

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*

Disertační práce

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

2017

Prohlášení

Prohlašuji, že jsem disertační práci napsal samostatně s využitím pouze uvedených a řádně citovaných pramenů a literatury a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 29. 3. 2017

.....
Pavel Pečený

Poděkování

Děkuji vedoucímu práce i všem konzultantům za cenné připomínky a laskavou podporu.

Abstrakt

V posledních několika letech vznikly dva korpusy autentických textů nerodilých mluvčích češtiny (korpus MERLIN a korpus CzeSL), díky nimž získali jazykovědci významný zdroj dat pro zkoumání češtiny jako cizího jazyka. Poprvé je tak možné provádět analýzu jazyka nerodilých mluvčích češtiny s pomocí nástrojů korpusové lingvistiky a formulovat empiricky podložené závěry výzkumu.

Toho využívá i předkládaná disertační práce, která se zaměřuje na zkoumání a popis užívání spojovacích prostředků v písemných produkcích nerodilých mluvčích češtiny. Opírá se především o doklady z žakovského korpusu MERLIN, který se oproti ostatním korpusům vyznačuje spolehlivým přiřazení písemných produkcí k úrovním podle *Společného evropského referenčního rámce pro jazyky (SERRJ)* a zahrnuje úroveň A2–B2. Zároveň obsahuje i chybovou anotaci textů, takže umožňuje zjistit, nejen jaký vliv má úroveň jazykové kompetence na rozsah repertoáru a frekvenci spojovacích prostředků, ale také, jak se úroveň odráží ve správnosti jejich užití.

Teoretická část práce shrnuje přístupy ke klasifikaci spojovacích prostředků v češtině a upozorňuje na některé potíže v jejich vymezení. Dále se věnuje popisům jazykové kompetence v SERRJ s ohledem na užití spojovacích prostředků. V neposlední řadě pak prezentuje základní principy chybové analýzy a teorie mezijazyka. Druhá část práce charakterizuje použité korpusy a vysvětluje postup vyhledávání a zpracování dat, která byla použita pro následnou analýzu. Ta je obsahem třetí části práce, v níž nejprve popisujeme repertoár spojovacích prostředků, které se v korpusu objevují, a následně zkoumáme tendence v jejich užití. S pomocí korespondenční analýzy zachycujeme vztah spojovacích prostředků k jednotlivým úrovním podle SERRJ. Popis na závěr doplňujeme základní kvantitativně-kvalitativní chybovou analýzou.

Klíčová slova:

spojovací prostředky, čeština jako cizí jazyk, mezijazyk, chybová analýza, úroveň podle SERRJ, korpus MERLIN

Abstract

Over the last few years two learner corpora of authentic texts of non-native speakers of Czech have originated (MERLIN Corpus and CzeSL Corpus), giving linguists an important source of data for researching Czech as a foreign language. Ergo, for the first time it is possible to carry out the language analysis of non-native speakers of Czech using tools of corpus linguistics to formulate evidence-based research findings.

The presented thesis uses that as well, focussing on the study and description of the use of connectors in the written text production of non-native speakers of Czech, being primarily based on evidence from the learner MERLIN Corpus, which as opposed to other corpora is characterized by linking written text production reliably to the proficiency levels of the *Common European Framework of Reference for Languages (CEFR)*, including the levels of A2–B2. At the same time, it also contains a text error annotation, thus enabling to ascertain what effect the language proficiency has on the extent of the repertoire and frequency of using connectors, as well as to determine how the proficiency is reflected in using them correctly.

The theoretical part of the thesis summarizes the approaches to the classification of connectors in Czech, pointing out some difficulties in their classification. Furthermore, it deals with the description of language proficiency within the CEFR with respect to using connectors. Last but not least, it presents the basic principles of an error analysis and the interlanguage theory. The second part of the dissertation characterises the used corpora, explaining the procedures for retrieving and processing data employed for the subsequent analysis, which is included in the third part of the thesis. Firstly, the repertoire of connectors that the corpus contains is described there. Subsequently, the tendencies in their use are researched. Using a correspondence analysis, the relationship of connectors to individual proficiency levels of the CEFR is shown. The description is in the end completed with a basic quantitative-qualitative error analysis.

Key words: connectors, Czech as a foreign language, interlanguage, error analysis, language proficiency level, CEFR, MERLIN corpus

OBSAH

Prohlášení.....	ii
Poděkování.....	iii
Abstrakt.....	iv
Abstract.....	v
OBSAH.....	vi
1 ÚVOD.....	1
1.1 Motivace a přínos práce.....	2
1.2 Cíl práce.....	4
1.3 Výzkumné otázky.....	4
1.4 Metody.....	5
1.5 Struktura práce.....	6
1.6 Terminologické poznámky, zkratky a značky.....	7
2 TEORETICKÁ VÝCHODISKA.....	9
2.1 SPOJOVACÍ PROSTŘEDKY: VHLED DO PROBLEMATIKY.....	9
2.1.1 Spojovací prostředky autosémantické.....	11
2.1.1.1 Výrazy tvořené od zájmenných tázacích základů -k-/-kd-, -c-/-č-, j-.....	12
2.1.1.2 Výrazy tvořené od netázacího základu j-.....	13
2.1.2 Spojovací prostředky sysémantické.....	14
2.1.2.1 Sysémantické spojovací prostředky a slovnědruhov \acute{e} zařazení.....	15
2.1.2.1.1 Spojky – relativa / interogativa.....	15
2.1.2.1.2 Spojky – příslovce.....	17
2.1.2.1.3 Spojky – částice.....	20
2.1.2.1.4 Spojky – předložky.....	21
2.1.2.2 Sysémantické spojovací prostředky a formálně syntaktické vztahy.....	22
2.1.2.2.1 Sysémantické spojovací prostředky souřadící.....	24
2.1.2.2.2 Sysémantické spojovací prostředky podřadící.....	25
2.1.3 Spojovací prostředky podle podoby a struktury.....	27
2.1.4 Shrnutí.....	30
2.2 PÍSEMNÁ PRODUKCE NERODILÝCH MLUVČÍCH A SPOJOVACÍ PROSTŘEDKY.....	32
2.2.1 Spojovací prostředky a úrovně SERRJ.....	32
2.2.2 Spojovací prostředky a popisy referenčních úrovní.....	36
2.2.2.1 Úroveň A1.....	38
2.2.2.2 Úroveň A2.....	38
2.2.2.3 Úroveň B1.....	39
2.2.2.4 Úroveň B2.....	40
2.2.3 Shrnutí.....	41
2.3 TEORIE MEZIJAZYKA A CHYBOVÁ ANALÝZA.....	43
2.3.1 Principy chybové analýzy.....	44
2.3.2 Shrnutí.....	46
3 MATERIÁLOVÝ PODKLAD – KORPUSY.....	47
3.1 CHARAKTERISTIKA KORPUSŮ.....	47
3.1.1 Korpus MERLIN.....	48

3.1.1.1	Charakter a velikost korpusu.....	48
3.1.1.2	Metadata a složení korpusu.....	48
3.1.1.3	Anotace	54
3.1.2	Korpus CzeSL-SGT.....	55
3.1.2.1	Charakter a velikost korpusu.....	56
3.1.2.2	Metadata a složení korpusu.....	56
3.1.2.3	Anotace	59
3.1.3	Korpus SKRIPT2012.....	59
3.1.3.1	Charakter a velikost korpusu.....	59
3.1.3.2	Metadata a složení korpusu.....	60
3.1.3.3	Anotace	60
3.1.5	Korpusy SYN.....	61
3.1.5.1	Charakter a velikost korpusu.....	61
3.1.5.2	Metadata a složení korpusu.....	61
3.1.5.3	Anotace	62
3.2	MOŽNOSTI VYHLEDÁVÁNÍ SPOJOVACÍCH PROSTŘEDKŮ	63
3.2.1	Autosémantické spojovací prostředky	63
3.2.2	Synsémantické spojovací prostředky.....	65
3.2.3	Shrnutí.....	68
3.3	PRÁCE S KORPUSEM MERLIN (VYHLEDÁVÁNÍ, ZPRACOVÁNÍ DAT...).....	69
3.3.1	Zjištění repertoáru spojovacích prostředků	69
3.3.2	Totožné výrazy se spojovací funkcí vs. bez spojovací funkce	70
3.3.3	Frekvence užití spojovacích prostředků.....	73
3.3.4	Práce s chybovou anotací	74
4	ANALÝZA TEXTŮ	75
4.1	SPOJOVACÍ PROSTŘEDKY V KORPUSU MERLIN.....	75
4.1.1	Repertoár spojovacích prostředků – globální pohled.....	75
4.1.2	Repertoár a četnost podle úrovně (celkové hodnocení)	80
4.1.3	Tendence v závislosti na úrovni (celkové hodnocení).....	85
4.1.3.1	Složení repertoáru	85
4.1.3.2	Frekvence spojovacích prostředků.....	88
4.1.3.3	Shrnutí.....	95
4.1.4	Korespondenční analýza spojovací prostředky – úrovně SERRJ	95
4.1.5	Korespondenční analýza spojovací prostředky – hodnocení koheze a koherence	99
4.1.6	Chyby v užití spojovacích prostředků.....	102
4.1.6.1	Záměna spojovacího prostředku.....	102
4.1.6.2	Vynechání spojovacího prostředku	108
4.1.6.3	Nadbytečné užití spojovacího prostředku	110
4.1.6.4	Pravopisné a morfologické chyby	111
4.1.6.5	Shrnutí.....	112
4.2	SONDA DO KORPUSU CZESL-SGT A SKRIPT2012 (A SYN2015).....	114
4.2.1	Poměr četnosti spojení podle syntaktické funkce.....	114
4.2.2	Pořadí spojovacích prostředků podle frekvence	116
4.2.3	Závislost užití spojovacích prostředků na typu korpusu	117
4.3	SHRUTÍ VÝSLEDKŮ	120
5	ZÁVĚR.....	123

LITERATURA	126
Seznam tabulek	131
Seznam grafů.....	132
Seznam obrázků	133
PŘÍLOHY	134

1 ÚVOD

Výuka češtiny pro cizince má u nás dlouholetou tradici a zabývají se jí s úspěchem významná univerzitní pracoviště. Zmínit můžeme kupříkladu Ústav bohemistických studií na Filozofické fakultě Univerzity Karlovy, Ústav jazykové a odborné přípravy na téže univerzitě nebo Kabinet češtiny pro cizince na Filozofické fakultě Masarykovy univerzity v Brně. Podpoře se těší i výuka češtiny v zahraničí, kde jsou provozovány lektoráty a výuka probíhá také v krajanských komunitách.

Přesto však chybí didaktice češtiny jako cizího jazyka teoretická opora, která je běžná u výuky jiných, celosvětově vyučovaných cizích jazyků, jako je angličtina nebo němčina. V češtině se často vychází z intuitivních nebo tradicí podmíněných postupů, jimž se nedostává náležitého ověření prostřednictvím zkoumání povahy a fungování osvojovaných jevů v autentických situacích či materiálech.

V posledních několika letech však lze pozorovat jistý posun, kdy se pozornost stále více obrací k procesu osvojování si češtiny nerodilými mluvčími, resp. k jeho popisu na základě shromážděných podkladových dat z reálné komunikace (např. Hudousková 2013, Bulejčíková 2014 nebo Mašín 2015). Tento obrat je podporován vznikem žákovských korpusů, které shromažďují texty nerodilých mluvčích češtiny a umožňují v nich snadno vyhledávat a přehledně zpracovávat získané výsledky (srov. Šebesta 2010, Štindlová 2013).

Spolupráce na tvorbě jednoho z žákovských korpusů (korpus MERLIN), také inspirovala autora k výběru zaměření této disertační práce. Ta se bude zabývat užíváním spojovacích prostředků v textech nerodilých mluvčích češtiny, které byly v uvedeném korpusu shromážděny. A protože se autor věnuje zároveň i jazykovému testování a validaci úrovní jazykové kompetence, mapuje práce užití spojovacích prostředků v souvislosti s hodnocením úrovně písemné produkce.

Předkládaná disertační práce se tedy týká češtiny jako cizího jazyka a částečně zasahuje také do korpusové lingvistiky. Není ale ve své podstatě prací didaktickou, neboť se primárně snaží o doplnění poznání z oblasti akvizice jazyka. Přímou didaktickou transformací získaných poznatků se práce nezabývá, ale jak doufáme, mohla by pro ni být vhodným východiskem.

1.1 Motivace a přínos práce

Kromě osobního zájmu a aktuálnosti daného tématu existují i další a v kontextu zkoumání češtiny jako cizího jazyka významnější důvody, které podnítily vznik této disertační práce.

(A) Prvním podnětem, který jsme naznačili již v úvodu, je nedostatek empiricky podložených popisů různých fází osvojování si češtiny jako cizího jazyka. V minulosti sice byly vytvořeny tzv. popisy referenčních úrovní pro češtinu, jež se vztahují k úrovním A1–B2 podle *Společného evropského referenčního rámce pro jazyky* (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 pokusí, byť na úzce vytknuté oblasti spojovacích prostředků, ukázat, jak by mohl popis založený na autentických datech vypadat.

(B) Poukázali jsme v úvodu na to, že se v posledních letech 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).¹ Je to o to překvapivější, uvědomíme-li 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

¹ 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*, jenž slouží k automatickému hodnocení povrchové koherence (koheze) textů v češtině psaných nerodilými mluvčími češtiny. Software hodnotí úroveň předloženého textu z hlediska jeho povrchové výstavby (zohledňuje např. frekvenci a rozmanitost užitých spojovacích prostředků, bohatost slovní zásoby apod.). Pilotní verze je dostupná on-line: <https://ufal.mff.cuni.cz/evald/evald-10-foreigners>. Autoři tohoto projektu cílí na predikaci úrovně na základě syntézy různých globálních kritérií. Naše práce je naproti tomu zaměřena mnohem úžeji na popis užití konkrétních spojovacích prostředků v závislosti na úrovni a zohledňuje mimo jiné také správnost a vhodnost jejich užití.

jazyka mohli opírat. Ve výsledku by se tak mohla nepřímo posílit i pozice, jakou zaujímá skladba v učebnicích češtiny jako cizího jazyka.²

(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–C1) 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, které může v budoucnu sloužit k dalšímu didaktickému zpracování a aplikaci ve výuce nebo při tvorbě učebních materiálů.

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

1.2 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.

1.3 Výzkumné otázky

Výše uvedené obecně formulované cíle jsme konkretizovali v podobě výzkumných otázek, jimiž zároveň udáváme směr analýze korpusového materiálu. Odpovědi na ně budou společně s úvahami, které nás během výzkumu provázely, součástí kapitoly 4.

1. Které spojovací prostředky užívají nerodilí mluvčí v textech zahrnutých do korpusu MERLIN?
2. Jak vypadá složení repertoáru spojovacích prostředků na jednotlivých úrovních podle 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?

1.4 Metody

V předkládané práci aplikujeme kvantitativně-kvalitativní analýzu korpusového materiálu. Jedná se tedy o smíšený výzkum, „který kombinuje kvalitativní i kvantitativní přístup, aby se navzájem kompenzovaly slabiny každého z nich [...]“ (Skutil 2011, s. 74)

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. S její pomocí lze popsat vztahy zadaných proměnných a získat grafické znázornění souvislostí ve vícerozměrném prostoru (srov. Ramos – Carvalho 2010). Výsledkem analýzy je tzv. korespondenční mapa představující osy redukovaného souřadnicového systému, ve kterém jsou graficky zobrazeny jednotlivé kategorie obou proměnných. Polohy bodů v mapě tak přímo vyjadřují asociaci mezi nimi.

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 kontrastivní a chybové analýzy mezijazyka (viz kap. 2.3). 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.

Jsme si zároveň vědomi omezení, která tuto práci provázejí. Jak upozorňují Biber – Xeppen (1998, s. 145), korpusová analýza například nevede k úplnému strukturnímu popisu systému, závisí totiž mimo jiné na složení a velikosti korpusu. Na tato omezení upozorňujeme v rámci charakteristiky použitých korpusů a nástinu práce s nimi (kap. 3). Na druhou stranu poskytuje korpusová analýza přesnější vhled do reálného užívání prostředků a není tak závislá na introspekci a intuici. Proto se také výsledky kontrastivních i chybových analýz opřených o korpus využívají při sestavování jazykových slovníků a tvorbě učebních materiálů (především v angličtině a němčině) (srov. Škodová – Štindlová – Hana – Rosen 2011).

1.5 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 budeme v této práci využívat. Součástí druhé kapitoly je také shrnutí charakteristiky úrovní 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.

1.6 Terminologické poznámky, zkratky a značky

Na tomto místě bychom rádi uvedli definice termínů, které v této práci budeme dále používat. Jde o termíny, které bývají v různých lingvistických koncepcích chápány odlišně nebo které se užívají pouze ve specifických oblastech jazykové výuky a testování, a proto nemusejí být široké lingvistické obci známy.

deskriptor SERRJ

Termínem deskriptor se míní popis dovednosti uživatele jazyka v konkrétní oblasti komunikativní jazykové kompetence na dané úrovni podle SERRJ, např. u koheze na úrovni A1: „[Uživatel jazyka na dané úrovni] dokáže propojit slova nebo skupiny slov pomocí nejzákladnějších lineárních spojovacích výrazů, jako jsou „a“ („and“) a „pak/potom“ („then“).“ (SERRJ 2001, s. 128).

stupnice SERRJ

Stupnice představuje posloupnost všech deskriptorů určité oblasti komunikativní jazykové kompetence uspořádanou od deskriptorů nejnižší úrovně po deskriptory nejvyšší úrovně. Příklady viz kap 2.2.

spojka = konjunkce

Neohebný sysémantický slovní druh vyznačující se syntaktickou funkcí spojovat slova, syntaktické slovní skupiny a věty, např. *a, že, a proto; bud' – anebo* atd.

spojovací výraz

Jednoslovná lexikální jednotka plnící syntaktickou funkci spojovací. Slovnědruhovou platností může, ale i nemusí přináležet ke spojkám, např. *a, co, kam, jehož* atd.

spojovací prostředek

Jednoslovná nebo víceslovná jednotka, která plní syntaktickou funkci spojovací. Slovnědruhovou platností může, ale i nemusí přináležet ke spojkám, např. *a, co, kam, jehož, a proto; bud' – anebo* atd.

mezijazyk, interlanguage

Jazyk nerodilých mluvčích, který je vnímán jako samostatný systémem, který má status přechodného systému mezi jazykem prvním a jazykem cílovým (srov. Selinker 1972).

žák, žákovský korpus

Termínem *žákovský korpus* označujeme korpus jazyka nerodilých mluvčích. Ve shodě se Štindlovou (2013) chápeme „žáka“ jako osobu učící se cizí jazyk, a to bez ohledu na věk, příp. jiné sociologické proměnné.

i.p.m. (z angl. instances per million)

Je jednotkou relativní frekvence. Vyjadřuje průměrný počet výskytů jednotky nebo slova v hypotetickém textu/korpusu o délce 1 milion slov/pozic.

- V příkladových větách značí možnost nahrazení spojovacího prostředku jiným konkurenčním prostředkem. Např. *Nevím, **jestli** (→zda) přijde.*
- ↓ V příkladových větách značí nesprávné užití spojovacího prostředku a uvádí možnou opravu. Např. *Chtěla bych se zeptat, **pokud** (↓jestli) máte volný pokoj.*

2 TEORETICKÁ VÝCHODISKA

V této kapitole věnujeme pozornost třem oblastem, které představují hlavní teoretická východiska našeho zkoumání. Pro analýzu a popis spojovacích prostředků vnímáme jako klíčové jejich vymezení, při němž se nelze vyhnout pohledu na dosavadní přístupy k jejich klasifikaci. Zároveň máme na zřeteli fakt, že se zabýváme jazykem nerodilých mluvčích. Shrnujeme proto charakteristiku jejich jazykové kompetence v oblasti užívání prostředků koheze a koherence podle SERRJ. Neobejdeme se ani bez dostatečného vhledu do teorie nabývání cizího jazyka a budování jazykové kompetence a bez orientace v pojmech, jako jsou chybová analýza nebo kontrastivní analýza mezijazyka.

2.1 SPOJOVACÍ PROSTŘEDKY: VHLED DO PROBLEMATIKY

V prvním z cílů této práce jsme si předsevzali zmapovat užití spojovacích prostředků v češtině nerodilých mluvčích s využitím korpusových dat. Jedním z předpokladů úspěšného naplnění tohoto cíle je definice toho, co pojmem *spojovací prostředky* rozumíme.

Pokusy o jejich vymezení ovšem narážejí na potíže spojené s tím, že představují poměrně širokou a rozmanitou skupinu výrazů zahrnující spojky, relativna (vztažná zájmena, vztažná zájmenná příslovce), interogativa (tázací zájmena a tázací zájmenná příslovce) a částice. Mezi lingvisty sice panuje shoda v tom, že jejich společným rysem je schopnost naplňovat syntaktickou funkci spojovací, ale kvůli množství specifických příznaků těchto výrazů se často rozcházejí při detailnějším třídění (srov. Doležalová 2008). Kromě potenciálu spojovat větné členy a věty totiž existuje řada příznaků, podle kterých je možné vydělovat a detailněji specifikovat různé podskupiny spojovacích prostředků. Zmíňme kupříkladu ostatní syntaktické funkce (např. relativna se mohou vyznačovat funkcí endeforickou) nebo jejich sémantické a morfologické příznaky (např. autosémantičnost, ohebnost atd.).

Z pohledu do význačných mluvnických prací vzniklých od poloviny minulého století je zřejmé, že každý pokus o klasifikaci spojovacích prostředků nutně odráží příslušnou koncepci zvolenou autory pro popis syntaktické roviny jazykového systému, resp. pro klasifikaci souvětí.

Zejména ve školské praxi je za tradiční považována klasifikace vedlejších vět podle větněčlenského hlediska (Šmilauer 1966 a Havránek – Jedlička, 1988 a 1996). Vedlejší věta představuje realizaci větného členu, který nahrazuje. Každému typu vedlejší věty jsou

přiřazeny spojovací prostředky sloužící k jejímu zapojení do věty hlavní, resp. souvětí. Jinými slovy, inventář spojovacích prostředků může být uspořádán na základě toho, kterou vedlejší větu uvozuje (např. pro vedlejší větu příslovečnou časovou jsou příznačné spojky *když, až, jak* atd.).³

Od konce 50. let se začala rozvíjet klasifikace souvětí podle meziprozodických vztahů, která se objevuje v pracích J. Bauera a M. Grepla (srov. Bauer 1958, Bauer – M. Grepl 1972, Grepl 1977–78 a Grepl – Karlík 1986). Podle jejich výkladu se souvětí třídí v prvním plánu podle toho, zda je propozice (obsah věty) zapojena do jiné propozice, nebo zda se jedná pouze o spojení (paralelnost) propozic. Teorie meziprozodických vztahů rezonuje také v díle F. Daneše (1977) a v duchu valenční teorie ji dále posouvá J. Hrbáček (1987) v akademické Mluvnici češtiny 3. Spojovací prostředky lze na základě tohoto přístupu třídit podle toho, jaký druh významového vztahu v souvětí vyjadřují (např. souvislosti časové vyjadřují spojky *když, až, dokud* atd.).⁴

S odlišným pojetím klasifikace souvětí přišel koncem 60. let K. Svoboda (1968, 1972). Představil koncepci, která je založena na formálně výrazovém kritériu a není valenční. Vycházel z toho, že tradiční klasifikaci souvětí souřadného (Havránek – Jedlička 1988, s. 403) by bylo možno po odstranění některých nedůsledností⁵ chápat jako „třídění podle hlavního gramatického rysu vět z hlediska souvětí stavby, tj. podle spojovacích výrazů“ (Svoboda 1972, s. 58). Nově rozšířil tento způsob klasifikace také na podřadné souvětí. Klíčovou je v jeho přístupu větněčlenská (a slovnědruhová) povaha spojovacích prostředků (věty uvozené výrazy autosémantickými vs. věty uvozené výrazy synsémantickými).⁶

V této disertační práci se do značné míry opíráme o poslední ze zmíněných teorií. Volba formálně výrazového kritéria klasifikace souvětí se nám jeví jako nejvhodnější východisko pro analýzu dat z korpusů umožňující efektivní vyhledání konkrétních výrazů nebo jejich skupin. Pro třídění spojovacích prostředků stejně jako Svoboda využíváme větněčlenskou charakteristiku výrazů a vydělujeme níže dvě skupiny. První skupinu tvoří spojovací

³ Srov. kritiku tohoto přístupu (Svoboda 1972, s. 58). „Ve výkladech z hlediska větněčlenského se spojovací výrazy pouze vypočítávají, popř. charakterizují po stránce stylové. Uvádí se např., že věty předmětové jsou připojeny spojovacími výrazy *že, aby, jak, kdo, co, kam, odkud*, tj. pouze se vypočítávají spojovací výrazy velmi různorodé. A obráceně se konstatuje, v kterých větách se užívá např. spojovacího výrazu *že, aby* atd., aniž se objasňuje, co tyto věty mají společného, co je např. společným sémantickým rysem každého jednotlivého spojovacího výrazu ve všech nebo alespoň v některých případech jeho užívání, v čem je rozdíl mezi spojovacími výrazy.“

⁴ K otázce významu spojek srov. Bauer (1967, s. 29n.).

⁵ Jako příklad uvádí Svoboda (Svoboda 1972, s. 58) neoddělování asyndetického souvětí od souvětí syndetického.

⁶ Na Svobodu navazuje Štěpán (2013), který svou teorii prohlubuje o třídění ze sémanticko-syntaktického hlediska a z hlediska komentování.

prostředky, které mají platnost větného členu, tj. výrazy autosémantické (zájmena, zájmenná příslovce, zájmenné číslovky)⁷, druhou skupinu tvoří věty uvozené spojovacími výrazy, které nemají platnost větných členů, tj. výrazy symsémantické (spojky a některé částice).

2.1.1 Spojovací prostředky autosémantické

Uvedli jsme výše, že mezi lexikálně plnovýznamovými slovy existuje skupina výrazů, které mohou sekundárně plnit spojovací funkci. Skupina zahrnuje některé druhy zájmen, příslovci a číslovek, které vykazují společné vlastnosti na různých rovinách popisu jazyka. Tyto vlastnosti na jedné straně tvoří společný jmenovatel výrazů zahrnutých do této skupiny a na straně druhé je vymezují vůči ostatním výrazům v rámci daného slovního druhu (např. spojovací funkci plní pouze některé druhy zájmen).

