

**Univerzita Karlova**

Filozofická fakulta

Ústav českého jazyka a teorie komunikace

Filologie – Český jazyk

Pavel Pečený

**UŽITÍ SPOJOVACÍCH PROSTŘEDKŮ  
V TEXTECH NERODILÝCH MLUVČÍCH  
ČEŠTINY**

*The Use of Connectors in the Texts  
of Non-Native Speakers of Czech*

Teze disertační práce

Vedoucí práce: doc. RNDr. Vladimír Petkevič, CSc.

2017

# 1 Cíl práce

Předkládaná práce se na obecné rovině věnuje češtině jako cizímu jazyku, konkrétněji je zaměřena na korpusovou analýzu spojovacích prostředků v textech nerodilých mluvčích češtiny. Klade si dva hlavní cíle:

1. Zmapovat užívání spojovacích prostředků v textech nerodilých mluvčích češtiny s pomocí korpusu MERLIN a s využitím sondy do korpusů CzeSL-SGT a SKRIPT2012.
2. Zjistit, zda/jaký vliv má na rozsah repertoáru, frekvenci a správnost užití spojovacích prostředků úroveň znalosti češtiny.

## 2 Výzkumné otázky

Výše uvedené obecně formulované cíle jsme konkretizovali v podobě výzkumných otázek, jimiž jsme zároveň udali směr analýze korpusového materiálu.

1. Které spojovací prostředky užívají nerodilí mluvčí v textech zahrnutých do korpusu MERLIN?
2. Jak vypadá složení repertoáru spojovacích prostředků na jednotlivých úrovních podle *Společného evropského referenčního rámce pro jazyky* (SERRJ)?
  - 2.1. Jaký je rozsah repertoáru a jeho rozmanitost?
  - 2.2. Liší se zastoupení spojovacích prostředků podle funkce (parataktická vs. hypotaktická spojení) a větněčlenské platnosti (autosémantické vs. synsémantické prostředky)? Pokud ano, jak?
3. Jaká je frekvence užití spojovacích prostředků na jednotlivých úrovních hodnocení písemné produkce podle SERRJ?
  - 3.1. Jak se liší frekvence užití spojovacích prostředků s ohledem na úroveň?
4. Jaké jsou tendence užití jednotlivých prostředků v závislosti na úrovni podle SERRJ?
  - 4.1. Které prostředky jsou pro danou úroveň typické?
5. Jaký je rozdíl v přesnosti a správnosti užití spojovacích prostředků na různých úrovních podle SERRJ?
  - 5.1. Jaké chyby v užití spojovacích prostředků se nejčastěji objevují?
  - 5.2. Jaký je vztah chybovosti k úrovni textu podle SERRJ?

### 3 Metody

V předkládané práci aplikujeme kvantitativně-kvalitativní analýzu korpusového materiálu. Kvantitativní analýza umožňuje přehledné a spolehlivé zpracování velkého množství dat a užíváme ji zejména pro analýzu počtu a frekvence spojovacích prostředků, tj. u výzkumných otázek 2, 3 a 4. U výzkumné otázky 4, tedy spojení úrovně a konkrétních spojovacích prostředků, jsme dále zvolili korespondenční analýzu (CA, *correspondence analysis*). Jedná se o techniku určenou k analýze vztahů mezi kategoriemi jedné či více proměnných v kontingenčních tabulkách.

Kvalitativní analýza se vedle kvantitativní uplatňuje při zpracování výzkumné otázky 5, tj. při klasifikaci a popisu chyb v užití spojovacích prostředků. Opíráme se při ní mimo jiné o principy chybové analýzy mezijazyka. Pracujeme jednak se srovnáním mezijazyka autorů textů na různých úrovních mezi sebou a jednak tyto mezijazyky srovnáváme s normou spisovné češtiny. Zajímá nás nejen nekorektní užití spojovacích prostředků, ale sledujeme okrajově také náznaky nadužívání nebo nedostatečného užívání jednotlivých prostředků.

### 4 Struktura práce

Disertační práce se skládá ze dvou částí. První představuje teoretické uvedení do problematiky a shrnuje dosavadní poznatky nutné k provedení analýzy textů. Tu obsahuje druhá, výzkumná část práce.

V kapitole 2 identifikujeme nezbytná teoretická východiska. Pozornost nejprve věnujeme různým přístupům k vymezení spojovacích prostředků v současném lingvistickém popisu češtiny (kap. 2.1). Zvláště se při tom zaměřujeme na problematická místa jejich klasifikace (např. slovnědruhové přiřazení). Následně prezentujeme zvolené třídění, které v práci využíváme. Součástí druhé kapitoly je také shrnutí charakteristiky úrovně jazykové kompetence s ohledem na užití spojovacích prostředků, které lze vyvodit ze SERRJ a popisů referenčních úrovní pro češtinu (kap. 2.2). V neposlední řadě se v této kapitole zabýváme teorií mezijazyka a základy chybové analýzy (kap. 2.3).

Kapitola 3 slouží především k představení korpusů, jejichž dat využíváme (kap. 3.1). Nejpodrobněji charakterizujeme složení a stavbu korpusu MERLIN, jenž představuje náš primární zdroj textů. Prostor je dále věnován korpusům CzeSL-SGT, SKRIPT2012 a korpusům řady SYN. Při jejich prezentaci se snažíme především o zdůraznění důvodů, kvůli nimž jsme je zvolili jako podpůrné zdroje. Nedílnou součástí této kapitoly tvoří rozbor

možností vyhledávání spojovacích prostředků v uvedených korpusech (kap. 3.2) a také popis vlastního postupu získání a zpracování dat (kap. 3.3).

Jádro disertační práce představuje kapitola 4. Věnuje se analýze a popisu užívání spojovacích prostředků ve shromážděných dokladech. Vycházíme v ní od globálního pohledu na rozsah a složení repertoáru v celém korpusu. Následně obracíme pozornost na jednotlivé úrovně podle SERRJ a snažíme se zachytit tendence ve frekvenci užívání konkrétních prostředků a/nebo jejich skupin. Vztah mezi úrovněmi a použitými spojovacími prostředky dále zkoumáme pomocí nástrojů korespondenční analýzy. Kapitola obsahuje také chybovou analýzu dokladů, jež slouží jako podklad pro stanovení závislosti četnosti chyb na dosažené úrovni jazykové kompetence. Výzkum je završen třemi sondami do užívání spojovacích prostředků v korpusech CzeSL-SGT a SKRIPT2012.

## 5 Motivace a přínos práce

Vznik této disertační práce podnítily vedle osobního zájmu a aktuálnosti daného tématu zejména následující skutečnosti.

