeSL-SGT jako součást Českého národního korpusu 11

1.2. Zařazení žákovských korpusů 11

1.2.1. Vznik a složení CzeSL 13

1.2.2. Velikost CzeSL-plain 13

1.3. CzeSL-SGT 14

1.3.1. Vznik CzeSL-SGT 14

1.3.2. Velikost korpusu CzeSL-SGT a problémy s ní spojené 14

1.3.3. Metadata 15

2. CHARAKTERISTIKA KORPUSŮ ČÍNSKÝCH STUDENTŮ ČEŠTINY 20

2.1. Metadata týkající se studentů 21

2.1.1. Zastoupení žen a mužů s ohledem na věkovou vyváženost 21

2.1.2. Věková kategorie 16+ 22

2.1.3. Zastoupené úrovně češtiny a jejich náročnost 23

2.1.4. Intenzita, délka a forma studia 24

2.1.5. Čeština v rodině a roky strávené v České republice 25

2.1.6. Česká učebnice 26

2.1.7. Bilingvismus a další jazyky studentů 26

2.2. Metadata týkající se pouze textu 27

2.2.1. Zadaný a reálný počet slov 27

2.2.2. Zadání a druh postupu psaní 28

2.2.3. Zkoušky a průběžné práce 29

2.2.4. Aktivita předcházející psaní textu 30

2.2.5. Stáří textů 30

3. VYBRANÉ CHYBY TYPICKÉ PRO ČÍNSKÉ STUDENTY ČESKÉHO JAZYKA 31
3.1. Přístupy k nabývání druhého jazyka 31
3.1.1. Pojetí chyby 31
3.1.2. Mezijazyk....32
3.2. Metodika 32
3.3. Výchozí a cílový jazyk....33
3.3.1. Čínština....33
3.3.2. Důležité faktory češtiny jako cizího jazyka....34
3.3.3. Vybrané rozdíly 35
3.4. Konkrétní případy chyb....37
3.4.1. Přísudek jmenný beze spony....37
3.4.2. Záměna přídavných a podstatných jmen 39
3.4.3. Duplikace pro zdůraznění slov 41
3.4.4. Množné číslo u podstatných jmen 42
3.4.5. Pomocné slovo s původním významem *moci*....42
3.4.6. Anteponovaný větný člen 44
ZÁVĚR....47
ZDROJE A POUŽITÁ LITERATURA 49

ÚVOD

Tato bakalářská práce se zabývá specifickou částí Českého národního korpusu, která doposud nebyla důkladně zkoumána. Jedná se o texty čínských mluvčích češtiny, tedy o druhou největší část lemmatizovaného žákovského korpusu CzeSL-SGT, který je složen z textů nerodilých mluvčích korspusu CzeSL-plain a dalších textů nerodilých mluvčích z roku 2013. První kapitola se soustředí na představení Českého národního korpusu jako takového, dále uvádí vznik korpusu CzeSL, CzeSL- plain a následně vysvětluje též podobu lemmatizovaného korpusu CzeSL-SGT, který je poté hlavním předmětem zkoumání. Druhá část je jádrem celé práce a soustředí se pouze na texty v korpusu, které napsali čínští studenti češtiny. Snaží se popsat charakter textů prostřednictvím metadat, která jsou součástí každého studentského textu. Metadata jsou údaje, které se v polovině počtu dělí na studentská a textová. To znamená, že informací, jež jsou nám sděleny o studentech, je stejně jako informací týkajících se pouze textu. Třetí část se zaměřuje na konkrétní typy chyb, které se pokouší uchopit pomocí kvalitativní analýzy, a to především se zaměřením na výběr chyb, které jsou typické pro mluvčí s mateřským analytickým jazykem při nabývání jazyka syntetického.

SLOVNÍČEK POJMŮ

K správnému terminologickému uchopení textu jsem se rozhodla využít terminologické poznámky Barbory Štindlové z knihy *Žákovský korpus češtiny a evaluace jeho chybové anotace*[1].

první jazyk = first language = L1

jazyk, který jedinec ovládá nejlépe, vyjadřuje se v něm nejčastěji a nejraději, obvykle shodný s mateřským jazykem

mateřský jazyk = mother tongue

jazyk, který si jedinec osvojí jako první v pořadí, obvykle shodný s prvním jazykem

cizí jazyk = foreign language = FL

jazyk nabývaný v prostředí, v němž se tímto jazykem nemluví (např. studium češtiny v Číně)

druhý jazyk = second language = L2

jazyk nabývaný v přirozeném prostředí (např. studium češtiny v České republice)

cílový jazyk = target language = TL

jazyk, který si mluvčí chce osvojit, může se shodovat s druhým i cizím jazykem

mezijazyk = interlanguage

žákovský jazyk, který má status přechodného jazyka mezi prvním a cílovým jazykem

žák, student = learner

jedinec, který se učí cizí jazyk

[1] ŠTINDLOVÁ, Barbora: *Žákovský korpus češtiny a evaulace jeho chybové anotace*, Karolinum, Praha 2013, s. 14-16.

žákovský korpus = learner corpus

korpus jazyka nerodilých mluvčích, někdy označován i jako studijní korpus (Čermák a Schmiedtová, 2004)

jazykový vstup/vklad = input/exposure

souhrně charakterizuje vlivy na žáka během studia cílového jazyka

jazykový výstup = output

žákova produkce cílového jazyka

osvojování jazyka = aquisition

učení se jazyku, které si učící se jedinec neuvědomuje, nejčastěji osvojování mateřského jazyka

učení se jazyku = learning

vědomý proces učení se jazyku za pomocí pravidel

nabývání jazyka = jazyková akvizice

zastřešující termín pro pojem učení i osvojování

1. ŽÁKOVSKÝ KORPUS CzeSL-SGT

1.1. CzeSL-SGT jako součást Českého národního korpusu

Nutnost rozsáhlé jazykové databáze vyvrcholila v České republice roku 1994, kdy byl založen Český národní korpus. Projekt měl sloužit především k didaktickým účelům a k výzkumu, a to ve formě elektronických korpusů, které se během 24 let vyšplhaly až k počtu tří miliard slov. Texty jsou dle svých parametrů rozděleny do korpusů různých druhů, jedná se například o synchronní a diachronní, mluvené a psané, nebo také paralelní či jednojazyčné. V současné době se o Český národní korpus starají především dva ústavy Filozofické fakulty Univerzity Karlovy, a to Ústav Českého národního korpusu a Ústav teoretické a komputační lingvistiky. Dále patří vděčnost všech uživatelů dalším externistům z celé republiky, kteří přispívají nejen velkým množstvím materiálů, ale jsou též pomocí při koordinaci potřebných činností.[2]

Krom výše uvedených příkladů rozdělení různých korpusů, jež jsou součástí velkého projektu, je třeba zmínit též obecné a specializované korpusy. Do obecných patří největší korpusy, které se utváří každých pět let – SYN2000, SYN2005, SYN 2010 a SYN 2015, které jsou žánrově vyvážené a jejichž výskyty svým počtem dosahují stamilionových řádů. Dále sem řadíme též publicistické korpusy SYN2006PUB, SYN2009PUB a SYN2013PUB, jejichž obsah dokonce ještě přerůstá výše zmíněné – žánrově vyvážené korpusy. Mezi specializované lze zařadit korpus soukromé korespondence – KSK-dopisy, dále například LINK aneb „Lingvistův narozeninový korpus“, který obsahuje lingvistické texty, anebo právě korpus češtiny jako druhého jazyka – CzeSL.[3]

1.2. Zařazení žákovských korpusů

Korpus češtiny nerodilých mluvčích s automaticky provedenou anotací (*Czech as a Second Language with Spelling, Grammar and Tags*) se řadí mezi akviziční korpusy češtiny, díky nimž lze pozorovat procesy osvojování cílového jazyka či jeho pozdější vývoj. Dále též slouží

[2] CVRČEK, Václav - RICHTEROVÁ, Olga (eds). "start". Příručka ČNK. 25 May. 2018. Web. 18 Dec. 2018.
[3] CVRČEK, Václav - RICHTEROVÁ, Olga (eds). "cnk:struktura". Příručka ČNK. 30 Jul. 2018. Web. 18 Dec. 2018.

ke studiu užívání češtiny mluvčími, již jazykem nedokáží komunikovat na úrovni dospělého rodilého mluvčího.[4] Akviziční korpusy lze snadno vymezit vzhledem k národním synchronním korpusům, které: „*se zaměřují na reprezentativnost ve vztahu k současnému jazyku, korpusy akviziční zaznamenávají jazyk předškolních dětí, jazyk školní mládeže či mezijazyk nerodilých mluvčích, tedy struktury velmi nestálé, proměnlivé, závislé na řadě vnějších faktorů, které se od běžně užívaného jazyka mohou někdy velmi podstatně lišit.*“[5]

Žákovské korpusy slouží primárně jako zdroj dat pro studium **mezijazyka**, který se ideálně vyvíjí a je proměnlivý. Vzhledem k ideálu co největší proměnlivosti a pozorovatelného vývoje je pro žákovské korpusy klíčové, aby obsahovaly co možná největší počet textů od co možná největšího počtu autorů.[6] Žákovské korpusy by dále měly obsahovat informace, které jsou významné pro další studium jako např. věk, pohlaví, první jazyk žáka, okolnosti osvojování druhého jazyka nebo délka formování jazykové výuky.[7]

Dalším důležitým rysem korpusů je autentičnost, která je ve své nejčistší formě jedním ze základních požadavků pro texty obecně lingvistických synchronních korpusů. Data neautentická, tedy ta, jež nevznikla v reálné komunikační situaci, do takových korpusů nepatří. U žákovských korpusů je situace složitější, z autentických dat totiž mohou čerpat velmi omezeně – především u žáků vysoké jazykové úrovně v přirozeném jazykovém prostředí. Za takových předpokladů je žák při produkci jazyka značně méně limitován kvantitativně, funkčně či tematicky. Přirozené jazykové prostředí mu poté umožňuje prožití reálných životních situací, s nimiž se musí (může) pomocí cílového jazyka vypořádat – nakupování, návštěva lékaře atd. Sběr dat pro žákovské korpusy se však většinou soustředí na školní situace, tudíž jde především o eseje psané v rámci výuky či materiály tvořené přímo pro korpus. [8]

[4] ŠEBESTA, Karel – ŠKODOVÁ, Svatava: Čeština – cílový jazyk a korpusy, s. 5.
[5] BEDŘICHOVÁ, Zuzanna – ŠEBESTA, Karel – ŠKODOVÁ, Svatava – ŠORMOVÁ, Kateřina: Podoba a využití korpusu jinojazyčných a romských mluvčích češtiny: CZESL a ROMi, in: Korpusová lingvistika Praha 2011, Lidové noviny, s. 94.
[6] ŠEBESTA, Karel – ŠKODOVÁ, Svatava: Čeština – cílový jazyk a korpusy, s. 17.
[7] BEDŘICHOVÁ, Zuzanna – ŠEBESTA, Karel – ŠKODOVÁ, Svatava – ŠORMOVÁ, Kateřina: Podoba a využití korpusu jinojazyčných a romských mluvčích češtiny: CZESL a ROMi, in: Korpusová lingvistika Praha 2011, Lidové noviny, s. 95.
[8] ŠEBESTA, Karel – ŠKODOVÁ, Svatava: Čeština – cílový jazyk a korpusy, s. 19.