Zaměříme se nejprve na rovinu syntaktickou. Zde stojí v popředí skutečnost, že autosémantické spojovací prostředky zastávají ve větné struktuře pozici větného členu. Tím se zásadně liší od spojovacích prostředků symsémantických. **Větněčlenské zapojení** má také vliv na to, jaké spojovací vztahy mohou realizovat. Obecně platí, že nenavazují spojovací relace mezi jinými větnými členy ve větě, ale uplatňují se pouze při uvozování vět v souvětí. Z hlediska formálně syntaktického se s jejich pomocí realizují hypotaktické mezivětné vztahy.⁸

Vezmeme-li v potaz funkčně sémantické charakteristiky, vystoupí do popředí další společný rys autosémantických spojovacích prostředků, kterým je **pronominálnost**, tzn. schopnost označovat skutečnost nikoliv pojmenováním, ale jiným způsobem. Jedná se o deixi, tj. „ukazování nebo odkazování“ (EŠČ 2002, s. 105). MČ2 označuje slova s touto schopností jako deiktická, resp. jako pronominalia (pro-nominalia, „zá-jmenná slova“) a řadí k nim zájmena, zájmenná příslovce a zájmenné číslovky (srov. MČ2 1986, s. 81). V rámci předkládané disertační práce je toto vymezení podstatné zejména pro příslovce a číslovky. A to z toho důvodu, že zatímco větněčlenská platnost představuje společnou vlastnost různých druhů příslovci a číslovek, pronominálnost, resp. deiktičnost je charakteristická pro ty z nich, které mají zároveň spojovací funkci.

⁷ I když si uvědomujeme problematičnost otázky významové samostatnosti u uvedených slovních druhů, držíme se v naší práci jejich zařazení k autosémantikům stejně jako Štěpán (2013) nebo PMČ (1995).

⁸ Jedná se o vztahy determinace, pouze v několika případech tzv. nepravých vedlejších vět vyjadřují vztah koordinační. Např. *K dohodě jsme nedospěli, což mě nepotěšilo.* → *K dohodě jsme nedospěli a to mě nepotěšilo.* (= význam slučovací. V syntaktických výkladech se v těchto případech mluví o pseudorelativách (zájmeno *což*, *čehož*, *čemuž*, ..., zájmenná příslovce *nacež*, *pročež*, *příčemž*, *začež*). Srov. MČ2 (1986, s. 220).

Deiktičnost, resp. její různé typy, s sebou zároveň přináší i bohaté spektrum funkčně sémantických a výrazových vlastností (určitost, kvantifikovatelnost apod.), podle nichž se autosémantické spojovací prostředky mohou zařadit do několika „tříd deiktických slov“ (MČ2 1986, s. 88). Najdeme je jednak ve třídě *interogativ* (tázacích výrazů) a jednak ve třídě *relativ* (vztažných výrazů). Do interogativ patří tázací zájmena, tázací zájmenná příslovce a tázací zájmenné číslovky, k relativům se řadí vztažná zájmena, vztažná zájmenná příslovce a vztažné zájmenné číslovky. Obě třídy různou měrou implikují neurčitost a nekonkrétnost, interogativa oproti relativům navíc vyjadřují „apel [...] k odstranění této neurčitosti“ (MČ2 1986, s. 96). Relativa se nadto vyznačují tím, že mají společný denotát s některým ze členů ve větě řídící (typicky jsou koreferenční). I když jsou jako větné členy zapojeny do závislé věty (např. získávají v ní pád), se členem věty řídící, k němuž se vztahují, sdílejí některé další gramatické kategorie (obvykle rod a číslo).⁹

Při bližším pohledu na repertoár autosémantických spojovacích prostředků, lze konstatovat, že se jedná o relativně uzavřenou skupinu výrazů majících několik společných slovtvorných základů a tvořených specifickými slovtvornými postupy (např. spřahováním). Vydělit můžeme:

2.1.1.1 Výrazy tvořené od zájmenných tázacích základů *-k/-kd-*, *-c/-č-*, *j-*

i) Ohebné¹⁰

(a) *kdo, co*¹¹, *jaký, který, čím*

(b) *kolik*

(c) *kolikátý, kolikery*

ii) Neohebné

(d) *kolikrát, kolikery, kolikero*

(e) *kde, odkud, kudy, kam*

(f) *kdy, odkdy, dokdy, nakdy*

(g) *jak*¹²

⁹ Užití interogativ a relativ pro spojování vět otevírá problematiku závislých tázacích vět (tzv. nepřímých otázek) a vztažných vět. Tato oblast je velmi rozsáhlá. Kupříkladu při popisu vztažných vět vycházejí čeští lingvisté z různých kritérií od toho, jaký větný člen nebo slovní druh vztažná věta zastupuje, přes povahu relativu až po povahu výrazu v řídící větě, k němuž se tato věta vztahuje. Blíže srov. NESČ 2017, heslo Vztažná věta.

¹⁰ Pro cizince, kteří si osvojují češtinu jako cizí jazyk, je navíc důležité, zda se dané výrazy skloňují, dělíme proto spojovací výrazy dále podle ohebnosti.

¹¹ Se zájmenem *co* je homonymní neskloňná spojka *co* (viz kap. 2.1.2.1.1).

(h) *proč, nač, zač*

Tradičně jsou tyto výrazy klasifikovány: (a) jako zájmena tázací a vztažná, (b)–(d) jako zájmenné číslovky, (e)–(h) jako tázací a vztažná zájmenná příslovce; a zahrnují se do skupiny interogativ, resp. relativ.

Výrazy (e)–(h) se podle okolnostního významu, který vyjadřují, dělí detailněji na zájmenná příslovce: (e) místa, (f) času, (g) způsobu, (h) příčiny.

Ze slovotvorných specifík této skupiny spojovacích prostředků zmiňme následující:

Výrazy (a), (b) a (e)–(h) se objevují s komponentem *-pak* (*kdopak, copak, jakýpak, kterýpak, čipak*), která zdůrazňuje význam otázky (srov. MSC 2010, s. 221). I když mohou i v této podobě spojovat věty, jejich výskyt ve spojovací funkci je minimální (např. *kdepak* uvozující závislou větu následující po větě řídicí – v SYN2015 pouze 2,88 výskytu na milion pozic, zatímco *kde* 804,63).

Výrazy (a), (e), (g) a *kdy* mohou přibírat *-koli(v)* (např. *kdokoli(v), cokoli(v), kdekoli(v), kamkoli(v)*). Frekvence jejich užití ve funkci spojovacích prostředků je obdobná jako u výrazů s *-pak*, takže v rámci všech autosémantických spojovacích prostředků zaujímají spíše periferní pozici (např. *kdekoli(v)* uvozující závislou větu následující po větě řídicí – v SYN2015 pouze 1,13 výskytu na milion pozic).

Zájmena *kdo, co, jaký* a *který* lze doplnit o postfix *-ž*. Vzniklé výrazy *kdož, což, kterýž* také mohou plnit spojovací funkci. Nejfrekventovanější je výraz *což*, naopak výrazy *kdož* a *kterýž* se vyskytují méně často a mají knižní příznak.

Zájmeno *co* tvoří ve spojení s akuzativními předložkami výrazy *proč, nač, zač, oč, več*. Tato spojení jsou homonymní se zájmennými příslovci vyjadřujícími příčinu, srov. (1) vs. (2).

- (1) *Student neví, nač by se zeptal nejdřív. (nač = na co, nelze nahradit výrazem proč)* (SYN 4)
- (2) *Nevidím důvod, nač (→proč) mluvit o smrti.*

2.1.1.2 Výrazy tvořené od netázacího základu *j-*

i) Ohebné

(a) *jenž*

(b) *jejíž*

¹² Se zájmenným příslovcem *jak* je homonymní spojka *jak* (viz kap. 2.1.2.1.1).

ii) Neohebné

(c) *jehož, jejichž*

Výrazy (a)–(c) se řadí k zájmenům vztažným, tj. mezi relativa; (b) a (c) mají navíc přivlastňovací význam.

2.1.2 Spojovací prostředky sysémantické

Druhou skupinu spojovacích prostředků, kterou jsme vyčlenili, jsou sysémantické spojovací prostředky. Oproti autosémantickým spojovacím prostředkům nemají větněčlenskou platnost. Jako takové k nim patří výrazy řazené k neohebným slovním druhům, které mohou mít spojovací funkci.¹³ Primárně ji plní spojky, které zároveň tvoří nejpočetnější skupinu. Zdálo by se proto jako výhodné použít pro vymezení těchto spojovacích prostředků právě jejich slovnědruhovou příslušnost. Zaměříme-li se však na konkrétní případy, zjistíme, že využití této charakteristiky je značně problematické.

Spojky totiž mohou vznikat z jiných slovních druhů (od zájmen, přes příslovce až po částice). U některých výrazů tento proces proběhl v minulosti a je již ukončen, u jiných můžeme jeho průběh pozorovat i v současnosti.¹⁴ Proto se při slovnědruhové klasifikaci setkáme s mnoha přechodnými případy. U výrazů, které nestojí v centru systému a spojovací funkci neplní primárně, často nepanuje shoda v přiřazení k určitému slovnímu druhu.¹⁵ Projevuje se zde navíc skutečnost, že se při třídění slovních druhů uplatňuje několik kritérií (sémantické, syntaktické a tvaroslovné), z nichž může v každém pojetí (a u každého slovního druhu) převládat jiné.¹⁶

¹³ Srov. Základy české skladby (1958), Česká mluvnice (1988), Nauka o českém jazyku (1982), PMČ (2000), ČŘaJ (2000) a EŠČ (2002). MČ2 (1987, s. 214) počítá jako výjimku v ohebnosti spojky *aby* a *kdyby*.

¹⁴ Srov. popis procesu konjunkcionalizace u Hošnové (2005, s. 37): „Jestliže se funkce spojovací stává základní funkcí daného výrazu, dochází ke ztrátě jeho původní sémantické platnosti, výraz se stává novým lexémem se samostatnou platností. Může dojít k naprosté konverzi z jiného slovního druhu; jestliže jde o autosémantický výraz, ztrácí platnost větného členu. Např. ve větě *Petr se vrátil domů až po půlnoci, proto se na něho rodiče rozzlobili* nelze výraz *proto* považovat za příslovce; nemá platnost adverbiale, lze jej beze změny významu nahradit jiným výrazem vyjadřujícím vztah důsledkový (*tak, tedy, tudíž* apod.).“

¹⁵ Jako příklad uveďme výraz *totiž*. Ten je ve starších pracích (srov. Šmilauer 1966, s. 314; SSJČ – *Znám ho dobře, je to totiž můj spolužák.*) klasifikován jako příslovce, zatímco nověji je řazen mezi spojky souřadící (srov. SSČ – *Nerozumím mu, mluví totiž anglicky.*).

¹⁶ Srov. Komárkovu kritiku teorie slovních druhů v časopise *Slovo a slovesnost*: „Soustava slovních druhů, s níž pracuje současná mluvnice, je nesourodá, postrádá jednotného principu. Potíže s roztříděním slovních druhů jsou způsobeny složitou povahou kategorie slovního druhu: slova jsou jednak výrazem různých kategorií myšlenkových, jednak mají své vlastnosti formální a své syntaktické využití. Tato tři hlediska

2.1.2.1 Syntaktické spojovací prostředky a slovnědruhové zařazení

Místa, kde se pokusy o tradiční slovnědruhovou klasifikaci v různých pojetích střetávají, nacházíme na hranici mezi (a) spojkami a relativy / interogativy, (b) spojkami a příslovci, (c) spojkami a částicemi, (d) spojkami a předložkami. V předkládané disertační práci si neklademe za cíl poskytnout přehled toho, jak jsou tyto slovní druhy vymezovány v rámci jednotlivých pojetí, ani nechceme tato pojetí mezi sebou porovnávat.¹⁷ Než přistoupíme k analýze zkoumaných textů, považujeme ale za důležité, upozornit na to, kde se zmíněné třecí plochy projevují. Zaměřujeme se především na ty oblasti nebo jevy, jež jsou relevantní pro korpusovou analýzu, kterou budeme provádět, tj. např. na případy, ve kterých je při analýze nutné oddělit od sebe totožné výrazy, z nichž jen jeden má v daném kontextu spojovací funkci a druhý nikoliv.

2.1.2.1.1 Spojky – relativy / interogativa

V předchozích výkladech zaznělo, že spojky a relativy (vztažná zájmena, vztažná zájmenná příslovce a vztažné zájmenné číslovky) sblíží jejich spojovací funkce, ale rozdílná je jejich větněčlenská platnost. To zmiňuje např. Hošnová (2005, s. 27): „[...] současné gramatiky se při stanovení hranice mezi relačními adverbii a zájmeny na straně jedné a spojkami na straně druhé opírají o to, zda daný výraz je začleněn do věty primárně jako její větný člen a funkci konektivní plní až sekundárně, nebo zda je pouze pomocnou gramatickou součástí věty, není jejím větným členem;...“. Větněčlenská platnost charakterizuje a do jisté míry omezuje uplatnění relativ ve spojovací funkci. Užívají se pouze pro vyjádření vztahů hypotaktických (a jen v omezené míře mohou vstupovat do vztahů koordinačních). Konkrétní výraz volíme podle toho, jakou větněčlenskou funkci má ve větě závislé a na jaké slovo ve větě řídící odkazuje (rozhodující je věcný význam tohoto slova) – (srov. Bauer 1962/63, s. 221n.).

nejdou paralelní a mnohdy se rozcházejí; dobře je to vidět např. na kategorii zájmen, číslovek i příslovcí, často nejednotných po stránce tvarové, syntaktické i významové.“ (Komárek 1965, s. 160)

¹⁷ O to se ve svých pracích pokusily např. Hošnová (2005, s. 26–33), Čermáková (2012, s. 18–25) nebo Doležalová (2008, s. 7–12).

Kritérium větněčlenské platnosti je důležité také pro odlišení některých výrazů na pomezí spojek a relativ / interogativ, které se vyznačují homonymií.¹⁸ Jedná se o frekventované dvojice totožných výrazů, z nichž každý plní jinou syntaktickou funkci. Podle toho, zda konkrétní výraz zaujímá ve větné struktuře větněčlenskou pozici, jej řadíme buď k autosémantickým spojovacím prostředkům (relativa), nebo ke spojovacím prostředkům synsémantickým (spojky). Demonstrovat lze tento případ na výrazu *co*. Srovnajme následující věty:

(3) *Zároveň netuší, **co** jí má říct.* (SYN 4)

(4) ***Co** (→*zatímco*) tady vedeme řeč, mohls spis už vyřídít, proklatý byrokrate!* (SYN 4)

Zatímco v první větě (3) má výraz *co* platnost vztažného zájmena a je větným členem závislé věty, v druhé větě (4) jde o spojku (nemá větněčlenskou platnost), která v tomto případě vyjadřuje současnost dějů věty řídicí a závislé a je ji možno nahradit spojkou *zatímco*.

Obdobně se s homonymií setkáme u výrazu *jak*.¹⁹ Ten může fungovat jednak jako příslovce (5), jednak jako spojka (6).

(5) *Přemýšlel jsem, **jak** (→*kerak*) ta pověra podzimního pouštění vznikla.* (SYN 4)

(6) *Že je ve správném pokoji, poznal hned, **jak** otevřel dveře.* (SYN 4)

Kromě toho, že dochází na hranici spojek a relativ k homonymii, pozorujeme u nich i jiný styčný bod. Na mysli máme případy konkurence spojek a víceslovných spojovacích výrazů složených ze zájmenného příslovce a spojky (*protože / proto, že; takže / tak, že* apod.). Kopečný (1962, s. 57) objasňuje jejich paralelní výskyt a uvádí, že má své kořeny v tom, že některé spojky vznikaly spojením příslovce ve větě řídicí se spojkou ve větě závislé (např. *proto, že > protože*).²⁰

¹⁸ „O homonymii, tj. o dvě slova, jde také tehdy, vyjadřují-li dva totožné slovní útvary sice významy, mezi nimiž je patrná spojitost, takže i u nich lze určit význam obecný, resp. základní, avšak mají platnost různých slovních druhů, např. týž slovní útvar je jednou příslovcem, podruhé spojkou, nebo jestliže se liší syntaktickou povahou, tj. jednou je souřadící, podruhé podřadící.“ (Svoboda, 1972, s. 50)

¹⁹ Srov. také Pečený (2009, s. 51n.).

²⁰ Svoboda (1972, s. 26) označuje zájmena a příslovce v takovýchto případech jako anticipační korelativní výrazy a považuje je mimo jiné za formální označení hypotaktické povahy souvětí. V jeho koncepci se korelativní výrazy stávají důležitou složkou pro třídění vedlejších vět. Podle toho, která korelativa a koreferenční výrazy se kombinují, definuje například věty zájmenně částicové (spojené např. pomocí *takový, že*) a příslovcově částicové věty (spojené např. výrazy *tak, že*). V tomto třídění na něj navazuje Štěpán (2013).

V historickém vývoji k sobě ostatně měly spojky a relativia blízko také tím, že sdílely některé společné zájmenné základy. Například spojka *že* má původ v praslovanském **ježe*, které je stejně jako zájmeno *jenž* tvořeno tvarem zájmena **jb* a navíc zesilovací částicí **-že*.²¹

U některých vztažných zájmenných příslovčí proběhl proces konjunkcionalizace. Bauer (1962, s. 35) uvádí jako příklad výrazy *dokud* a *pokud*, které připojují věty „určující trvání hlavního děje, a to jednak uvedením souběžného děje [(7)], jednak děje, který hlavní děj ukončí (zde je v časové větě vždy zápornka *ne*) [(8)]“. O přehodnocení slovnědruhové platnosti těchto výrazů svědčí podle Bauera mimo jiné to, že ztratily větněčlenskou platnost v závislé větě.

(7) *Pak jsem si řekl, že pojedou, **dokud** (→do té doby, co) to půjde.* (SYN 4)

(8) *Bojuje, **dokud** nepadne (→tak dlouho, než padne).* (SYN 4)

2.1.2.1.2 Spojky – příslovce

V předchozím oddílu, který pojednával o vztahu spojek a relativ / interogativ, jsme se dotkli oblasti vztažných a tázacích zájmenných příslovčí. Ukázali jsme, že jejich odlišení od spojek je klíčové pro zařazení výrazu buď k autosémantickému, nebo sysémantickému spojovacímu prostředkům. Zbývá ovšem celá řada dalších příslovčí (zájmenná příslovce ukazovací, příslovce modální, kauzální aj.), která mohou být homonymní se spojkami (nebo v ně přecházet). Při jejich rozlišení není otázkou pouze větněčlenské zapojení do syntaktické struktury věty, ale to, zda/jak plní spojovací funkci. Proces ztráty větněčlenské platnosti na jedné straně a nabývání (posilování) spojovací funkce na straně druhé, je aktuální u nemalé skupiny výrazů.²² O jaké výrazy se jedná a v jakých kontextech se s nimi můžeme setkat, ukážeme na několika příkladech.

V podřadném souvětí se často objevují ve větách příslovečných a mohou signalizovat rozmanité významové vztahy. Obvykle se také stávají součástí skupin spojovacích prostředků, které si v daném významovém okruhu konkurují. Kupříkladu věta

²¹ Srov. Rejzek (2012, s. 787). Kopečný (1962, s. 305) vznik *že* osvětluje příkladem: „*vidím, že prší* může spočívat na původním pojetí: *vidím, hle, prší!*“

²² Některým problematickým skupinám, kde není rozlišení jednoznačné nebo kde dochází k přechodu výrazu od příslovce ke spojce, se věnoval Bauer (1962). Ve svých úvahách se zabíral i významovou stránkou zkoumaných výrazů. Srov. Bauer (1962, s. 29): „Z definic vyplývá, že mezi oběma slovními druhy je především rozdíl v jejich syntaktické platnosti: adverbium je větným členem, spojka nikoli. K tomu se druží rozdíl sémantický: příslovce jsou slova lexikální, s vlastním věcným významem, spojky jsou slova mluvnická, bez vlastního věcného významu.“

se spojovacím *jen* (9) vyjadřuje, že děj závislé věty proběhl bezprostředně před dějem věty řídicí (srov. také to, že věta závislá předchází větě řídicí). Spojovací *jen* můžeme v tomto případě nahradit výrazy *sotvaže* nebo *jakmile*. Stojí-li *jen* ve funkci příslovce²³ (10), má význam omezovací.

(9) **Jen** (→*hned když / sotvaže / jakmile*) přišel po střídání na hřiště, skóroval. (SYN 4)

(10) Poštovní zásilky se nakonec ve voze našly, zmizely **jen** (→*pouze*) peníze. (SYN 4)

U některých výrazů je užití ve spojovací funkci méně frekventované. Za příklad může sloužit výraz *třeba*, který v souvětí signalizuje význam přípustky (11). Mnohem častěji se ale vyskytuje jako příslovce vyjadřující nutnost (12).

(11) Udělal to, **třeba** (→*třebaže / i když*) měl jinou práci. (SSJČ)

(12) Přepravu nadměrného nákladu je **třeba** (→*nutno*) projednat a ověřit. (SYN 4)

Na příkladech výše jsme ukázali užití spojky homonymní s příslovcem v hypotaktickém souvětí. Výjimkou ale není ani u parataktických vztahů v koordinačně spojených souvětích. V souřadném souvětí s vylučovacím vztahem se objevuje například výraz *jinak* (13). Zde je sice patrná významová souvislost s adverbiálním užitím (14), ale není *jinak* příslovečným určením, které by popisovalo okolnosti slovesa *končit*.²⁴

(13) Maloobchodní firmy na to musí rychle reagovat, **jinak** (→*nebo*) končí. (SYN 4)

(14) Do hotelového apartmánu vtančila, **jinak** (→*jiným způsobem*) se to nedá nazvat. (SYN 4)

Frekventovaná je skupina souřadících spojovacích výrazů s původem v příslovcích, které mohou stát jednak samostatně: *proto*, *tak*, *tedy*, *tudíž*, jednak v kombinaci se spojkou *a* (*a proto*, *a tak*, *a tedy*, *a tudíž*). Objevují se jako spojovací prostředky v souvětích s důsledkovým významovým vztahem (15) a (16). Bauer objasňuje (1962, s. 30), že se u nich

²³ V SSČ, z něhož zde vycházíme, se uvádí také užití *jen* jako částice s významem zdůraznění, srov. *Z koní se jen kouřilo*. V některých mluvnických pracích se *jen* pokládá za částici i tam, kde je SSČ charakterizuje jako příslovce, srov. MČ2 (1986, s. 235) nebo PMČ (1995, s. 361).

²⁴ Srov. Bauer (1962, s. 31): „Kdybychom v něm [ve výrazu *jinak*] chtěli vidět příslovečné určením, museli bychom velmi obecně říci, že odkazuje na okolnosti, za nichž nastal uvedený děj.“

spojovací funkce vyvinula z funkce anaforické a vyjadřovala navázání na to, co již bylo řečeno nebo známo z kontextu.²⁵

(15) *Sleva byla poměrně vysoká, **proto** (→tak) uvedený portál navštívil.* (SYN 4)

(16) *Vánočky obsahují spoustu másla, **a proto** (→a tak) je obtížné je skladovat.* (SYN 4)

U parataktického souvětí stojí za povšimnutí také výrazy vyjadřující časovou následnost (*nejdřív, potom, pak, nakonec* apod.). Jedná se o časová příslovce, která naznačují postupný sled dějů a ve větě fungují jako příslovečná určení času (17). Samy o sobě tyto výrazy nemají spojovací funkci, takže pokud se zároveň nevyskytují např. se spojkou *a*, jedná se asyndetické koordinační spojení.

(17) ***Nejprve** nalepíte jeden bok, **pak** druhý a **nakonec** přilepíte chlopně čelního skla.*
(SYN 4)

Homonymní jsou spojovací výrazy *pak, potom, tedy*, které připojují větu řídicí následující za větou podmínkovou (18).

(18) *Pokud chceme nechat mléko zkysnout, **pak** to musíme provést při pokojové teplotě a ne v chladničce.* (SYN 4)

V podobném duchu bychom mohli pokračovat s výkladem i u dalších výrazů jako *sice, sotva, přece* atd. To ovšem není cílem této části práce. Zde jsme se pouze pokusili pomocí vybraných příkladů upozornit na několik oblastí, kde může vyvstávat potřeba ujasnění syntaktické funkce dvou totožných výrazů, z nichž jeden je tradičně chápán jako příslovce a druhý by mohl být spojkou. Ukázali jsme, což je pro náš další výzkum relevantní, že homonymní výrazy se vyskytují jak v hypotaktickém, tak v parataktickém souvětí. U něho jsme se zvláště věnovali výrazům s původem v příslovcích, jež se vyskytují v důsledkových souvětích a mohou být doprovázeny spojkou *a*. Nakonec jsme se zastavili u výrazů časových, které *sice* naznačují posloupnost dějů v souvětí, ale nemají spojovací funkci.

²⁵ Bauer (1962, s. 30) ale zároveň sám problematizuje slovnědruhové zařazení. Na jednu stranu tvrdí, že výrazy nelze chápat jako příslovečné určení, protože jako slova zájmenné povahy sama o sobě pouze odkazují. Na druhou stranu podle něho proti spojkové platnosti svědčí právě to, že se kombinují se spojkami *a* (*ale, však*). V takovém spojení jako příslovečný výraz modifikují vztah vyjádřený danou spojkou. Argumentuje také tím, že k hromadění spojek v dnešní češtině nedochází a také v minulosti, kdy vznikaly složené souřadící spojky typu *avšak, alevšak*, měl druhý komponent povahu příslovce.

2.1.2.1.3 Spojky – částice

Přesuneme-li se ke vztahu spojek a částic, dostáváme se do patrně nejproblematictější oblasti. V předcházejících hraničních případech (relativa/interogativa, příslovce) jsme se vždy opírali o větněčlenskou platnost konkrétních výrazů v daných kontextech. Spojky a částice ale v této charakteristice vykazují shodu. Ani jeden z těchto slovních druhů není do větné struktury jako větný člen zapojen. Jsou považovány za nezákladní, nesyntagmatické a nesamostatné.

Pokud se částice vydělují jako samostatný slovní druh, je to obvykle na základě funkčního kritéria. V úvahu se bere to, že uvozují věty a udávají jejich modalitu (např. náladový ráz) – (srov. Trávníček 1949, s. 181 nebo ČŘaJ 2000, s. 90). Zároveň jsou ale často homonymní s jinými slovními druhy a z tohoto důvodu je při jejich identifikaci nezbytné vycházet z kontextu.

Příkladů homonymie bychom skutečně mohli uvést celou řadu, pro ukázkou vybereme ale pouze dva. První je výraz *vždyť* ze skupiny apelových částic²⁶. Kromě toho, že funguje jako částice (19), může také spojovat věty v parataktickém souvětí s determinací a významem příčinným (20).

(19) *Já nic nechápu, strýčku Marcusi. **Vždyť** (→přece) jsem nic špatného neudělal.*
(SYN 4)

(20) *Výhru jsme nutně potřebovali, **vždyť** druhá polovina tabulky je pořádně našlapaná.*
(SYN 4)

V hypotaktickém souvětí je častým spojovacím prostředkem v časových větách výraz *když* (21). Jako částice vyjadřuje zklamání, lítost nebo povzdech (22).