(A) Prvním podnětem, je nedostatek empiricky podložených popisů různých fází osvojování češtiny jako cizího jazyka. V minulosti sice byly vytvořeny tzv. popisy referenčních úrovní pro češtinu, jež se vztahují k úrovním A1–B2 podle SERRJ, při jejich tvorbě ale nebyly k dispozici žádné rozsáhlejší (elektronické) korpusy autentických projevů nerodilých mluvčích češtiny. Navíc z těchto popisů u konkrétních dovedností často není patrné, jak přesně se od sebe jednotlivé úrovně odlišují, a tak ani nelze odvodit, jak vývoj jazykové kompetence probíhá. Předkládaná disertační práce se snaží, byť na úzce vytknuté oblasti spojovacích prostředků, ukázat, jak by mohl popis založený na autentických datech vypadat.

(B) V posledních letech se někteří autoři pokusili o zachycení a analýzu vybraných gramatických jevů (především z oblasti morfologické) v jazyce nerodilých mluvčích češtiny, přičemž pracovali s korpusovými daty (např. Hudousková 2013). Jedinečnost naší práce ovšem spočívá v zaměření na oblast syntaxe. Doposud totiž existuje pouze jedna práce, resp. projekt, který se syntaktickému zkoumání písemné produkce nerodilých mluvčích ve větším rozsahu věnuje (srov. Rysová a kol. 2016).<sup>1</sup> Je to o to překvapivější, uvědomíme-li

---

<sup>1</sup> Jedná se o právě probíhající projekt Ústavu formální a aplikované lingvistiky Matematicko-fyzikální fakulty Univerzity Karlovy. V rámci tohoto projektu vzniká počítačový nástroj *EVALD 1.0 pro cizince*, jež slouží

si, že výstavba textu (propojování větných členů a vět, komplexnost struktur atd.) je oblastí, ve které se rozvoj jazykové kompetence velmi silně odráží. A právě teoretický popis toho, jak nerodilí mluvčí spojovací prostředky užívají, by měl být jedním ze základů didaktického uchopení a zpracování jejich prezentace ve výuce. Rádi bychom proto disertační prací přispěli k vytvoření takového teoretického podkladu, o který by se didaktici a autoři učebních materiálů češtiny jako cizího jazyka mohli opírat. Ve výsledku by se tak mohla nepřímo posílit i pozice, jakou zaujímá skladba v učebnicích češtiny jako cizího jazyka.<sup>2</sup>

(C) Již několikrát jsme odkazovali k úrovním jazykové kompetence a je třeba uvést, že předkládaná disertační práce se v neposlední řadě dotýká i samotné podstaty jejich vymezení. Na počátku tohoto století došlo zveřejněním SERRJ k zásadnímu přehodnocení ve vnímání úrovní znalosti cizího jazyka, resp. úrovní komunikativní jazykové kompetence. Do jazykové výuky, jejího plánování a zjišťování výsledků (testování) se začaly promítat referenční úrovně SERRJ (A1–C2) založené na obecném vymezení pomocí deskriptorů společných pro všechny evropské jazyky. Tím bylo mimo jiné umožněno porovnání dosažené úrovně v různých cizích jazycích. SERRJ se stal důležitým nástrojem standardizace a popisu postupu osvojování si cizího jazyka. Přesto existuje jeden aspekt, který stále není dostatečně uchopen, a tím je empirická validita stupnic SERRJ, jinými slovy to, jak odrážejí skutečnou úroveň žákovského jazyka. V této oblasti bylo zatím provedeno jen velmi málo výzkumů (srov. např. Alderson 2007, Carlsen 2010, Wisniewski 2014). Naše práce si sice primárně neklade za cíl validovat konkrétní deskriptory/stupnice SERRJ, ale může se stát podnětem k otevření diskuse na toto téma také v kontextu češtiny jako cizího jazyka. Dotýká se totiž distribuce spojovacích prostředků v závislosti na úrovni písemných produkcí podle SERRJ. Z výsledků tedy bude například možno odvodit, které prostředky jsou typické pro kterou z úrovní, a konfrontovat toto zjištění s očekáváním vycházejícím z vybraných stupnic SERRJ, jež se týkají písemné produkce.

Můžeme tedy shrnout, že disertační práce vychází z aktuální potřeby vzniku empiricky podloženého popisu vývoje jazykové kompetence nerodilých mluvčích češtiny a inspiruje se současnou diskusí o nutnosti validace úrovní vymezených v SERRJ. Její přínos spatřujeme v obohacení poznání o užívání spojovacích prostředků v textech nerodilých mluvčích češtiny,

---

k automatickému hodnocení povrchové koherence (koheze) textů v češtině psaných nerodilými mluvčími češtiny.

<sup>2</sup> Srov. Hrdlička (2009, s. 12): „Je známou skutečností, že je syntaxi [ve výuce] v porovnání s morfologií vyhrazen omezenější prostor (toto konstatování se netýká pouze způsobu prezentace spojek, ale také výkladů slovosledu v češtině aj.).“

kteří může v budoucnu sloužit k dalšímu didaktickému zpracování a aplikaci ve výuce nebo při tvorbě učebních materiálů.

## 6 Shrnutí výsledků

Výzkumná část práce se věnovala analýze autentických textů s cílem zmapovat, jak nerodilí mluvčí s různou úrovní jazykové kompetence užívají spojovací prostředky v češtině. První část byla určena obecné charakteristice složení repertoáru podle formálně syntaktických vztahů parataxe a hypotaxe a větněčlenské platnosti spojovacích prostředků. Byla zjištěna celková početní převaha hypotaktických spojovacích prostředků (47 z 81) a také se ukázalo, že se repertoár se stoupající úrovní rovnoměrně rozšiřuje jak o hypotaktické, tak o parataktické spojovací prostředky. Frekvence jejich užití ovšem nestoupá rovnoměrně, a tak zatímco na úrovni A2 jsou čteněji zastoupena parataktická spojení, na úrovni B2 je již poměr téměř vyrovnaný.

Co se větněčlenské platnosti týče, můžeme konstatovat, že se v repertoáru celkově nachází více synsémantických než autosémantických spojovacích prostředků. Synsémantické prostředky obohacují repertoár každé úrovně, naproti tomu počet autosémantických výrazů v repertoáru se od úrovně B1 téměř nemění. Se stoupající úrovní se ale zvyšuje jejich frekvence (především u zájmen *který, jaký* a *co*). S úrovní B1 do repertoáru významněji pronikají také víceslovné spojovací prostředky (srov. tabulka 1).