1.2.1. Vznik a složení CzeSL

CzeSL vznikl jako první žákovský korpus pod záštitou projektu *Inovace vzdělávání v oboru čeština jako druhý jazyk*, v rámci programu Vzdělávání pro konkurenceschopnost s finanční podporou Strukturálních fondů EU a státního rozpočtu České republiky, ve spolupráci TU v Liberci, UK v Praze a AUČCJ. Projekt dále podpořilo množství dalších institucí, organizací a jednotlivců. Primárně pedagogická funkce korpusu CzeSL je pro výběr jazykových dat a pro jejich zpracování směrodatná.[9] Kromě textů nerodilých mluvčích se CzeSL (dnes CzeSL-plain) skládá též z korpusu romských žáků, u nichž nelze jednoznačně rozhodnout, zda je čeština v roli prvního či druhého jazyka, data romského korpusu se formou sběru i povahou dat samých od textů nerodilých mluvčích v mnohém liší. Slohové práce jsou získávány z českých škol základních, praktických i speciálních, a to v Praze i mimo Prahu. V současné době texty dohledáme pod korpusem CzeSL-plain, konkrétně část s metadatem *rom*, která nyní obsahuje 428 161 slov.

1.2.2. Velikost CzeSL-plain

Velikost korpusu se již ve startu projektu plánovala na 2 miliony slov, tento záměr byl splněn a CzeSL-plain nyní čítá 2 320 678 výskytů. Zveřejněn byl roku 2012. Pomineme-li anglické korpusy, patří CzeSL-plain k největším žákovským korpusům vůbec. Zaznamenaná data měla být mluvené i písemné povahy, tento záměr se však bohužel nesplnil a všechny texty jsou psané, a to většinou rukopisně, z části na počítači. Cizinecká část korpusu tvoří polovinu celého korpusu a tvoří ji 1 160 701 slov. Právě tato část se stala základem pro korpus CzeSL-SGT, kterému se v práci budeme věnovat.

[9] ŠEBESTA, Karel – ŠKODOVÁ, Svatava: Čeština – cílový jazyk a korpusy, s. 29.

1.3. CzeSL-SGT

1.3.1. Vznik CzeSL-SGT

CzeSL-SGT je složen z přepisů písemných prací studentů češtiny – tzn. jednotlivců s jiným mateřským (prvním) jazykem, než je čeština. CzeSL-SGT navazuje na korpus CzeSL-plain, který z části taktéž obsahuje texty nerodilých mluvčích češtiny, avšak CzeSL-SGT k tomu obsahuje ještě další texty z roku 2013. Všechny texty jsou navíc opatřeny jazykovými daty. Lingvistická analýza korpusu zahrnuje označení slovních druhů, zařazení do morfologických kategorií nebo také určení základních tvarů (lemmat).

Lemmatizace u korpusu CzeSL-plain neproběhla, a proto je v ní mnohem obtížnější vyhledávat. Chceme-li například v korpusu najít všechny tvary slovesa být, je potřeba je všechny vypsat. CzeSL-SGT je v těchto případech mnohem praktičtější, jelikož stačí vyhledat pouze lemma, tzn. základní tvar slovesa. Slovnědruhové zařazení poté usnadní vyhledávání například v případě, že potřebujeme zjistit frekvenci užívání spojek. Tehdy stačí vyhledat pouze slovní druh spojky, aniž bychom museli složitě vypisovat všechny české spojky jako v případě korpusu CzeSL-plain.

Jelikož se jedná o žákovský korpus, lze zde více než kdekoli jinde očekávat jisté množství chyb. Tyto chyby jsou povětšinou opraveny autokorekcí a na základě správné podoby tvaru je chyba zařazena k určitému chybovému typu, resp. označena příslušným chybovým kódem. Je třeba zdůraznit, že z důvodu užívání automatického určování chyb a jazykových dat, nemůžeme v těchto ohledech očekávat stoprocentní úspěšnost.[10]

1.3.2. Velikost korpusu CzeSL-SGT a problémy s ní spojené

CzeSL-SGT dohromady obsahuje 8 617 textů, jež jsou tvořeny 1 147 477 slovy. To však zdaleka neznamená, že na jejich tvorbě se podílel stejný počet autorů. Studentů, kteří texty tvořily, je zhruba čtvrtina – 1 966, což znamená, že na každého z nich průměrně připadají

[10] CVRČEK, Václav - RICHTEROVÁ, Olga (eds). "cnk:czesl-sgt". Příručka ČNK. 24 Mar. 2015. Web. 10 Dec. 2018.

čtyři texty korpusu. S ohledem na tuto skutečnost je třeba býti na pozoru ve chvíli, kdy vyhledáváme specifické jevy v menší části korpusu. Pracujeme-li totiž s menším množstvím výskytů, spokojí se náš kvalitativní výzkum i s menším množstvím výsledků (například hledaný jev se objeví desetkrát v pěti různých textech). Tehdy je nezbytné zkontrolovat autorství všech pozitivních textů, a vyloučit tak možnost, že hledaný jev se objevuje pouze v textech jednoho autora.

1.3.3. Metadata

Téměř všechny texty korpusu CzeSL-SGT jsou vybaveny metadaty, které nám mohou sdělit o textu i autorovi důležité informace. Každý text je vybaven vlastním id (*id_t*), aby nebyl zaměnitelný s žádným z ostatních. Protože však neplatí pravidlo; co text, to jiný autor, taktéž tvůrci prací mají vlastní id (*id_a*). V korpusu je tedy možné dohledat, zda jsou jednotlivci autorem jednoho či více textů. Tato skutečnost je velmi vítaná například v chybové analýze, či jiném výzkumu, jak je již zmíněno v kapitole 1.2.2.

Maximální počet metadat může vystoupat až na třicet. Není však u všech textů stejný, ne vždy totiž známe potřebné informace pro vyplnění všech nabízených políček. Metadata se dělí přesně na poloviny podle toho, zda nám sdělují informace o textu, či o studentovi. Mezi textovými metadaty najdeme například údaje o tom, zda se jedná o rukopis či počítačový zápis textu, dále například zda byla práce předmětem zkoušky, či pouze průběžného testování, jaký byl slovní rozsah práce, co bylo povoleno za pomůcku či zda předcházela psaní textu nějaká aktivita. Metadata týkající se studentů obsahují informace o genderu, o věkovém zařazení, délce studia češtiny, či mateřském jazyku studenta.

Na základě posledního zmíněného údaje, tedy mateřského jazyka jednotlivců, byly již při tvorbě korpusu projeveny snahy zařadit texty studentů ze tří jazykových skupin:

(a) *jazyků slovanských, tedy blízce příbuzných. Výrazně mezi nimi převažují mluvčí s ruštinou nebo jiným východoslovanským jazykem, významněji jsou zastoupeni rovněž Poláci, další slovanské jazyky jen minimálně;*

(b) jiných jazyků indoevropských. V této skupině jsou podle očekávání převážně mluvčí s francouzštinou, němčinou, angličtinou, španělštinou jako prvními jazyky;

(c) jazyků nepříbuzných; mezi nimi převažují zejména jazyky dálněvýchodní, především čínština a vietnamština, a arabština.[11]

V tabulce T1 můžeme pozorovat, jak jsou v korpusu zastoupeny skupiny autorů různých mateřských jazyků. Druhý sloupec uvádí jazykovou zkratku, třetí sloupec informuje, kolik slov připadne dané jazykové skupině a poslední sloupec nese celé názvy jazyků. Porovnáme-li hned zezačátku první a druhou pozici, zjistíme, že se liší o obrovský rozdíl, a to o 622 282 slov. Texty ruského původu totiž tvoří téměř 60 % korpusu, z 8 617 textů je ruských hned 5 061 a z celkových 1 966 autorů je 1036 Rusů. Druhá příčka statistiky patří čínským mluvčím, kteří vytvořili cca 5,4 % korpusu. Podrobně se jim budeme věnovat v hlavní části této práce. Třetí místo poté náleží Ukrajincům, jejichž texty tvoří asi 4,3 % korpusu.

pořadí	jaz. zkratka	Frek.	počet studentů	% z počtu studentů (pořadí)	jazyk
1	ru	684177	1036	53 (1)	ruština
2	zh	61895	116	6 (2)	čínština
3	uk	49436	76	3,9 (3)	ukrajinština
4	neurčen	42039	16	0,8 (11)	jazyk neurčen
5	ko	27198	59	3 (6)	korejština
6	ja	26150	42	2,1 (8)	japonština
7	en	26022	64	3,3 (5)	angličtina
8	de	25379	65	3,3 (4)	němčina
9	kk	25166	39	2 (9)	kazaština
10	pl	19304	37	1,9 (10)	polština
11	vi	16859	57	2,9 (7)	vietnamština

[11] BEDŘICHOVÁ, Zuzanna – ŠEBESTA, Karel – ŠKODOVÁ, Svatava – ŠORMOVÁ, Kateřina: Podoba a využití korpusu jinojazyčných a romských mluvčích češtiny: CZESL a ROMi, in: Korpusová lingvistika Praha 2011, Lidové noviny, s. 97.

12	ar	13436			arabština
13	fr	13428			francouzština
14	es	13225			estonština
15	uz	11920			uzbečtina
16	it	8240			italština
17	tr	7700			turečtina
18	hu	7572			maďarština
19	bg	5541			bulharština
20	mo	5144			moldavština
21	ky	5136			kyrgyzština
22	ro	4502			rumunština
23	fi	4125			finština
24	mn	3979			mongolština
25	be	3624			běloruština
26	th	3021			thajština
27	az	2739			azerbajdžánština
28	nl	2623			nizozemština
29	mk	2534			makedonština
30	ba	2119			baškirština
31	he	2113			hebrejština
32	pt	1758			portugalština
33	hy	1731			arménština
34	sq	1691			albánština
35	ka	1512			gruzínština
36	sk	1504			slovenština
37	el	1487			řečtina
38	sr	1397			srbština
39	sv	1344			švédština
40	lv	1227			lotyština
41	fa	1135			perština
42	da	1087			dánština

43	hr	819			chorvatština
44	tl	730			tagalština
45	cs	597			čeština
46	tg	588			tádžičtina
47	sl	470			slovinština
48	kg	386			konžština
49	xal	378			kalmyčtina
50	sh	320			srbochorvatština
51	ne	288			nepálština
52	id	265			indonéština
53	hi	226			hindština
54	la	96			latina
55	ms	95			malajština

Tabulka T1 Statistika obsazenosti studentů v CzeSL-SGT podle prvního jazyku

U prvních jedenácti zastoupených jsem v Tabulce T1 uvedla též přesný počet studentů, následně pak procento, které tento počet představuje z celkového počtu studentů (1966), již se na tvorbě korpusu CzeSL-SGT podíleli. Tato zastoupení jsem se rozhodla porovnat se statistikou národnostních menšin, kterou poskytuje Ministerstvo vnitra: *Cizinci 3. zemí se zaevidovaným povoleným pobytem na území České republiky a cizinci zemí EU + Islandu, Norska, Švýcarska a Lichtenštejnska se zaevidovaným pobytem na území České republiky k 31. 10. 2018.*[12]

Příjemným překvapením je skutečnost, že mluvčí s jazyky všech prvních dvaceti států máme obsažené v korpusu CzeSL-SGT. Případné odchylky nastávají ve chvíli, kdy se podíváme na procenta. Největší zastoupenou menšinou v Česku je nyní Ukrajina, jejíž zástupci tvoří téměř čtvrtinu (22,935 %), v korpusové Tabulce T1 se ukrajinština ocitla na třetím místě s pouhými 3,9 %. Slováci žijící v České republice se z historických důvodů za menšinu mnohdy nepovažují. Pro nás je však nejdůležitější fakt, že se pro život v Česku nepotřebují učit česky, protože zásluhou velké podobnosti jazyků jim ke všem rovinám komunikace v Česku stačí rodný jazyk.