(21) ***Když** já byl kvestorem, připadalo mi, že mí předchůdci na hlavní cesty snad úplně zapomněli.* (SYN 4)

(22) ***Když** já nikdy nepoznám, kdy si ze mě utahuješ.* (SYN 4)

V příkladech (19) a (22) je patrná hlavní úloha částic, kterou je naplnění aktualizační funkce. Tím je myšleno to, že částice umožňují mluvčím záměrně zvýraznit některé jazykové prostředky (srov. MČ2 1986, s. 228). O této funkci částic není pochyb, mezi lingvisty je ale předmětem diskuse otázka toho, zda plní sekundárně i funkci spojovací. Zastáncem názoru, že částice by mohly mít i spojovací funkci, je Bauer (1964, s. 132n). Svou domněnku ukazuje na částicích uvozujících nepřímé otázky (*-li, zda, zdali, jestli*) a nepřímé práci a žádací věty

²⁶ Srov. klasifikace částic podle MČ2 (1986, s. 230n.).

(*aby, ať*) (24). V těchto souvětích podle něho částice nemají citové zbarvení (na rozdíl od užití typu věty 23), ale vyjadřují modální a obsahový význam mezi větami (např. signalizují, že jde o tázací význam).

(23) *Ať je zatracený a jde k čertu!* (SYN 4)

(24) *On mi vždycky říkal, ať skáču vpravo.* (SYN 4)

Bauerovu teorii dále rozpracoval Svoboda (1972, s. 64–66). V jeho pojetí mají spojovací funkci kromě spojek i některé částice. Zatímco částice kromě syntaktické závislosti signalizují jen modální význam vedlejší věty, spojky kromě syntaktické závislosti a modálního významu signalizují druh okolnostního významu (typicky u vět příslovečných). Na Svobodu navazuje v tomto třídění také Štěpán (2013, s. 40n.) a mezi částice tak počítá i výrazy, které se tradičně pojímají jako spojky (např. *že, aby, ať, necht*).

Jak už jsme několikrát naznačili výše, v předkládané disertační práci nehraje slovnědruhovému třídění spojovacích prostředků hlavní roli. Z diskusí o spojovací funkci částic však využijeme některé myšlenky, které formulovali Svoboda a Štěpán v souvislosti s vyjadřováním okolnostního významu při spojování vět v souvětí.

2.1.2.1.4 Spojky – předložky

Stejně jako částice patří i předložky k neohebným slovním druhům, nezákladním a nesamostatným, což je sblíží se spojkami. Avšak vzhledem k tomu, že se v předkládané práci zabýváme spojovacími prostředky, které navozují také mezivětné relace, není pro nás diskuse o podobnostech mezi spojkami a předložkami tak zásadní, jako tomu bylo u relativ, příslovcí a částic v předchozích třech oddílech. Předložky totiž vstupují pouze do vztahů mezi slovy, nikoliv mezi větami. Také co do vyjadřování sémanticko-syntaktických vztahů je funkce předložek oproti spojkám omezená. Zatímco u spojek přichází v úvahu jak vztahy koordinace, tak determinace, částečně i apozice, u předložek pouze vztahy determinační a okrajově koordinační.²⁷

V kontextu předkládané práce ale stojí za pozornost případy, kdy sekundární předložky tvoří spolu se spojkou několikaslovné výrazy. Zde vyvstává otázka, jestli jde o spojení samostatných jednotek s uvedenou slovnědruhovou platností, anebo o nový lexém s platností

²⁷ Resp. koordinačně sociativní vztah dvou výrazů označujících živé bytosti, např. *Před půlnocí mě Anna s Olinou vytáhly na panáka.* (SYN 4)

spojky. Hošnová (2005, s. 30) za nové lexémy považuje kombinace výrazů, u kterých jednak mizí (primární) lexikální platnost báze výrazu a zároveň je není možné od sebe oddělit. Sekundární spojku tak například vidí ve spojení *vzhledem k tomu, že* – srov. (25) a (26).²⁸

(25) *Bylo nutno nakonec dodávku proudu přerušit vzhledem k tomu, že mnozí občané nedodržovali ustanovení k odběru.*

(26) *→ *Vzhledem k tomu bylo nutno nakonec dodávku proudu přerušit, že mnozí občané nedodržovali ustanovení k odběru.*

2.1.2.2 Syntaktické spojovací prostředky a formálně syntaktické vztahy

Předchozí oddíl upozornil na komplikovanou problematiku rozhodování o slovnědruhovém platnosti spojovacích prostředků. V předkládané práci se o toto kritérium třídění primárně opírat nebudeme. Jednak se tak vyhneme rozhodování o zařazení každého jednotlivého výrazu, jednak, což je hlavním důvodem, vycházíme z předpokladu, že nerodilí mluvčí češtiny při osvojování a užívání spojovacích výrazů nemusí nutně reflektovat slovnědruhovému zařazení, ale měli by si být vědomi právě jejich spojovací funkce a toho, jaký vztah mezi větami mohou signalizovat.²⁹

Tím se dostáváme k další obecně přijímané charakteristice spojek, a sice že se kromě syntaktické funkce spojovací vyznačují také schopností vyjadřovat vztahy mezi větnými členy a mezi větami. Domníváme se, že charakter myšlenkově obsahových vztahů, jež konkrétní výraz vyjadřuje, ovlivňuje jeho pozici v rámci posloupnosti osvojování spojovacích prostředků nerodilými mluvčími. SERRJ (2001, s. 128) předpokládá nejprve ovládnutí lineární (parataktické) stavby souvětí a následně složitějších víceúrovňových (hypotaktických) struktur. Z toho důvodu v navrženém třídění zohledňujeme druh formálně syntaktického vyjádření vztahů mezi spojovanými větami. Jde o vztahy parataxe a hypotaxe, na jejichž základě můžeme vymezit syntaktické spojovací prostředky **sořadící**

²⁸ Příklady převzaty, viz Hošnová (2005, s. 30).

²⁹ Srov. např. formulace deskriptorů v SERRJ. Do popředí staví formálně syntaktický charakter spojovacích prostředků, které uživatel jazyka na dané úrovni užívá. Například na úrovni A1 se hovoří o lineárních spojovacích výrazech, tj. parataktických spojeních: „[Uživatel jazyka na úrovni A1] Dokáže propojit slova nebo skupiny slov pomocí nejzákladnějších lineárních spojovacích výrazů, jako jsou „a“ („and“) a „pak/potom“ („then“).“ (SERRJ 2006, s. 128) Na úrovni B2 lze uvažovat o rozlišování vztahů mezi větami, které signalizují spojovací prostředky hypotaktické. „Dokáže efektivně používat rozmanité spojovací výrazy tak, aby výstižně postihl(a) vztahy mezi myšlenkami.“ (SERRJ 2006, s. 128)

(parataktické – význam řazení a souřadnosti) a **podřadící** (hypotaktické – význam závislosti).³⁰

Otázkám parataxe a hypotaxe byla věnována pozornost v mnoha pracích předních českých lingvistů,³¹ odkazujeme tedy na tomto místě na ně a zmíníme pouze několik hlavních bodů, ke kterým budeme dále odkazovat. Co se týče fungování spojovacích prostředků ve větné struktuře, budeme sledovat, zda fungují pouze při mezivětném spojování nebo zda mohou spojit i větné členy.

„Parataktické (souřadící) spojky ve své velké většině slouží jak spojování vět, tak i spojování jednotlivých výrazů, popř. i připojování samostatnějších větných částí (různých dodatečných vyjádření apod.). Vnitrovětně lze sice v zásadě užít všech parataktických spojek, v praxi se však všechny neuplatňují. Naopak jsou některé parataktické spojky (a spojkové výrazy) omezeny víceméně jen na užívání vnitrovětné v několikanásobných větných členech: *jakož i, jak – tak (i), i – i*. Hypotaktické (podřadící) spojky v zásadě spojují pouze věty, s výjimkou dvou případů: jednak jde o srovnávací (komparační) spojky – *jako, než* [(27) a (28)] a další, které spojují jak věty, tak i jednotlivé výrazy, jednak o spojky a spojkové výrazy přípustkové (*ačkoliv/i když/přestože/třebaže*) a příčinnou spojku *protože*, které vedle hypotaktického spojování vět vyjadřují koordinační vztahy vnitrovětné.“ (MČ2 1986, s. 217n.)

(27) *Čechům je mnohem bližší mobil než internet.* (SYN 4)

(28) *Chceme mít více obchodních míst, než máme dnes, ale předpokládáme, že počet našich vlastních poboček bude klesat.* (SYN 4)

Vzácněji se objevuje funkční homonymie, tedy to, že totožné výrazy vyjadřují jednou hypotaktický a jednou parataktický vztah. Kupříkladu výraz *aniž* v koordinačním spojení parataktickém (knižně – 29) a v přípustkových větách (30).

(29) *Kovářka Viktorce přikázala, aby to od sebe nedávala, aniž komu ukazovala (→a ani to nikomu neukazovala).* (Němcová, cit. podle SSJČ)

(30) *Obrátila se a vrátila se k vozu, aniž by vkročila na písek.* (SYN 4)

³⁰ K rozlišení vztahů sémantických od jejich formálního vyjádření prostřednictvím parataxe a hypotaxe viz např. Hošnová (2005, s. 41–48). Štěpán v souvislosti s těmito vztahy definuje obecnější rys třídění, a sice tzv. začleněnost a nezačleněnost vět, které jsou jimi připojeny. „Začleněnost je v poměru k determinaci, popř. přirovnání, a k apozici, nezačleněnost v poměru ke koordinaci a parentezi, přičemž pojmy koordinace a apozice překračují již sémanticko-syntaktické vztahy do vztahů referenčních (denotačních):...“ (Štěpán 2013, s. 19)

³¹ Např. Kopečný (1958), Bauer – Grepl (1972), Daneš (1985), Svoboda (1972), Grepl – Karlík (1986, 1998), MČ2 (1986), MČ 3 (1987), ČRaJ (2000).

2.1.2.2.1 Syntaktické spojovací prostředky souřadící

Jak jsme naznačili výše, souřadící spojovací prostředky v souvětí nejsou ze syntaktického hlediska součástí žádné ze spojených vět, protože vyjadřují vztah mezi souvětými členy rovnocennými (na lineární úrovni). Stojí mezi spojovanými větami a až na výjimky (*však*, *totiž* a někdy *ale*) je klademe hned za čárku (pokud jsou jí věty odděleny).³² Signalizují „jistý významový vztah následující syntaktické jednotky k jednotce předcházející“ (Bauer 1967, s. 30), o němž se tradičně hovoří jako o poměru mezi hlavními větami.³³

Podle poměru, který naznačují, se v gramatikách vydělují souřadící spojovací prostředky (srov. MČ2 1986, s. 225n.):

(a) slučovací³⁴

jednoslovné: *a, i, ani, nebo...*

víceslovné: *jak – tak, at' – at'...*

(b) odporovací

jednoslovné: *ale, avšak, však, leč...*

víceslovné: *ovšem – ale, sice – ale...*

(c) stupňovací

jednoslovné: *ba, nejen, dokonce...*

víceslovné: *ba dokonce, nejen – ale i*

(d) vylučovací

jednoslovné: *nebo, anebo, či...*

víceslovné: *bud' – (a)nebo, at' – či...*

(e) příčinné a důvodové – vysvětlovací

jednoslovné: *neboť, vždyť, totiž...*

víceslovné: *však také...*

³² Opakované a dvojité spojovací výrazy (viz níže) se kladou vždy na začátek příslušných vět.

³³ Tyto poměry však vyjadřují i mezi souřadně spojenými větnými členy nebo mezi vedlejšími větami ve složeném souvětí.

³⁴ V naší práci nezahrnujeme mezi spojovací prostředky interpunkci, resp. čárku mezi větami. Je to především z toho důvodu, že její užití je v textech nerodilých mluvčích do značné míry záležitostí pravopisnou, a nikoliv projevem naznačení typu syntaktického vztahu mezi spojovanými větami.

(f) důsledkový³⁵

jednoslovné: *proto, tedy, tudíž, tak...*

víceslovné: *a proto, a tedy, a tak...*

Některé ze souřadících spojovacích prostředků mohou být vícefunkční a uplatňují se při vyjadřování různých vztahů (např. *a*). Pro každý z uvedených vztahů však vždy existuje skupina spojovacích prostředků základních (typických, frekventovaných), další konkurenční prostředky k nim pak přistupují obvykle v různých modifikacích (např. jako víceslovné spojovací prostředky).

Dodejme na okraj, že se souřadící spojky také často objevují i na začátku samostatných vět nebo souvětí (31).

(31) *Nedokážeme dělat zázraky, mladíku. **Ale** udělali jsme všechno, co je v našich silách.*

(SYN 4)

Vyjadřují zde vztah k předcházejícímu kontextu a mají funkci navazovací. I zde jde ve své podstatě o spojovací funkci, i když v rámci útvarů nadvětých.³⁶

2.1.2.2.2 Syntaktické spojovací prostředky podřadící

Na rozdíl od souřadících spojovacích prostředků se syntaktické spojovací prostředky užívají jen v závislých větách a tyto věty typicky začleňují do nějaké syntaktické struktury jako větný člen. Kladou se na začátek závislé věty. Ta ovšem může předcházet větě řídicí, takže spojovací výraz nemusí stát mezi větami, ale může zaujímat pozici na začátku souvětí.

Mezi lingvisty se vedou diskuse o tom, jaký je podíl hypotaktických spojovacích prostředků na konstituování významového vztahu řídicí a závislé věty.³⁷ Ve funkčně pojatých gramatikách se hypotaktické spojky běžně klasifikují podle toho, jakou okolnost platnosti

³⁵ Štěpán (2013) v souvislosti se sémanticko-syntaktickými vztahy definuje věty v poměru příčinném a důsledkovém jako začleněné, zatímco nezačleněné jsou typicky věty v poměru slučovacím a odporovacím.

³⁶ Hrbáček (1994, s. 56) v této souvislosti odlišuje spojovací prostředky (junktory) a připojovací prostředky (konektory). „... spojovací prostředky vyjadřují sémantické vztahy mezi výpověďmi uvnitř výpovědního celku (uvnitř výpovědního spojení), kdežto připojovací prostředky vyjadřují sémantické vztahy mezi výpověďními celky („přes tečku“) nebo mezi výpovědním celkem a větším úsekem textu...“

³⁷ Emblematickými se v tomto názorovém nesouladu staly spojky *že, aby, zda* aj., tj. spojky tradičně označované jako obsahové. Hrbáček (1967, s. 139) tvrdí: „Podřadící spojky jsou specializovány buď na vyjadřování vztahu syntaktické závislosti (*že*), nebo naopak vyjadřují tuto závislost jen nepřímo tím, že vyjadřují určité druhy významového vztahu mezi větami, např. příčinný (*protože, poněvadž, jelikož*), časový (*když*) atd.“ Proti tomuto názoru se staví Bauer (1967, s. 30n.), který tvrdí, že „i věty se spojkou *že* mají jistý společný význam, který nelze oddělovat od spojky – je spojkou *že* vyjadřován...“. Pokud by tomu tak nebylo, nemohla by spojka *že* uvozovat kromě vět obsahových také věty příslovečné. MČ2 (1986, 216) hovoří o míře významové specifikovanosti, která je v případě spojky *že* nulová.

řídící věty vyjadřuje věta závislá s daným spojovacím prostředkem. Např. v MČ2 je dělení provedeno na spojky:

(a) obsahové

jednoslovné: *že, aby, zda(-li), -li, jestli, ať...*

(b) časové

jednoslovné: *když, až, dokud, jakmile, než(li)...*

víceslovné: *hned když, hned jak, ještě ne – a už*

korelativa: *tak dlouho, dokud; tehdy, když...*

(c) příčinné

jednoslovné: *protože, poněvadž, jelikož...*

víceslovné: *zvlášť když...*

korelativa: *proto, že; vzhledem k tomu, že...*

(d) účelové

jednoslovné: *aby, ať, zda...*

víceslovné: *jen aby, ...*

korelativa: *proto – aby...*

(e) podmínkové

jednoslovné: *jestliže, kdyby, jestli, pokud...*

víceslovné: *kdyby tak, když by...*

korelativa: *pokud – pak...*

(f) přípustkové

jednoslovné: *ačkoliv, třebaže, přestože, ač...*

víceslovné: *třeba i, byť jen...*

(g) způsobové

jednoslovné: *jak (relativum)*

korelativa: *tak, že; tím, že...*

(h) srovnávací

jednoslovné: *jako, jak, jakoby, než...*

víceslovné: *jako když, jako že, jinak než...*

korelativa: *stejně jako – tak; více – než; čím – tím...*

(i) účinkové

jednoslovné: *takže, až, divže...*

víceslovné: *než aby...*

korelativa: *tak, že; tak až...*

(j) výjimečné

jednoslovné: *leda(že), leda...*

víceslovné: *leda by, leda když...*

korelativa: *až na to, že...*

U pořadických spojovacích výrazů jsme upozornili na jejich možnou polyfunkčnost. Stejnou vlastnost vykazují i některé výrazy pořadicích. Například *když* ve větách časových (32), příčinných/důvodových (33) i podmínkových (34).

(32) *Málem jsem zakopla, **když** mi Jerry začal povídat o té šťávě z řepy.* (SYN 4)

(33) *Dívka se zdráhala, ale když jí Štefan Bošňák vysvětlil, že ji chce obdarovat, **když** (→protože) doma trou bídu s nouzí, souhlasila a vykročila po jeho boku na faru.* (SYN 4)

(34) ***Když** (→pokud) to nejde jinak, tak zloději auto odtáhnou.* (SYN 4)

2.1.3 Spojovací prostředky podle podoby a struktury

V příkladových výčtech, které jsme uvedli výše, je možno postřehnout, že vedle jednoslovných existují i víceslovné spojovací prostředky. Jeden spojovací prostředek někdy představuje i kombinaci několika slov. Tím se konkretizuje a specifikuje jeho sémantika. Víceslovné spojovací prostředky – typicky synsémantické – díky tomu mohou vyjádřit významové rysy, které jednoslovné vyjádřit nedokážou, srov. (35) a (36).

(35) ***Teprve až když** máte děti, musíte přestat brát ohledy sami na sebe a začít se zajímat taky o život někoho jiného.* (SYN 4)

(36) ***Když** máte děti, musíte přestat brát ohledy sami na sebe a začít se zajímat taky o život někoho jiného.*

Podle MČ2 (1987, s. 216n.) se tak děje dvěma způsoby:

(a) Jednoduché spojky paratactické se zmnožují a vznikají tak spojovací prostředky vícedílné, které nemají navzájem kontaktní pozici. Například spojení *bud' – (a)nebo* má oproti samotnému výrazu *nebo* zdůrazněný vylučovací význam.

(b) Spojky se sdružují s jinými spojkami nebo výrazy dalších slovních druhů (určité druhy příslovcí a částic, případně předložek) a vznikají spojovací prostředky, v nichž stojí všechny členy v kontaktní pozici. U spojovacích prostředků tohoto typu se často vede debata o tom, do jaké míry ve funkci spojovací zdomácněly a jaká je míra jejich konjunkcionalizace (srov. Hošnová 2005, s. 36n.).

Postup tvoření spojek popsáný v bodě (b) je stále progresivní. „Někdy takto nastává i jakési „přerozdělování“ významu. Např. k méně specifikované spojce *jak* se přidružuje adverbium *hned* a vzniká spojkový výraz *hned jak*, který se významem přibližuje speciální časové spojce *jakmile* a stává se jednotkou její synonymní řady. Takovýto pohyb, přibývání synonymních spojovacích prostředků, se někdy děje i přes hranice parataxe a hypotaxe. Srov. paratactický přípustkově odporovací vícedílný spojkový výraz *sice – ale přesto* a hypotactickou spojku přípustkovou *přestože* (vzniklou z korelativa *přesto, že...*).“ (MČ2 1986, s. 217)

Pro účely této práce třídíme spojovací prostředky na jednoslovné (např. *a, že, když*), či víceslovné (např. *a proto, i když, ani – ani*). Víceslovné podrobněji klasifikujeme podle postavení výrazů na víceslovné kontaktní (např. *a proto, i když*) a víceslovné distantní (např. *ani – ani, nejen – ale i*). U víceslovných kontaktních spojovacích prostředků můžeme rozlišit, zda se skládají ze dvou spojek³⁸ (např. *i když*) – kombinovaný spojovací prostředek, nebo ze spojky a jiného slovního druhu (např. *a proto*) – složený spojovací prostředek. Srov. schéma 1.³⁹

³⁸ Případně ze dvou částic, např. *ba dokonce*.

³⁹ V současných českých lingvistických pracích narazíme při pohledu na třídění spojek, resp. spojkových výrazů podle počtu slov, z nichž sestávají, na různé klasifikace. I když se v nich řada pojmů (např. „spojovací výraz“, „spojka dvojitá“ apod.) opakuje, často se jimi označují výrazy odlišné. Hlavní příčina této nejednotnosti tkví v tom, zda za spojku považují pouze jednoslovnou jednotku, nebo i jednotky víceslovné. Podrobněji k této problematice srov. Doležalová (2008, s. 12–17).

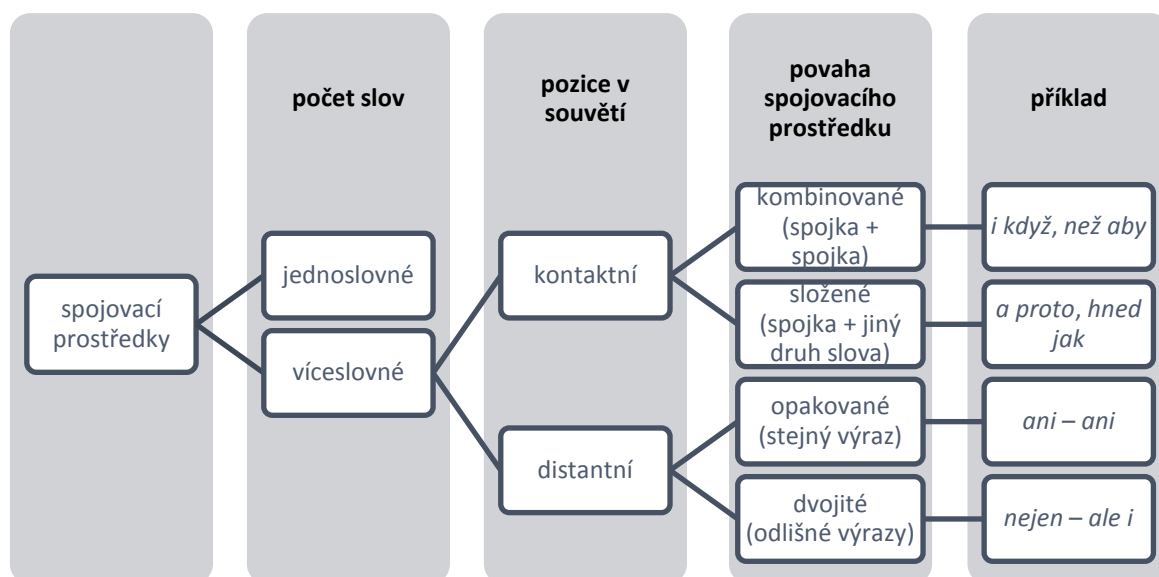


Schéma 1: Třídění spojovacích prostředků podle počtu slov

Ve zvláštních případech se mohou dvě spojky vyskytnout v kontaktní pozici, aniž by tvořily jeden spojovací prostředek. Je to jednak u koordinačních spojení závislých vět (37) a také tam, kde se různovětné spojky dostaly vedle sebe následkem změny pořadí vět (38) a (39).

(37) *Myslím si, že je to věc trhu a že časem se ceny nájmu ustálí.* (SYN 4)

(38) *Řekli mi, že když se něco stane, mám sebou praštit na zem a nasadit si přilbu.*
(SYN 4)

(39) *Chlapci však cestou usnuli, a když se probudili, byli bez peněz.* (SYN 4)

Dodejme ještě, že v některých koncepcích (srov. Štěpán 2013, s. 18) můžeme pozorovat náznak pojetí spojovacích prostředků v širším smyslu. Za spojovací prostředek se v něm počítají i spojení několika lexémů skládající se „z podřadicích spojovacích výrazů a zájmenných t-výrazů [...] zapojených do řídicí věty bezpředložkovým pádem zájmena, multiverbizačními jednotkami a předložkami, např. *tím, že; v tom smyslu, že; vzhledem k tomu, že*, a z podřadicích spojovacích výrazů a příslovcových a číslovkových t-výrazů, např. *tak, že; tolik, že*.“⁴⁰

Jak je vidno, tato problematika zasahuje do několika oblastí. Jednak jsme se s ní setkali již výše u vztahu spojek a předložek a konjunkcionalizace (srov. kap. 2.1.2.1.4), jednak souvisí

⁴⁰ Právě toto zapojení rozhoduje o příslovecné sémantice vedlejší věty.

s otázkou korelativních spojení řídicí a závislé věty. Druhé jmenované oblasti se věnoval mimo jiné Hrbáček (1964). V jeho pojetí je t-výraz v rámci korelativního spojení nazýván jako odkazovací. Formálně signalizuje přítomnost závislé věty a naznačuje druh této závislosti.⁴¹ T-výraz sám nemá lexikální význam, a není proto ve větě řídicí větným členem. Umožňuje ale „pevněji“ a „pregnantněji“ vyjádřit syntaktický vztah nebo i obsahovou souvislost vět, než kdyby závislou větu připojoval pouze spojovací prostředek uvozující závislou větu. Někdy signalizuje vazbu predikátu a může být i nutným doplněním větné struktury.⁴²

2.1.4 Shrnutí

V této kapitole jsme přiblížili různé přístupy ke třídění spojovacích prostředků, resp. typů souvětí. Zmínili jsme větněčlenský přístup, založený na pojetí vedlejší věty jako rozvinuté realizace příslušného větného členu, valenční teorii meziprozodických vztahů akcentující významové vztahy mezi větami a nakonec také nevalenční koncepci založenou na formálně výrazovém kritériu třídění. Poslední jmenovaný koncept se nejvíce promítá do našeho výzkumu, vychází totiž od formy povrchové realizace spojovacích vztahů, tj. od spojovacích prostředků samotných, a tak se dá vhodně aplikovat při korpusové analýze, kterou budeme provádět.

Představili jsme dále základní dělení spojovacích prostředků, které při našem výzkumu uplatníme. V souladu s přístupem, k němuž se přikláníme, se přidržíme klasifikace podle větněčlenské povahy a budeme vydělovat prostředky autosémantické (s větněčlenskou platností) a synsémantické (bez větněčlenské platnosti). U autosémantických prostředků navíc zohledňujeme jejich ohebnost, což má své nesporné výhody pro zamýšlenou chybovou analýzu, neboť se tak jasně vymezí skupina prostředků, které mohou být postiženy morfologickými chybami. Do předkládané práce jsme se rozhodli zavést také třídění podle formálně syntaktických vztahů, tj. na spojovací prostředky parataktické a hypotaktické, které jsme v této kapitole blíže popsali. Činíme tak na základě předpokladu, že se vývoj

⁴¹ T-výraz může přímo signalizovat i druh následující závislé věty. Typické je to pro věty příslovečné, např. (*tak, že; proto, že*).