Tabulka 1: Globální charakteristika úrovní podle spojovacích prostředků

	A2	B1	B2
<b>frekvence užití</b>	kontinuálně roste společně se zvyšující se úrovní		
<b>rozsah repertoáru</b>	podstatně se rozšiřuje		rozšiřuje se
<b>typy prostředků podle podoby</b>	zvyšuje se poměr zastoupení víceslovných prostředků		
<b>formálně syntaktické vztahy</b>	repertoár se s rostoucí úrovní rozšiřuje stejnou měrou o parataktické i hypotaktické spojovací prostředky		
	četností převažují parataktická spojení	poměr počtu parataktických a hypotaktických spojení je vyrovnaný	četností převažují hypotaktická spojení
<b>funkčně syntaktické hledisko</b>	přibývají synsémantické a v menší míře také autosémantické prostředky		přibývají hlavně synsémantické prostředky
	roste frekvence užití autosémantických spojovacích prostředků (vztažných zájmen)		

Na vyšších úrovních se však repertoár spojovacích prostředků nejen rozrůstá o spojovací prostředky vyjadřující nové mezivětné vztahy, ale také se rozšiřují řady konkurenčních prostředků pro vyjádření již zastoupených mezivětných vztahů. Tento jev probíhá nejrozsáhleji na úrovni B1, a to při vyjadřování odporovacích, přípustkových a časových vztahů, na úrovni B2 pak již jen v menší míře, hlavně v hypotaktických vztazích přípustkových a důsledkových (srov. tabulka 2 a 3).

Tabulka 2: MERLIN: Tendence v rozšiřování skupin konkurenčních prostředků

	B1	B2
<b>nově v repertoáru</b>	srovnání (víceslovné) nepřímý rozkaz vztažné věty s <i>což</i>	vztažné věty s <i>jenž</i> přípustkové věty s kondicionálem
<b>více konkurenčních prostředků</b>	odporovací spojení vylučovací spojení přípustková spojení důsledková spojení srovnávací spojení časové věty	přípustková spojení důsledková spojení

Tabulka 3: MERLIN: Skupiny konkurenčních prostředků

významový vztah	A2	B1	B2
<b>odporovací</b>	<i>ale</i>	<i>však</i> <i>avšak</i> <i>přesto</i>	<i>nýbrž</i>
<b>vylučovací</b>	<i>nebo</i> <i>anebo</i> <i>bud'to – (a)nebo</i>	<i>či</i> <i>bud' – (a)nebo</i>	
<b>stupňovací</b>	<i>nejen – ale (i)</i>	<i>nejenom – ale (i)</i>	
<b>přípustkový</b>	<i>přestože</i>	<i>i když</i> <i>ačkoliv</i> <i>aniž</i>	<i>i kdyby</i> <i>třebaže</i> <i>ačkoli</i> <i>byť</i>
<b>důsledkový</b>	<i>proto</i> <i>tak</i> <i>a proto</i> <i>a tak</i>	<i>tedy</i> <i>a tedy</i>	<i>tudiž</i> <i>pak</i>
<b>důvodový/příčinný</b>	<i>totiž</i> <i>protože</i>	<i>jelikož</i>	
<b>podmínkový</b>	<i>jestli</i> <i>pokud</i>	<i>jestliže</i>	
<b>srovnávací</b>	<i>jako</i>	<i>čím – tím</i> <i>jak – tak</i>	
<b>časový</b>	<i>když</i> <i>až</i>	<i>jakmile</i> <i>kdykoliv</i>	<i>zatímco</i>

V druhé části výzkumu jsme se pokusili zachytit tendence užívání spojovacích prostředků detailněji a zjistit jejich závislost na úrovni podle SERRJ. Nejprve jsme identifikovali relativně uzavřenou skupinu výrazů, které jsou nejpočetněji zastoupeny na všech úrovních (*a, ale, když, který, nebo, protože, že*) – srov. tabulka 4.

Tabulka 4: 10 nejčtenějších spojovacích prostředků

	MERLIN						SYN 2015
	A1	A2	B1	B2	C1	celkově	
1	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
2	<i>ale</i>	<i>že</i>	<i>že</i>	<i>že</i>	<i>že</i>	<i>že</i>	<i>že</i>
3		<i>ale</i>	<i>ale</i>	<i>který</i>	<i>který</i>	<i>ale</i>	<i>který</i>
4		<i>protože</i>	<i>který</i>	<i>ale</i>	<i>ale</i>	<i>který</i>	<i>i</i>
5		<i>nebo</i>	<i>když</i>	<i>když</i>	<i>ani; i</i>	<i>když</i>	<i>ale</i>
6		<i>který</i>	<i>nebo</i>	<i>nebo</i>	<i>aby; totiž</i>	<i>protože</i>	<i>jako</i>
7		<i>kdy</i>	<i>protože</i>	<i>i</i>	<i>co; když</i>	<i>nebo</i>	<i>když</i>
8		<i>kde</i>	<i>co</i>	<i>co</i>	<i>jak; kdy; proto; tak</i>	<i>co</i>	<i>aby</i>
9		<i>když</i>	<i>aby</i>	<i>aby</i>	<i>kdo; protože</i>	<i>aby</i>	<i>nebo</i>
10		<i>proto</i>	<i>jak</i>	<i>jako</i>	<i>a ne; až; jak – tak; jaký; kdyby</i>	<i>i</i>	<i>co</i>

Následně jsme porovnáním relativního pořadí podle frekvence všech spojovacích prostředků z korpusu dospěli ke zjištění, že u většiny z nich se projevuje stoupající tendence, tzn. že jejich frekvence užití roste spolu se zvyšující se jazykovou kompetencí autorů textů. U více než poloviny prostředků zastoupených na všech úrovních se dokonce jedná o kontinuální nárůst frekvence od úrovně A2 až po úroveň B2 (srov. tabulka 5 a 6). Jako nejprogresivnější se v tomto trendu jeví paratactické spojovací prostředky *i* a *dokonce* a hypotactické *aby, totiž, -li*. U zbytku se projevuje hlavní nárůst mezi úrovní A2 a B1 a posun mezi B1 a B2 již není tak výrazný. Pouze u malé skupiny (zájmených příslovcí *kdo, kdy, kam, číslovky kolik* a spojky *protože*) byl zjištěn výraznější pokles frekvence na vyšších úrovních, což by mohlo svědčit o tom, že je uživatelé s nižší jazykovou kompetencí nadužívají.