[12] Ministerstvo vnitra: https://www.mvcr.cz/clanek/cizinci-s-povolenym-pobytem.aspx?q=Y2hudW09MQ%3d%3d

Vietnamští obyvatelé České republiky činí téměř 11 % menšin, v tabulce českého korpusu se však množstvím výskytů umístili až na 11. místě, počet studentů (57) jim ovšem stačil na sedmé místo, stále však nedosahují například na počet korejských studentů, kteří jsou ve statistice menšin až na 25. místě. S pouhými 6,75 % přichází ruské obyvatelstvo. Tak malá část je samozřejmě velmi kontrastní ke korpusové nadpoloviční míře procent (53 %). Čína, která v korpusové tabulce zabírá druhé místo (stejně jako Rusko díky počtu výskytů i obyvatel) se ve statistice menšin přesouvá až na 11. příčku (rozdíl: cca 4,5 %), stále je však nutno poznamenat, že k tak velkému rozptylu procentových mír jako u Ruska rozhodně nedochází.

Nejmenší zastoupení v korpusové tabulce mají z menšinové statistiky Indové. Ti tvoří třetí nejmenší část korpusu, přičemž její tři texty mají pouze jednoho autora.

pořadí	země	%
1	Ukrajina	22,935
2	Slovensko	20,881
3	Vietnam	10,946
4	Rusko	6,754
5	Německo	3,820
6	Polsko	3,815
7	Bulharsko	2,762
8	Rumunsko	2,585
9	Spojené státy	1,600
10	Mongolsko	1,585
11	Čína	1,331
12	Velká Británie	1,253
13	Maďarsko	1,162
14	Kazachstán	1,085
15	Bělorusko	1,083
16	Moldavsko	1,035
17	Itálie	0,938
18	Indie	0,754
19	Francie	0,744
20	Srbsko	0,711

Tabulka TS Statistika národních menšin (prvních 20 příček)

2. CHARAKTERISTIKA KORPUSŮ ČÍNSKÝCH STUDENTŮ ČEŠTINY

Jak už jsme zmínili v předchozí kapitole, Číňané jsou druhou nejsilnější skupinou mezi autory korpusu CzeSL-SGT. Jejich část tvoří 519 textů celkem od 116 studentů. Dálněvýchodní jazyky mají v roli prvního jazyka v žákovských korpusech jednoznačnou převahu. Čínština je mezi nimi nejsilnější, s velkým odstupem pak následuje japonština a korejština (ve statistice CzeSL-SGT jde o jazyky, které následují hned po třetí ukrajinštině).

„Od čínských mluvčích pochází téměř třetina dat v dosud známých nekomerčních žákovských korpusech. Převaha čínštiny jako prvního jazyka je dána především díky velkému korpusu angličtiny čínských studentů HKUST (Hong Kong University of Science and Technology Learner Corpus)o udávané velikosti 25 milionů slov,ale i řadě korpusů dalších, méně objemných, jako je SWECCL (The Spoken and Written English Corpus of Chinese Learners, cca 2 miliony slov), CLEC (Chinese Learner English Corpus, cca 1 milion slov), MSEE (Corpus for Middle School English Education, 2,3 milionu slov), TLCE (The Taiwanese Corpus of Learner English, cca 2 miliony slov) aj.

Situace na poli žákovských korpusů se samozřejmě rychle mění a přehled, o nějž tyto informace opíráme, nemusí být zcela spolehlivý, masivní převaha angličtiny jako cílového a čínštiny jako jazyka výchozího je však faktem těžko zpochybnitelným.“[13]

Jak je řečeno v citaci, čínština je jedním z nejočekávatelnějších výchozích jazyků, což splňuje i její druhé umístění v naší statistice korpusu CzeSL-SGT. Aby však mohla být čínská část korpusu co nejadekvátnější pro další zkoumání či pro chybovou analýzu, je nutno doplnit ji příslušnými texty. Ke zjištění, jaké typy studentů je nutno hledat a jaké typy prací jim zadávat tak, abychom docílili co nejvyváženějšího korpusu, nám pomůžou již zmíněná metadata.

[13] ŠEBESTA, Karel – ŠKODOVÁ, Svatava: Čeština – cílový jazyk a korpusy, s. 10.

2.1. Metadata týkající se studentů

2.1.1. Zastoupení žen a mužů s ohledem na věkovou vyváženost

Co se týče cíle vyváženosti studentů na základě pohlaví, můžeme mluvit o úspěchu. V poměru 1:1 tvoří korpus 57 žen a 59 mužů. Ženy jsou však v porovnání s muži o něco sdílnější a při 36 441 slovech jsou autorkami 286 textů. Muži, ač početně silnější skupina, mají na svědomí pouze 233 textů dohromady o 25 454 slovech.

Podíváme-li se na věkové údaje čínských mužů a žen zastoupených v korpusu, ženy dosahují na nejpočetnějších příčkách lehce vyššího věkového rozptylu než muži. Vezmeme-li v potaz prvních pět příček Tabulky T2, zjistíme, že u žen lze pátrat v textech od autorů ve věku 21–38 let, u mužů pak pouze v rozsahu 18–23 let. Z tohoto závěru vyplývá, že čínské ženy byly v průměru o něco starší než čínští muži.

Protože studenti zapojení do tvorby korpusu v souhrnu odpovídají velmi pestré škále věkových hodnot, kromě metadat o přesném věku studentů, existují i metadata, jež studenty řadí do příslušné věkové kategorie. Kategorie pro nejmenší děti ve věku 6–11 let v našem případě zahrnula pouze jednoho jedenáctiletého chlapce a jeho práci o 28 slovech, které ve výsledku činí zanedbatelných 4,5 setin procenta čínského korpusu. Druhá kategorie 12–15 let je v obou skupinách téměř stejně zastoupena a tvoří asi 3,2 % všech textů.

	Ž		M	
pořadí	**věk**	**Frek.**	**věk**	**Frek.**
1	21	7762	21	8869
2	25	5160	22	3000
3	38	3973	20	2846
4	23	3408	18	1588
5	22	2573	23	1238
6	20	2534	17	1141
7	26	1922	25	948

8	24	1750	26	893
9	19	1514	34	774
10	18	1188	28	716

Tabulka T2 Srovnání deseti nejpočetnějších věkových pozic u mužů a žen

2.1.2. Věková kategorie 16+

Jak dokládá Tabulka T3, nejsilnější kategorií je poslední skupina *16+*, která by s ohledem na různost věku studentů určitě zasloužila rozdělit na více kategorií. Tato skupina je 97% částí korpusu, přičemž tři nejpočetnější věkové pozice jsou *21*, *22* a *25*. S ohledem na nejvyšší věkovou pozici – *38* – lze navrhnout členění na kategorie *16–20*, *21–25*, *26–30* a jako poslední *31–40*. Věkové kategorie by měly být vždy určitým způsobem ohraničeny, aby bylo jasné, jaký je pravděpodobný nejnižší i nejvyšší věk v nich obsažený. Co se týče nejslabších věkových pozic statistiky, jedná se o již zmíněného jedenáctiletého studenta, dále o třináctiletého studenta, který přispěl též velmi krátkou prací – 66 slov – a o poslední případ šestatřicetiletého studenta, který se umístil třetí od konce se svými čtyřmi kratičkými texty o celkovém rozsahu 279 slov.

pořadí	věk	Frek.	% z počtu výskytů
1	21	16631	27
2	25	6108	10
3	22	5573	9
4	36	279	1 student
5	13	66	1 student
6	11	28	1 student

Tabulka T3 Tři nejsilnější a tři nejslabší věkové pozice

2.1.3. Zastoupené úrovně češtiny a jejich náročnost

Podle SERR (Společného evropského referenčního rámce) jsou v korpusu obsaženy čtyři základní skupiny (A1; A2; B1 a B2), které doplňují dvě další méně obsažené skupiny (A1+; A2+). Korpus bohužel neobsahuje žádné texty od studentů na úrovni vyšší než B2, proto by bylo velmi vítané rozšířit jej též o úroveň C1 a C2. Zastoupení A skupin v poměru k B skupinám je téměř vyvážené. B1 a B2 dohromady tvoří 47 % korpusu, ostatní čtyři A skupiny poté zbylých 53 %. Jak ukazuje tabulka T4, všechny hlavní jazykové úrovně můžeme z hlediska počtu označit za dostatečně reprezentativní.

pořadí	jaz. úroveň	Frek.	%
1	B1	23155	37,4
2	A1	15942	26
3	A2	14806	24
4	B2	5745	9,3

Tabulka T4 Hlavní jazykové úrovně a jejich procentuální podíl

Mezi metadata, která se v ohledu na úroveň mluvčího často liší, patří jednoznačně limit slov prací studentů. Jak můžeme vidět v Tabulce T5, B skupiny jsou předvídatelně náročnější než A skupiny. Znepokojující je fakt, že v obou skupinách tvoří velkou část texty, u nichž limit slov není vyznačen. Jen těžko lze počítat s variantou, že si studenti mohli rozsah určit naprosto libovolně.

	B1, B2		A1, A1+, A2, A2+	
pořadí	limit slov	Frek.	limit slov	Frek.
1	150	14091		8784
2		8172	40-	5925
3	50-	1607	40	3804
4	90	1214	30	3577
5	50	1064	20	2819

Tabulka T5 Srovnání pěti nejčastějších limitů slov u studentů B skupin a A skupin

Dalším aspektem, v němž by se úrovně mohly lišit, jsme předpokládali povolenou pomůcku. Zjistili jsme, že zatímco studenti A úrovní neměli možnost použít pomůcku v téměř 84 % výskytů, studenti úrovně B1, či B2 naopak ve většinové části (64 %) pomůcku používat mohli, ve všech případech se jednalo o slovník.

2.1.4. Intenzita, délka a forma studia

Téměř 91 % studentů studiem češtiny tráví více než 15 hodin týdně, což vypovídá o skutečnosti, že korpus téměř postrádá texty studentů, kteří by navštěvovali méně intenzivní kurz. Od studentů, kteří studují méně intenzivně; 5–15 hodin týdně, máme v korpusu 4064 slov a studenti, již češtině věnují méně než 3 hodiny týdně, utvořili pro korpus texty o hodnotě 1705 výskytů. Zvláštní je, že mezi kategoriemi úplně chybí skupina pro studenty, kteří češtině věnují 3–5 hodin za týden.

Údaje o délce studia češtiny jsou rozděleny na skupiny podle počtu měsíců. Abychom potvrdili předpoklad, že studenti, kteří se češtině věnovali vyšší počet měsíců, budou pravděpodobně na vyšší úrovni, rozdělili jsme si tabulku opět na tři nejobsazenější pozice pro A úrovně a B úrovně češtiny. U skupiny A úrovní se potvrdily častější nižší počty měsíců studia, u druhého nejvyššího počtu textů údaj o délce studia bohužel chybí. Co se týče B úrovní, na první tři příčky statistiky vyšly nejvyšší počty měsíců.

Tabulka T6 nám dokládá, že kategorie 24–36 potom činí cca 28 % korpusu. Druhou největší skupinou jsou studenti, kteří se učí česky 6–12 měsíců, v rámci korpusu jde o 24 %. Nejmenší kategorii „do tří měsíců“ tvoří jediná studentka se dvěma texty o celkové hodnotě 366 slov. Korpus CzeSL-SGT nabízí v případě tohoto metadata dvě nejsilnější kategorie, které však v rámci části čínských mluvčích nejsou vůbec obsazeny. Jedná se o kategorii *48–60* a *60+* měsíců. Do budoucna by tedy bylo vhodné zaměřit se na studenty, kteří češtinu studují více než dva a více než tři roky.