⁴² U vět nepříslovečných „signalizuje vazbu predikátu, valenci (*ptal se na to, kdo; vyprávěl o tom, že*), nebo příslovečných, u kterých signalizuje nevazebné spojení (*nepracoval od té doby, co; udělal to díky tomu, že; napsal to tak dobře, že; vyšel těsně před tím, než*). Užití t-výrazů je závazné u predikátů se vztahovým významem (*tkví v tom, že; závisí na tom, zda*) a u syžetových predikátů *jde/jedná se o to, že/aby*.“ (Štěpán 2013, s. 23)

jazykové kompetence projeví v linearitě, resp. hierarchičnosti souvětných struktur, které nerodilí mluvčí produkují. A tu lze právě díky této klasifikaci postihnout.

V našem výzkumu sice primárně nebudeme pracovat se slovnědruhovým tříděním spojovacích prostředků, přesto jsme v této kapitole nastínili, jaké komplikace s sebou přináší stanovení hranice mezi zájmeny, příslovci, spojkami a částicemi. Tyto teoretické úvahy jsme využili k identifikaci problematických jevů, které se mohou předkládané práce dotýkat. Narazili jsme zaprvé na nutnost zohlednění polyfunkčnosti, resp. funkční homonymie některých prostředků (např. *co, jen, třeba*), jež mohou ale i nemusejí plnit spojovací funkci. Zadruhé jsme pojednali o případech, kdy se totožné výrazy vyskytují jednou v platnosti autosémantických spojovacích prostředků a jindy jako prostředky sysémantické (např. *jak*).

Dále jsme podali přehled tradičního třídění sysémantických spojovacích prostředků podle poměrů mezi větami a podle okolnostních vztahů, které vyjadřují. Toto třídění budeme využívat pouze okrajově v kvalitativních oddílech výzkumu. O něco podrobněji jsme naopak rozpracovali třídění spojovacích prostředků podle podoby, resp. počtu slov, protože tento formální rys hraje důležitou roli pro výběr adekvátního postupu vyhledávání v korpusech. Spojovací prostředky dělíme na jednoslovné a víceslovné, víceslovné mohou stát buď v pozici kontaktní (*a proto*), nebo distantní (*bud' – anebo*).

2.2 PÍSEMNÁ PRODUKCE NERODILÝCH MLUVČÍCH A SPOJOVACÍ PROSTŘEDKY

V předchozí kapitole jsme pojednali o spojovacích prostředcích v češtině prostřednictvím přístupů současných českých lingvistických prací. Dalo by se tedy říci, že jsme je prezentovali pohledem rodilého mluvčího češtiny. Klasifikaci i problematické oblasti vymezení jednotlivých podskupin, které jsme v uvedené kapitole naznačili, využijeme zejména při zadávání vyhledávacích dotazů v korpusu (viz kap. 3.2) a při třídění a ověřování vyhledaných dokladů. Naše analýza ale bude dále spočívat ve zmapování repertoáru a užití spojovacích prostředků nerodilými mluvčími na jednotlivých úrovních osvojení si češtiny. Z toho důvodu pokládáme za důležité představit charakteristiku úrovní s ohledem na dovednost psaní (konkrétně užití spojovacích prostředků) podle SERRJ (kap. 2.3.1) a také podle popisů referenčních úrovní v češtině (2.3.1).

2.2.1 Spojovací prostředky a úrovně SERRJ

Společný evropský referenční rámec pro jazyky – jak se učíme jazykům, jak je vyučujeme a jak v jazycích hodnotíme (SERRJ) je všeobecně známým dokumentem, a proto není potřeba ho blíže představovat. Od svého publikování v roce 2001 se totiž stal nejdůležitějším nástrojem pro plánování a realizaci výuky a pro jazykovou certifikaci v podobě standardizovaných testů.

Vývoj jazykové znalosti uživatele cizího jazyka se podle SERRJ dělí do 6 úrovní (od nejnižší po nejvyšší – A1, A2, B1, B2, C1 a C1). Jazykovou kompetenci v různých dovednostech popisují tzv. deskriptory sdružené do stupnic/škál zachycujících vývoj od nejnižšího stupně až po nejvyšší. Tyto škály mají být obecně použitelné pro všechny evropské jazyky, a proto bylo jejich znění formulováno na obecném základě (srov. tabulka 1).

Tabulka 1: Společné referenční úrovně: globálně pojatá stupnice (SERRJ 2001, s. 24)

	úroveň	Klíčové pojmy ⁴³	Znění deskriptoru
Zkušený uživatel	C2 Zvládnutí	<i>rozumí téměř všemu shrne informace, polemizuje – spontánně, přesně, plynule složité situace</i>	Snadno rozumí téměř všemu , co si vyslechne nebo přečte. Dokáže shrnout informace z různých mluvených a psaných zdrojů a přitom dokáže přednést polemiku a vysvětlení v logicky uspořádané podobě. Dokáže se spontánně , velmi plynule a přesně vyjadřovat a rozlišovat jemné významové odstíny dokonce i ve složitějších situacích.
	C1 Účinná operační způsobilost	<i>náročné, dlouhé texty implicitní významy efektivní komunikace složité témata produkuje kohezí texty</i>	Rozumí širokému rejstříku náročných a dlouhých textů a rozpozná implicitní významy textů. Umí se plynule a pohotově vyjadřovat bez zjevného hledání výrazů. Umí jazyka užívat pružně a efektivně pro společenské, akademické a profesní účely. Umí vytvořit srozumitelné, dobře uspořádané , podrobné texty na složité témata , čímž prokazuje ovládnutí kompozičních útvarů, spojovacích výrazů a prostředků koheze .
Samostatný uživatel	B2 Rozhled	<i>hlavní myšlenky – konkrétní i abstraktní témata (+ svůj obor) běžný rozhovor podrobný text vysvětlí, argumentuje</i>	Dokáže porozumět hlavním myšlenkám složitých textů týkajících se jak konkrétních, tak abstraktních témat včetně odborně zaměřených diskusí ve svém oboru . Dokáže se účastnit rozhovoru natolik plynule a spontánně, že může vést běžný rozhovor s rodilými mluvčími, aniž by to představovalo zvýšené úsilí pro kteréhokoliv účastníka interakce. Umí napsat srozumitelné podrobné texty na širokou škálu témat a vysvětlit své názorové stanovisko týkající se aktuálního problému s uvedením výhod a nevýhod různých možností.
	B1 Práh	<i>hlavní myšlenky běžná témata + cestování jednoduchý souvislý text popis zážitků, událostí vysvětlí, odůvodní názory</i>	Rozumí hlavním myšlenkám srozumitelné spisovné vstupní informace (input) týkající se běžných témat , se kterými se pravidelně setkává v práci, ve škole, ve volném čase atd. Umí si poradit s většinou situací, jež mohou nastat při cestování v oblasti, kde se tímto jazykem mluví. Umí napsat jednoduchý souvislý text na témata, která dobře zná nebo která ho/ji osobně zajímají. Dokáže popsat své zážitky a události, sny, naděje a cíle a umí stručně vysvětlit a odůvodnit své názory a plány .
Uživatel základů jazyka	A2 Na cestě	<i>věty + časté výrazy potřeby, známé skutečnosti „já + rodina + nakupování + místopis + zaměstnání“ popis rodiny + okolí</i>	Rozumí větam a často používaným výrazům vztahujícím se k oblastem, které se ho/jí bezprostředně týkají (např. základní informace o něm/ní a jeho/její rodině, o nakupování, místopisu a zaměstnání). Dokáže komunikovat prostřednictvím jednoduchých a běžných úloh, jež vyžadují jednoduchou a přímou výměnu informací o známých a běžných skutečnostech. Umí jednoduchým způsobem popsat svou vlastní rodinu, bezprostřední okolí a záležitosti týkající se jeho/jejích nejnaléhavějších potřeb .
	A1 Průlom	<i>výrazy + zákl. fráze každodenní, konkrétní potřeby osobní otázky + odpovědi „já + moje“</i>	Rozumí známým každodenním výrazům a zcela základním frázím , jejichž cílem je vyhovět konkrétním potřebám , a umí tyto výrazy a fráze používat. Umí představit sebe a ostatní a klást jednoduché otázky týkající se informací osobního rázu , např. o místě, kde žije, o lidech, které zná, a věcech, které vlastní , a na podobné otázky umí odpovídat . Dokáže se jednoduchým způsobem domluvit, mluví-li partner pomalu a jasně a je ochoten mu/jí pomoci .

⁴³ Do tabulky jsme oproti originálu v SERRJ doplnili sloupec s klíčovými pojmy, kde uvádíme hlavní charakteristiky v oblasti situační, tematické i vyjadřovací, jimiž se od sebe jednotlivé úrovně odlišují.

Každá úroveň je definována průnikem charakteristik obecné a jazykové kompetence z několika úhlů pohledu. Zohledňují se společenské funkce (soukromá, veřejná, profesní atd.), tematické okruhy komunikačních situací, jazykové funkce, rozsah a komplexnost produkovaných struktur apod. Kombinací těchto charakteristik se od sebe také jednotlivé úrovně odlišují. Zatímco na úrovni A1 je situační sféra omezena na osobní a nejnnutnější potřeby, směrem k vyšším úrovním se rozšiřuje o běžné každodenní společenské situace (představení se, zdvořilostní výměny informací apod.) a interakce v nich (A2), o zvládání situací každodenního života, např. při cestování (B1), o sféru studijní/profesní (B2) až po všeobecnou zběhlost v různých i ne příliš častých a obvyklých situacích (C1 a C2).

Společné referenční úrovně lze prezentovat mnoha různými způsoby a s různou mírou detailního propracování. Náš výzkum se věnuje psaným textům, a proto dále upozorňujeme na místa, resp. stupnice, o něž se můžeme při analýzách opírat. První z nich představuje výňatek z tabulky sebehodnocení v oblasti písemného projevu (tabulka 2). Následuje všeobecná stupnice k písemnému projevu (tabulka 3) a stupnice věnovaná koherenci (tabulka 4).

Tabulka 2: Společné referenční úrovně: sebehodnocení (SERRJ 2001, s. 26n.)

PSANÍ – PÍSEMNÝ PROJEV	
C2	Umím napsat jasný plynulý text vhodným stylem . Dokážu napsat složité dopisy, zprávy nebo články a vystavět text logicky tak, aby pomáhal čtenáři všimnout si důležitých bodů a zapamatovat si je. Umím psát resumé a recenze odborných nebo literárních prací.
C1	Umím se jasně vyjádřit, dobře uspořádat text a podrobně vysvětlit svá stanoviska. Umím psát podrobné dopisy, pojednání nebo zprávy o složitých tématech a zdůraznit to, co považuji za nejdůležitější . Umím zvolit styl textu podle toho, jakému typu čtenáře je určen.
B2	Umím napsat srozumitelné podrobné texty na širokou škálu témat souvisejících s mými zájmy. Umím napsat pojednání nebo zprávy, předávat informace, obhajovat nebo vyvracet určitý názor . V dopise dovedu zdůraznit, čím jsou pro mě události a zážitky osobně důležité.
B1	Umím napsat jednoduché souvislé texty na témata, která dobře znám nebo která mě osobně zajímají. Umím psát osobní dopisy popisující zážitky a dojmy.
A2	Umím napsat krátké jednoduché poznámky a zprávy týkající se mých základních potřeb. Umím napsat velmi jednoduchý osobní dopis , například poděkování.
A1	Umím napsat stručný jednoduchý text na pohlednici, například pozdrav z dovolené. Umím vyplnit formuláře obsahující osobní údaje, například své jméno, státní příslušnost a adresu při přihlašování v hotelu.

Tabulka 3: SERRJ: Písemný projev – všeobecná stupnice (SERRJ 2001, s. 63)

PÍSEMNÝ PROJEV – VŠEOBECNÁ STUPNICE	
C2	Dokáže napsat jasné a plynulé složité texty , které mají vhodný a účinný styl a jsou logicky vystavěny tak, aby pomáhaly čtenáři najít základní body.
C1	Dokáže napsat jasné, dobře uspořádané texty týkající se složitých témat a zdůraznit důležité otázky. Přitom dokáže za pomoci doplňujících argumentů a vhodných příkladů podrobně rozvést a podepřít svá názorová stanoviska a vhodně zakočit .
B2	Dokáže napsat srozumitelné podrobné texty týkající se různých témat z oblasti jeho/jejich zájmů, přičemž dokáže shrnout, skloubit a zhodnotit informace a argumenty z velkého počtu zdrojů.
B1	Dokáže napsat velmi jednoduše členěné souvislé texty týkající se okruhu známých témat z oblasti jeho/jejího zájmu, a to tak, že spojuje řadu kratších úseků do lineárního sledu .
A2	Dokáže napsat řadu jednoduchých frází a vět spojených jednoduchými spojkami, jako jsou „a“, „ale“ a „protože“ .
A1	Dokáže psát jednoduché izolované fráze a věty .

Tabulka 4: SERRJ: Koherence (SERRJ 2001, s. 128)

KOHERENCE	
C2	Dokáže vytvořit koherentní a kohezní texty a přitom plně a vhodně použít rozmanité slohotvorné útvary a širokou škálu kohezních prostředků .
C1	Dokáže vypracovat věcnou, plynulou a správně strukturovanou řeč a přitom prokazuje ovládnutí slohotvorných útvarů, spojovacích a kohezních prostředků .
B2	Dokáže efektivně používat rozmanité spojovací výrazy tak, aby výstižně postihl(a) vztahy mezi myšlenkami. Dokáže užívat omezené množství kohezních prostředků tak, že umí ze svých výpovědí vytvořit věcnou a koherentní promluvu, ačkoliv v dlouhém textu nedokáže občas postihnout myšlenkovou návaznost.
B1	Dokáže spojit sérii kratších, jednotlivých bodů v propojený a lineární sled určitých myšlenek .
A2	Dokáže používat nejběžněji se vyskytující spojovací výrazy k propojení jednoduchých vět, aby mohl vyprávět příběh nebo něco popsat prostým vyjmenováním určitých myšlenek. Dokáže propojit skupiny slov pomocí jednoduchých spojovacích výrazů , jako jsou „a“ („and“), „ale“ („but“) a „protože“ („because“).
A1	Dokáže propojit slova nebo skupiny slov pomocí nejzákladnějších lineárních spojovacích výrazů , jako jsou „a“ („and“) a „pak/potom“ („then“).

V každé stupnici jsme tučně zvýraznili klíčové pojmy, které mají vztah ke struktuře a uspořádání textu. Z nich můžeme mimo jiné odvozovat, jak komplexní texty/struktury uživatelé na dané úrovni produkují, a usuzovat, jaké spojovací prostředky k tomu lze využít. Například na nejnižší úrovni A1 se víceméně počítá pouze s řazením naučených frází, a tak můžeme očekávat, že repertoár spojovacích prostředků bude velmi omezený. Se stoupající úrovní roste komplexnost očekávaných textů i jejich obsahová a propoziční náročnost. Kupříkladu na úrovni B2 se očekává schopnost argumentovat (obhajovat nebo vyvracet názory). Můžeme si tedy dovodit, že se repertoár spojovacích prostředků na úrovni B2 neobejde bez výrazů, které tyto významové vztahy mohou navozovat, např. parataktické

spojky důvodové nebo hypotaktické spojky ve větách příčinných a přípustkových. Syntézou všech těchto klíčových pojmů z různých částí SERRJ jsme dospěli ke speciální stupnici (tabulka 5) zachycující vztah struktury textu, mezivětných vztahů a spojovacích prostředků na jednotlivých úrovních.

Tabulka 5: Syntaktická kompetence podle SERRJ

	charakter textu	vyjadřované vztahy	užívané spojovací prostředky
C2	složitě rozvětvená struktura s řazením i podřazováním	všechny parataktické i hypotaktické vztahy	i málo frekventované spojovací prostředky
C1	rozvětvená struktura s řazením i podřazováním	všechny parataktické i hypotaktické vztahy	i méně frekventované spojovací prostředky
B2	jednoduše rozvětvená struktura s řazením i podřazováním	parataktické vztahy koordinační i determinační (slučovací, vylučovací, odporovací, důvodový, důsledkový, stupňovací) většina hypotaktických vztahů (obsahové, časové, příčinné, účelové, podmínkové, přípustkové)	frekventované spojovací prostředky
B1	lineární struktura s řazením i podřazováním	parataktické vztahy koordinační a omezeně determinační (slučovací, vylučovací, odporovací, důvodový) některé hypotaktické vztahy (obsahové, časové, příčinné, účelové)	nejfrekventovanější spojovací prostředky
A2	lineární struktura převážně s řazením	parataktické vztahy koordinační (slučovací, vylučovací, odporovací) některé hypotaktické vztahy (obsahové)	nejfrekventovanější spojovací prostředky
A1	izolované fráze a věty	parataktické vztahy, asyndetická spojení	minimum spojovacích prostředků

Tato stupnice „syntaktické kompetence“ je stejně jako ostatní stupnice SERRJ zatížena několika nedostatky. Jednak je velmi obecná a jednak značně spoléhá na intuici. Chybí jí tedy konkretizace pro daný národní jazyk, v našem případě pro češtinu a také ilustrace na autentickém materiálu (srov. kap. 1.1).

2.2.2 Spojovací prostředky a popisy referenčních úrovní

V předchozí podkapitole jsme vytýkali stupnicím SERRJ jistou míru vágnosti a nespecifikovanosti některých deskriptorů. Ta, jak už víme, pramení z toho, že SEERJ představuje obecnou koncepční a metodickou bázi pro všechny evropské jazyky. Proto byl SERRJ v mnoha zemích konkretizován v podobě tzv. popisů referenčních úrovní. Ty by měly výstižněji vymezit příslušnou „úroveň zběhlosti v cizím jazyce, která je dána mírou osvojení

základních jazykových dovedností a související úrovni komunikativní kompetence mluvčího“.⁴⁴

Pro češtinu existují popisy čtyř úrovní (B1, A1, A2 a B2 – pořadí podle vzniku), jejichž struktura je obdobná. Vychází se od obecných charakteristik, postupuje se přes jejich konkretizaci, dále přes výčet jazykových funkcí dané úrovně, přes popis jednotlivých dovedností, kompenzačních strategií a uzavírá se dodatky, které obsahují poznámky k výslovnosti a přehled jazykových prostředků a lexika.

Popisy by měly, jak už bylo řečeno, sloužit ke konkretizaci obsahu SERRJ. Autorské kolektivy si při jejich tvorbě kladly cíl zmapovat češtinu pro dané úrovně z funkčního i sémantického hlediska. A domnívaly se, že budou moci sloužit „k sestavování výukových programů v rámci jazykových kursů češtiny pro cizince, ale též k tvorbě učebních materiálů i dalších pomůcek a v neposlední řadě jako referenční rámec pro potřeby zavedení standardního systému zkoušek“.⁴⁵ (PopisA2 2005, s. 2)

Otázka naplnění těchto ambiciózních záměrů je mezi odborníky předmětem diskuse. Slabinou popisů je bezesporu to, že všechny přehledy vznikaly bez širšího podkladového materiálu v podobě autentických textů nerodilých mluvčích na dané úrovni. Jejich tvorba různými autorskými týmy se také promítla do toho, že popisy sice podávají charakteristiku dané úrovně, ale neakcentují to, co ji odděluje od úrovní ostatních, resp. sousedních.

Předkládaná práce však nemá za cíl popisy kritizovat. Obdobně jako u SERRJ chceme pouze shrnout, co se z nich dozvíme o užívání spojovacích prostředků na jednotlivých úrovních a co nám eventuálně může posloužit jako opora při našem vlastním výzkumu. Přehled relevantních informací shromážděných z různých míst uvedených dokumentů podáváme v následujících oddílech v pokud možno co nejkompaktnější formě.

⁴⁴ Webové stránky MŠMT: <http://www.msmt.cz/mezinarodni-vztahy/referencni-urovne-pro-cestinu-jako-cizi-jazyk>, [cit. 15. 3. 2017]

⁴⁵ Obdobně i v popisu úrovně B2: „Čeština jako cizí jazyk disponuje slušnou řádkou učebnic (hlavně pro začátečníky), ale až dosud chyběl explicitně formulovaný základ, o který by se mohli jednotliví autoři opřít a z něj vycházet. Každý „na vlastní pěst“ řešil problémy východisek, cílů a přístupů a pojetí popisovaného materiálu. Takové otázky se většinou výslovně nekladly a postupovalo se víceméně „tušivě“. Obvyklým důsledkem byla improvizace (jejíž větší či menší úspěšnost závisela na pedagogické zkušenosti autora). Vlastně ani jinak se postupovat nedalo: praxe naléhala a nedávala příliš prostoru na promyšlení strategie. Tato překážka je už tedy odstraněna. Řada A 1, A 2, B 1, B 2 nabízí spolehlivou oporu a navíc zaručuje, že uživatelé nebudou zůstávat izolováni od svých kolegů v jiných evropských zemích, a mohou tak snadněji čelit bludům o „specifice“ a jedinečné nepřístupnosti českého jazyka. Uživatel-autor může kompetentně zvolit, kde na této škále leží jím zpracovaný úsek, a plně se soustředit na jeho didaktickou prezentaci a plně rozvinout svou vlastní vynalézavost.“ (PopisB2 2005, s. III)

2.2.2.1 Úroveň A1

Popis úrovně A1 charakterizuje písemný projev uživatele na dané úrovni takto: „Charakter syntaxe na úrovni A1 je dán požadavky negramatičnosti, redukce jmenných a slovesných tvarů a limitované slovní zásoby. Ústní i písemný projev studenta na úrovni A1 bude složen zejména z jednoduše strukturovaných vět a nevětných výpovědí. Výraznou menšinu v něm budou tvořit souvětí (většinou souřadná), v nichž vztahy mezi větami budou vymezeny malým počtem spojovacích výrazů s pokud možno jednoznačným významem.“ (PopisA1 2005, s. 258)

Tomu odpovídá i přehled spojovacích prostředků, které by měl uživatel aktivně ovládat, resp. měl by umět vyjádřit významové vztahy jimi reprezentované (tabulka 6).

Tabulka 6: Spojovací prostředky očekávané na úrovni A1 (srov. PopisA1 2005, s. 262n.)

spojovací prostředky parataktické = celkem 5	
<i>a</i>	Slučovací spojka ve významu „plus“
<i>ani</i>	Slučovací spojka užívaná s větnou negací
<i>ale</i>	Odporovací spojka kontrastující protikladné jevy, vlastnosti a děje
<i>nebo</i>	Slučovací (slabě disjunktivní) spojka s významem blížícím se „plus“. Vylučovací (silně disjunktivní) spojka, volba jen jedné z více alternativ
<i>a tak</i>	Spojka pojící příčinu s důsledkem. Spojuje jen věty. Spojované věty jsou si rovnocenné pouze formálně
spojovací prostředky hypotaktické	
autosémantické	Uvozují obsahové věty.
vztažná zájmena a příslovce	Přepokládá se aktivní užívání pouze po slovesech
syntémantické = celkem 1	<i>vím/nevím a myslím/nemyslím.</i>
<i>že</i>	

2.2.2.2 Úroveň A2

Na rozdíl od úrovně A1 nenajdeme v popisu A2 podrobnější obecnou charakteristiku písemného projevů uživatele s ohledem na syntax. Uvádí se pouze deskriptor, se kterým se setkáme v SERRJ, rozšířený o spojku *že*: „Student dokáže napsat řadu jednoduchých frází a vět. Je schopen použít i souvětí s jednoduchými spojovacími výrazy (např. „*a*“, „*ale*“, „*(a) proto*“, „*protože*“, „*že*“ atd.).“ (PopisA2 2005, s. 111)

A doplňuje se doporučení: „Na úrovni A2 se doporučuje zabývat se pouze jednoduchými konstrukcemi vedlejší věty předmětné, přívlastkové, důvodové, podmínkové, způsobové a případně i podmětné a přípustkové.“ (PopisA2 2005, s. 229)

Dále se pak můžeme opírat pouze o výčet spojovacích prostředků a mezivětných vztahů, které by měl uživatel zvládnout vyjádřit (tabulka 7). Ten ale dle našeho názoru rozsahem značně překračuje uvedenou charakteristiku „jednoduchých spojovacích výrazů“.

Tabulka 7: Spojovací prostředky očekávané na úrovni A2 (srov. PopisA2 2005, s. 214–229)

spojovací prostředky parataktické = celkem 18	
<i>a, (a) i, (a) také, (a) pak, (a) potom</i>	poměr slučovací
<i>i; dokonce (i); nejen(že) – ale i; dokonce ani</i>	poměr stupňovací
<i>ale, avšak, však, ovšem aj.</i>	poměr odporovací
<i>(a)nebo; buď – (a)nebo</i>	poměr vylučovací
<i>neboť</i>	poměr důvodový
<i>(a) proto, a tak</i>	poměr důsledkový
spojovací prostředky hypotaktické	
autosémantické = 13	
<i>kdo, co, jaký, který, či kde, kam, odkud, kudy, kdy, odkdy, dokdy, jak</i>	„Ostatní zájmena vztažná včetně zájmena <i>jenž</i> se na úrovni A2 neprobírají.“ (s. 171)
synsémantické = celkem 20	
Spojky uvozující vedlejší věty příslovečné	
<i>i když; ačkoli(v) / ač aj.</i>	přípustka
<i>když, až</i>	čas
<i>protože</i>	příčina – důvod
<i>aby</i>	účel
<i>jestliže, pokud, kdyby aj.</i>	podmínka
<i>že, tak – že, takže aj.</i>	účinek
<i>jak, jako, než aj.</i>	srovnání
Spojky uvozující vedlejší věty předmětné, přívlastkové, doplňkové a podmětné	
<i>že, ať, jestli</i>	předmětná, přívlastková, podmětná
<i>jak</i>	předmětná, přívlastková, podmětná, doplňková

2.2.2.3 Úroveň B1

Zatímco jsme u předchozích úrovní zaznamenali vždy alespoň stručnou zmínku o charakteru produkovaného textu s ohledem na jeho syntaktickou stavbu, v popisu úrovně B1 takováto obecná informace chybí. Jestliže si tedy chceme udělat představu o písemné produkci uživatele češtiny na této úrovni, nezbude nám než vycházet pouze ze soupisu prostředků

uvedeného slovy: „Prahová úroveň počítá s následujícími výrazy:“ (tabulka 8). (PopisB1 2001, s. 315)

Tabulka 8: Spojovací prostředky očekávané na úrovni B1 (PopisB1 2001, s. 315–318)

spojovací prostředky parataktické = celkem 21	
<i>a, ani; ani – ani; i; i – i; jak – tak; jednak – jednak</i>	slučování
<i>ale, (a)nebo, a přece, avšak; sice – ale, však</i>	odporování
<i>(a) dokonce (i); nejen – ale i</i>	stupňování
<i>bud' – (a)nebo, nebo</i>	vylučování
<i>neboť, proto</i>	důvod
<i>(a) proto, tedy</i>	následek – důsledek
<i>čím – tím</i>	
spojovací prostředky hypotaktické	
autosémantické = 12	
<i>kdo, co, jaký, který, čím kde, kam, odkud, kudy, kdy, jak, proč</i>	„Prahová úroveň nepočítá se vztažným zájmenem <i>jenž</i> , to je knižním synonymem zájmena <i>který</i> .“ (s. 286) Prahová úroveň nevyžaduje aktivní znalost struktur s korelativem (např. <i>ten, kdo</i>). (s. 333)
syntaktické = celkem 29	
<i>ačkoli(v), i když, přestože, třebaže</i>	přípustka
<i>až, dokud; dokud ne; jak, jakmile, když, mezitímco, než, zatímco</i>	čas
<i>jelikož, ježto, poněvadž, protože</i>	příčina – důvod
<i>aby</i>	účel
<i>-li, jestli(že), kdyby, když, pokud</i>	podmínka
<i>takže</i>	následek – důsledek
<i>jako, než</i>	srovnání
<i>ať, jestli, že</i>	obsahové věty

2.2.2.4 Úroveň B2

Popis úrovně B2 se ve zpracování dodatku o syntaxi drží stejně jako popis B1 valenční teorie a taktéž předkládá přehled větných struktur tříděný na základě formálních a sémantických vztahů mezi členy, které struktury tvoří. Na rozdíl od všech předchozích úrovní se ale u B2 důsledně vyskytuje také informace o tom, kterým strukturám má uživatel pouze porozumět v rámci receptivních dovedností a které má sám aktivně produkovat.