Tabulka 5: MERLIN: Tendence v pořadí podle relativní frekvence A2–B2

rozdíl relativní frekvence A2–B2		
stoupající	stoupavě-vyrovnaný	klesající
<i>jak</i>	<i>tak</i>	<i>kde</i>
<i>jako</i>	<i>jestli</i>	<i>kdy</i>
<i>proto</i>	<i>ani</i>	<i>kolik</i>
<i>než</i>		<i>kam</i>
<i>kdo</i>	<b>stoupavě-klesající</b>	
<i>jaký</i>	<i>a tak</i>	
<i>a proto</i>	<i>zda</i>	



<i>až</i>	<i>proč</i>
<i>nejen – ale (i)</i>	<i>takže</i>
<i>dokonce</i>	<i>kdyby</i>
<i>pokud</i>	
<i>-li</i>	
<i>totiž</i>	<b>klesavě-stoupající</b>
<i>přestože</i>	<i>anebo</i>

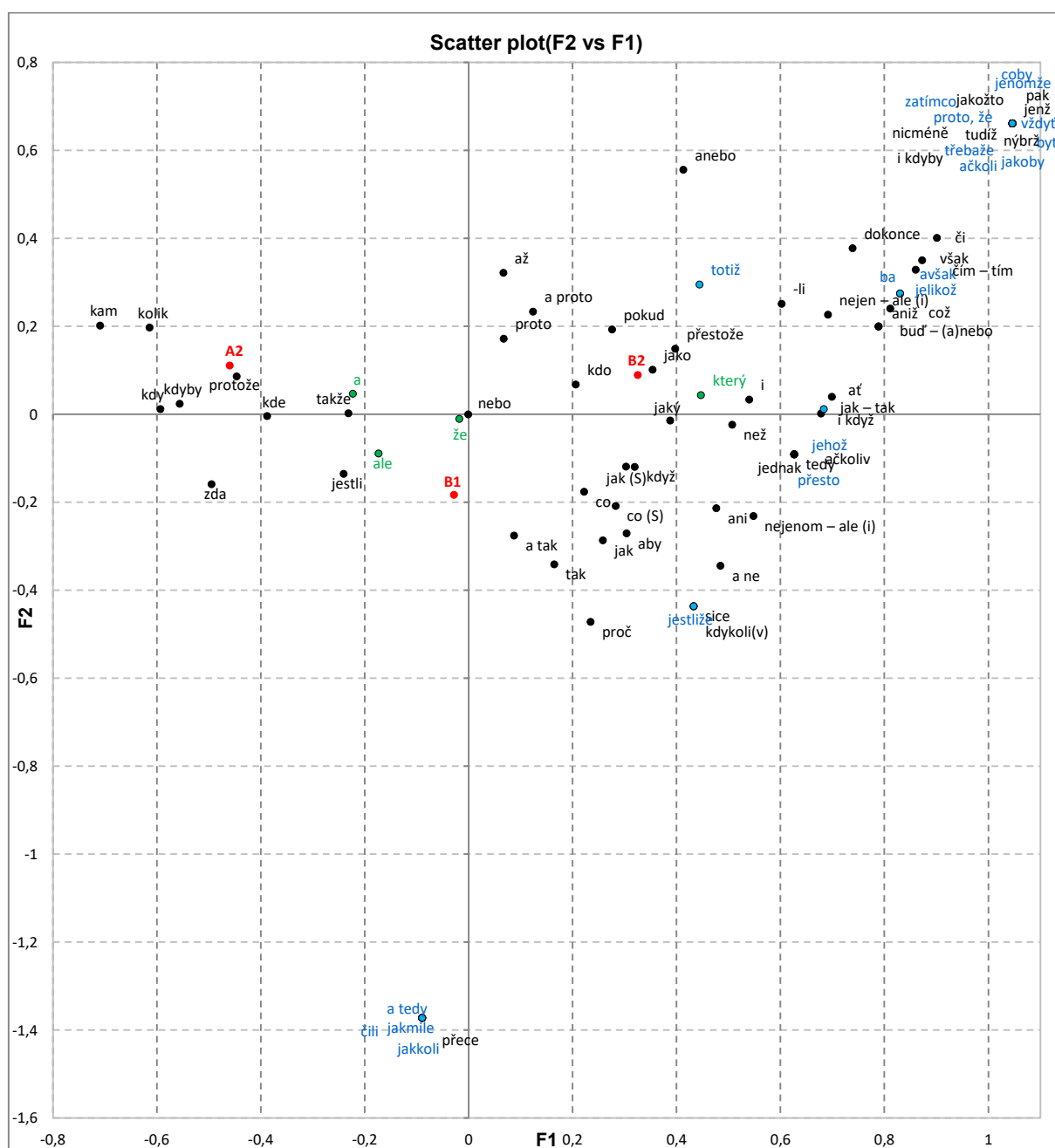
Tabulka 6: MERLIN: Tendence v pořadí podle relativní frekvence B1–B2

rozdíl relativní frekvence B1–B2		
stoupající	obdobná	klesající
<i>či</i>	<i>přesto</i>	<i>a ne</i>
<i>i když</i>	<i>jednak</i>	<i>nejenom – ale (i)</i>
<i>však</i>	<i>jelikož</i>	<i>sice</i>
<i>což</i>		<i>jestliže</i>
<i>jak – tak</i>		<i>kdykoliv</i>
<i>avšak</i>		
<i>čím – tím</i>		
<i>ať</i>		
<i>bud' – (a)nebo</i>		
<i>jehož</i>		
<i>ačkoliv</i>		
<i>tedy</i>		
<i>aniž</i>		
<i>ba</i>		

Ve třetí části jsme vztah užívání konkrétních spojovacích prostředků a úrovně podle SERRJ podrobili korespondenční analýze. Ta prokázala vzájemnou závislost obou proměnných a potvrdila, že uživatelé jazyka na nižších úrovních spoléhají pouze na omezené množství spojovacích prostředků, zatímco uživatelé vyšších úrovní mají v tomto ohledu variabilnější a širší repertoár vyjadřovacích prostředků. Po tom, co jsme v analýze zohlednili také frekvenci spojovacích prostředků z korpusu SKRIPT2012, výsledky podpořily hypotézu, že na vyšších úrovních lze (na rozdíl od nižších úrovní) očekávat užívání spojovacích prostředků, jež jsou v cílovém jazyce méně frekventované.