	A1, A1+, A2, A2+		B1, B2	
pořadí	**měsíce studia**	**Frek.**	**měsíce studia**	**Frek.**
1	6-12	11466	24-36	17535
2		9988	36-48	3973
3	3-6	7642	12-24	3365

Tabulka T6 Tři nejčastější délky studia u studentů A úrovní a B úrovní

Další údaje se soustředí na formu studia češtiny. Bylo možné podat informace jak o absolvovaném studiu, tak o studiu právě probíhajícím. Studenti mohli označit i více hodnot, na výběr měli například individuální či komerční výuku, zda studium probíhalo pod nějakou základní, střední či vysokou školou, popřípadě v zahraničí, nebo zda je student samouk. Jak už mohla napovědět věková statistika, převážná část se se studiem češtiny setkala během studií vysoké školy. Jedná se o celých 90 % studentů. Téměř 25 % Číňanů se česky učilo při jiné příležitosti, než je v možnostech uvedeno. Třetí nejsilnější pozici obsadili studenti, kteří se ke studiu češtiny dostali v zahraničí – cca 22 %.

2.1.5. Čeština v rodině a roky strávené v České republice

Studentů, kteří označili někoho ze svých blízkých jako českého mluvčího, bylo pouhých sedm. Přihlédneme-li k celkovému počtu čínských studentů – 116 – tvoří tato část jen pouhých 6 % všech studentů – jejich texty jen 3 % výskytů korpusu. Jediné vybrané možnosti navíc byly *matka*, *otec* a *partner*. Nikdo nevyužil možnosti *sourozence*, *oba dva rodiče*, *tři rodinní příslušníci* ani jinou – vlastní možnost. Pro doplnění korpusu, tedy bude vhodné hledat čínské studenty, kteří již mají v rodině (nebo mezi blízkými) mluvčí češtiny.

U otázky, jak dlouhý čas studenti strávili v České republice, se bohužel největší část vůbec nevyjádřila. Zhruba polovinu výskytů tedy vytvořili studenti, o nichž nevíme, jak dlouho v Česku pobyli. Studenti, kteří v přirozeném prostředí češtiny strávili méně než jeden rok, vytvořili okolo 30 % korpusu a jsou druhou nejpočetnější skupinou. Dalšími volbami

bylo *1–2 roky* a *2+ let*. Vzhledem k početnější skupině „méně než 1 rok“ (počet výskytů je možný nahlédnout v Tabulce T7) a s přihlédnutím k množství studentů, kteří pravděpodobně Česko navštěvují v rámci studijního pobytu, by bylo vhodné přidat například kategorii *6 měsíců*.

Filter	Roky strávené v ČR	Frek.
1		31593
2	Méně než 1	19323
3	2+	6749
4	1-2	4230

Tabulka T7 Počet let, který studenti strávili v ČR

2.1.6. Česká učebnice

V otázce, jakou studenti používají učebnici, bylo možné vybírat ze sedmi možností, a to následující: *BC – Basic Czech; CC – Communicative Czech; CE – Ceština pro ekonomy; CMC – Chcete mluvit cesky?; CpC – Ceština pro cizince; ECE – Easy Czech Elementary; NCSS – New Czech Step by Step*. Mezi uvedenými možnostmi nenašlo svou učebnici hned 36 studentů, jednalo se o větší část z hlediska počtu výskytů v korpusu, a proto by bylo do budoucna jistě lepší nabídnout studentům více možností. Z nabídky Číňané využili pouze dvě učebnice. V převaze stojí texty studentů, kteří se učili z *New Czech Step by Step*, tři studenti potom uvedli jako svou učebnici *Communicative Czech*.

2.1.7. Bilingvismus a další jazyky studentů

V otázce bilingvismu bohužel značná část studentů neodpověděla. S jistotou však můžeme říci, že téměř 60 % korpusu vytvořili Číňané, kteří bilingvismus popřeli. V této otázce odpověděla kladně pouze jedna šestadvacetiletá studentka, která jako svůj paralelní jazyk k čínštině uvedla angličtinu. U otázky dalších jazyků studentů jsme kromě výše zmíněné

bilingvní studentky zaregistrovali pouze jediného autora, a to studentku, která krom čínštiny taktéž ovládá japonštinu.

2.2. Metadata týkající se pouze textu

2.2.1. Zadaný a reálný počet slov

Jak už bylo řečeno, čínskou část korpusu CzeSL-SGT tvoří 519 různých textů, které dohromady obsahují 61 895 slov. Průměrný text by tedy měl mít cca 120 slov. Vzhledem k faktu, že některé texty mají pouze něco okolo 20 slov (viz následující citace) a delší texty například 300 slov, jsou texty rozřazeny podle počtu slov do kategorií.

„*Prodam byt 5 + 1 , bilzko centra , ma velkou zaradu a novy kuchyňsky kout , velké prokoje a garage . Tel QQQ nebo E-mail QQQ* .“

Každý text by měl obsahovat tři různé typy metadat týkajících se počtu slov. Prvním je *limit slov*, který byl studentům zadán, dalším údajem je *reálný počet slov*, který práce obsahovala a posledním pak *reálný rozsah slov*, který pomůže text zařadit do příslušné kategorie dle reálného počtu slov. Práce s přesnými reálnými počty nejsou pro tuto práci rozhodující, proto budeme srovnávat pouze limity a reálné rozsahy slov.

Texty, u kterých se informace o limitu slov neobjevila, je nejvíc – cca 27 % korpusu. Všechny ostatní limity jsme přiřadili k odpovídajícím slovním rozsahům, aby bylo možné porovnat, s jakou úspěšností studenti slovní limity dodržovali. Z Tabulky T8 je patrné, že zatímco u prvního slovního rozsahu *50-99* není frekvence až tolik odlišná, poslední rozsah *150-199* slov Frekvenci příslušného zadání vůbec neodpovídá.

pořadí	rozsah slov	Frek.	limit slov	Frek.
1	50-99	21894	50, 70, 80, 90, 40+, 50+, 60+, 70+, 90+	17584
2	200-	20196	40+, 50+, 60+, 70+, 90+	12228
3	100-149	9189	100, 120, 40+, 50+, 60+, 70+, 90+	14453
4	-50	5653	20, 25, 30, 40, 40+	11039
5	150-199	4963	150, 40+, 50+, 60+, 70+, 90+	26319

Tabulka T8 Srovnání reálného rozsahu slov a zadaných limitů slov prací

Pro srovnání těchto dvou skupin by bylo mnohem příhodnější, kdyby kategorie limitu a reálného rozsahu slov byly totožné.

2.2.2. Zadání a druh postupu psaní

Znovu se dostáváme k problému, na nějž už jsme bohužel narazili i u jiných kategorií. Nejobsazenější pozicí ve statistice je prázdné, tedy nevyplněné, místo – cca 63 %. Druhé místo statistiky uvádí texty, u nichž byl zadaný postup práce – téměř 33 %. Pouze o necelých 4 % korpusu víme jistě, že postup práce měli studenti volný.

Reálnému postupu práce potom odpovídají 4 kategorie plus část, která postup nevyplnila. První čtyři takřka vyrovnané pozice (kolem 23 %) tvoří *informace*, *vyprávění*, *popis* a neznámý postup. Poslední nejmenší skupinkou jsou texty psané *úvahou*. Úvahy dávají v rámci korpusu dohromady pouhých 5 %, zatímco v celém CzeSL-SGT tvoří 18 %. Doplnění o více textů psaných v úvahovém postupu by tedy bylo jedině vítáno.

	INFORMACE		VYPRÁVĚNÍ		POPIS		ÚVAHA	
pořadí	počet slov	Frek.	počet slov	Frek.	počet slov	Frek.	počet slov	Frek.
1	50-99	10621	50-99	4855	50-99	5138	150-199	992
2	100-149	2437	100-149	3630	200-	3185	200-	837
3	-50	1857	200-	3201	-50	2888	50-99	713
4	150-199	438	150-199	2149	100-149	1815	100-149	619
5	200-	261	-50	699	150-199	958	-50	54

Tabulka T9 Srovnání rozsahu prací s ohledem na postup psaní

V kapitole 2.2.1 o počtu slov jsme zjistili, že rozsah *50-99 slov* je mezi čínskými studenty nejběžnější. Tabulka T9 však prozradila, že tomu tak není u všech postupů práce. Nejčastější rozsah úvah je totiž *150-199 slov*. Texty s *více než 200 slovy* si drží vysoká čísla v rámci všech postupů až na informace, kde se objevily až na posledním místě, a to pouze s jediným textem. Po kontrole limitu minut a jeho případných odchylek v závislosti na postupu psaní, vyšlo najevo, že všechny čtyři kategorie měly na sepsání práce nejčastěji 45 minut. Ověřili jsme též, zda některé postupy práce nebyly upřednostněny k sepsání na počítači, zjistili jsme však, že všechny texty, u nichž byl postup práce určen, byly napsány rukopisně.

2.2.3. Zkoušky a průběžné práce

Práce, které byly předmětem zkoušky, zaujímají v korpusu největší část – téměř 53 %. Všechny tyto texty byly psané rukopisně a časové limity na jejich tvorbu nepřesáhly 45 minut. Krom tohoto časového údaje měl menší počet textů limit 30 minut. Více než tři čtvrtiny zkouškových textů tvoří práce studentů na úrovni A1 či A2. Proto by bylo vhodné vyvážit tento typ textů i pracemi od více pokročilých studentů. Zhruba 40 % textů bylo součástí průběžné činnosti během vyučování.

2.2.4. Aktivita předcházející psaní textu

Texty, u kterých nebyla informace o předešlé aktivitě uvedena, nebo u nichž víme s jistotou, že studenti před jejich psaním žádnou související aktivitu neabsolvovali, tvoří víc než tři čtvrtiny korpusu (cca 76 %). *Práce s obrázkem* předcházela písemným pracím, které obsahují zhruba 7 % všech výskytů. A studenti, kteří před psaním zpracovávali *cvičení*, *diskutovali* nějaké téma, či trénovali *slovní zásobu*, dali dohromady pouze 4,6 % celého korpusu.

pořadí	aktivita	Frek.
1		33489
2	ne	13580
3	jine	7473
4	obrazek/film	4506
5	cviceni	1385
6	diskuse	932
7	slovni zasoba	530

Tabulka T10 Počet výskytů podle druhu aktivity předcházející psaní

Jak tedy dokládá Tabulka T10, korpus z většinové části postrádá texty, jimž by předcházely uvedené aktivity. Častěji dochází k práci s obrázkem, avšak je třeba doplnit kategorii, jež uvádí řešení cvičení, diskutování či aktivitu se slovní zásobou.

2.2.5. Stáří textů

Jak ukazují data konkrétních textů, veškeré výskyty korpusu pocházejí pouze z let 2009–2011, což znamená, že mezi texty z roku 2013, které byly při vzniku korpusu CzeSL-SGT k pracím nerodilých mluvčích přiloženy, nebyly žádné od čínských studentů.