Porovnáme-li přehled spojovacích prostředků uvedený u B2 (tabulka 9) a B1, zjistíme, že je téměř totožný. Úrovně se ovšem liší právě rozsahem-počtem větných struktur,

kteře dokáže uživatel na úrovni B2 s danými spojovacími prostředky vytvořit. Například dokáže oproti nižším úrovním produkovat věty spojkové věty s *že, aby, jestli* závislé na substantivu ve větě řídicí.⁴⁶

Tabulka 9: Spojovací prostředky očekávané na úrovni B2 (PopisB2 2005, s.314–317)

spojovací prostředky parataktické = celkem 21	
<i>a, ani; ani – ani; i; i – i; jak – tak; jednak – jednak</i>	slučování
<i>ale, (a)nebo, a přece, avšak; sice – ale, však</i>	odporování
<i>(a) dokonce (i); nejen(že) – ale i</i>	stupňování
<i>bud' – (a)nebo, nebo</i>	vylučování
<i>neboť, proto</i>	důvod
<i>(a) proto, tedy</i>	následek – důsledek
<i>čím – tím</i>	
spojovací prostředky hypotaktické	
autosémantické = 12	
<i>kdo, co, jaký, který, či</i> <i>kde, kam, odkud, kudy, kdy, proč</i>	
syntaktické = celkem 30	
<i>ačkoli(v), i když, přestože, třebaže</i>	přípustka
<i>až, dokud; dokud ne; jak, jakmile, když, mezitímco, než, zatímco</i>	čas
<i>jelikož, ježto, poněvadž, protože</i>	příčina – důvod
<i>aby</i>	účel
<i>-li, jestli(že), kdyby, když, pokud</i>	podmínka
<i>takže</i>	následek – důsledek
<i>jako, než</i>	srovnání
<i>ať, jak, jestli, že</i>	obsahové věty

2.2.3 Shrnutí

Tato kapitola poskytla vřled do toho, jakým způsobem je ve SERRJ – stěžejním dokumentu evropské jazykové politiky – charakterizována písemná produkce nerodilých mluvčích s ohledem užívání spojovacích prostředků. Ukázali jsme, že přestože je dnes popis úrovní dle SERRJ velmi rozšířený, v mnoha případech je velmi obecný. Koneckonců už od roku 2001 samotná Rada Evropy podporovala myšlenku vytvoření doplňkových nástrojů k SERRJ, které by lépe demonstrovaly situaci v jednotlivých jazycích. Jednou z takových iniciativ bylo

⁴⁶ Např. *Mám strach, aby mě nepřejelo auto.* vs. na nižší úrovni *Mám strach, aby mě nepřejelo auto. / Mám strach, že mě přejede auto.*

i vydání publikace *Reference Level Descriptions* (Popisy referenčních úrovní) pro národní a regionální jazyky. Pro češtinu vznikly v období 2001–2005 popisy úrovní A1–B2. Bohužel se stejně jako SERRJ dostatečně neopírají o příklady z reálných dat uživatelů, kteří se učí češtinu jako cizí jazyk.⁴⁷ Uživatelé obou dokumentů (SERRJ i popisů) jsou tak při hodnocení dosažené úrovně odkázáni na své vlastní představy a zkušenosti. A není jasné, do jaké míry tato představy korespondují s popisem úrovní škály (srov. např. Arras 2010). Během vytváření SERRJ ani národních popisů pro češtinu totiž neproběhly analýzy jazyka nerodilých mluvčích, které by umožnily jejich empirickou validaci. Z těchto důvodů může být zajímavé sledovat a konfrontovat výsledky provedených analýz autentických materiálů s výše uvedenými stupnicemi a přehledy spojovacích prostředků se SERRJ i popisů referenčních úrovní pro češtinu. Předkládanou prací, jak věříme, bychom mohli přispět k jejich validaci nebo případné úpravě.

⁴⁷ Srov. v současnosti převládající silnou tendenci zakládat tyto popisy na korpusech žákovského jazyka, jako např. korpus angličtiny (www.englishprofile.org), ale také itaštiny (Spinelli/Parizzi 2010) a norštiny (Carlsen 2013).

2.3 TEORIE MEZIJAZYKA A CHYBOVÁ ANALÝZA

V úvodu předkládané práce jsme si jako jeden z dílčích cílů stanovili prozkoumání vztahu úrovně jazykové kompetence a správnosti užívání spojovacích prostředků. Výzkum, který budeme provádět, se bude opírat o chybovou analýzu. Proto se v této kapitole zaměříme na shrnutí hlavních principů, které jsou předpokladem pro její aplikaci, a pokusíme se o její zasazení do širšího kontextu teorií osvojování cizího jazyka (FLA – foreign language acquisition).

V posledních několika desetiletích prošly pokusy o teoretické uchopení procesu osvojování cizího jazyka značným vývojem. Objevila se řada přístupů, z nichž některé na tomto poli získaly silnou pozici a řadu stoupců (např. behaviorální, generativistické, kognitivní). V souvislosti s těmito přístupy byly rozpracovány různé metody zkoumání toho, jak osvojování cizího jazyka probíhá a jaké faktory jej ovlivňují. V současné době zaujímá klíčové postavení teorie mezijazyka. V češtině o přístupech k osvojování cizího jazyka a pojetí chyby zevrubně pojednala Štindlová (2013), odkazujeme tedy na tomto místě na její studii a omezíme se v následujícím textu pouze na základní přehled relevantní pro tuto disertační práci.

Pojem mezijazyk představil Selinker (1972), který ho vymezil jako svébytný systém jazyka s vlastní strukturou a organizací vznikající a proměňující se během osvojování cizího jazyka.⁴⁸ Proměnlivost spočívá v tom, že se jedná o systém přechodový, který je sice oddělený jak od výchozího, tak cílového jazyka, ale k oběma má vztah. Dynamickým vývojem a neustálým „přebudováním“ se výchozímu jazyku vzdaluje a blíží se jazyku cílovému. Tento vývoj se snaží popsat a vysvětlit teorie mezijazyka.

Před tím, než se teorie mezijazyka dostala do povědomí a získala si své uznání, se ve vědeckých kruzích prosazovala kontrastivní analýza, která vnímala žakovský jazyk pouze jako výsledek transferu z jazyka mateřského. Její zastánci se domnívali, že na základě důkladné kontrastivní analýzy výchozího a cílového jazyka je možné predikovat všechny obtíže a chyby, které jsou s osvojováním cizího jazyka spojeny (srov. Ellis 1994, s. 306–308). Často to ale dokládali pouze na příkladech, jež nebyly získány na základě systematického studia žakovského jazyka.

⁴⁸ Zhruba ve stejné době přišli s podobným konceptem také Nemser (1971), který hovoří o tzv. aproximativním systému, a Corder (1967), jenž zavádí termín idiosynkratický dialekt. Oba dva akcentují různé rysy mezijazyka. U Nemsera stojí v popředí proměnlivost a stálé směřování mezijazyka ke sblížení s mateřským jazykem. Corder zdůrazňuje, že jde o systém jedinečný vzhledem k jednomu určitému individu, tj. jeho struktura je typická pouze pro daného jedince. Srov. Štindlová 2013, s. 15 a 25n.

Pochybnosti o dostatečnosti tohoto přístupu vedly k pokusům o validaci výsledků kontrastivní analýzy pomocí analýz většího množství shromážděných žákovských produkcí. Na jejich základě se ukázalo, že se v produkcích nerodilých mluvčích vyskytuje řada chyb, které nelze vysvětlit transferem z mateřského jazyka. Tím byla zpochybněna platnost hypotéz formulovaných na základě kontrastivní analýzy, která sice obrátila pozornost na důležitou otázku toho, čím je formován žákovský jazyk, ale jak se ukázalo, nebyla schopna ji uspokojivě zodpovědět.

Se zpochybněním kontrastivní analýzy a jejího výlučného spojení chyb s transferem z mateřského jazyka došlo k celkovému obrátu ve vnímání chyb v žákovském jazyce. Corder (1967) přišel s pojetím chyby v intencích vznikající teorie mezijazyka. Ukázal, že chyby nemusí být znakem neúspěšné produkce, ale mohou naopak sloužit jako sledovatelný projev vývoje žákovského jazyka na cestě od výchozího jazyka k jazyku cílovému. Tento přístup akcentuje také chybová analýza, kterou v této práci využijeme ke zkoumání textů nerodilých mluvčích.

2.3.1 Principy chybové analýzy

Jak jsme naznačili, chybová analýza se na rozdíl od kontrastivní analýzy nezaměřuje na srovnání žákovského jazyka s jazykem výchozím, ale zkoumá systém mezijazyka ve vztahu k jazyku cílovému. Tím také poskytuje větší prostor pro popis a vysvětlení chyb, které se v mezijazyce objevují. Chybou je v tomto případě myšlena jakákoliv odchylka od stanovené normy cílového jazyka. Corder (1967, s. 163n.) rozlišuje chyby nesystémové (mistakes) a chyby systémové (errors). Nesystémové chyby jsou většinou jednorázové a mluvčí jsou schopni je sami rozpoznat a případně korigovat. Naproti tomu systémových chyb se nerodilí mluvčí dopouštějí opakovaně a nevědomě. Chybová analýza se zaměřuje na druhý zmíněný typ chyb, protože pouze ten lze považovat za projev mezijazyka.

Proces chybové analýzy popisují např. Corder (1974) nebo Ellis (1994) jako posloupnost několika dílčích kroků. Jsou jimi (1) sběr dat; (2) identifikace chyb; (3) popis chyb prostřednictvím jejich klasifikace a kategorizace a (4) explanace chyb. Nad rámec chybové analýzy se někdy připojuje také (5) evaluace chyb.

(1) Sběr dat spočívá ve shromáždění dokladů produkce nerodilých mluvčích. Jeho zásadám je věnována značná pozornost při tvorbě žákovských korpusu (srov. např. Šebesta – Škodová 2012).

(2) Identifikace chyb probíhá jako souhra detekce (detection), tj. odhalení přítomnosti chyby a lokalizace (locating), při níž se určí přesné místo chyby a rozsah jednotek, které zasahuje – slovo, skupina slov, věta atd. (srov. James 1998, s. 93). Obojí probíhá na základě srovnání produkce nerodilého mluvčího s rekonstrukcí daného vzorku rodilým mluvčím.⁴⁹

(3) Popis chyb se opírá o různé systémy klasifikace chyb (tzv. chybové taxonomie), které se liší přístupem k organizaci jevů, jež mají být kategorizovány (srov. James 1998, s. 102). Rozlišují se čtyři typy taxonomií – (a) taxonomie odrážející povrchovou realizaci; (b) taxonomie založená na lingvistických kategoriích; (c) komparační chybová taxonomie a (d) taxonomie na základě komunikačního efektu (srov. Dulay – Burt – Krashen 1982, s. 146n.). Zatímco první dva typy taxonomie jsou svou povahou deskriptivní, dalšími dvěma typy prostupují také explanace a evaluace chyb.

(a) Taxonomie na základě povrchové realizace klasifikuje chyby podle toho, k jakému typu strukturální deformace dochází. Jinými slovy podle toho, jakou změnu by provedl rodilý mluvčí při rekonstrukci bezchybně realizované struktury.⁵⁰ Běžně se vydělují tyto typy chyb: (i) vynechání (omission), (ii) přidání (addition), (iii) užití chybné formy (misformation) a (iv) chybný slovosled (misordering) (srov. Dulay – Burt – Krashen 1982, s. 146n.).

(i) K vynechání dochází nezařazením obligatorního členu do větné konstrukce.

(ii) Přidání představuje nadbytečné užití prvku, který se v daném kontextu nesmí/nemá vyskytnout. Může být způsobeno například přílišným zobecněním pravidel a jejich aplikací i v případě, kdy to není adekvátní, např. užití pomocného slovesa *být* u forem 3. os. pl.

(iii) Užití chybné formy se týká chyb ve formě slovního tvaru nebo syntaktické struktury. Opět může být následkem přílišného „zpravidelnění“, např. u tvarů nepravidelných sloves. Objevuje se ale také ve spojitosti se záměnou prvků stejné třídy, např. záměna různých tvarů osobních zájmen. Patří sem i tzv. chybný výběr, např. záměna časových spojek *když* a *až*. Někdy se do této skupiny zařazují ještě případy kontaminace (srov. James 1998, s. 111), a to proto, že se v podstatě jedná o chybu ve výběru jedné ze dvou sémanticky obdobných struktur, např. **Mám rád čtu*.

⁴⁹ Tento krok může působit značné problémy, protože neexistuje pravidlo, které by určovalo, zda je provedená rekonstrukce adekvátní.

⁵⁰ Srov. označení této taxonomie jako taxonomie podle typu úpravy, které prosazuje James (1998, s. 106).

(iv) Poslední typ chyby – špatné uspořádání (misordering) je záležitostí nesprávného slovního pořádku větné struktury žakovského jazyka.

(b) Druhý typ deskriptivní taxonomie je založen na klasifikaci chyb na základě lingvistických kategorií. Při jejich vymezení se vychází z popisu cílového jazyka a uplatňuje se hierarchická kategorizace. Nejčastěji se zohledňuje jazyková rovina, které se chyba týká (fonetická, gramatická apod.), dále slovní druh a gramatická kategorie (pád, číslo apod.). Jak uvádí Štindlová (2012, s. 42n.), mohou při užití tohoto typu taxonomie nastat problémy např. s nejednotným vymezením některých subkategorií v rámci daných domén. Problematickou může být i volba rozsahu (hloubky) této taxonomie, která by ideálně měla vycházet ze samotných dat (tj. data-driven přístup) a reflektovat konkrétní chyby, které se v nich objevují (srov. Ellis – Barkhuizen 2009, s. 60).

(4) Identifikace a popis chyb tvoří nezbytný předstupeň pro explanaci, tj. určení zdroje chyby. Běžně se rozlišují (a) interlingvální chyby na základě mezijazykového transferu; (b) intralingvální chyby na základě cílového jazyka; (c) učební chyby (tj. plynoucí z kontextu výuky) a (d) chyby vyplývající z uplatňovaných komunikačních strategií (srov. např. James 1998, s. 173n.).

(5) Zmínili jsme výše, že někdy na chybovou analýzu navazuje ještě evaluace chyb. Ta se na základě vlastních metod a evaluačních nástrojů zabývá určením závažnosti chyb. Předkládaná disertační práce se explanací ani evaluací chyb zabývat nebude, a proto tyto dva kroky podrobněji nerozebíráme.

Dodejme závěrem, že ani chybová analýza se nevyhnula kritice. Vytýká se jí například, že nedokáže dostatečně postihnout strategii vyhýbání, tj. případy, kdy se mluvčí v produkci záměrně vyhýbají jazykovým jevům, kterými si nejsou jistí (srov. Schachter 1974).

2.3.2 Shrnutí

Tato kapitola představila hlavní principy chybové analýzy. Zmínili jsme její kroky a podrobněji jsme se věnovali otázce identifikace a popisu chyb. Naopak stranou jsme ponechali fázi explanace a evaluace chyb, které nebudou předmětem této disertační práce. Chybovou analýzu jsme zasadili do kontextu teorie mezijazyka, kde je chyba vnímána jako projev vývoje systému žakovského jazyka.

3 MATERIÁLOVÝ PODKLAD – KORPUSY

Praktická část našeho výzkumu spočívá v analýze autentických textů, ve kterých budeme mapovat užití spojovacích prostředků. Taková práce by v minulosti nebyla možná bez časově náročné excerpce textů. Čeština až do nedávna nedisponovala korpusovými soubory textů nerodilých mluvčích, které by navíc byly opatřeny přiřazením k úrovním podle SERRJ. Takzvané žákovské korpusy začaly vznikat až v posledních letech. Korpus MERLIN, který představuje náš primární zdroj dat, byl na veřejnost uveden teprve na začátku roku 2015. Vzhledem k tomu, že se jedná o relativně nový a unikátní nástroj, věnujeme v první podkapitole (3.1) hlavní pozornost právě jemu. V některých dílčích sondách jsme využívali také další korpusy, a proto se následně stručněji zmíníme i o nich. Půjde o další dva žákovské korpusy: jeden s texty nerodilých mluvčích – CzeSL-SGT a jeden s texty rodilých mluvčích češtiny – SKRIPT2012. Závěrem krátce připojíme charakteristiky korpusů SYN, jež jsme také příležitostně využívali.

Druhá podkapitola je věnována obecným úvahám o způsobech vyhledávání různých druhů spojovacích prostředků v korpusech, které ilustrujeme na příkladech z korpusů SYN a MERLIN.

V podkapitole třetí popisujeme vlastní postup, který jsme uplatnili při práci s korpusem MERLIN a při analýze dokladů. Uvádíme, jak jsme dospěli k soupisu repertoáru spojovacích prostředků, a vysvětlujeme způsob provedené analýzy jejich chybného užití.

3.1 CHARAKTERISTIKA KORPUSŮ

Cílem této podkapitoly je zprostředkovat nejdůležitější informace o využitých korpusech. Soustředíme se na uvedení údajů o charakteru, rozsahu a složení korpusu, o obsažených metadatech, o způsobu anotace a o možnosti vyhledávání. Vždy ujasníme také roli, kterou hraje daný korpus v našem výzkumu.

3.1.1 Korpus MERLIN⁵¹

Na několika místech jsme již uvedli, že korpus MERLIN představuje náš hlavní zdroj dat. Je to totiž jediný žákovský korpus s texty nerodilých mluvčích češtiny, které byly zároveň spolehlivě přiřazeny k úrovním podle SERRJ.

Korpus vznikl v rámci mezinárodního projektu *Merlin: Multilingvální platforma pro evropské referenční úrovně: Výzkum jazyka studentů v kontextu*. Tento projekt se snažil reagovat na chybějící ukázky písemné produkce, které by ilustrovaly úrovně SERRJ (srov. kap. 2.2.3). Specifičnost korpusu MERLIN spočívá jednak v tom, že se jedná o trojjazyčný korpus (čeština, němčina a italština), a jednak, jak jsme již zmínili, v tom, že jako jeden z mála žákovských korpusů podstatným způsobem reflektuje SERRJ. Jak uvádí Štindlová – Čurdová (2015, s. 191), „česká část korpusu Merlin je vedle žákovského korpusu CzeSL v současné době jediným relevantním zdrojem projevů nerodilých mluvčích češtiny. Svým charakterem se zároveň řadí k souboru akvizičních korpusů AKCES.“⁵²

3.1.1.1 Charakter a velikost korpusu

Texty v korpusu pocházejí ze subtestů psaní standardizovaných zkoušek vysoké kvality vztahujících se k SERRJ.⁵³ Počet shromážděných textů a celkový počet slov v rámci jednotlivých subkorpusů je uveden v tabulce 1. Jádrem korpusu pro češtinu tvoří texty reprezentující úrovně A2, B1 a B2. To byly také úrovně, na základě jejichž testování texty vznikly. Korpus obsahuje celkem 441 textů, přes 64 tisíc slov a přes 76 tisíc pozic (tokenů).

3.1.1.2 Metadata a složení korpusu

Spolu s texty byla sbírána i běžná metadata, tj. informace o mateřském jazyce kandidáta, pohlaví, věku, typu zpracovávaného úkolu a úrovni podle zadání a hodnocení textu.

Autoři korpusu MERLIN si zakládají na tom, že všechny texty byly v rámci projektu ohodnoceny profesionálními hodnotiteli podle kritérií vztahujících se k SERRJ.⁵⁴ Písemná

⁵¹ Korpus je přístupný pomocí platformy: <http://merlin-platform.eu>.

⁵² Soubor akvizičních korpusů AKCES viz <http://akces.ff.cuni.cz>.

⁵³ Jmenovanými institucemi jsou telc Frankfurt nad Mohanem (italské a německé testy, www.telc.net) a z Ústavu jazykové a odborné přípravy Univerzity Karlovy v Praze (české testy, www.ujop.cuni.cz). Obě zkušební instituce prošly auditem asociace mezinárodní jazykových testerů ALTE (www.alte.org). Testové úlohy byly používány do roku 2013.

⁵⁴ Hodnocení probíhalo podle přesných pravidel a pod striktním dohledem. Reliabilita opětovného hodnocení byla ověřena klasickou metodologií pomocí teorie CTT (Classical Test Theory) a analýzou

produkce byla vztažena ke stupnicím pro komunikativní jazykovou kompetenci z 5. kapitoly SERRJ (viz kap. 2.3.1). Použita byla holistická škála, založená na škále SERRJ pro rozsah všeobecných znalostí jazyka, a také analytická hodnotící tabulka s následujícími kritérii hodnocení: pravopis; gramatická správnost; rozsah slovní zásoby; ovládání slovní zásoby; koherence a koheze; sociolingvistická přiměřenost; a to v rozpětí úrovní A1 až C2.

Při hodnocení se podle očekávání projevilo to, že výkony některých kandidátů neodpovídaly úrovni zkoušky, kterou skládali. Kandidáti buď podali výkon horší a hodnotitelé jejich texty zařadili k úrovni nižší, nebo naopak napsali text, který absolvovanou úroveň překračoval a byl při hodnocení zařazen k úrovni vyšší. Z toho také pramení diskrepance mezi počtem textů podle úrovně zkoušky a podle jejich skutečného hodnocení (tabulka 10). V korpusu je možno vyhledávat jak podle úrovně zadání testu, tak podle úrovně hodnocení. Kupříkladu je možné vyhledat české texty z testů na úrovni B2, které ale získaly pouze hodnocení úrovně B1.

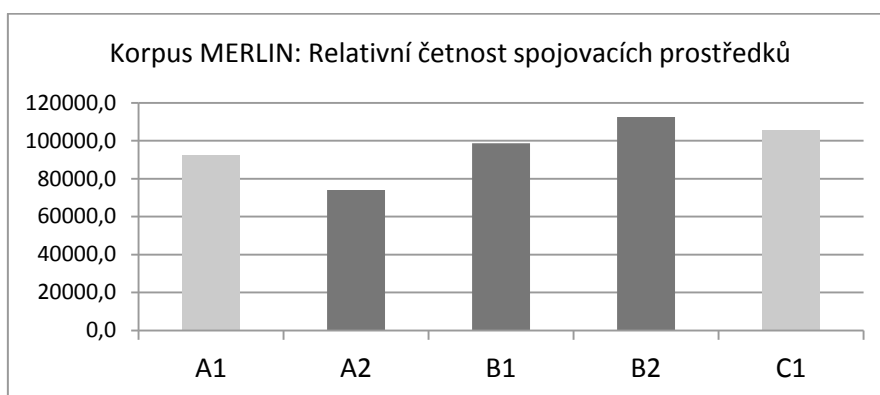
Tabulka 10: Složení a rozsah korpusu MERLIN

jazyk	počet textů				počet slov	počet pozic (tokenů)
	podle úrovně zkoušky		podle hodnocení			
čeština			A1	1	50	54
	A2	111	A2	188	17 828	21 505
	B1	143	B1	167	28 591	34 007
	B2	186	B2	81	16 891	19 953
			C1	4	987	1 167
CELKEM		441		441	64 297	76 632

italština	A1	207	A1	29	92 359
	A2	202	A2	378	
	B1	201	B1	394	
	B2	201	B2	2	
němčina	A1	206	A1	57	125 927
	A2	209	A2	297	
	B1	210	B1	331	
	B2	204	B2	293	
	C1	204	C1	42	
			C2	4	

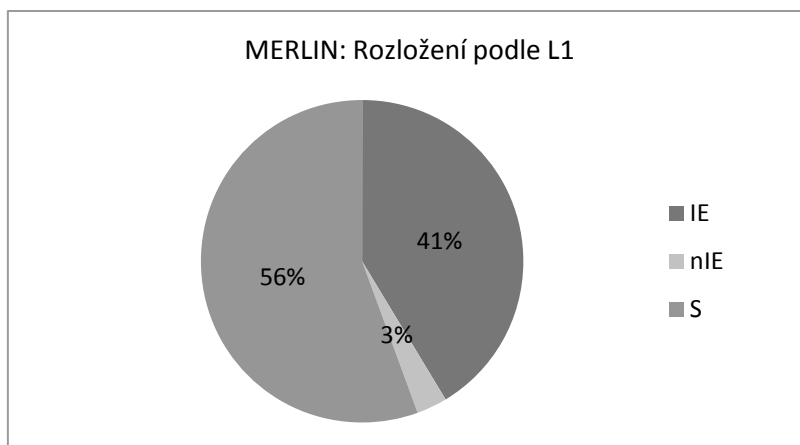
multiparametrickým Raschovým modelem. Druhá ze zmíněných metod je pravděpodobnostní statistická metoda, která se často používá pro koncipování jazykových testů a která umožňuje korekce tendencí hodnotitelů, např. shovívavost/přísnost a díky níž je možné pro každý text vypočítat adekvátní průměr hodnocení. Jak uvádějí autoři, vnitřní konzistence hodnotitelů (intra-rater reliabilita) i shoda mezi nimi (inter-rater reliabilita) byly v platformě MERLIN obecně na vysoké úrovni kromě pár výjimek u italských textů. Srov. Merlin project 2014, s. 8.