Výsledky zobrazuje graf 1. Na vodorovné ose vidíme znázorněno užívání spojovacích prostředků na jednotlivých úrovních. Směrem zleva doprava se úroveň jazykové kompetence zvyšuje. Hlavní svislá osa tvoří předěl mezi úrovněmi A2/B1 a signalizuje pomyslný vstup na „prahovou“ úroveň B1. Nalevo od této osy najdeme spojovací prostředky typické pro nižší úrovně a napravo od ní jsou zachyceny výrazy užívané ve spojeních na úrovních vyšších. Ze vzdálenosti spojovacího prostředku od červeně značených bodů úrovní A2–B2 můžeme odečíst četnost jeho užívání na dané úrovni. Díky tomuto prostorovému zobrazení lze sledovat, v jakých skupinách se spojovací prostředky sdružují a pro kterou úroveň jsou

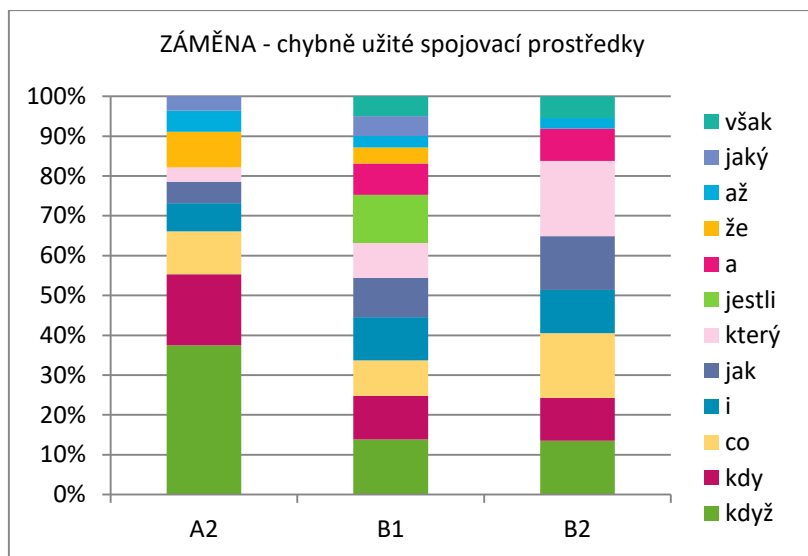
typické. Čím blíže se spojovací prostředek k označení úrovni nachází, tím je na této úrovni frekventovanější. Spojovací prostředky na průsečíku mezi úrovněmi (*že, nebo, ale*) jsou běžné na všech úrovních a neodlišují nižší od vyšších úrovní. Podle barevného kódování v korespondenční mapě tak můžeme sledovat, jakou frekvenci mají dané výrazy v akvizitivním korpusu rodilých mluvčích češtiny. Pro přehlednost jsme vydělili pouze tři skupiny: zelené popisky mají nejfrekventovanější spojovací prostředky (i.p.m > 10 000), černé jsou středně frekventované (10 000 > i.p.m > 1 000) a modré značí málo frekventované výrazy (1 000 > i.p.m > 100).



Graf 1: MERLIN: korespondenční analýza relativní frekvence a úrovně SERRJ + barevně znázorněná frekvence výrazů v L1

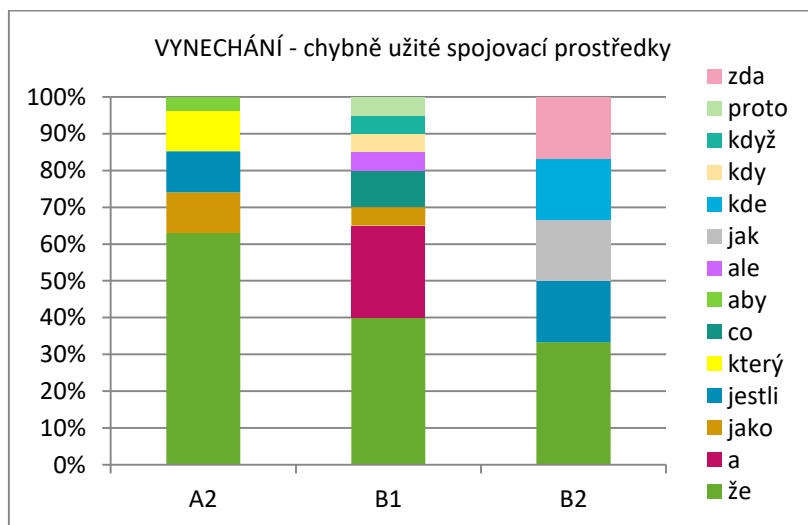


Poslední část výzkumu spojenou s korpusem MERLIN jsme zaměřili na chybovou analýzu získaných dokladů. Z ní vyplynulo, že v produkcích nerodilých mluvčích nejčastěji dochází k záměnám jednoho spojovacího výrazu za jiný (3,31 %). Následně se také potvrdila pozitivní korelace stoupající úrovně podle SERRJ a bezchybného ovládnutí spojovacích prostředků. K záměnám dochází nejčastěji mezi synsémantickými spojovacími prostředky v hypotaktických spojeních. Záměny parataktických spojovacích prostředků jsou méně časté a vyznačují se jen omezenou variabilitou (srov. graf 3).



Graf 3: Nejčastěji zaměňované spojovací prostředky podle úrovně (i.p.m.)

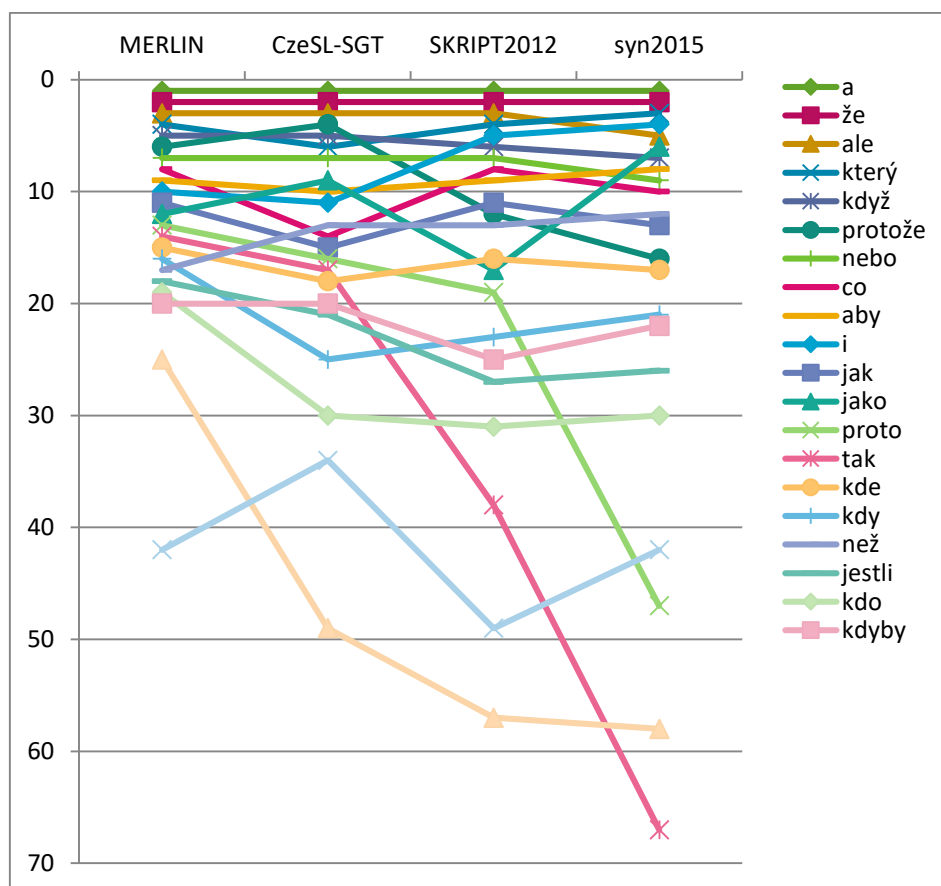
Další typy chyb – vynechání a nadbytečné užití – byly zaznamenány pouze okrajově u omezeného množství výrazů. Vynechání se nejsilněji projevuje u spojek *že* a *jestli/zda* v nepřímých otázkách a u vztažných zájmen *co* a *který* v komplexních souvětých strukturách s korelativním spojením.