3. VYBRANÉ CHYBY TYPICKÉ PRO ČÍNSKÉ STUDENTY ČESKÉHO JAZYKA

3.1. Přístupy k nabývání druhého jazyka

Dnes víme, že přístupy k definování učení se druhému jazyku jsou z historického hlediska nesourodé, přitom lze prakticky zmínit tři základní koncepce. První vidí v učení se druhému jazyku stejný či obdobný proces jako osvojování mateřského jazyka. Druhá koncepce upozorňuje na podstatné odlišnosti mezi nabýváním prvního a druhého jazyka a konečně třetí – teorie interlanguage.[14] S myšlenkou mezijazyka během nabývání druhého jazyku přišel poprvé Corder a to roku 1967. Jedná se o teorii přechodového jazyka, která první dvě koncepce odráží jako nesprávné především ve vnímání chyby jako nežádoucího jevu během učení. Pro interlanguage je chyba přirozenou součástí nabývání druhého jazyka, na jejímž základě můžeme vyhledat jistý charakter. Podle daných charakterů lze poté chyby rozdělit na jednotlivé typy a podle nich pak určovat příčiny vzniku či stádium procesu učení, ve kterém se student nachází.[15] Na základě pojetí chyby dle třetí koncepce je zpracována i tato kapitola.

3.1.1. Pojetí chyby

Jak už bylo zmíněno výše, chyby jsou dle teorie mezijazyka jevy, které nastávají při každém nabývání druhého jazyka, detailní vymezení chyby však s ohledem na jazykovou akvizici zůstává otázkou značně problematickou, a to především s ohledem na nejasnosti ohledně cílové formy češtiny – má-li se jednat o mluvenou nebo psanou podobu spisovného jazyka, či považovat za chybu specifické hovorové tvary. „*Definice chyby v současných výzkumech nabývání cizího jazyka tedy odkazují většinou na produkci nerodilého mluvčího, která se odchyluje od ‚správné' verze cílového jazyka, jejíž normou je nejednoznačně vymezená tzv. ‚norma (dospělého) rodilého mluvčího'.*"[16]

[14] HRDLIČKA, Milan: Kapitoly o češtině jako cizím jazyku, Fakulta pedagogická, Plzeň 2010, s. 143.
[15] HRDLIČKA, Milan: Kapitoly o češtině jako cizím jazyku, Fakulta pedagogická, Plzeň 2010, s. 149.
[16] ŠTINDLOVÁ, Barbora: Žákovský korpus češtiny a evaulace jeho chybové anotace, Karolinum, Praha 2013, s. 23.

3.1.2. Mezijazyk

Jinak lze označit též jako žákovský jazyk či interlanguage. Corder v roce 1967 označil mezijazyk jako idiosynkratický dialekt, čímž chtěl zdůraznit jedinečnost každého žákovského jazyka jakožto specifického vyjadřovacího systému jednoho konkrétního studenta. Takový přístup způsobuje vnímání studenta nejen jako pasivní bytost, která během učení přijímá poznatky, ale též jako aktivní jednotku, jež sama vytváří jazyková pravidla. Učení se tedy začíná vnímat jako proces, který vytváří metajazyk, pomocí něhož lze nahlížet vývoj charakteru žákova systému. Nejdůležitějším aspektem takového pozorování je samozřejmě studium chyby.[17]

3.2. Metodika

Zdrojem chyb byly texty napříč celým čínským korpusem v CzeSL-SGT, to znamená, že nedošlo k pozorování různosti chyb vzhledem k jazykové úrovni či k věku studenta a nevybírali jsme ani texty, které by byly specifické svojí délkou či tématem. Protože se v tomto případě jedná o kvalitativní analýzu vybraných typů chyb, které jsou typické buď přímo pro čínské studenty či pro mluvčí analytického mateřského jazyka, postup práce začal v důkladném nastudování gramatik čínského jazyka, a to především *Gramatiky současné čínštiny* (Lingea, Brno 2018). Následně jsme si určili pravděpodobné chyby, které by mohly být v korpusu obsaženy, a na základě této domněnky poté hledali chybové typy v textech korpusu. Výsledky vyhledávání potvrdily šest typů chyb, které jsme se pomocí pravidel čínské gramatiky pokusili analyzovat.

[17] ŠTINDLOVÁ, Barbora: Žákovský korpus češtiny a evaulace jeho chybové anotace, Karolinum, Praha 2013, s. 26.

3.3. Výchozí a cílový jazyk

3.3.1. Čínština

Čínštinu řadíme do sinické větve sinotibetské jazykové rodiny, přičemž existují dvě cesty, jak k tomuto jazyku přistupovat. První teorie zastává nynější charakteristiku, jež vidí čínštinu jako jeden jazyk s mnoha dialekty. Druhý přístup naopak považuje za vhodnější rozdělit ji do autonomních jazyků. Čínský jazyk prochází vývinem více než tři tisíce let, proto se v literatuře dochovala spousta různých forem. Pro tuto práci je nejvýznamnější moderní hovorová čínština, protože právě touto formou jsou čínští studenti češtiny ovlivněni nejvíce. Všechny známé podoby čínštiny byly a jsou zapisovány znakovým písmem, které primárně neodráží výslovnost. Tato skutečnost je výrazným argumentačním prvkem pro prvně zmíněnou teorii, která zastává jednolitost čínštiny. Čínské dialekty se totiž liší především z hlediska výslovnosti, naopak znakový zápis je napříč čínským územím mnohem méně odlišný. I z tohoto důvodu je běžnou praxí, že Číňané, kteří si v mluvené konverzaci těžko rozumí, mohou komunikovat pomocí dialogového zápisu.[18]

Co se týče typologického zařazení, většina lingvistů se shoduje na tom, že moderní čínština vykazuje největší množství charakteristických prvků typu izolačního jazyka. Jedná se o podtyp analytických jazyků, u nějž gramatické kategorie vyjadřují pomocná slova – neboli volné morfémy. Základ čínské věty tvoří pevný slovosled – hlavní uspořádání představuje podmět, přísudek a předmět (v tomtéž pořadí). Protože mohou morfémy v čínské větě vystupovat samostatně, není většinou nutné připojovat žádné afixy. Tato skutečnost se týká i moderní čínštiny, v níž se slov složených ze dvou morfémů objevuje stále více. Nejedná se totiž o doplňování významových morfémů afixy, nýbrž dalšími významovými morfémy. Afixy však v čínštině přece jen existují, i když jen ve velmi omezeném množství – např. afix pro plurál zájmen 们 „men“. Na rozdíl od českých složenin, jsou v čínštině u slov tvořených

[18] *Gramatika současné čínštiny*, Lingea, Brno 2018, s. 10.

více než jedním morfémem stále jasná ohraničení morfémů a morfémy lze v jakémkoli případě vydělit ze slova a užít je v jiném významu.[19]

Čínština se vyznačuje též minimální flexí a aglutinací, tudíž musí gramatické kategorie vyjadřovat pomocí volných morfémů (pomocných slov), či pomocí slovosledu – např. spojení *pojď se mnou* 跟我去 „gēn wǒ qù“, kdy první morfém vyjadřuje instrumentál.[20] Ačkoli se čínština nejpřirozeněji řadí mezi izolační typy, stále vykazuje různé znaky jiných jazykových typů. Malé množství sufixů svědčí o drobném zastoupení typu aglutinačního, pro převod sloves do kategorie podstatných jmen slouží původně lexikální morfémy jako např. podstatné jméno *malba* 画儿 „huàr“, které vzniklo spojením *malovat* 画 „huà“ a sufixu 儿 „er“. V čínském jazyce lze dokonce najít i jisté stopy flexe. Ta je především v podobě introflexe uskutečněna změnou tónu. Slovní zásobu bychom mohli označit jako rys polysyntetický, a to hlavně popisná pojmenování – např. podstatné jméno *velikost* 大小 „dàxiǎo“, které tvoří slova *(být) velký* 大 „dà“ a *(být) malý* 小 „xiǎo“.[21]

3.3.2. Důležité faktory češtiny jako cizího jazyka

Jak je tomu u většiny jazyků, ani český jazyk nemá pouze jednu podobu. Student, který se ho učí v přirozeném prostředí, se denně setkává hned s několika jeho podobami. Ráno při sledování nejnovějších zpráv slyší publicistický spisovný jazyk, cestou do školy může při koupi kávy narazit na prodavače, který na něj bude mluvit prostřednictvím moravského dialektu, přednáška ve škole poté probíhá ve spisovém jazyce akademickém, čeští kolegové ze školy komunikují pomocí obecné nespisovné formy českého jazyka a tyto všechny podoby ještě může završit četbou některého z Vančurových děl s velmi květnatou češtinou překypující archaismy.

Z těchto důvodů je stále velkým otazníkem, jakým způsobem k těmto jednotlivým podobám češtiny během výuky přistupovat. Hlavními dvěma favority můžeme s jistotou jmenovat spisovný jazyk a obecnou češtinu. V takové situaci narážíme na soupeře zcela

[19] PACKARD, Jerome: *Chinese as an Isolating Language*, University of Illinois, 2006, s. 355–358.
[20] POPELA, Jaroslav: *Skaličkova jazyková typologie*, Masarykova univerzita, Brno 2006, s. 18–19.
[21] POPELA, Jaroslav: *Skaličkova jazyková typologie*, Masarykova univerzita, Brno 2006, s. 26.

odlišného charakteru.[22] V případě, že učitel pracuje se studenty, již jsou pravými začátečníky, měl by výuku vést skrze spisovnou verzi češtiny. Jsou samozřejmě známy i případy, kdy se vyučující rozhodnou volit též tvary obecné češtiny, v této situaci je však nezbytné seznámit studenta s charakterem obecné češtiny komplexně a systematicky. Student nesmí mít pocit, že je možné volně zaměňovat prvky obecné a spisovné češtiny, aniž by to mělo vliv na sémantiku sdělení. V případě, že jsou studenti s charakteristickými prvky obeznámeni, není třeba považovat obecně české prostředky za chybné.[23]

Z typologického hlediska češtinu – stejně jako další slovanské jazyky – řadíme mezi flexivní typ syntetických jazyků. Tato skupina je charakteristická především značením jedné gramatické kategorie různými afixy: „*např. v češtině nominativ plurálu u substantiv může být vyjádřen mnoha různými koncovkami: muž-i, měst-a, žen-y atd*.“ Stejně tak platí pravidlo, že pro různé gramatické kategorie může fungovat pouze jeden afix – můžeme vidět například u jednotlivých slovních základů: „*např. koncovka –e u slova stroje označuje buď singulár, genitiv a maskulinum nebo plurál a nominativ nebo akuzativ*.“ [24]

3.3.3. Vybrané rozdíly

Čeština je jazyk syntetický, tudíž se v ní gramatické významy vyjadřují předponami či příponami, to znamená v rámci slova, které nese věcný význam. Čínština se naopak řadí do druhé skupiny – analytických jazyků, v níž se gramatické významy tvoří prostřednictvím slovosledu, či užitím pomocných slov. Je důležité zmínit, že slovosled hraje v čínštině obrovskou roli, a proto je na rozdíl od češtiny pevně stanoven.[25]

Nejvýznamnější rozdíl mezi češtinou a čínštinou pro chybovou analýzou je rozdílnost morfologických kategorií. Čínština totiž nemá totožné rozdělení slovních druhů, proto působí čínským žákům při studiu češtiny obrovské potíže: „*…čínská slova nevyjadřuj druh svým tvarem jako na př. slova česká. Tato okolnost způsobila, že existence druhů slov v čínštině není tak nesporná jako v jazycích flektujících, kde se přijímá jako samozřejmá skutečnost*.“[26]

[22] ČERMÁK, František: *Jazyk a jazykověda*, Univerzita Karlova v Praze, Karolinum, Praha 2004, s. 43.
[23] HRDLIČKA, Milan: *Cizí jazyk čeština*, ISV, Praha 2002, s. 105.
[24] ČERNÝ, Jiří: *Úvod do studia jazyka*, Rubico, Olomouc 1998, s. 62–63.
[25] KRUPA, Viktor – GENZOR, Jozef – DROZDÍK, Ladislav: *Jazyky světa*. Bratislava 1983, s. 19.
[26] KALOUSOVÁ, Jarmila: *Vybrané kapitoly z gramatiky moderní čínštiny*, Karlova universita v Praze, Praha 1954, s. 3.