Z přehledu je patrné, že úrovně A1 a C1 v češtině, jejichž zkoušky nebyly součástí sběru dat, jsou zastoupeny pouze prostřednictvím jednoho textu, který neuspěl u úrovně A2 a dvou textů, které převyšovaly úroveň B2. Jedná se tedy o zastoupení marginální, které musíme brát v potaz pokaždé, když budeme interpretovat výsledky našich analýz. U úrovní A1 a C1 nelze na základě jednoho, resp. dvou textů vyvozovat dalekosáhlejší závěry. V následujících kapitolách budeme tuto skutečnost připomínat také vizuálně. Ve všech grafech, kde budou uvedena data zahrnující dotčené úrovně, odlišíme, bude-li to možné, světle šedou barvou výsledky úrovní A1 a C1 od výsledků ostatních úrovní (např. graf 1).

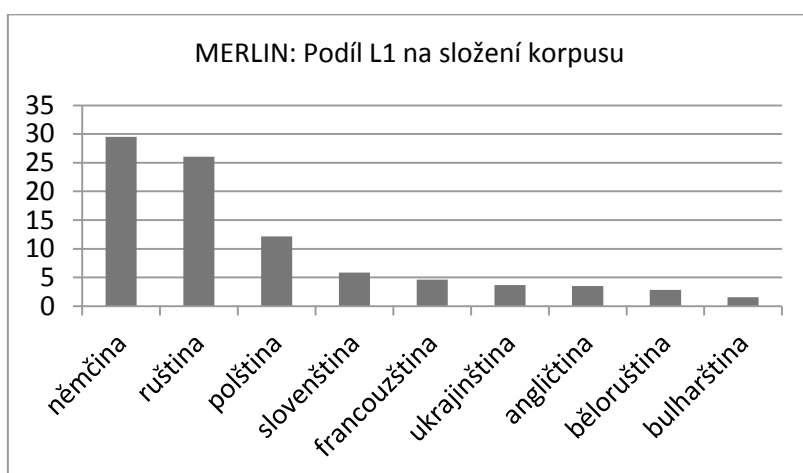


Graf 1: Ukázka vizuálního odlišení úrovní A1 a C1 (i.p.m. podle úrovní)

Ani rozložení korpusu podle mateřského jazyka není zcela vyvážené. Větší podíl mají texty autorů se slovanským mateřským jazykem (56 %). Ostatní texty napsali autoři s neslovanským mateřským jazykem (44 %), většinou indoevropským (41 %) a se zlomkem neindoevropských jazyků (3 %) – viz graf 2. Zastoupeno je celkem 28 různých jazyků, prvních 10 z nich podle počtu tokenů ukazuje graf 3.



Graf 2: MERLIN: Rozložení podle L1 (% podle počtu tokenů)



Graf 3: MERLIN: Podíl L1 na složení korpusu (% podle počtu tokenů)

Zastavme se nyní ještě u další informace obsažené v metadatech a vysvětleme blíže, jakým způsobem texty shromážděné v korpusu vznikly. Víme již, že byly produkovány jako součásti jazykových zkoušek v subtestech psaní. To znamená, že na základě stejného zadání (tabulka 11) vzniklo vždy několik textů od různých autorů.

Tabulka 11: MERLIN: složení korpusu podle typu a tématu zadání textů

Úroveň testu	Typ produkováného textu	Specifikace/téma zadání	Počet textů
A2	e-mail – formální	požádat o informace k ubytování	17
	e-mail – neformální	reagovat na pozvání na oslavu narozenin	20
	popis fotografie	žena sedí u okna a čte knihu	25
	popis fotografie	rodina je na dětském hřišti	19
	popis fotografie	otec s malým synem se koupou v moři	30
B1	e-mail – neformální	odpovědět Aleně a zeptat se na letní setkání a letní školu češtiny	86
	e-mail – neformální	odpovědět Janě a napsat o plánech ve studiu a plánech na léto	50
	e-mail – formální	odpověď agentuře Tandem v reakci na zájem o letní školu češtiny	7
B2	esej/zamyšlení	Všude dobře doma nejlépe	29
	esej/zamyšlení	Škola základ života	38
	esej/zamyšlení	Víc hlav víc ví	37
	esej/zamyšlení	Šaty dělají člověka	22
	esej/zamyšlení	Bez práce nejsou koláče	32
	esej/zamyšlení	V nouzi poznáš přítele	29
			441

Na úrovni A2 převažují texty s popisem fotografie podle bodů v zadání (ukázka 1), kromě nich se objevují dva typy písemné interakce (formální – ukázka 2 – a neformální), které jsou zastoupeny v přibližně stejném poměru mezi sebou. Můžeme tedy konstatovat, že se zde objevují hlavní typy textů, u kterých se podle SERRJ očekává, že je bude uživatel schopen produkovat (srov. kap. 2.2.1).

Úloha 2 – Popis

Napište do záznamového archu popis fotografie. Napište minimálně 55 slov.

Prohlédněte si fotografii a odpovězte na všechny otázky.



- *Kdo je na fotografii?*
- *Kde jsou?*
- *Proč tam jsou?*
- *Co tam dělají?*
- *Co mají na sobě?*
- *Jak se cítí?*
- *Co říkají?*
- *Které je roční období?*

Ukázka 1: Zadání testu A2 – popis fotografie

Úloha 1 – E-mail



Úloha 1 – E-mail

Napište do záznamového archu e-mail. Napište minimálně 35 slov.

Kamarád/ka vás pozval/a na oslavu svých narozenin. Napište mu/jí e-mail.

V e-mailu musíte:

- *poděkovat,*
- *zeptat se na:*
 - *místo oslavy*
 - *začátek oslavy*
 - *další hosty*

Ukázka 2: Zadání testu A2 – neformální e-mail

Úroveň B1 bohužel neposkytuje tak rozmanitá zadání jako úroveň A2. Obsahuje totiž pouze texty písemné interakce (ukázka 3) a zcela zde chybí texty samostatného písemného projevu, které bychom dle SERRJ mohli očekávat (popis událostí, zážitků, pocitů apod.).

Beantworten Sie Janas Email (45 min / min. 100 slov). Gehen Sie auf die gestellten Fragen ein. Es soll außerdem Folgendes dargestellt bzw. erfragt werden:

- Fragen Sie,
 - welches Fach Jana in Prag studieren will
 - wie ihr der Englischkurs gefällt und ob es dort im Sommer auch Tschechischkurse gibt. Sie würden gern im Sommer nach Brunn kommen. Dann können Sie sich öfter treffen.
- Bitten Sie sie, dass sie Ihnen einen Link zu (odkaz na) der Webseite der Schule schickt.
- Sagen Sie auch, dass Sie einige Fotos vom letzten Ausflug mit Thomas mitschicken.

Finden Sie einen geeigneten Abschluss.

Odesílatel <jana-neni-doma@seznam.cz>
Komu <petra-in-dresden@bmx.de>

Ahoj Peťo,

díky moc za Tvůj mail, který mě fakt potešil! Promiň, že reaguju až teď. Byla jsem na Slovensku na semináři a nesledovala jsem e-maily.

Mám novinku: já jsem se právě dneska rozhodla, že budu studovat dál. Chci studovat dálkově¹ v Praze postgradual. Jenom nevím, jestli dám dohromady včas všechny papíry. Termin je totiž už 15.7! **Jaké plány máš po ukončení studia Ty?**

Jinak se mi tady v Brně celkem líbí. Konečně zase učím každý den a někdy i dva lidi denně. ☺ A sama se taky ještě učím. Chodím na kurz angličtiny.

Peťo, pošli nějaké fotky. My jsme se tak dlouho neviděly ;-). Snad se brzy zase uvidíme. **UŽ máš nějaké plány na léto?**

Promiň, musím teď končit. Sedím v knihovně, tak nechci dlouho blokovat počítač s internetem ...

Měj se dobře a zase se ozvi.
Jana
P.S. Zdraví Tě taky Pavel.

¹ Im Fernstudium

Ukázka 3: Zadání testu B1 – neformální e-mail

U úrovně B2 se variabilita co do formy/žánru textu vyskytuje. Zadání jsou postavena na modelu výběru ze dvou přísloví, která udávají tematický rámec buď pro úvahový text, nebo vyprávění (ukázka).⁵⁵

Úloha 2 – Přísloví

Vyberte si jedno ze 2 témat A nebo B a napište na něj vyprávění nebo úvahu (zamyšlení) v rozsahu minimálně 150 slov.

Text musí mít:

- úvod,
- vyjádření vlastního názoru,
- vysvětlení,
- závěr.

Témata:

A. V nouzi poznáš přítele.
B. Bez práce nejsou koláče

Ukázka 4: Zadání testu B2 – vyprávění / úvaha (zamyšlení)

⁵⁵ Testologickou otázkou zůstává, do jaké míry je vhodné zařazení vyprávění do úlohy na úrovni B2. Očekáváme-li podle SERRJ, že uživatel na této úrovni dokáže mimo jiné obhajovat nebo vyvracet určitý názor nebo shrnout, skloubit a zhodnotit informace a argumenty z velkého počtu zdrojů, je malá pravděpodobnost, že tuto dovednost v rámci vyprávění prokáže. Srov. SERRJ 2006, s. 26 a 63.

Pohledem do zadání úloh, na jejichž základě texty sebrané v korpusu vznikly, jsme chtěli upozornit na to, že některé produkce sdílejí jisté společné rysy, zejména tematický okruh (a tím potažmo okruh lexika), a podle toho, jak podrobné/návodné zadání kandidáti obdrželi, také strukturu a typ/žánr textu. To je patrné zejména na úrovni A2 a B2. Tento fakt budeme mít na paměti při interpretaci výsledků našich korpusových analýz, neboť se může odrážet například v rozmanitosti lexika, tj. eventuálně i spojovacích prostředků.

3.1.1.3 Anotace

Anotace korpusu MERLIN je založena na rozsáhlé škále indikátorů, které umožňují zachytit komplexní charakter žákovského jazyka. Indikátory se vztahují ke všem třem jazykům v korpusu a popisují jak jejich společné rysy (např. použití spojek, verbonominální shoda, kolokace apod.), tak jevy jazykově specifické (např. čeština: dvojí negace, posesivní reflexíva; italština: lexikalizovaná klitika, němčina: modální částice, italština+němčina: členy apod.). Dotýkají se oblastí ortografie, gramatiky, lexika i sociolingvistiky a zohledňují standardní i nestandardní aspekty jazyka. Celkem jich bylo vymezeno více než 70, a to na základě charakteristik úrovní uváděných v SERRJ i v sekundární literatuře, podle předběžné analýzy písemných projevů studentů, ale také v souvislosti s dotazníkovým šetřením mezi budoucími uživateli korpusu, tj. učiteli, hodnotiteli, studenty. Následně byly těmto indikátorům přiřazeny tagy shrnuté v anotačním schématu (Abel a kol. 2014).

Anotace korpusu MERLIN představuje kombinaci značkování založeného na formálních typech alternace zdrojového textu (chybějící element, přebývající element, chybně spojené elementy apod.) a hierarchicky strukturované lingvistické klasifikace.⁵⁶ Hlavní anotace, které lze použít pro téměř celý korpus MERLIN, jsou cílové hypotézy (cílová hypotéza 1 – TH1)⁵⁷ a anotace gramatických a pravopisných rysů žákovského jazyka (anotace chyb 1 – EA1). V menším, tzv. jádrovém korpusu⁵⁸ je anotace rozšířena o druhou cílovou hypotézu (cílová hypotéza 2 – TH2), která má za cíl vytvářet (pro rodilého mluvčího) přijatelnou verzi původního žákovského textu. Zahrnuje proto i aspekty týkající se slovní zásoby, pragmatiky

⁵⁶ Blíže k této problematice např. Štindlová (2013, s. 79n).

⁵⁷ Cílová hypotéza 1 (TH1) je verze žákovského textu s minimální intervencí. Jedná se o převedení textu do podoby, která je sice pravopisně a gramaticky správná, ale na jiných rovinách může obsahovat odchylky vůči přirozenému užívání jazyka rodilého mluvčího, např. na rovině lexikální. TH1 byly vytvořeny pro celý korpus MERLIN. Srov. Merlin project 2014, s. 8.

⁵⁸ Jádrový korpus se skládá ze dvou skupin textů, které byly ohodnoceny na úrovni A2 nebo B2 (pro italštinu: A2 a B1/B1+).

a sociolingvistické přiměřenosti, které jsou zachyceny na druhé úrovni anotace (anotace chyb 2 – TH2). Struktura anotací je znázorněna ve schématu níže (schéma 2).

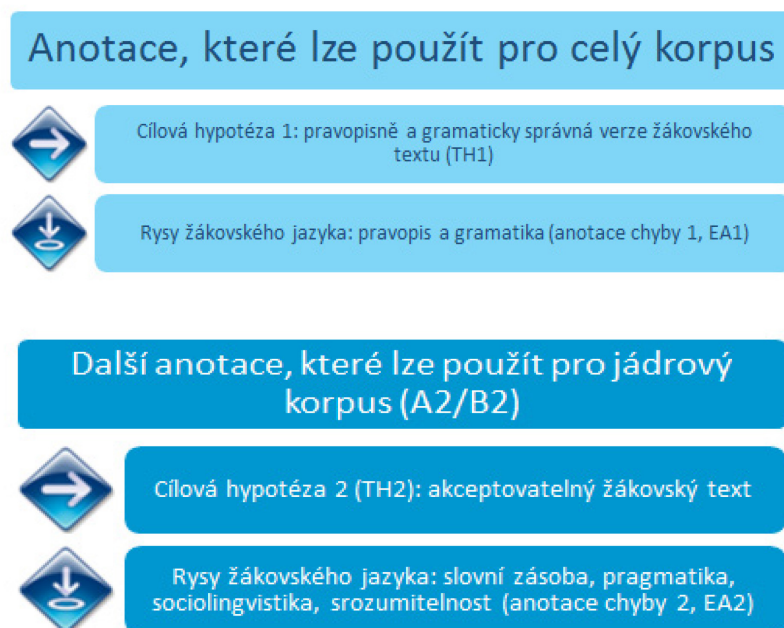


Schéma 2: *Strukturální výstavba anotací v platformě MERLIN (Merlin project 2014, s. 9)*

3.1.2 Korpus CzeSL-SGT

Žákovský korpus CzeSL-SGT (*Czech as a Second Language with Spelling, Grammar and Tags*) patří do skupiny akvizičních korpusů AKCES (Akviziční korpusy češtiny),⁵⁹ které vznikaly na Filozofické fakultě Univerzity Karlovy od roku 2005. Byl vytvářen tak, „aby byl v maximální míře kongruentní s ostatními korpusy této skupiny, tj. především s korpusy SCHOLA a SKRIPT.“⁶⁰ (Štindlová 2011, s. 99)

Stejně jako korpus MERLIN obsahuje CzeSL-SGT přepisy písemných prací nerodilých mluvčích češtiny, proto jsme si jej vybrali jako zdroj pro dílčí sondy, kterými chceme podpořit závěry z analýz textů v korpusu MERLIN. Oba korpusy se od sebe ale v některých parametrech liší. Níže uvedeme rozdíly, které se týkají kontextu našeho výzkumu. Jde například o rozličnou velikost korpusů, rozmanitější zastoupení respondentů podle mateřského jazyka, ale hlavně o způsob hodnocení úrovně písemné produkce podle SERRJ.

⁵⁹ Viz také <http://akces.ff.cuni.cz/>.

⁶⁰ Korpus SCHOLA je mluveným korpusem vyučovacích hodin, korpus SKRIPT je souborem školních písemných prací. Ke korpusu SKRIPT viz kap. 3.1.3.

Zatímco korpus MERLIN disponuje spolehlivým a detailním přiřazením všech textů k úrovním SERRJ (viz 3.1.1.1), u korpusu CzeSL-SGT se o hodnocení uvedené u textů v rámci metadat se stejnou důvěrou opírat nemůžeme. Informace o úrovni znalosti češtiny nebyla nijak ověřována (viz níže 3.1.2.1). To je také hlavní důvod, proč jsme v předkládané práci využili jako primární zdroj dat rozsahem menší korpus MERLIN.

3.1.2.1 Charakter a velikost korpusu

CzeSL-SGT navazuje na korpus CzeSL-plain, který vznikl před ním. Obsahuje jazyková data z jeho subkorpusu CIZ (přepisy písemných prací /esejů/ nerodilých mluvčích, které vznikly souvislosti s jazykovým vyučováním v kurzech různého druhu a úrovně) a k tomu i další texty sebrané později. Na rozdíl od verze plain je anotovaný (viz níže 3.1.2.3). Obsahuje celkem 8 617 textů, 958 tisíc slov, 1 148 tisíc pozic (tokenů). Oproti korpusu MERLIN je jeho rozsah zhruba 15krát větší a také rozmanitost textů podle zadání je širší (přes 8 500 zadání oproti 15 v korpusu MERLIN).

3.1.2.2 Metadata a složení korpusu

Rozsah shromážděných metadat je velmi bohatý. CzeSL-SGT eviduje celkem 30 metadatových parametrů, 15 z nich se týká respondenta a dalších 15 proměnných upřesňuje charakter materiálu, především podmínky jeho vzniku (např. časový limit, možnost použít slovník apod.). Nicméně ne všechny texty disponují všemi atributy – např. první jazyk je zjištěn přibližně u 86 % textů.⁶¹

Zmínili jsme již, že se uvádí také úroveň znalosti češtiny podle SERRJ. Přehled složení podle úrovní podáváme v tabulce 12. Rozdíl oproti korpusu MERLIN spočívá v tom, že se charakteristika úrovně vztahuje přímo k autorovi, a ne k produkovanému textu. Udává se tedy úroveň jazykové znalosti autora v době vzniku textu. Nebere se tedy v potaz, že i autor s jistou znalostí může (např. v závislosti na zadání úkolu) vyprodukovat text, jehož hodnocení nebude dané úrovni odpovídat. Může být horší, ale i lepší. Za podstatnější slabinu, která využitelnost těchto metadat limituje, považujeme ale to, že informaci o úrovni uváděli

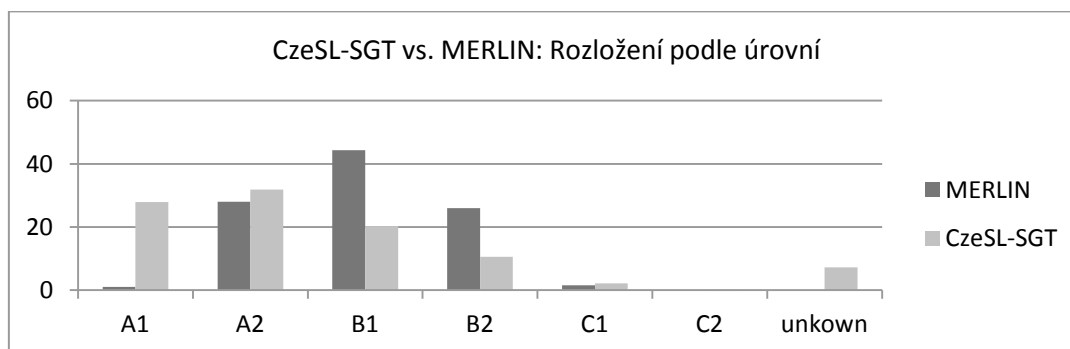
⁶¹ Podrobnosti lze najít zde: <http://wiki.korpus.cz/doku.php/cnk:czesl-sgt>, [cit. 15. 3. 2017].

sami sběrači textů (učitelé, lektori apod.)⁶² a hodnocení nebylo žádným způsobem ověřeno. To budeme mít také na paměti při práci s daty z tohoto korpusu.

Tabulka 12: Složení a rozsah korpusu CzeSL-SGT

úroveň	počet textů	počet tokenů
A1	2 609	82 839
A1+	315	277 532
A2	2 098	42 341
A2+	570	271 735
B1	1 481	94 342
B2	745	232 675
C1	123	120 795
C2	1	24 711
unknown	675	507
celkem	8 617	1 147 477

Srovnáme-li rozložení korpusu MERLIN a CzeSL-SGT podle úrovní (srov. graf 4), můžeme identifikovat jádro korpusu MERLIN v úrovních A2–B2 (s vrcholem u B1) a jádro korpusu CzeSL-SGT v úrovních nižších (A1–B1). Druhý zmíněný korpus tedy pokrývá úroveň A1, jež v korpusu MERLIN chybí.

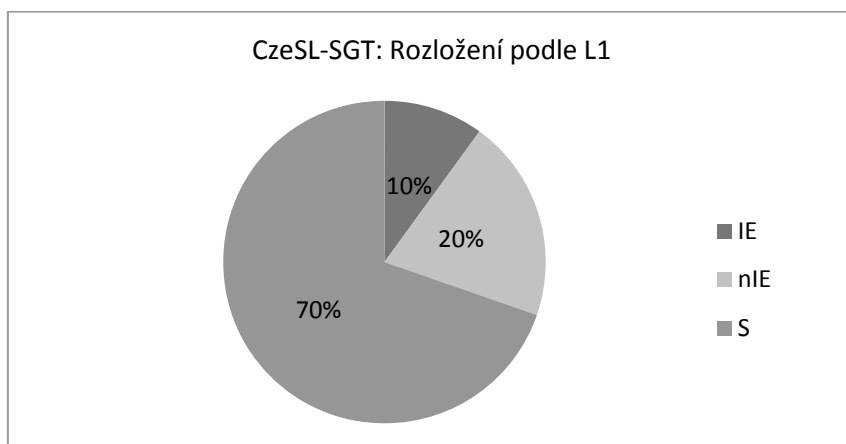


Graf 4: CzeSL-SGT vs. MERLIN: Rozložení podle úrovní (% podíl v celém korpusu na základě počtu tokenů)

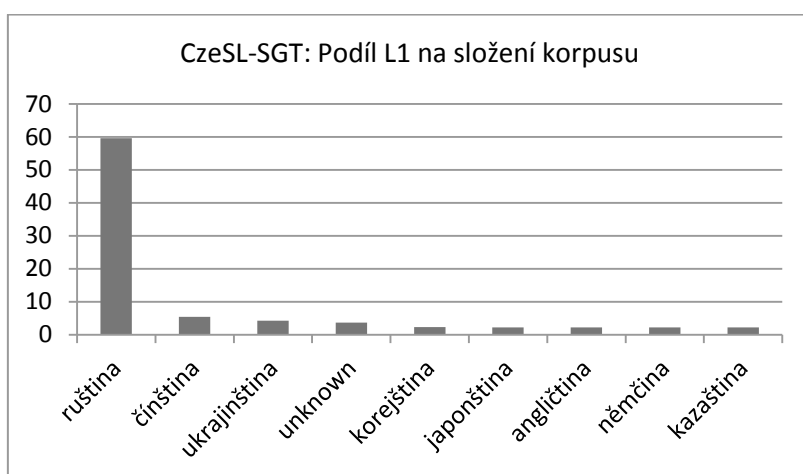
Co se mateřského jazyka týče, převahu tvoří texty autorů se slovanským mateřským jazykem (70 % – podle počtu tokenů), nejvíce ze všech je zastoupena ruština (srov. graf 6). Více než v korpusu MERLIN se sice objevují také texty autorů s neindoevropským jazykem (graf 5), ale i v souhrnu s texty autorů s indoevropským mateřským jazykem jsou neslovanské

⁶² Diskutabilní je jejich poučenost o úrovních podle SERRJ i přístup, kterým k rozhodování o úrovni přistupovali. Mohli se rozhodovat například podle zařazení do kurzu, podle užívané učebnice, absolvovaných testů apod.

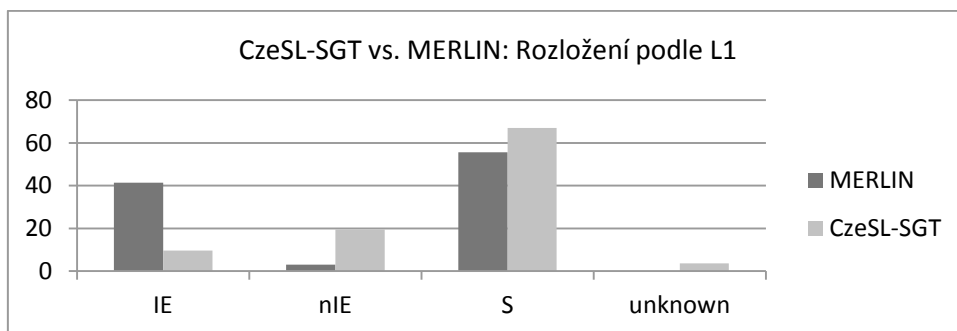
jazyky zastoupeny pouze 30 %. Po celkovém zhodnocení tak můžeme konstatovat, že složení korpusu CzeSL-SGT z hlediska L1 je méně vyvážené než v korpusu MERLIN (graf 7).



Graf 5: CzeSL-SGT: Podíl L1 na složení korpusu (% podle počtu tokenů)



Graf 6: CzeSL-SGT: Podíl L1 na složení korpusu (% podle počtu tokenů)



Graf 7: CzeSL-SGT vs. MERLIN: Rozložení podle L1 (% podle počtu tokenů)

3.1.2.3 Anotace

Anotace korpusu CzeSL-SGT není tak dalekosáhlá a propracovaná jako u korpusu MERLIN.⁶³ Slovní tvary jsou označeny slovním druhem, morfologickými kategoriemi a základním tvarem (lemmatem). Některé tvary jsou opraveny a výsledná podoba textu znovu slovnědruhově a morfologicky označena. Na základě porovnání původní a opravené podoby tvarů je stanoven druh chyby. Identifikace chyb zahrnuje pouze ortografickou rovinu (srov. Rosen 2014). Jak upozorňují autoři, „všechny tyto údaje jsou určeny automaticky, je tedy třeba počítat s nepřesnostmi a omyly“. (Rosen 2015, s. 1)

3.1.3 Korpus SKRIPT2012

Další korpus, který jsme se rozhodli využít pro dílčí sondy našeho výzkumu je SKRIPT2012. Ten obsahuje přepisy slohových prací českých žáků na různých úrovních školní docházky. Jedná se tedy o další korpus akviziční povahy, neboť obsahuje projevy mluvčích, kteří si češtinu osvojují. Na rozdíl od předchozích dvou popsanych korpusů jsou ale autory rodilí mluvčí češtiny. Nabízí se tedy možnost srovnávacího pohledu na užití spojovacích prostředků během vývoje jazykové kompetence u rodilých a nerodilých mluvčích.⁶⁴

3.1.3.1 Charakter a velikost korpusu

SKRIPT2012 obsahuje celkem 708 668 pozic (bez komentářů v závorce 707 511, bez interpunkce 587 857) a zahrnuje přepisy písemných prací českých žáků, které vznikly v souvislosti s jazykovým vyučováním na základních a středních školách od 5. třídy ZŠ (resp. ekvivalentních ročníků víceletých gymnázií) po poslední ročník gymnázií.