Graf 4: Chybějící spojovací prostředky (podíl podle i.p.m.)

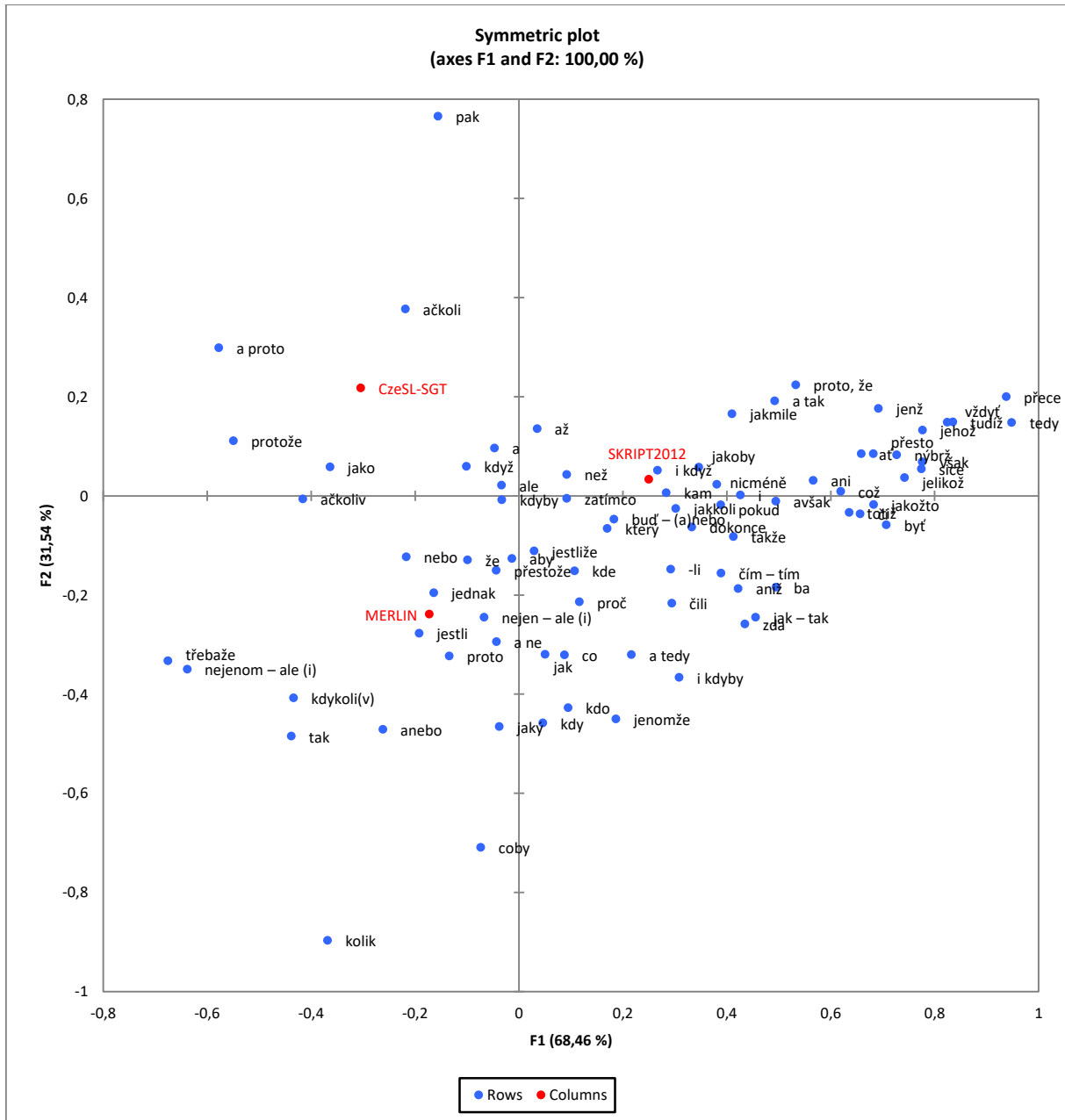
Nadbytečné užití bylo doloženo pouze u *a*, *ale*, *že* a *aby*, a to zejména na úrovních A2 a B1. Také morfologické a ortografické chyby, které jsme vyčlenili zvlášť, se spojovacích prostředků v dokladech dotýkají pouze minimálně (0,22 % a 1,39 %). Formální nedostatky tvaroslovné postihují pouze autosémantické spojovací prostředky, hlavně zájmena *který*, *co*, *kdo*, *jaký* a *jenž*. Pravopisné chyby se projevily v diakritice a také v nerespektování hranice slova u víceslovných prostředků.

Na analýzu materiálu z korpusu MERLIN navazovaly 3 sondy do korpusu CzeSL-SGT a SKRIPT2012, které měly za úkol podpořit závěry, které byly doposud formulovány. Zajímal nás nejprve celkový poměr četnosti spojení podle syntaktické funkce (paratactická vs. hypotactická a autosémantická vs. synsémantická spojení). Výsledek sondy nás přesvědčil o tom, že se u průnikové množiny spojovacích prostředků nevyskytují žádné zásadní signály, které by nasvědčovaly tomu, že korpus MERLIN obsahuje abnormální složení dat. Druhou sondou jsme podrobili zkoumání srovnání pořadí spojovacích prostředků podle frekvence v jednotlivých korpusech. Potvrdilo se očekávání, že výsledky z obou korpusů nerodilých mluvčích budou bližší frekvenci prostředků v korpusu SKRIPT2012 než výsledkům z korpusu SYN2015 (graf 5).