Většinu slovních druhů sice lze pojmenovat stejně jako v češtině, charakteristiky jednotlivých kategorií se však značně liší.

Pokud se na kontrast češtiny s čínštinou podíváme z hlediska výslovnosti, vyvstane nám nejvýraznější rozdíl mezi systémem českých a čínských samohlásek, a to absence krátkých a dlouhých samohlásek v čínštině. Délku samohlásky v čínštině ovlivňují jiné faktory, a to například počet hlásek ve slabice, či přízvuk. Délka samohlásky ale nikdy nemá významotvorný charakter. Proto je velmi očekávatelné, že čínským studentům bude rozlišování dlouhých a krátkých samohlásek studium ztěžovat.[27]

Další problematickou oblastí pro čínské mluvčí češtiny je určitě otázka přísudku. Skutečnost, že kromě sloves může na pozici přísudku stát taktéž samotné adjektivum, může čínské studenty vést ke stejnému postupu při tvorbě české věty. Jak adjektivní přísudek beze spony funguje v rámci čínštiny, jsme se rozhodli ukázat na příkladu: *Miminko je roztomilé.* 宝贝很可爱。„Bǎobèi hěn kě'ài.“ – *Miminko (moc) roztomilé.* Ve větě nedochází k užití spony být (是 „shì“), proto je pravděpodobné, že Číňan vynechá sponu i v češtině a výsledný překlad poté může vypadat následovně: *Miminko roztomilé*.

Dalším významným rysem je čínská adjektivizace, k níž v čínštině dochází zejména u sloves a podstatných jmen, a to nejčastěji bez formálních změn. Může se tedy stát, že při tvorbě české věty se na pozici shodného přívlastku ocitne podstatné jméno. Taktéž dochází k záměnám podstatných jmen a přídavných jmen na pozicích podmětu, předmětu či neshodného přívlastku.

V čínštině je často užívaná duplikace různých slovních druhů za různými účely. Duplikování vždy pozměňuje původní význam slova. Například duplikací pravých adjektiv se zdůrazní jejich vlastnost, při duplikaci některých podstatných jmen či měrových slov, dojde k zobecnění jejich původního významu – např. *každý rok* 念念 „niàn niàn“ (= *rok rok*). Proto jsme se rozhodli prozkoumat korpus a zjistit, zda se v korpusu čínských mluvčích češtiny některé duplikované tvary nevyskytují.

[27] JAKUBŠE, Karolína: Problémy Číňanů při nácviku české výslovnost, in: *Sborník Asociace učitelů češtiny jako cizího jazyka (AUČCJ) 2012*, Akropolis, Praha: 2012, s. 156.

Dále bychom se rádi vyjádřili ke kategorii čísla. Čínská substantiva většinou svým tvarem nerozlišují rod ani číslo, existuje však několik způsobů, jak číslo v čínštině vyjádřit. První možností je připojení deiktického slova či číslovky, a to pomocí měrového slova – např. *tři lidé* 三个人 „sān gè rén“ (*tři, numerativ, člověk*). Jednotné číslo se obyčejně vyjadřuje pomocí číslovky *jedna* 一 „yī“, takové spojení by se dalo přirovnat k principu anglického neurčitého členu *a, an*, v praxi může vypadat například následovně: *(jedno) dítě* 一个孩子 „yī gè háizi“ (*jedno, numerativ, dítě*). Poslední možností, jak vyjádřit číslo je použití morfému „men“ 们 u zájmen. Naprosto jednoduše tak lze vytvořit množné číslo od prvních tří osobních zájmen:

já, ty, on	我ᵒ你ᵒ他	„wǒ“, „nǐ“, „tā“
my, vy, oni	我们ᵒ你们ᵒ他们	„wǒmen“, „nǐmen“, „tāmen“

Na základě těchto skutečností je očekávatelné, že Číňané během užívání českých (především) podstatných jmen budou tvary množného a jednotného čísla zaměňovat.

3.4. Konkrétní případy chyb

Následující kapitola vycházela především ze znalostí gramatických pravidel z knihy *Gramatika současné čínštiny*, Lingea, Brno 2018 a všechny uvedené příklady českých vět čínských studentů jsou získány z korpusu CzeSL-SGT[28].

3.4.1. Přísudek jmenný beze spony

V následujícím oddílu se budeme věnovat přídavným jménům. V čínském kontextu je pravděpodobně lepší tuto kategorii nazývat spíše jako adjektiva, protože v čínštině jsou si na základě svých gramatických funkcí bližší spíše se slovesy nežli s podstatnými jmény. Adjektiva v čínštině tedy stejně jako slovesa mohou být například blíže určována příslovci, nebo stát v pozici přísudku bez použití sponového slova. Taková situace nastává například ve větě: *Tato kniha je drahá*. 这本书很贵。 „Zhè běn shū hěn guì.“ – *tato (pomocné sloveso) kniha (moc) drahá*. Zde je tedy zřetelně viditelné, že adjektivum *drahá* 贵 „guì“ zaujalo

[28] *CzeSL-SGT: korpus češtiny nerodilých mluvčích s automaticky provedenou anotací.* Ústav Českého národního korpusu FF UK, Praha 2013. Dostupný z WWW: http://www.korpus.cz

pozici přísudku. Čínský student by tedy v možných verzích takových vyjádření mohl mít větší sklony k chybám, například přeložit uvedenou větu jako: *Tato kniha drahá*. Na základě takových předpokladů jsme korpus prohledali a narazili na několik situací.

„*Babička je veselá , i když často nemocná* ."

„*Když jsem byla malá , se mnou chodila jedna holka spolu do třídy . a moc osamělá . Jsem jedina ní bavila . Aproto její duševní charakter zlepšila* ."

V těchto souvětích vypráví studenti o třetích osobách. V první větě je osoba (*babička*) popisována zprvu pomocí jmenného se sponou a v následující větě souvětí je jako přísudek uvedeno pouze přídavné jméno beze spony. V tomto případě je sporné, zda se jedná o vynechání z důvodu převedení pravidla z čínské gramatiky, anebo zda se jedná o elipsu, která by za jistých okolností byla možná i v českém jazyce (*Je chudý, ale šťastný*.). Vzhledem k jazykové úrovni studenta (A2) ale předpokládáme, že o možnosti elipsy v češtině neví, a proto chybu přisuzujeme zamýšlenému použití samotného adjektiva jakožto přísudku a správná verze věty tedy měla znít: *Babička je veselá, i když je často nemocná*. Ve druhé větě vypráví studentka o své spolužačce, o které následně sděluje informaci, že byla osamělá, ovšem opět bez použití spony. Protože spona nebyla užita ani v předchozích větách souvětí, jasně se jedná o chybné užití adjektiva, které je orientované čínskou gramatikou. Větu bychom tedy opravili následovně: *Když jsem byla malá, chodila se mnou do třídy jedna holka, která byla moc osamělá*.

„*Myslím , že lepší v pátek , protože celé odpoledne mám čas…*"

Prostřednictvím této věty se student v dopise snaží domluvit s adresátem na schůzku. Přísudkovým adjektivem je nyní *lepší*, které postrádá sponu a v tomto případě i potřebný podmět. Studenti češtiny zpočátku studia často slýchají, že podmět v češtině nemusí být vyjádřený, a kvůli těmto upozorněním poté dochází k vypouštění podmětů i na místech, kde je jejich užití zapotřebí. Souvětí bychom tedy opravili takto: *Myslím, že v pátek to bude lepší, protože mám čas celé odpoledne*.

„*Praha nemá mnoho vysokých budov , a to je velmi jedinečný charakter . Naše hlavní město*

Peking je plné různých vysokých budov a proto vypadá jako les betonů a aut . Opačně , odlišná od Pekingu . V Praze téměř nemůžu najít budovy , které jsou vyšší než pět pater .“

V tomto případě dochází k podobnému případu jako v předešlé situaci. Student hovoří o Praze a přirovnává ji k Pekingu. Nejdříve popisuje Prahu, poté charakterizuje Peking a následně se opět vrací k Praze pomocí věty *Opačně , odlišná od Pekingu .* Adjektivum *odlišná* je zde chybně užitým jmenným přísudkem beze spony. I v této větě navíc chybí podmět. Přihlédneme-li totiž k významové stránce, jasně se mluví o Praze, jejíž téma již bylo mezitím přebito tématem Pekingu, a proto bylo na místě použít opět podmět vztahující se k Praze. Navrhovali bychom tedy znění: *Kdežto Praha je od Pekingu odlišná*. V ideálním případě upustit od užití adjektiva a vyjádřit význam následujícím způsobem: *Kdežto Praha se od Pekingu liší*.

3.4.2. Záměna přídavných a podstatných jmen

Kvůli jasně nedefinovaným hranicím mezi čínskými slovními druhy, dochází u čínských mluvčích češtiny k častým záměnám adjektiv a substantiv. Čínská gramatika navíc uvádí pojem *adjektivizace*, u něhož dochází k užití sloves či podstatných jmen na pozicích typických pro přídavná jména. Takové přeměny nejčastěji probíhají bez formálních změn, a tak je význam sdělení patrný jedině z kontextu. Stejně tak může dojít i k užití adjektiv na místech, jež jsou typická pro substantiva, a právě na tento případ jsme se soustředili během výzkumu.

„*…rodiče mi často říkali " porzo , tam je nebezpečná…*“

„*Ti , kteří mají rádi lyžování , nemusejí se bát laviny nikdy na českých horách . V ČR neexistuje . Někdy je sněhový vichr . Je nebezpeční kvůli spatně viditelnosti .Je nebezpeční kvůli spatně viditelnosti .*“

V tomto případě víme, že slovo *nebezpečí* 危险 „wéixiǎn“ je v čínštině užíváno jako podstatné jméno – *Mám rád nebezpečí.* 我喜欢危险。„Wǒ xǐhuān wéixiǎn.“ – *já (mám) rád nebezpečí* – i jako přídavné jméno – *nebezpečný člověk* 危险的人 „wéixiǎn de rén“ – *nebezpečný (numerativ) člověk*. Proto je pravděpodobné, že čínský student se může spokojit pouze se znalostí jedné z forem překladu čínského výrazu (s přídavným či podstatným

jménem) a poté jí využívá ve větách bez ohledu na potřebu užití určitého druhu gramatické kategorie.

„…*Samozřejmě , všechno je v pořádku , protože jsme nedělali žádné zlé*…“

Podobná situace nastává u slov *zlý* a *zlo*, kdy byl taktéž nalezen výskyt s tímto typem záměny. Student zde popisoval zážitek, během něhož byli s kamarádem obviněni z trestné činnosti, nakonec však bylo vše v pořádku, protože neudělali nic špatného. Z významu sdělení vyplývá, že chtěl zřejmě použít slova *zlo*, ale právě z důvodu relativního zařazení čínského překladu slova použil formu přídavného jména. Zamýšlená verze tedy měla být: *...protože jsme nedělali žádné zlo*. Avšak správná verze zní: …*protože jsme nedělali nic špatného*.