⁶³ Ve fázi před zveřejněním je nyní další verze korpusu CzeSL s označením CzeSL-MAN. Ta operuje stejně jako korpus MERLIN s vícerovinnou anotací. Pro anotaci bylo celkem definováno třídvačacet manuálních značek a devět chybových značek přiřazovaných automaticky. „V prvním kroku, tj. při opravě z tzv. roviny 0 (R0), která reprezentuje transkribovanou podobu originálního textu bez chybových značek, na rovinu 1 (R1), se opravují chyby primárně ortografické a morfologické, resp. řeší se opravy izolovaných tvarů, neexistující formy se opravují na formy existující a výsledkem je řetěz existujících českých forem. Druhým krokem, tzn. při emendaci z R1 na rovinu 2 (R2), je oprava chyb, které plynou ze syntaktických vztahů, chyb lexikálních, v užití jazykového jevu, chyb slovosledných, stylových apod., jejímž výsledkem je gramaticky korektní věta.“ (Petkevič a kol., 2012, s. 64–65)

⁶⁴ Ostatně s využitím korpusu SKRIPT2012 ke srovnání tohoto typu počítali i sami jeho tvůrci. Srov. Šebesta (2010, s. 20.).

3.1.3.2 Metadata a složení korpusu

Korpus SKRIPT2012 obsahuje díky svému zaměření řadu sociolingvistických i didaktických informací: o sondě, škole, třídě, textu, žáku a učiteli nebo o hodnocení.⁶⁵ Pokud bychom se stejně jako u korpusů nerodilých mluvčích chtěli nějakým způsobem vztahovat k úrovni jazykové kompetence, přicházely by v úvahu informace o stupni školní docházky (třídě, ročníku) autora textu a o hodnocení konkrétní písemné práce.⁶⁶ Z tabulky 13 je patrné, že převažují práce středoškolských studentů. Rozložení mezi jednotlivými třídami a ročníky na ZŠ a SŠ je rovnoměrné.

Tabulka 13: SKRIPT2012: Složení korpusu podle stupně výuky

		Počet textů	Počet slov	Počet tokenů
ZŠ	5. tř.	197	25 208	30 084
	6. tř.*	125	20 148	24 185
	7. tř.*	92	21 229	25 267
	8. tř.*	120	20 865	24 777
	9. tř.*	113	20 368	24 775
SŠ	1. roč.*	300	121 746	147 678
	2. roč.*	286	120 435	144 380
	3. roč.*	299	117 377	140 909
	4. roč.*	162	120 481	146 613

* Zahrnuje texty z příslušného ročníku víceletých gymnázií.

3.1.3.3 Anotace

Texty v korpusu SKRIPT2012 jsou lemmatizovány a opatřeny morfosyntaktickou anotací běžnou v korpusech ČNK. Prozatím neobsahují chybovou anotaci (automatickou ani manuální).

⁶⁵ Podrobnosti lze nalézt zde: http://wiki.korpus.cz/doku.php/cnk:statistiky_ke_korpusu_skript2012, [cit. 15. 3. 2017].

⁶⁶ Ani jedna informace ale není směrodatná pro náš výzkum. Zařazení do třídy/ročníku je víceméně formální na základě věku a pokroku ve výuce podle školních vzdělávacích programů. Hodnocení konkrétních prací je zase do značné míry subjektivní, nemá společná kritéria a nebylo standardizováno. V naší práci se tedy budeme vztahovat pouze ke dvěma souhrnným skupinám – texty žáků ZŠ a texty studentů SŠ.

3.1.5 Korpusy SYN

Za účelem získání příkladových vět pro určité oddíly teoretické části předkládané práce a také pro účely doplňujících frekvenčních analýz jsme využívali korpusy SYN. Volili jsme reprezentativní korpusy SYN2005, SYN2010 a SYN2015 a pro některé příklady jsme sáhli také do nereprezentativního korpusu SYN 4. Domníváme se, že korpusy SYN není třeba podrobněji představovat, uvedeme tedy níže pouze údaje relevantní pro tuto práci a v ostatních oblastech odkazujeme na přístupně zpracovanou dokumentaci jednotlivých korpusů na webových stránkách ČNK.⁶⁷

3.1.5.1 Charakter a velikost korpusu

Na rozdíl od předchozích popsanych korpusů (kap. 3.1.1–3) nemají korpusy SYN akviziční charakter. Představují soubory psaných textů autorů-profesionálů (spisovatelů, publicistů, akademiků atd.). Každý z korpusů řady SYN zachycuje především jazyk posledních pěti let, jež předcházely jeho zveřejnění (tabulka 14).

Tabulka 14: Korpusy SYN

korpus	velikost (počet slov)	charakteristika korpusu
SYN (verze 4)	3,626 mld.	verzovaný korpus, spojující synchronní psané korpusy řady SYN (SYN2000, SYN2005, SYN2006PUB, SYN2009PUB, SYN2010, SYN2013PUB, SYN2015) a další dosud nezveřejněné texty
SYN2015	100 mil.	referenční reprezentativní korpus, převažují texty z let 2010–2014
SYN2010	100 mil.	referenční reprezentativní korpus, převažují texty z let 2005–2010
SYN2005	100 mil.	referenční reprezentativní korpus, převažují texty z let 2000–2004

3.1.5.2 Metadata a složení korpusu

Metadata korpusů SYN se týkají především původu textů (žánr, zdroj, tematický okruh...). Na nejobecnější rovině stojí vždy poměr zastoupení tří hlavních textových typů – beletrie (zahrnující krásnou literaturu v nejširším pojetí: prózu, poezii a drama), oborová literatura (zahrnující odborné texty vědecko-naučné i popularizující, učební texty a profesní literaturu)

⁶⁷ Podrobnosti lze dohledat zde: <http://wiki.korpus.cz/doku.php/cnk:syn>, [cit 15. 3. 2017].

a publicistika (zahrnující denní tisk a další periodika a neperiodické publicistické texty). Poměrem mezi nimi se od sebe korpusy SYN liší.

Tyto rozdíly máme na paměti při srovnávání frekvence spojovacích prostředků a upozorňujeme na ně. Nevyužíváme však srovnávací frekvenční seznamy,⁶⁸ a to proto, že musíme operovat i s víceslovnými spojovacími výrazy, které seznamy neobsahují.

3.1.5.3 Anotace

Korpusy SYN jsou lemmatizovány a morfologicky tagovány. Jednotlivé korpusy se od sebe ovšem liší pokročilostí některých nástrojů automatického zpracování textu a také způsobem různých postupů s ním spojených. Jedná se např. o vylepšení disambiguace⁶⁹ nebo odlišný způsob lemmatizace.⁷⁰ U korpusu SYN2015 se poprvé v korpusech řady SYN objevuje automatická anotace syntaktická vycházející z koncepce *Pražského závislostního korpusu* (PDT)⁷¹. Nutno podotknout, že předkládané práce se žádná z vývojových úprav v anotaci korpusu SYN zásadnějším způsobem nedotýká.

⁶⁸ Srovnávací frekvenční seznamy uvádějí mimo jiné celkovou přepočítanou frekvenci. Ta udává, jaká by byla frekvence daného slova ve 100milionovém srovnávacím korpusu, ve kterém by byly všechny tři hlavní textové skupiny (beletrie, oborová literatura, publicistika) zastoupeny rovnoměrně (tj. každý přesně jednou třetinou) při zachování takové průměrné četnosti v rámci každé textové skupiny, která odpovídá příslušnému korpusu. Srovnávací frekvenční seznamy jsou dostupné zde: http://wiki.korpus.cz/doku.php/seznamy:srovnavaci_seznamy, [cit 15. 3. 2017].

⁶⁹ Disambiguace (zjednoznačnění) představuje proces, při kterém se homonymním výrazovým prostředkům přiřazuje na základě kontextu jedna konkrétní interpretace, např. *jí* – tvar zájmena *ona* vs. tvar slovesa *jíst*. K této problematice viz např. Petkevič (2006).

⁷⁰ Vyvíjel se například způsob lemmatizace interpunkčních znamének – od přístupu, který sjednocoval zápis různých druhů uvozovek, závorka apod., k přístupu, jež tuto variabilitu zachovává v rámci atributu *word* a sjednocenou podobu uvádí až v atributu *lemma*. (srov. <http://wiki.korpus.cz/doku.php/cnk.syn2015>, [cit 15. 3. 2017])

⁷¹ Pro více informací o anotacích na analytické rovině PDT viz <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/layer/html/index.html>, [cit 15. 3. 2017].

3.2 MOŽNOSTI VYHLEDÁVÁNÍ SPOJOVACÍCH PROSTŘEDKŮ

V teoretické části předkládané práce jsme ukázali, jak rozmanitou skupinu spojovací prostředky tvoří (kap. 2.1) a v předchozí kapitole (3.1.) jsme také představili různorodost korpusů, s nimiž budeme pracovat. Obojí dohromady nás vede k zamyšlení nad tím, jakým způsobem můžeme spojovací prostředky vyhledávat. Je zcela zřejmé, že bude třeba uplatnit promyšlený přístup k volbě a zadávání vyhledávacích dotazů.

V jednotlivých oddílech této kapitoly ukážeme, jak lze pro zjištění repertoáru prostředků využít korpusový dotazovací jazyk a sestavovat dotazy s využitím pozičních atributů, konkrétně tagů, jež zachycují gramatické informace o hledaném slovu v konkrétním kontextu.⁷² Zároveň upozorníme na některá omezení, se kterými jsme se museli v průběhu prací vypořádat.

3.2.1 Autosémantické spojovací prostředky

V případě autosémantických spojovacích prostředků, přichází v úvahu využití následujících morfologických značek (tabulka 15).⁷³ S jejich pomocí lze vyhledat všechny výskyty zájmen, která vždy plní spojovací funkci.

Tabulka 15: Tagy pro vtažná zájmena

tag	popis
[tag=".1.*"]	vztažné přivlastňovací zájmeno <i>jehož, jejíž, ...</i>
[tag=".J.*"]	vztažné zájmeno <i>jenž, již, ...</i> ne po předložce
[tag=".9.*"]	vztažné zájmeno <i>jenž, již, ...</i> po předložce („ <i>n-</i> “: <i>něhož, níž, ...</i>)
[tag=".E.*"]	vztažné zájmeno <i>což</i>

Omezeně lze využít také další značky (tabulka 16), které identifikují vybraná relativa a interogativa.

⁷² V korpusech ČNK se jako dotazovací jazyk užívá CQL (Corpus Query Language – více informací viz http://wiki.korpus.cz/doku.php/pojmy:dotazovací_jazyk, cit. 1. 2. 2017). Korpus MERLIN je provozován v rozhraní ANNIS a pro tvorbu dotazů je k dispozici AQL (ANNIS Query Language – <http://corpus-tools.org/annis/aql.html>, cit. 1. 2. 2017). Oba jazyky jsou založeny na dotazech na jednu pozici (slovo) v korpusu. Dotazy mají podobu [atribut="hodnota"], resp. atribut=/hodnota/, kde atributem se myslí atribut poziční (word, lemma, tag apod.), hodnotou je pak hledaný výraz sám, nebo vzor specifikovaný pomocí regulárního výrazu. Součástí dotazu mohou být i omezení na strukturní atributy (např. věta, dokument apod.).

⁷³ Podrobně viz http://wiki.korpus.cz/doku.php/seznamy:tagy#popis_jednotlivych_pozic_znacky, [cit. 1. 2. 2017].

Tabulka 16: Tagy pro vztažná nebo tázací zájmena

tag	popis
[tag=".4.*"]	vztažné nebo tázací zájmeno s adjektivním skloňováním (obou typů: <i>jaký, který, čím, ...</i>)
[tag=".K.*"]	zájmeno tázací nebo vztažné <i>kdo</i> , vč. tvarů s „-ž“ a „-s“
[tag=".Q.*"]	zájmeno tázací nebo vztažné <i>co, copak, cožpak</i>

Ostatní autosémantické spojovací prostředky nemají přiřazen speciální morfologický tag, ale můžeme je získat vyhledáním konkrétního výrazu jako lemmatu⁷⁴ (se zohledněním varianty s velkým i malým počátečním písmenem). Eventuálně můžeme operovat s dotazem v CQL pro zahrnutí variant rozšířených o *-pak* (např. [lemma="kde" | lemma="kdepak"]).

Je třeba mít na paměti, že do výsledků jsou (kromě dotazů v tabulce 15) zahrnuty jak doklady, kde výraz v daném kontextu plní spojovací funkci, tak doklady, které nepředstavují spojovací relaci. Specifikací dotazu bohužel nedojdeme k uspokojivému omezení výsledků pouze na doklady se spojovací funkcí. S vědomím tohoto nedostatku, resp. nepřesnosti se pro dílčí sondy doplněné o ruční selekci dá uvažovat o následujících variantách dotazu.

(a) Vyhledáváním daného výrazů po čárce, např. [word=","][lemma="kde"] získáme doklady typu (40) (naš příklad = 97 158 dokladů v SYN2015).

(40) *Všude v kraji, **kde** jsem se narodil, je kvůli únikům z nádrží otrávená pitná voda.*
(SYN2015)

Chceme-li zohlednit i případy, kdy po čárce před daným výrazem stojí jiný výraz/y (nejčastěji spojka souřadící, příslovce nebo částice), můžeme zvolit například:

(b) vyhledávání daného výrazu po čárce včetně dokladů s jedním libovolným výrazem na pozici mezi nimi, např. [word=","][{0,1}[lemma="kde"]. V SYN2015 získáme 98 305 dokladů typu (41), ale narazíme také na doklady jako (42) nebo (43), které spojovací funkci nemají.

(41) *„Nemáme tušení, co se s ní stalo, **ani kde** jsou její ostatky,“ zabručel Banville.“*
(SYN2015)

(42) *„Ale prosím vás,“ řekl, „**kde** byste byl?“* (SYN2015)

(43) *Každý den na vás čekají souhrny dění, dozvíte se tak například, **kolik kde** stojí skipasy, jaké jsou podmínky pro lyžování v zahraničí a kam se vydat.* (SYN2015)

⁷⁴ Tím zohledníme i ostatní tvary ohebných spojovacích prostředků.

(c) Následně by bylo třeba postihnout také případy, kdy věta závislá uvozená autosémantickým spojovacím prostředkem předchází větě řídicí. Formulovat by se dal například dotaz `[word="Kde"][]*[word=","] within <s />`, kterým získáme doklady *kde* ve spojovací funkci (44), ale opět i některé bez ní (45). V SYN2015 je dohromady celkem 3 802 dokladů.

(44) *Od západu sem míří chladné mořské proudy omývající úpatí sopečných břehů.*

Kde se objeví, tam se díky přísunu živin explozivně rozrůstá plankton. (SYN2015)

(45) *Kde jsi, Bobillote! Savorgnane! Chanoine! Pomoc!* (SYN2015)

Bylo by jistě možné vytvářet i mnohé další restriktce nebo různé kombinace dotazů ve snaze co nejvíce eliminovat v dokladech vyhledaných na základě lemmatu nespojovací struktury (např. přímé řeči). Avšak v této práci se o to pokoušet nebudeme. Vyhneme se tak možností (b), (c) i možným dalším, kdy vyhledané doklady mohou častěji zahrnovat i užití nespojovací, a spokojíme se s využitím výše zmíněné varianty (a), tj. vyhledáním daného spojovacího výrazu po čárce. S vědomím jisté nedostatečnosti pro nás budou doklady získané v korpusech rodilých mluvčích na základě tohoto typu reprezentovat užití daného autosémantického spojovacího prostředku, o němž víme, že nutně nemusí plnit spojovací funkci.⁷⁵ Považujeme to za přijatelné řešení, které nám v případě konzistentní aplikace umožní mimo jiné i srovnání frekvence dotčených spojovacích prostředků v různých korpusech. Budeme tímto způsobem vždy srovnávat stejně definované podmnožiny se spojovací funkcí (čárka + daný výraz) ze skupiny všech výskytů daného výrazu.⁷⁶

3.2.2 Syntaktické spojovací prostředky

Vzhledem k tomu, že jádro syntaktických spojovacích prostředků tvoří spojky, lze k jejich vyhledání použít několik hlavních tagů (tabulka 17).

Tabulka 17: Tagy pro spojky jednoslovné

tag	popis
<code>[tag="J.*"]</code>	spojka
<code>[tag=".\^.*"]</code>	spojka souřadící

⁷⁵ V korpusech nerodilých mluvčích je třeba tento způsob vyhledávání uplatnit na rovině, kde už proběhla oprava základní oprava chyb. Jinak hrozí zkreslení výsledků chybami v užití interpunkce, kterých se nerodili mluvčí často dopouštějí.

⁷⁶ Kupříkladu v SYN2105 činí celkový počet výskytů výrazu *kde* 115 261, při omezení na doklady daného výrazu po čárce (`[word=","][lemma="kde"]`) jich získáme 97 158.

[tag=".,.*"]	spojka podřadicí (vč. <i>aby, kdyby</i> ve všech tvarech)
--------------	---

Uvedené tagy vyhledávají pouze jednoslovné spojky, pro vyhledání víceslovných prostředků je třeba dotazy upravit a zohlednit také to, že spojky se mohou v kontaktní pozici kombinovat jak s částicemi (a), tak příslovci (b) a zájmeny (c).

(a) Pro vyhledávání víceslovných spojovacích prostředků v kontaktní pozici je možné užít kombinaci výše uvedených tagů (spojka + spojka), např. [tag="J.*"] [tag="J.*"]. V korpusu SYN2015 takto získáme 1 259 různých spojovacích prostředků, prvních 10 nejfrekventovanějších uvádí tabulka 18. Nejhojněji zastoupené (ze 60 kombinací) v korpusu MERLIN jsou uvedeny v tabulce 19. V obou korpusech se však v první desítce objevují kromě víceslovných spojovacích prostředků také kombinace se spojkou *a* v koordinačním vztahu (*a že, a když...*) a různovětné spojky (*že když*).

Tabulka 18: SYN2015: Nejfrekventovanější víceslovné kombinace spojek

pořadí	kombinace spojek
1	<i>i když</i>
2	<i>ale i</i>
3	<i>a že</i>
4	<i>a když</i>
5	<i>a proto</i>
6	<i>že když</i>
7	<i>ale když</i>
8	<i>že i</i>
9	<i>a pokud</i>
10	<i>a i</i>

Tabulka 19: MERLIN: Nejfrekventovanější víceslovné kombinace spojek

pořadí	kombinace spojek
1	<i>a proto</i>
2	<i>i když</i>
3	<i>a že</i>
4	<i>a když</i>
5	<i>ale když</i>
6	<i>ale i</i>
7	<i>že když</i>
8	<i>a nebo</i>
9	<i>a i</i>
10	<i>proto že</i>

U následujících dotazů se již více projevuje problematika slovnědruhového zařazení komponentů víceslovných spojovacích prostředků, o níž jsme pojednali v kapitole (2.1.2.1). Výsledky vyhledávání na základě dotazů specifikujících pouze slovní druh výrazu obsahují i kombinace slov, které dohromady spojovací prostředek netvoří.

(b) Víceslovné spojovací prostředky mohou být kombinací spojky a částice – [tag="T.*"] [tag="J.*"]. Vedle typických případů souřadících spojovacích prostředků stupňovacích a podřadicích prostředků srovnávacích vět narazíme také na případy, kdy se nejedná o víceslovný spojovací prostředek, ale o spojení částice a spojky, kde je více či méně

zachovaná aktualizací funkce částice. V korpusu SYN2015 najdeme mezi prvními 10 nejfrekventovanějšími kombinacemi (tabulka 20) tři takové příklady jako (46)–(48).

Tabulka 20: SYN2015: Nejfrekventovanější kombinace částice a spojky

pořadí	kombinace částice a spojky
1	<i>dokonce i</i>
2	<i>jako kdyby</i>
3	<i>jako když</i>
4	<i>no a</i>
5	<i>dokonce ani</i>
6	<i>ne že</i>
7	<i>jen aby</i>
8	<i>jen jako</i>
9	<i>až když</i>
10	<i>spíš jako</i>

Tabulka 21: MERLIN: Nejfrekventovanější kombinace částice a spojky

pořadí	kombinace částice a spojky
1	<i>Jen když</i>
2	<i>jenom abychom</i>
3	<i>nejen že</i>
4	<i>ne když</i>
5	<i>nejen proto</i>
6	<i>nejen protože</i>
7	<i>zejména když</i>
8	<i>ba ani</i>
9	<i>jenom jestli</i>

(46) *Prostě se tam furt něco láme a kazí. **No a** letos v létě odešla ventilace.* (SYN2015)

(47) *Své bratrance a sestřenice navštěvovali rádi, proto je nebylo těžké pro celý nápad získat, **ne že** by skutečně měli na výběr.* (SYN2015)

(48) *Vlastně tady nejsou skoro žádní a někdy si připadám **spíš jako** noční hlídač než prodavač.* (SYN2015)

(c) Některé souřadící spojovací prostředky důsledkově představují kombinaci spojky a příslovce (např. *a tak*) nebo spojky a zájmena (*a tím*). Jejich vyhledávání pomocí tagů ale není příliš efektivní, protože výsledky obsahují i velké množství kombinací, které víceslovné spojovací prostředky nepředstavují. Srov. například 10 nejfrekventovanějších kombinací ve výsledcích na dotaz [tag="J.*"] [tag="D.*"] v korpusu SYN2015 (tabulka 22).

Tabulka 22: SYN2015: Nejfrekventovanější kombinace spojky a příslovce

pořadí	kombinace spojky a příslovce
1	<i>a pak</i>
2	<i>a tak</i>
3	<i>a také</i>
4	<i>ale také</i>
5	<i>a jak</i>
6	<i>a teď</i>
7	<i>že už</i>
8	<i>a ještě</i>

9	<i>a zároveň</i>
10	<i>a už</i>

Víceslovné spojovací prostředky v distantním postavení lze vyhledáváním pomocí morfologických tagů postihnout ještě obtížněji. Možným pokusem by mohlo být vyhledávání dvou spojek v rámci jedné věty s minimálním odstupem jedné pozice mezi nimi – `[tag="J.*"][]*[tag="J.*"] within <s/>`. To ovšem nepřinese uspokojivý výsledek, protože existuje mnoho kombinací, kdy se nebude jednat o souvztažná spojení výrazů, srov. vedle souvztažného (49) také nesouvztažné (50).

(49) *Ve volných chvílích si **bud'** vyšla do Apostolisovy kavárny **nebo** jen tak bloumala a snila.* (SYN2015)

(50) *Postupně žehnal **a** podával ruce poutníkům, **až** došel dopředu.* (SYN2015)

V případě již předem vytipovaných víceslovných spojovacích výrazů však lze aplikovat hledání kombinace konkrétních výrazů v rámci jedné věty, např. `[lemma="bud'"][]*[lemma="nebo"] within <s/>`.

3.2.3 Shrnutí

V této kapitole jsme ukázali některé z postupů, které je možno použít k vyhledávání spojovacích prostředků v korpusech ČNK a MERLIN. Zabývali jsme se hlavně možnostmi vyhledávání pomocí tagů – morfologických značek, jež primárně vede ke zjištění repertoáru užitých výrazů. Na příkladech jsme ilustrovali, kdy může být vyhledávání pomocí obecné slovnědruhové charakteristiky nedostatečné. Problematizovali jsme tím vyhledávání vztažných a tázacích zájmen a vztažných zájmených příslovcí ve skupině autosémantických spojovacích prostředků. U sysémantických prostředků jsme se zamýšleli především nad vyhledáváním víceslovných prostředků v kontaktní i distantní pozici.

3.3 PRÁCE S KORPUSEM MERLIN (VYHLEDÁVÁNÍ, ZPRACOVÁNÍ DAT...)

Poté, co jsme v předešlé kapitole nastínili, ale i problematizovali možnosti vyhledávání spojovacích prostředků v korpusech, popíšeme nyní postup, který jsme použili pro účely našeho výzkumu. Ten byl založen na kombinaci metody vyhledávání s využitím korpusového jazyka a ručního zpracování žákovských textů. Podrobněji o aplikaci obou metod pojednáme v oddílech uspořádaných podle toho, co bylo cílem vyhledávání.

3.3.1 Zjištění repertoáru spojovacích prostředků

Abychom co nejjednodušeji získali přehled o širší repertoáru spojovacích prostředků v žákovských textech, tedy o tom, které různé spojovací prostředky jsou v textech zastoupeny, snažili jsme se v maximální míře využít vyhledávání s pomocí morfologických tagů uvedené v kapitole 3.2.

U synsémantických spojovacích prostředků jsme získali výpis výrazů označených jako spojky, který jsme filtrovali tak, abychom odstranili opakující se výrazy. Při pohledu na výsledný seznam, bylo jasné, že se projevila slabina automatické morfologické anotace, kterou tvůrci korpusu na žákovské texty aplikovali. Téměř třetina z vyhledaných výrazů ve skutečnosti nebyla spojkami (např. *blízko, často, Němečko, odpočíváme, photo*). Některé výrazy byly také kvůli pravopisné chybě zahrnuty do výpisu dvakrát (např. *zatimco / zatímco*). Z těchto důvodů byla nutná ruční korekce, kterou jsme došli ke konečné podobě soupisu zastoupených jednoslovných spojovacích prostředků v tradiční platnosti spojky (viz kap. 4.1.1).

Víceslovné synsémantické spojovací prostředky kontaktní jsme vyhledávali kombinacemi dotazů s tagy pro spojky, částice a příslovce (viz kap. 3.2.2) a následným filtrováním jedinečných výrazů. Následovalo rovněž i ruční vyloučení kombinací, které neodpovídaly požadované charakteristice. Šlo zejména o případy koordinačně spojených hypotaktických prostředků (51) nebo různovětných spojek (52).

(51) *Těším se, že máte docela dobře a že mi pozváš!* (MERLIN)

(52) *Ale práce musí být zajímavá pro člověka, který pracuje, protože když dlouho pracuješ na neoblíbené práci, mužiš být neshastný.* (MERLIN)

Zastoupené distantní spojovací prostředky jsme získali manuální analýzou všech dokladů obsahujících spojku. U nich jsme zjišťovali, zda je tato spojka součástí souvztažného spojení dvou výrazů. Na základě dotazů obsahujících konkrétní výrazy jsme nakonec zvlášť vyhledali očekávaná souvztažná spojení ve srovnávacích strukturách *čím – tím, jak – tak*.

Úplnost souhrnu prostředků získaných na základě uvedených operací jsme ověřili ruční analýzou všech textů. Během ní jsme také doplnili sysémantické spojovací prostředky, jež nejsou tradičně pojímány jako spojky (např. *sice, -li*). Tím jsme získali kompletní souhrn sysémantických spojovacích prostředků pro další analýzy.