Graf 5: Srovnání korpusů: pořadí podle frekvence

Poslední sonda cílila na zjištění závislosti užití spojovacích prostředků na typu korpusu (rodilí mluvčí vs. nerodilí mluvčí). Na základě interpretace jejích výsledků se posílila domněnka, že v textech nerodilých mluvčích dochází k nadužívání (např. *protože*, *tak*, *jestli*, *kdo*), anebo naopak opomíjení (např. *i*) některých spojovacích prostředků (srov. graf 6).



## 7 Závěry

Předkládaná disertační práce si v úvodu vytkla dva hlavní výzkumné cíle. Prvním z nich bylo zmapovat a popsat užívání spojovacích prostředků v textech nerodilých mluvčích češtiny s pomocí korpusu MERLIN a s využitím sondy do korpusu CzeSL-SGT a SKRIPT2012. Druhým cílem bylo na základě provedeného popisu zjistit, zda a jaký vliv má na rozsah repertoáru, frekvenci a správnost užití spojovacích prostředků úroveň jazykové kompetence. Oba stanovené cíle se nám povedlo naplnit, i když v některých dílčích oblastech stále zůstává prostor pro další detailnější analýzy.

Na základě výsledků výše popsaných dílčích výsledků můžeme na tomto místě formulovat nejzásadnější závěry našeho výzkumu:

(1) Prokázalo se, že se stoupající úrovní podle SERRJ se rozšiřuje repertoár spojovacích prostředků, které nerodilí mluvčí užívají. Přibývají v něm především prostředky pro vyjadřování hypotaktických mezivětných vztahů a v menší míře víceslovné spojovací prostředky parataktické. Platí také, že zatímco nové synsémantické spojovací prostředky obohacují repertoár každé úrovně, počet autosémantických výrazů v repertoáru se od úrovně B1 téměř nemění, roste pouze frekvence užívání některých z nich (hlavně výrazů  *který, jenž a jehož*).

(2) Byla zjištěna nepřesnost v rozložení stupnice SERRJ pro ovládání koheze a koherence. V ní je hranice mezi užíváním omezeného a širšího repertoáru prostředků postavena až na úroveň B2+. Z naší analýzy ale vyplývá, že již uživatelé na úrovni B1+/B2 disponují relativně širokým repertoárem, který obsahuje všechny základní spojovací prostředky, a u některých z nich dokonce i několik konkurenčních výrazů.

(3) U většiny spojovacích prostředků byla zaznamenána stoupající (nebo stoupavě-vyrovnaná) tendence ve frekvenci užívání se zvyšující se úrovní. Klesající tendence se objevila pouze u skupiny zájmených příslovcí a číslovek a u spojky *protože*. Tyto výrazy se na základě konfrontace s frekvencí v češtině rodilých mluvčích jeví především na nižších úrovních jako nadužívané.

(4) Korespondenční analýza potvrdila, že nerodilí mluvčí užívají na nižších úrovních pouze spojovací prostředky, které jsou v písemné produkci rodilých mluvčích více nebo středně frekventované. Teprve na vyšších úrovních se repertoár rozšiřuje o prostředky rodilými mluvčími méně užívané.

(5) Prokázala se také závislost počtu chyb v užití spojovacích prostředků na rovní SERRJ, a to i přesto, že celkový počet chyb tvoří pouze necelá 4 % ze všech výskytů spojovacích prostředků. Nejčastěji se objevují záměny spojovacích prostředků, méně časté je vynechání a k nadbytečnému užití téměř nedochází. Produktivní jsou záměny podle vzorců: *jak > když <> jestli < že <> co < který <> jaký <> jako <> jak > kdyby < když <> kdy; proto <> protože > ale < a <> i > ale < však <> avšak*; kde X > Y značí, že výraz X se chybně užívá místo výrazu Y. Všem typům chyb je společné, že se zvyšující úrovní ustupují.

Předkládaná disertační práce představuje první takto pojatou studii syntaxe písemné produkce nerodilých mluvčích, a tak s sebou přináší samozřejmě i řadu dalších podnětů jak ve výzkumu pokračovat a popis prohlubovat. Zmínit můžeme kupříkladu možné analýzy vzorců užívání konkrétních skupin spojovacích prostředků (např. časových, příčinných a přípustkových) nebo bližší pohled na nadužívání nebo nedostatečné užívání některých mezivětných vztahů.

## 8 Literatura (výběr)

Abel, A. a kol. (2014): A Trilingual Learner Corpus Illustrating European Reference Levels. *Ricognizioni – Rivista di Lingue, Letterature e Culture Moderne*, 2 (1), s. 111–126. [cit. 15. 3. 2017] Dostupné z: <http://www.ojs.unito.it/index.php/ricognizioni/article/view/702>

Alderson, J. C. (2007): The CEFR and the need for more research. In: *The Modern Language Journal* 91, s. 658–662.

Bauer, J. (1958): Parataxe a hypotaxe při studiu souvětí. In: *O vědeckém poznání soudobých jazyků*. Praha: ČSAV, s. 270–278.

Bauer, J. (1967): K otázce významu a funkcí spojek. In: *Sborník prací Filozofické fakulty brněnské univerzity*, A 15, s. [27]–36.

Bauer, J. – Grepl, M. (1972): *Skladba spisovné češtiny*. Praha: SPN.

Carlsen, C. (2010): Discourse connectives across CEFR levels: A corpus-based study. In: Bartning, I. – Martin, M. – Vedder, I. (eds.): *Communicative Proficiency and Linguistic Development: intersections between SLA and language testing research (Eurosla)*, s. 191–210. [cit. 15. 3. 2017] Dostupné z: <http://eurosla.org/monographs/EM01/EM01tot.pdf>

Corder, S. P. (1981): *Error Analysis and Interlanguage*. Oxford: OUP.

Corder, S. P. (1967): The Significance of Learner's Errors. *IRAL*, vol. 5, no. 4, s. 161–170.

Čermáková, K. (2012): *Souvětí kombinující spojovací prostředky. Popis a analýza na základě korpusových dat*. Disertační práce. Olomouc: FF UP Katedra bohemistiky.