„*Brno je nejdůležitější město na Moravě , které má dost vyznamých . Přiklady jsou Milan Kundera , Gregor Mendel , Leoš Janáček Ernst Mach atd* .“

V této situaci chtěl student popsat město Brno jako místo, které je význačné spoustou důležitých osobností. Použil pro tento záměr přídavného jména *významných*, zatímco zamýšleným významem bylo nejspíš *osobností*. Ideální by však v této větě bylo využít obou těchto prostředků a navrhujeme zde též změnu přísudku, výsledek opravy na správné vyjádření zamýšleného významu by tedy byl pravděpodobně tento: …*které je známé spoustou významných osobností*.

„*Já osobně si myslim , že češti studenti mají více výhody než ostatní studenti . Protože lety povinná školné docházky . Ale hlavni je to zdarma* .“

U posledního příkladu tohoto oddílu dochází ke komplikované situaci. Význam prostřední věty *Protože lety povinná školné docházky*. může být interpretovaný dvěma způsoby. První způsob umožňuje naše zařazení příkladu a říká, že správná verze věty byla myšlena takto: *Protože léta je povinnost školní docházky*. Při druhém uchopení významu by došlo k následující úpravě: *Protože školní docházka je léta povinná.* První teorii bychom rádi obhájili tvarem slova *docházky*, který je v pozici neshodného přívlastku a též skutečností, že v čínštině slovo povinné a povinnost opět splývají v jeden tvar 义务 „yìwù“. Proto považujeme za pravděpodobnější, že student chtěl vyjádřit spíše význam, pro nějž je v češtině nutné využít podstatného jména. Pokud bychom se přiklonili spíše k druhé teorii, museli

bychom tuto větu přesunout do prvního oddílu této podkapitoly, kde jsme se zabývali přídavnými jmény na místě přísudku, protože v této větě taktéž nebylo užito potřebné spony.

3.4.3. Duplikace pro zdůraznění slov

Pro hledání duplikací v korpusu bylo nutné vymezit, na jaký konkrétní typ se zaměříme. Vzhledem k malému množství nálezů duplikací jsme se nakonec rozhodli zvolit druh, který zdůrazňuje vlastnost slova. Při výzkumu jsme narazili na tři výskyty slovního spojení *moc moc*. Případy byly nalezeny ve třech různých textech, které napsali tři různí studenti. V těchto situacích se bezpochyby jedná o vyzdvihnutí původního významu slova. Uvádíme všechny tři nalezené případy:

„…*Číšník je moc moc pomalý jako moji babička*…“
„…*Podíváš historický domy a hrady . Rokoko , Cotický … jsou moc moc hezký*…“

U těchto vět došlo ke zdůraznění slova *moc*, které zároveň akcentuje následující adjektivum. V první větě se tedy sémanticky podtrhl též význam slova *pomalý*, sdělení bychom mohli náležitě opravit například tímto způsobem: *Číšník je nehorázně pomalý jako moje babička*. Druhá věta potom vyzdvihla též význam slova *hezký*, správně by mohla znít takto: *Prohlédneš si historické domy a hrady, rokoko i gotické styly jsou převelice hezké*.

„…*nechci jenom v restauraci , nemam rad restauraci . Protože , teď ďela činsXXX restauraci moc moc lidi*…“

Třetí část sdělení se opět týká dopisu. Student v něm oznamuje, že nechce pracovat v restauraci, protože v ní pracuje spousta lidí. Pomocí duplikace slova *moc* akcentuje jeho význam a dává tím najevo, že se jedná o opravdu velký počet. Větu bychom proto navrhovali opravit například na toto znění: *Nechci pracovat jen v restauraci, nemám je rád, protože nyní v čínských restauracích pracuje nespočet lidí.*

Z dalších případů duplikace uvádíme obdobný příklad.

„*Vojací si vzali hodně hodně mužů na praci na vesnici* .“

Jedná se o esej, v níž student líčí průběh stavby Velké čínské zdi. Duplikace slova *hodně* pomohla záměru zdůraznit velké množství mužů, které si vojáci vzali na pomoc. Pro téměř shodnou volbu postupu jako u minulého příkladu, uvádíme stejnou možnost překladu opravy duplikace: *Vojáci si vzali na práci nespočet mužů na vesnici*.

3.4.4. Množné číslo u podstatných jmen

Jak už bylo zmíněno v předchozí podkapitole, čínská podstatné jména svým tvarem číslo primárně nevyjadřují. Číslo lze určit pomocí číslovky nebo deiktického slova, dále také můžeme číslo určit pomocí čísla podmětu. Nejčastěji však Číňané vyjadřují číslo prostřednictvím kontextu, a proto se při tvorbě českých vět stává, že tento zvyk si z mateřštiny přinášejí.

V otázce množného čísla jsme se zaměřili především na postavení podstatných jmen jednotného čísla na pozici ve větě, kde podle kontextu mělo být podstatné jméno čísla množného. Osoba a číslo totiž v čínštině nemusí být vyjadřovány právě z toho důvodu, že jsou dané osobou a číslem podmětu. „…*Podstatné mená v čínštine často skôr označujú celý druh jako jediný předmět, napr. kung-žen môžeme v rozličných súvislostiach preložiť nielen jako* ‚*robotník*‘*, ale aj jako* ‚*robotníci*‘ *alebo* ‚*robotníctvo*‘…“[29]

K aplikování takovéhoto pravidla v čínské větě se dostáváme například ve větě: „…*Měl jsem hodně dárek*…“ (správně *Měl jsem hodně dárků*.) Pro čínského mluvčího je zde míra jasně daná pomocí neurčité číslovky, a tak už pro něj není nutné měnit tvar substantiva s ní spojeného. Obdobných příkladů je v korpusu možno dohledat mnoho.

3.4.5. Pomocné slovo s původním významem *moci*

Čínština má velké rezervy ve způsobu vyjadřování času. Běžné kategorie času jsou určovány pomocí tvarů sloves, díky nimž poté poznáme, zda daný děj již proběhl, probíhá, či zda teprve

[29] KRUPA, Viktor – GENZOR, Jozef – DROZDÍK, Ladislav: *Jazyky světa*. Bratislava 1983, s. 294.

probíhat bude. U čínských sloves však ke změně tvaru nedochází, vždy tedy mohou vyjadřovat minulost, přítomnost i budoucnost. Časový rámec věty je v čínštině nejčastěji vyjadřován pomocí *jmen času* a *příslovcí času*. Všechna pomocná slova, která pomáhají přiblížit časové zařazení, však působí spíše jako ukazatele slovesného vidu. Samotná slovesa jsou tedy vnímána jako nedokonavá, to znamená, že nevíme, kdy děj začal, kdy skončil, ani zda byl dosažen cíl takového děje. Slovesa tohoto typu se užívají především u dějů všeobecně probíhajících, charakteristických, anebo vyjadřují spíše budoucnost. Dokonavost poté lze dosáhnout pomocí užití různých pomocných slov, mezi něž patří například vidočasové slovesné ukazatele, větné částice či výsledkové modifikátory.

Jak bylo řečeno výše, budoucnost lze nejjednodušeji vyjádřit použitím holého slovesa bez připojení jakéhokoli z pomocných slov či značení jakýchkoli gramatických prostředků. Další cestou může být využití různých výrazů, jež se prostřednictvím vlastního významu jasně vážou k budoucímu času. Takovouto funkci ukazatele budoucnosti mohou ve větě plnit například modální slovesa *muset, chtít* 要 „yào“ a *moci, umět* 会 „huì“. První sloveso pomáhá formulovat oznámení budoucích plánů či předpovědí, které jsou jasně dané, neoddiskutovatelné, či podložené důkazy. Využití v praxi si ukážeme na větě: *Dnes večer jdu na večírek.* 我今晚要去参加派对。„Wǒ jīn wǎn yào qù cānjiā pàiduì.“ – *já dnes večer (chci/musím) odejít účastnit se večírku*. V případě, že by čínský student chtěl vyjádřit jasný plán, kde večer bude a místo věty *Večer tam jdu.* by použil větu *Večer tam chtít jít*. Je zajisté jasné, že by došlo k jistému sémantickému rozdílu.

Modální sloveso *moci, umět* 会 „huì“ naopak vyjadřuje budoucí plány, které jsou spíše návrhem, než aby byly podloženy důkazy. Jedná se tedy o soukromé domněnky toho, co se bude v budoucnosti dít. Toto sloveso však může být využito i při výhružkách či vlastních předpokladech. Určitě se shodneme na faktu, že věta *Budete toho litovat.* 你们都会后悔。„Nǐmen dōu huì hòuhuǐ.“ – *vy všichni moci litovat* – obsahuje jasný záměr, vyvolat v recipientovi pocit, že platnost předpovědi je jistá. Přesto je zde využito slovesa, které má původní význam nejistoty. Dojde-li tedy v češtině na chybně doslovný překlad takové věty (například *Můžete toho litovat.*), je významový posun velmi znatelný.

V korpusu jsme se proto rozhodli hledat podobné výrazy, u kterých jsme se pokusili odhadnout zamýšlený význam.

„*Slyšel jsem že budeš přijet do prahy . V sobotu . Můžeme se sejít na letiště v 18 hodin . Budu čekat na tebe před východem…*“

Prvním příkladem bylo dopisní sdělení plánů, které obsahuje tři přísudkové pozice, přičemž dvakrát bylo použito vyjádření pomocí složených tvarů *budeš přijet* (správně *přijedeš*) a *budu čekat* a jednou došlo k použití slovesa *můžeme*. Předpokládáme, že student užitím slovesa *můžeme* zamýšlel význam, který se shoduje s užitím modálního slovesa *moci, umět* 会 „huì“ v čínštině. Tento předpoklad podporuje především význam následující věty, kdy student adresátovi svého dopisu oznamuje, že na něj bude na letišti čekat, proto se zdá býti nelogické, že by mu sraz na letišti v předchozí větě pouze navrhoval. Správné znění prostřední věty by tedy mohlo být: *Sejdeme se na letišti v 18 hodin*.

„*Pojedu vlakem tři hodiny . Budu spát v hotelu dvě dny . Můžeme nakupovat v supermaketu . V poledne budeme vařit jidlo . Po oběd , budeme hrát basketbal .*“

Tento příklad z korpusu je také součástí dopisového formátu a student v něm sděluje svému adresátovi jasné a podrobné plány o tom, co budou společně dělat. Na poli pěti krátkých jednoduchých vět dochází k užití pěti přísudků. Čtyři z nich jsou v budoucím tvaru (*pojedu*, *budu spát*, *budeme vařit*, *budeme hrát*) a pátý je vyjádřen pomocí slovesa *můžeme*. I zde se domníváme, že jde o vyjádření budoucnosti pomocí odkazu na čínské modální sloveso, a to z hlediska výčtu naplánovaných událostí, které nelze měnit. Z významového hlediska je též pravděpodobné, že jídlo se student chystá vařit ze surovin nakoupených v supermarketu, proto by nebylo logické uvádět nákup pouze jako možnost.

3.4.6. Anteponovaný větný člen

V čínštině existuje málo věcí, které mohou změnit pevnou formu slovosledu a právě téma věty je jednou z nich. Pokud je předmět hlavním sdělením celé věty čínská gramatika jej dovoluje přenést na začátek věty. V takovém případě anteponovaným větným členem nazýváme přímý předmět, který se většinou pro lepší orientaci v čínské větě odděluje čárkou. Jedná se o obvykle známý a již určitý předmět slovesa. V některých situacích je dokonce

možné vynechat podmět pro ještě větší zdůraznění tematického jádra. Takový postup je však možný jen za okolnosti, že nemůže dojít k záměně anteponovaného větného členu za podmět.

Jak už bylo zmíněno v předešlých kapitolách, běžné postavení předmětu v čínském slovosledu je až za podmětem a přísudkem. Pro příklad si můžeme uvést typickou čínskou větu: *Četl jsi už noviny?* 你看完报了吗？„Nǐ kànwán bào le ma?“ – *ty přečíst noviny (vidočasový slovesný ukazatel), (tázací slovo)*. Záměrně jsme uvedli příklad, na kterém je vidět, že pevný slovosled je v čínštině zachován i v otázce. Odpovědí na zmíněnou otázkou by mohla být například věta: *Přečetl jsem dnešní noviny.* 今天的报我看完了！ „Jīntian de bào wǒ kànwánle!“ – *dnešní (přivlastňovací slovce) noviny já přečíst (vidočasový slovesný ukazatel)*. Jak je tedy vidět, v odpovědi se tématem věty staly noviny s důrazem na doplnění, že jde o noviny dnešní. Došlo tedy k použití anteponovaného větného členu, kdy byl přímý (někdy též určitý) předmět přesunut na začátek věty. V češtině by překlad takovéto věty nemusel působit jako chybný – *Dnešní noviny jsem přečetl.* – to ovšem platí pouze do chvíle, kdy předmět není oddělen čárkou, jak bývá v čínštině zvykem v zájmu lepší orientace.

Anteponovaný větný člen často bývá předmět, který už známe z kontextu vyprávění, právě z tohoto důvodu bývá nejčastěji vyjádřen pomocí ukazovacího výrazu. Dva takové případy uvádíme z čínského korpusu:

„*Tu je jedna můj nejraději čast v té knize . Ale <u>ta kniha</u> přečetl asi před sedmi lety . už hodně zapoměl .*“

Student hovořil o jeho oblíbené části v knize, načež chtěl upozornit na skutečnost, že ji už dlouho nečetl. Téma věty (předmět *kniha*) tedy přesunul na začátek věty a poté teprve doplnil přísudek.

„*Co potřebuju , když vařím . Já mám ráda nudle . Když vařím nudle , potřebuju půl kil čínské těstoviny . <u>Toho</u> těstoviny musí vařit . Ještě potřebuju čtvrt kilo masa*…“

K podobnému případu došlo u druhého příkladu, v němž student popisoval svůj styl vaření. Nejprve uvedl, jaké suroviny potřebuje (*těstoviny*) a poté chtěl navázat, co s již zmíněným předmětem bude dělat. V další větě tedy opět pomocí ukazovacího zájmena

dosadil přímý předmět hned na začátek věty a teprve následně doplnil přísudek. V tomto příkladu jistě vyvstane otázka, zda se namísto přímého předmětu nejedná o podmět. Celý postup vaření je však nahlížen pouze z první osoby a současně je také využito ukazovacího zájmena ve 4. pádu (*někdo* vaří – koho? co?), proto si dovolujeme dát přednost první variantě.

V dalších případech se jednalo o anteponované větné členy, které byly odděleny čárkou.

„*Proti nám , lidé v České Rebuplice si víc váží historií .*“

V tomto případě se jedná o srovnání čínské a české populace. Zdůrazněno je téma kontrastu, a to pomocí vazby předmětu a dativní předložky.

„*Výsledkem toho , byl jsem nervóznější a nervóznější .*“

U posledního příkladu se dostáváme k poněkud komplikovanější situaci. Domníváme se, že počáteční slovo věty *výsledkem* je zde použito jako předložka, věta by tedy správně mohla znít například: *Kvůli tomu jsem byl nervóznější a nervóznější*. Nastává zde tedy stejný stav jako u předchozího příkladu. Tematický předmět je přesunut na začátek věty a oddělen čárkou. K posledním dvěma ukázkovým větám je třeba poznamenat, že z pohledu českého syntaxe se jedná spíše o příslovečné určení, pro pochopení chyby ze strany čínského studenta je však potřeba přistupovat k těmto větným členům jako k předmětům.

ZÁVĚR

Ve druhé kapitole práce jsme se pokusili analyzovat čínský korpus prostřednictvím metadat, která charakterizovala jak studenty, tak samotné texty a podmínky jejich vzniku. Metadata tak s ohledem na množství výskytů v celém korpusu tvoří přesné statistiky, na jaké části lze korpus rozdělit podle jednotlivých parametrů. Přesné statistiky se však nepodařilo získat vždy, protože většina parametrů nebyla uvedena u všech textů. Některé metadatové typy dokonce zaznamenaly prázdné políčko u nadpoloviční většiny. K takovému případu došlo například u metadata *aktivity předcházející psaní textu*, kdy počet výskytů bez žádné odpovědi na tento parametr utvořil víc než 54,1 % korpusu. Podobný případ se objevil u metadata *let strávených v ČR*, kdy počet výskytů bez odpovědi taktéž překonal polovinu.

S absolutními čísly jsme naopak pracovali například v případě genderového zastoupení, kdy jsme zjištili, že korpus tvoří přesně 57 žen a 59 mužů (a to beze zbytku, u něhož by pohlaví nebylo určeno). I věková otázka byla plně zastoupena, a tak bylo možné zjistit, že čínské studentky byly v průměru o něco starší než čínští studenti. Pravdou však zůstává, že zastoupení jednotlivých věkových kategorií bylo velice nevyrovnané. Kategorie dětí a mládeže do 15 let nedosáhly ani na 4 %, zatímco skupina *16+*, která už dále není členěna, obsahovala relativně pestrou škálu věků, kterou ohraničil nejstarší osmatřicetiletý student. Pro lepší orientaci bychom navrhovali členění na kategorie *16–20*, *21–25*, *26–30* a jako poslední *31–40*. Další sběr dat by se potom měl zaměřit na starší studenty.

Mezi úrovněmi češtiny čínský korpus naprosto postrádá mluvčí úrovně C1 a C2. Ostatní úrovně jsou poměrně vyváženy, proto by bylo vhodné doplnit korpus především o texty studentů s nejvyššími úrovněmi SERR. Dalším zajímavým poznatkem s ohledem na úroveň jazyka bylo, že studenti A úrovní neměli téměř v 84 % možnost využít jakékoli pomůcky, a to na rozdíl od studentů B úrovní, kteří měli v 64 % případů k dispozici slovník. Proto by bylo zajímavé při příštích zadáváních prací, které budou zamýšleny k zařazení do korpusu, povolit pomůcku i studentům slabší úrovně.

Při zkoumání intenzity studia nás překvapilo, že 91 % studentů tráví více než 15 hodin týdně studiem češtiny. Při takovém množství je tedy otázkou, zda takováto statistika opravdu odráží možnou skutečnost. Dále bychom také rádi upozornili na skutečnost, že chybí

kategorie pro studenty, kteří češtinu studují *3–5* hodin týdně. S ohledem na délku studia češtiny bylo zjištěno, že čínský korpus zcela postrádá kategorie *48–60* a *60+* měsíců, proto je třeba se nyní zaměřit na mluvčí, kteří se češtině věnují více než 2 nebo více než 3 roky. Dále je potřeba zmínit, že 97 % výskytů je tvořeno studenty, kteří neuvedli nikoho ze svých blízkých jako mluvčího českého jazyka, proto je nutné vyhledat zástupce, kteří by např. v rodině měli někoho, kdo umí česky. Při již zmíněném metadatu *let strávených v ČR* bylo navíc zaznamenáno dle nás nedostatečné členění, navrhujeme rozdělit kategorii *Méně než 1 rok* na *do 6 měsíců* a *6–12 měsíců*.

Při srovnávání skupin zadaného a reálného počtu slov jsme zjistili nevyhovující členění. U každého metadata je totiž způsob rozdělení odlišný, což ztěžuje jejich srovnání, proto navrhujeme jednotlivé podskupiny sjednotit. Dokonce 63 % výskytů bez odpovědi se objevilo u metadata *zadání textu*. Po prověření metadata *postupů práce* jsme zjistili, že *informace*, *vyprávění* i *popis* jsou značně vyrovnané. Jediný podtyp, který by byl potřeba doplnit jistým množstvím textům, byl postup *úvahy*, která v rámci korpusu tvoří pouze 5 %, zatímco v průměru celého korpusu CzeSL-SGT sahá až na 18 %. U již zmíněné skupiny aktivit předcházejících psaní můžeme doporučit doplnění kategorie cvičení, diskuze či práce se slovní zásobou.

U poslední kapitoly práce jsme se věnovali vybraným chybám v korpusu, které jsme si předem vytipovali na základě znalostí čínské gramatiky. Celkem bylo zaznamenáno šest druhů chyb, které byly popsány pomocí kvalitativní chybové analýzy. Do chyb byl zařazen případ jmenného přísudku beze spony, záměny přídavných a podstatných jmen, dále duplikace pro zdůraznění významu slova, neutvoření množného čísla u podstatných jmen, pomocné slovo s původním významem moci a jako poslední kategorie anteponovaného větného členu.

ZDROJE A POUŽITÁ LITERATURA

CzeSL-SGT: korpus češtiny nerodilých mluvčích s automaticky provedenou anotací. Ústav Českého národního korpusu FF UK, Praha 2013. Dostupný z WWW: http://www.korpus.cz

Cizinci s povoleným pobytem - Ministerstvo vnitra České republiky. *Úvodní strana - Ministerstvo vnitra České republiky* [online]. Copyright © 2018 Ministerstvo vnitra České republiky, všechna práva vyhrazena [cit. 20.12.2018]. Dostupné z: https://www.mvcr.cz/clanek/cizinci-s-povolenym-pobytem.aspx?q=Y2hudW09MQ%3d%3d

Gramatika současné čínštiny, Lingea, Brno 2018.

Sborník Asociace učitelů češtiny jako cizího jazyka (AUČCJ) 2012, Akropolis, Praha 2012.

ČERMÁK, František: *Jazyk a jazykověda*, Univerzita Karlova v Praze, Karolinum, Praha 2004.

ČERMÁK, František (ed.). *Korpusová lingvistika Praha 2011* – 1 InterCorp, Nakladatelství Lidové noviny, Praha 2011.

ČERMÁK, František (ed.). *Korpusová lingvistika Praha 2011* – 2 Výzkum a výstavba korpusů, Nakladatelství Lidové noviny, Praha 2011.

ČERNÝ, Jiří: *Úvod do studia jazyka*, Rubico, Olomouc 1998.

HRDLIČKA, Milan: *Kapitoly o češtině jako cizím jazyku*, Fakulta pedagogická, Plzeň 2010.

HRDLIČKA, Milan: *Cizí jazyk čeština*, ISV, Praha 2002.

KALOUSOVÁ, Jarmila: *Vybrané kapitoly z gramatiky moderní čínštiny*, Karlova universita v Praze, Praha 1954.

KRUPA, Viktor – GENZOR, Jozef – DROZDÍK, Ladislav: *Jazyky světa*. Bratislava 1983.

PACKARD, Jerome: *Chinese as an Isolating Language*, University of Illinois, 2006.

PETKEVIČ, Vladimír; ROSEN, Alexandr (eds): *Korpusová lingvistika Praha 2011* – 3 Gramatika a značkování korpusů, Nakladatelství Lidové noviny, Praha 2011.

POPELA, Jaroslav: *Skaličkova jazyková typologie*, Masarykova univerzita, Brno 2006.

SKALIČKA, Vladimír: *Typ češtiny*, Slovanské nakladatelství, Praha 1951.

ŠEBESTA, Karel; ŠKODOVÁ, Svatava: *Čeština - cílový jazyk a korpusy*, TUL, Liberec 2012.

ŠTINDLOVÁ, Barbora: *Žákovský korpus češtiny a evaluace jeho chybové anotace*, Karolinum, Praha 2013.

ŠULC, Michal: *Korpusová lingvistika: první vstup*, Karolinum, Praha 1999.