- Daneš, F. (1977): Zpracování souvětí v nové mluvnici češtiny. *Slavica Slovaca*, 12, s. 244–252.
- Doleželová, B. (2008): *Popis spojovacích prostředků v mluvnických pracích od 2. poloviny 20. století*. Diplomová práce. Praha: FF UK Ústav českého jazyka a teorie komunikace.
- Ellis, R. (1994): *The Study of Second Language Acquisition*. Oxford: OUP.
- Ellis, R. – Barkhuizen, G. (2009): *Analysing learner language*. Oxford: OUP.
- Grepl, M. (1977–8): Souvětí v sémanticky orientovaném popisu syntaxe. In: *Sborník prací Filozofické fakulty brněnské univerzity*, A 25/26, s. 7–29.
- Hošnová, E. (2005): *Studie z vývoje novočeské syntaxe (konjunkce, pronominalizace)*. Praha: Karolinum.
- Hrbaček, J. (1964a): O syntaktické funkci korelativního spojení řídící a závislé věty. *Slovo a slovesnost*, 25, s. 81–95.
- Hrdlička, M. (2009): *Gramatika a výuka češtiny jako cizího jazyka*. Praha: Karolinum.
- Hudousková A. (2013): The Corpus CzeSL in the Service of Teaching Czech for Foreigners – Errors in the Use of the Pronoun který. In: Gajdošová, K. – Žáková, A.: *Slovko 2013 – Natural Language Processing, Corpus Linguistics, E-learning*. Lüdenscheid: RAM-Verlag, s. 100–107.
- James, C. (1998): *Errors in language learning and use: exploring error analysis*. London: Longman.
- MERLIN project (2014): *User Manual*. [cit. 15. 3. 2017] Dostupné z: [http://www.merlin-platform.eu/docs/MERLIN\\_user-manual-CZ.pdf](http://www.merlin-platform.eu/docs/MERLIN_user-manual-CZ.pdf)
- Petkevič, V. a kol. (2012): Anotace chybových textů v českém žákovském korpusu. In: Šebesta, K. – Škodová, S. (eds.): *Čeština – cílový jazyk a korpusy*. Liberec: TUL, s. 61–87.
- PopisA1 = Hádková, M. – Línek, J. – Vlasáková, K. (2005): *Čeština jako cizí jazyk – Úroveň A1*. Praha: MŠMT ČR.
- PopisA2 = Čadská, M. – Bidlas, V. – Confortiová, H. – Turzíková, M. (2005): *Čeština jako cizí jazyk – úroveň A2*. Praha: MŠMT ČR.
- PopisB1 = Šára, M. a kol. (2001): *Prahová úroveň – čeština jako cizí jazyk*. Strasbourg: Council of Europe.
- PopisB2 = Holub, J. a kol. (2005): *Čeština jako cizí jazyk: Úroveň B2*. Praha: MŠMT ČR.
- Ramos, M. – Carvalho, H. (2010): *Perceptions of quantitative methods in higher education: mapping student profiles*. *Higher Education*, [cit. 15. 3. 2017] Dostupné z: <http://dx.doi.org/10.1007/s10734-010-9353-3>
- Rosen, A. (2014): *Korpus CzeSL – automaticky identifikované chyby*. [cit. 15. 3. 2017] Dostupné z: <http://utkl.ff.cuni.cz/~rosen/public/SeznamAutoChyBR0R1.html>

Rosen, A. (2015): *CzeSL-SGT – korpus češtiny nerodilých mluvčích s automaticky provedenou anotací*. [cit. 15. 3. 2017] Dostupné z: <http://utkl.ff.cuni.cz/~rosen/public/2014-czesl-sgt-cs.pdf>

Rysová, K. – Rysová, M. – Mírovský, J. (2016): Automatic evaluation of surface coherence in L2 texts in Czech. In: *Proceedings of the 28th Conference on Computational Linguistics and Speech Processing ROCLING XXVIII. The Association for Computational Linguistics and Chinese Language Processing (ACLCLP)*. Taipei, Taiwan, s. 214–228. [cit. 15. 3. 2017] Dostupné z: <http://aclweb.org/anthology/O/O16/O16-1021.pdf>

SERRJ = *Společný evropský referenční rámec pro jazyky: jak se učíme jazykům, jak je vyučujeme a jak v jazycích hodnotíme* (2001). Council of Europe. [cit. 15. 3. 2017] Dostupné z: [http://www.msmt.cz/uploads/soubory/mezinarodni\\_vztahy/evropskyreferencniramec.zip](http://www.msmt.cz/uploads/soubory/mezinarodni_vztahy/evropskyreferencniramec.zip)

Selinker, L. (1972): Interlanguage. *IRAL*. 1972, vol. 10, no. 3, s. 209–231.

Schachter, J. (1974): An Error in Error Analysis. *Language Learning*. 24, no. 2, s. 205–214.

Shea, M. (2009): *A corpus-based study of adverbial connectors in learner text*. Michigan State University Working Papers in SLS 1 (1), s. 1–13.

Svoboda, K. (1968): K třídění vedlejších vět podle spojovacích výrazů. *Slovo a slovesnost*, 29, s. 142–151.

Svoboda, K. (1972): *Souvětí spisovné češtiny*. Praha: Univerzita Karlova.

Šebesta, K. (2010): Korpusy češtiny a osvojování jazyka. *Studie z aplikované lingvistiky/Studies in Applied Linguistics*, roč. 1, č. 2, s. 11–34. [cit. 15. 3. 2017] Dostupné z: [http://akces.ff.cuni.cz/system/files/SALi\\_2\\_2010.pdf](http://akces.ff.cuni.cz/system/files/SALi_2_2010.pdf)

Štěpán, J. (2013): *Souvětí současné češtiny*. Liberec, TUL.

Šebesta, K. – Škodová, S. a kol. (2012): *Čeština – cílový jazyk a korpusy*. Liberec: TUL.

Škodová, S. – Štindlová, B. – Hana, J. – Rosen, A. (2011): Víceúrovňová anotace českého žakovského korpusu. In: Petkevič, V. – Rosen, A. (eds.). *Korpusová lingvistika Praha 2011: 3 – Gramatika a značkování korpusů, Studie z korpusové lingvistiky 16*, s. 208–225. Praha: Ústav Českého národního korpusu, NLN.

Štindlová, B. (2013): *Žakovský korpus češtiny a evaluace jeho chybové anotace*. Varia, FF UK: Praha.

Štindlová, B. – Čurdova, V. (2015): MERLIN: Multilingvální platforma pro evropské referenční úrovně. *Časopis pro moderní filologii*, 97, 2, s. 190–202.

Wisniewski, K. (2014): *Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen. Eine empirische Untersuchung der Flüssigkeits- und Wortschatzskalen des GeRS am Beispiel des Italienischen und des Deutschen*. Language Testing and Evaluation Series, 33. Frankfurt: Peter Lang.