Ruční analýzu textů, při níž jsme kontrolovali, zda nebyl opomenut některý ze sysémantických spojovacích prostředků, jsme využili také pro vyhledání spojovacích prostředků autosémantických. Vycházeli jsme tedy z ručně sestaveného souboru zastoupených výrazů, který jsme následně porovnali se soupisem očekávaných autosémantických spojovacích prostředků (srov. kap. 2.1.1.1). Pokud některý z očekávaných prostředků zastoupen nebyl, ověřili jsme vyhledáváním na základě konkrétního slova, zda nebyl při ruční analýze nedopatřením vynechán.

Podíváme-li se na doposud popsaný postup pro stanovení repertoáru spojovacích prostředků, je patrné, že i přes maximální snahu využít možnosti vyhledávání v korpusu bylo nutné texty procházet i ručně. Situaci totiž zkomplikoval fakt, že automatická morfologická anotace provedená na žákovských textech, vykazovala nedostatky. Manuální práce s texty v korpusu byla možná pouze díky jejich omezenému množství a rozsahu, a i přesto představovala značnou časovou investici.

Jakmile byl soupis všech spojovacích prostředků, jež se v korpusu vyskytují, hotový, bylo možné provést další (již většinou automatizované) operace. Definováním subkorpusů jsme docílili sestavení dílčích soupisů použitých spojovacích prostředků podle různých kritérií založených na metadatech v korpusu. Zajímal nás repertoár prostředků podle:

- (a) celkové úrovně hodnocení podle SERRJ
- (b) úrovně hodnocení dílčích kritérií podle SERRJ
- (c) hodnocení koheze a koherence

3.3.2 Totožné výrazy se spojovací funkcí vs. bez spojovací funkce

Když jsme se věnovali teoretickým východiskům předkládané práce (kap. 2.1.), poukázali jsme na případy, kdy se dva totožné výrazy liší svou syntaktickou funkcí. Máme na mysli

zejména ty případy, kdy je třeba odlišit výraz se spojovací funkcí a výraz, který spojovací funkcí nedisponuje. Z toho důvodu jsme v další fázi zpracování dokladů jednotlivých výrazů prováděli ruční selekci těchto případů.

V korpusu MERLIN se objevily tyto skupiny:

(a) U výrazů tvořených od zájmenných tázacích základů bylo nutno od sebe odlišit otázky zjišťovací (53) a užití ve funkci synsémantického spojovacího prostředku připojujícího závislé věty – zejména nepřímé otázky (54), ale i věty příslovečné nebo vztahné (55) atd.

(53) **Kde** bydlíš? (MERLIN)

(54) *Můžeš mi psát, **kde** přesně bydlíte?* (MERLIN)

(55) *Ubytovali jsme se v malém obytnem domů, **kde** jsme bydleli s paní, která nám to přenajala na 2 týdni.* (MERLIN)

(b) Další skupinu tvořily výrazy, které mohou být součástí víceslovných spojovacích prostředků synsémantických, např. příslovce *nejen, tak...* Zde jsme zkoumali, zda se skutečně jedná o užití v tomto souvztažném distantním spojovacím prostředku (56), či nikoliv (57).

(56) *To přísloví můžeme slyšet **nejen** v České republice **ale i** taky v Německu...* (MERLIN)

(57) *Dnes a také v minulosti **nejen** šaty dělali člověka.* (MERLIN)

(c) Třetí skupina obsahovala výrazy, které mohou plnit aktualizační funkci částice (58), ale někdy také funkci spojovací (59), např. *ovšem, přece, tedy*.

(58) *A co **tedy** dostaneme?* (MERLIN)

(59) *Co se týče prvního bodu, **tedy** vzhledu a jeho vlivu na celkový dojem, tak samozřejmě nemůžeme to vůbec zapírat.* (MERLIN)

(d) V další skupině se vyskytly některé výrazy, které stojí na pomezí příslovcí a spojek (srov. 2.1.2.1.2), např. *pak* (60) jako příslovce, nebo jako spojka v důsledkových souřadných vztazích (61).

(60) *První, samozřejmě, dříve musíme vařit **pak** můžeme jíst.* (MERLIN)

(61) *Evropa zásadně podporuje vzdělání a především vyuku cizích jazyků, když se zamýšlím nad samotným významem tohoto druhu vzdělání pro život člověka ve 21. století, **pak** si -unreadable- vspomenu na poslední výsledky eurovýzkumů, které*

ukazují, že nejlépe jsou vzděláni a jazykově vybaveni obyvatelé malých evropských států... (MERLIN)

(e) Objevila se také homonymie spojky *bud'* (62) s tvarem slovesa *být* v rozkazovacím způsobu (63).

(62) *Myslím, že ten stesk má každý, který bud' často cestuje nebo je v zahraničí delší dobu.* (MERLIN)

(63) *Mohla bys poslat inzerát? Bud' tak lásková.* (MERLIN)

(f) V neposlední řadě bylo třeba myslet na to, že pracujeme s texty nerodilých mluvčích, které obsahují různé typy chyb. V důsledku ortografických nepřesností vznikla také záměna tvaru slovesa *ztát se* ve 3. os. sg. (64) se spojovacím *zda*.

(64) *Samozřejmě život v bohatším místě se zda (= zdá) jako příjemnější.* (MERLIN)

Ze všech výše zmíněných případů byly nejpočetněji zastoupeny výrazy ze skupiny (a), tj. výrazy tvořené od zájmenných tázacích základů. U nich také (až na výrazy *kdy* a *kolik*) převažoval počet výskytů ve spojovací funkci nad výskyty ve funkci tázacích výrazů ve zjišťovacích otázkách. Obdobně tomu bylo s poměrem ve skupině (b), kde převažovalo užití ve víceslovných spojovacích prostředcích. Ve skupině (c) a (d) převládalo užití nespojovací, tj. ve funkci částic, resp. příslovcí. Kompletní přehled dotčených výrazů a jejich distribuce s konkrétní funkcí je patrná z tabulky 23.

Tabulka 23: MERLIN: Totožné výrazy se spojovací funkcí a bez ní

	nespojovací	spojovací	nespojovací / spojovací
bud'	1	6	_____ ■■■■■
co	108	214	_____ ■■■■■
jak	66	162	_____ ■■■■■
jaký	24	44	_____ ■■■■■
kde	25	88	_____ ■■■■■
kdo	24	49	_____ ■■■■■
kdy	111	79	■■■■■ _____
kolik	92	32	■■■■■ _____
který	10	394	_____ ■■■■■
nejen	3	25	_____ ■■■■■
nejenom	2	7	_____ ■■■■■
ovšem	4	0	■■■■■ _____
pak	64	1	■■■■■ _____
potom	59	0	■■■■■ _____
proč	9	19	_____ ■■■■■
přece	9	1	■■■■■ _____
tak	140	116	■■■■■ _____
tedy	9	4	■■■■■ _____
zda	2	9	_____ ■■■■■

3.3.3 Frekvence užití spojovacích prostředků

Vedle rozsahu repertoáru spojovacích prostředků nás samozřejmě také zajímá, jaké je četnost jejich zastoupení. Sledování frekvence užití jednotlivých prostředků nám pomůže zodpovědět otázku, jež jsme formulovali v úvodu této práce (viz 1.4).

Při výpočtu frekvence vycházíme z počtu výskytů spojovacího prostředku v daném souboru textů stanoveném na základě výše zmíněných kritérií (úroveň, mateřský jazyk autora apod.). Vzhledem k nestejně velikosti a rozsahu těchto subkorpusů (rozdílného počtu textů, slov a pozic) operujeme při srovnávání s relativní frekvencí, tzv. i.p.m (z angl. instances per million). Ta vyjadřuje průměrný počet výskytů spojovacího prostředku v hypotetickém textu/korpusu o délce 1 milion slov. Ukázkou výsledků takového přepočtu pro spojku *a* v korpusu MERLIN rozděleném na subkorpusy podle úrovně uvádíme v tabulce 24.

Tabulka 24: MERLIN: Výpočet i.p.m pro spojku *a*

úroveň	A1	A2	B1	B2	C1	celkem
počet tokenů	54	21 505	34 007	19 953	1 167	76 686
počet textů	1	188	167	81	4	441
počet výskytů	3	680	1004	559	44	2 290
i.p.m.	55 555,56	31 620,55	29 523,33	28 015,84	37 703,51	29 862,03

Na tomto místě považujeme za důležité upozornit na způsob, jakým jsou v našich přehledech započítávány výskyty výrazů, které se objevují jednak jako jednoslovné spojovací prostředky, jednak v kombinaci s jinými výrazy ve víceslovných spojovacích prostředcích (např. *a, a proto, a tedy; nebo, buď – nebo,...*). U příslušného víceslovného spojovacího prostředku jsou uvedeny všechny výskyty dané kombinace výrazů, ze kterých se skládá. Zároveň jsou výskyty jednotlivých výrazů z víceslovného prostředku započítány u dotčených jednoslovných prostředků. Kupříkladu u jednoho započteného výskytu *a proto* je zároveň započítán jeden výskyt u výrazu *a* i u výrazu *proto*.

Pouze zdánlivě se jedná o nelogický krok, který jsme ale zvolili s ohledem na zamýšlené porovnávání frekvence výskytů spojovacích prostředků v korpusu MERLIN s frekvencemi získanými v dílčích sondách podniknutých v jiných korpusech (CzeSL-SGT, korpusy SYN, atd.). Pokud bychom totiž chtěli důsledně oddělit počet výskytů dotčených výrazů stojících samostatně (jedna množina) od výskytů v rámci víceslovných jednotek (druhá množina), museli bychom u každého z využitých korpusů pořídit úplný soupis všech spojovacích prostředků. Jen na základě něho bychom mohli od sebe obě množiny výskytů oddělit. A to,

jak jsme viděli u korpusu MERLIN (viz 3.3.1), je velmi náročný úkol, který nebyl předmětem této práce.

3.3.4 Práce s chybovou anotací

V teoretické části práce (kap. 2.3.1) jsme představili principy chybové analýzy. Běžně se provádí ve čtyřech, resp. pěti krocích, jimiž jsou sběr dat, resp. výběr vzorku pro analýzu, identifikace chyby, popis chyby, její vysvětlení a zhodnocení (srov. Ellis 1994, s. 48). Již jsme také uvedli, že se předkládaná práce omezuje pouze na první tři, respektive dva z nich.⁷⁷

Identifikace chyb bude provedena smíšenou metodou. U spojovacích prostředků se slovnědruhovým zařazením ke spojkám lze využít chybovou anotaci korpusu MERLIN. V chybové taxonomii je zohledněno chybné užití spojky, kterému byl přiřazen tag „G_Conj“. Zbývající případy, tj. chybné užití všech autosémantických spojovacích prostředků a sysémantických prostředků jiného slovního druhu než spojka budou vyhledány ručně.

Pro popis chyb jsme se rozhodli využít taxonomii odrážející povrchovou realizaci (srov. kap. 2.3.1), tj. konfrontaci textů nerodilých mluvčích s jejich rekonstruovanou variantou. Zvolili jsme ji především s ohledem na očekávané typy chyb, které mohou být s užitím spojovacích prostředků spojeny. V úvahu přichází vynechání, přidání nebo záměna v rámci stejné skupiny prostředků. U autosémantických spojovacích prostředků (zájmen a zájmených číslovek) se může projevit také chybná forma. Lingvistickou taxonomii chyb předem nezavádíme, i když teoreticky by se mohla uplatnit pro doplňkové třídění (např. chybné užití parataktické vs. hypotaktické spojky). Pokud by se její užití ukázalo jako opodstatněné, přistoupíme k navržení speciální varianty parciální taxonomie na základě získaných dat (srov. tzv. data-driven přístup podle Ellis – Barkhuizen 2009, s. 60).⁷⁸

⁷⁷ Sběrem dat nebylo nutné se zabývat, protože využíváme již dostupný korpus MERLIN.

⁷⁸ Parciální taxonomie je zaměřena na specifické typy chyb podle deklarovaného výzkumného záměru.

4 ANALÝZA TEXTŮ

Tato kapitola tvoří stěžejní část předkládané práce, budeme se v ní věnovat analýze dokladů z autentických textů nerodilých mluvčích češtiny, které jsme získali na základě postupu popsaného v předchozí kapitole (kap. 3). Při třídění a zpracování se přidržujeme zásad a principů, o nichž jsme pojednali v kapitole o teoretických východiscích této práce (kap. 2).

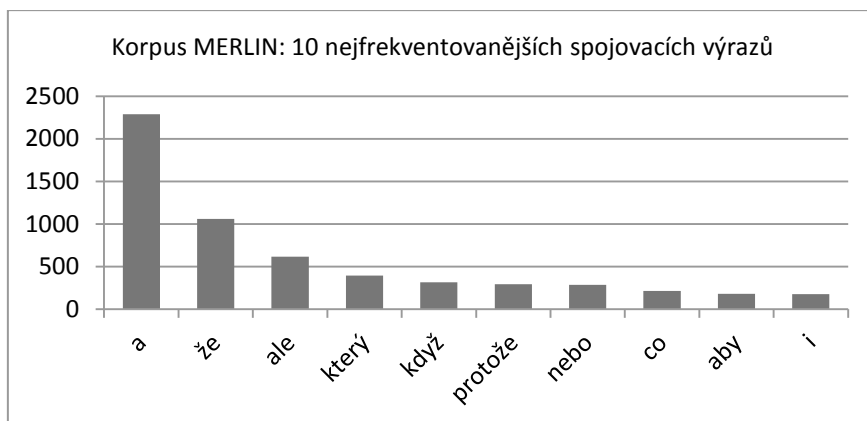
Vycházet budeme od globálního pohledu na rozložení repertoáru a na frekvenci užití konkrétních výrazů i jejich skupin. Následně obrátíme pozornost také na jednotlivé úrovně podle SERRJ, abychom zjistili, jak se v užívání spojovacích prostředků odlišují a které prostředky jsou pro ně typické. V neposlední řadě provedeme chybovou analýzu dokladů, která nám poslouží ke stanovení závislosti četnosti chyb na dosažené úrovni jazykové kompetence. Třemi doplňkovými sondami se nakonec dotkneme vybraných oblastí také v korpusech CzeSL-SGT a SKRIPT2012.

4.1 SPOJOVACÍ PROSTŘEDKY V KORPUSU MERLIN

Jak již bylo řečeno, jako hlavní zdroj materiálu jsme zvolili korpus MERLIN. V následujících podkapitolách se budeme zabývat interpretací získaných dat. Avšak ještě než k ní přistoupíme, je třeba znovu připomenout, že zatímco v globálních analýzách využíváme celý rozsah korpusu, při popisu úrovní se omezujeme pouze na úrovně A2–B2, a to proto, že ostatní úrovně nejsou reprezentovány dostatečným počtem textů (srov. kap. 3.1.1).

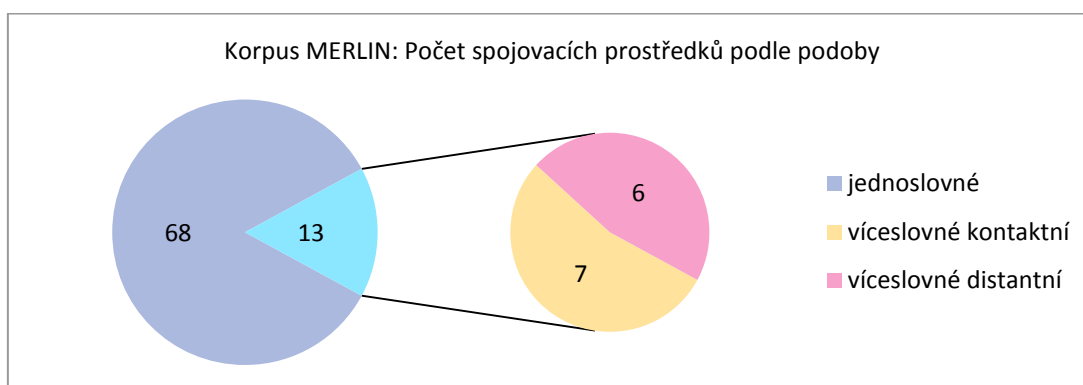
4.1.1 Repertoár spojovacích prostředků – globální pohled

Spojovací prostředky najdeme v korpusu MERLIN celkem v 7 333 výskytech. Z tohoto počtu je téměř 80 % reprezentováno pouhými 10 různými spojovacími výrazy (graf 8). Mezi nimi je 8 výrazů synsémantických (*a, že, ale, když, protože, nebo, aby, i*) a 2 jsou autosémantické (*který, co*).



Graf 8: Korpus MERLIN: 10 nejfrekventovanějších spojovacích výrazů (podle počtu)

Zbýlých 20 % výskytů se dělí mezi dalších 71 spojovacích prostředků. Korpus tedy obsahuje celkem 81 různých spojovacích prostředků.⁷⁹ Z toho je 68 jednoslovných a 13 víceslovných (7 v kontaktní pozici a 6 v distantní pozici) – graf 9.



Graf 9: Korpus MERLIN: Počet spojovacích prostředků podle podoby

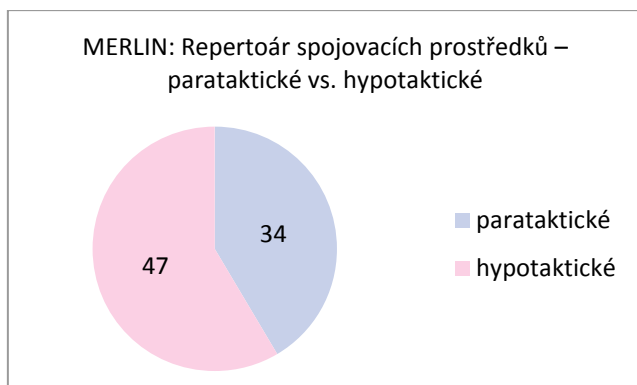
V tabulce 25 uvádíme kompletní přehled řazený podle frekvence (i.p.m.) v celém korpusu, kde je zjevné, že víceslovné spojovací prostředky se objevují v pásmu středně a méně frekventovaných.

⁷⁹ Protože vycházíme od formy spojovacího výrazu (srov. kap. 2.1), započítáváme v celkovém přehledu totožné spojovací prostředky, jež mohou být svou funkcí homonymní (např. autosémantické i sysémantické výrazy *jak* a *co*), pouze jednou. Pokud však uvádíme složení repertoáru a/nebo četnost s ohledem na funkci spojovacích prostředků, je vždy zohledněna funkce výrazu v daném kontextu.

Tabulka 25: MERLIN: Pořadí spojovacích prostředků podle frekvence

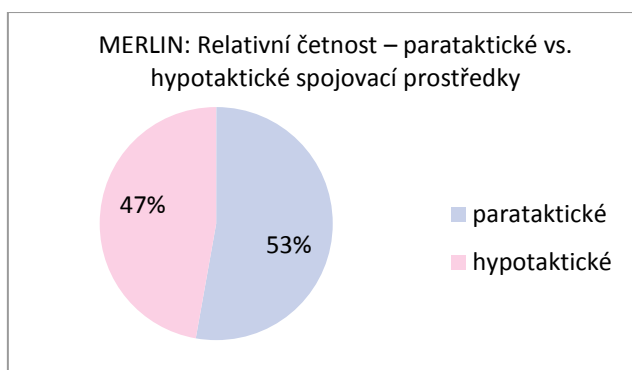
pořadí	i.p.m.	jednoslovné	víceslovné	
			kontaktní	distantní
1	29 888,1	<i>a</i>		
2	13 835,6	<i>že</i>		
3	8 032,8	<i>ale</i>		
4	5 137,8	<i>který</i>		
5	4 107,7	<i>když</i>		
6	3 833,8	<i>protože</i>		
7	3 716,5	<i>nebo</i>		
8	2 790,6	<i>co</i>		
9	2 347,2	<i>aby</i>		
10	2 321,2	<i>i</i>		
11	2 112,5	<i>jak</i>		
12	1 617,0	<i>jako</i>		
13	1 551,8	<i>proto</i>		
14	1 512,7	<i>tak</i>		
15	1 147,5	<i>kde</i>		
16	1 030,2	<i>kdy</i>		
17	1 004,1	<i>než</i>		
18	991,1	<i>jestli</i>		
19	639,0	<i>kdo</i>		
20	599,8	<i>kdyby</i>		
21	586,8	<i>ani</i>		
22	573,8	<i>jaký</i>	<i>a proto</i>	
23	430,3	<i>až</i>		
24	417,3	<i>kolik</i>		
25	326,0			<i>nejen – ale (i)</i>
26	260,8	<i>či</i>	<i>i když</i>	
27	247,8	<i>proč</i>		
28	234,7	<i>dokonce</i>		
29	221,7	<i>pokud</i> <i>takže</i> <i>však</i>		
30	195,6		<i>a tak</i>	
31	169,5	<i>což</i> <i>-li</i>		
32	156,5	<i>totiž</i>		
33	130,4		<i>jak – tak</i>	
34	117,4	<i>anebo</i> <i>zda</i>	<i>a ne</i>	
34	104,3	<i>kam</i> <i>přestože</i>		
36	91,3	<i>at'</i> <i>avšak</i> <i>jenž</i>		<i>čím – tím</i> <i>nejenom – ale (i)</i>
37	78,2	<i>přesto</i> <i>sice</i>		<i>bud' – (a)nebo</i>
38	52,2	<i>ačkoliv</i> <i>jelikož</i> <i>tedy</i>		
39	39,1	<i>aniž</i> <i>ba</i> <i>jakmile</i> <i>jestliže</i> <i>kdykoliv</i> <i>vždyť</i> <i>zatímco</i>	<i>i kdyby</i>	
40	26,1	<i>jednak</i> <i>jehož</i> <i>jenomže</i> <i>nicméně</i>	<i>proto, že</i>	
41	13,0	<i>ačkoli</i> <i>byť</i> <i>coby</i> <i>čili</i> <i>jakoby</i> <i>jakkoli</i> <i>jakožto</i> <i>nýbrž</i> <i>pak</i> <i>přece</i> <i>třebaže</i> <i>tudíž</i>	<i>a tedy</i>	<i>bud' to – (a)nebo</i>

V repertoáru prostředků jsou početněji zastoupeny spojovací prostředky hypotaktické. V uvedeném seznamu všech spojovacích prostředků z korpusu najdeme více než polovinu těch, které připojují větu závislou k větě řídicí (graf 10).



Graf 10: MERLIN: Repertoár spojovacích prostriedkú – počet parataktické vs. hypotaktické

Podíváme-li se ale na to, jak jsou tyto dva typy spojovacích prostriedkú zastoupeny počtem výskytú, které daný typ obsahují, vyjví se nám poměr opačný (graf 11). V korpusu je sice obsaženo více různých výrazú pro vyjádření hypotaktických vztahú, ale počet dokladú, ve kterých se objevují, je nižší než počet dokladú, které reprezentují parataxi. Na počtu parataktických spojení má totiž významný podíl spojka *a*, která je v korpusu zastoupena nejpočetněji ze všech spojovacích prostriedkú. Objevuje se ve více než 30 % všech dokladú spojovacích relací. Toto zjištění ovšem není nijak překvapující, protože spojka *a* je nejfrekventovanějším slovem v češtině vůbec. V korpusech řady SYN zaujímá v žebříčku všech slov řazených podle frekvence shodně první pozici. Její odstup od nejbližšího spojovacího prostriedku (spojka *že*), který v pořadí následuje na sedmé pozici, navíc představuje víc než trojnásobek počtu výskytú tohoto výrazu.



Graf 11: MERLIN: Relativní četnost – parataktické vs. hypotaktické spojovací prostriedky

Z hlediska větněčlenské platnosti převažují spojovací prostriedky sysémantické nad autosémantickými (graf 12). Během analýzy dokladú se potvrdil předpoklad formulovaný v teoretické části předkládané práce (kap. 2.1.2.1.1) a u dvou spojovacích prostriedkú bylo zjištěno, že se v korpusu vyskytují ve funkci autosémantické i sysémantické. Jde o výrazy *co*

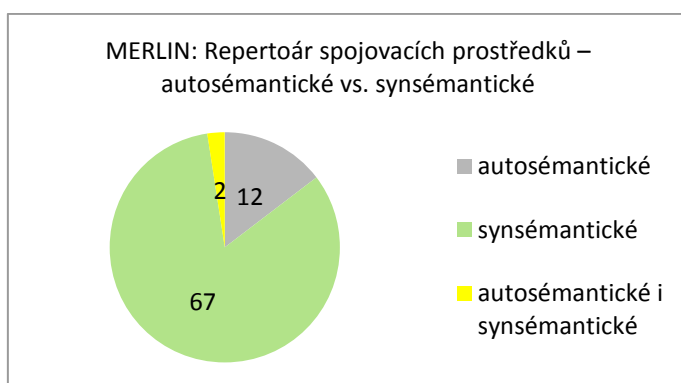
a *jak* stojící někdy jako relativa ve funkci větného členu (65)–(66) a jindy jako spojky bez větněčlenské platnosti (67)–(68).⁸⁰

(65) *V posledním emaile jsi se zeptala **co** dělám.* (MERLIN)

(66) *Obraz ukazuje pěkným způsobem, **jak** (→jakým způsobem) si užívat léto.* (MERLIN)

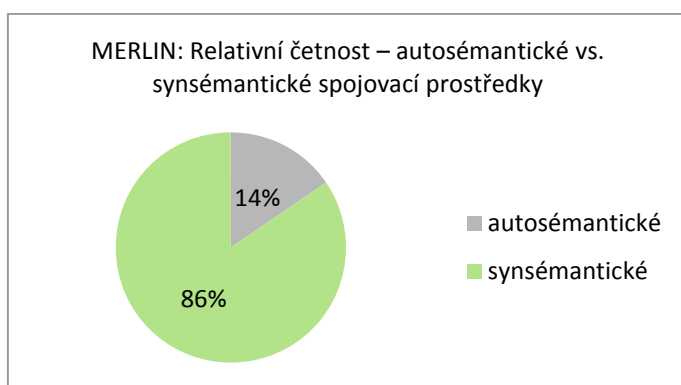
(67) *Rozdíl mezi minulostí a současností je v tom, že současná situace, **co** se manuální práce týče, hodně změnila kvůli počítačům [...].* (MERLIN)

(68) ***Jak** (→jakmile) budeš už v Draždanech a budeš mít čas, zatelefonuj mi a setkáme se na městě nebo mě domů.* (MERLIN)



Graf 12: MERLIN: Repertoár spojovacích prostředků – autosémantické vs. synsémantické

Synsémantické spojovací prostředky mají v korpusu nejen širší repertoár, ale jsou také celkově frekventovanější než výrazy autosémantické (graf 13). To koresponduje s jejich převažujícím zastoupením v první desítce nejčastěji užívaných prostředků (8 ku 2).



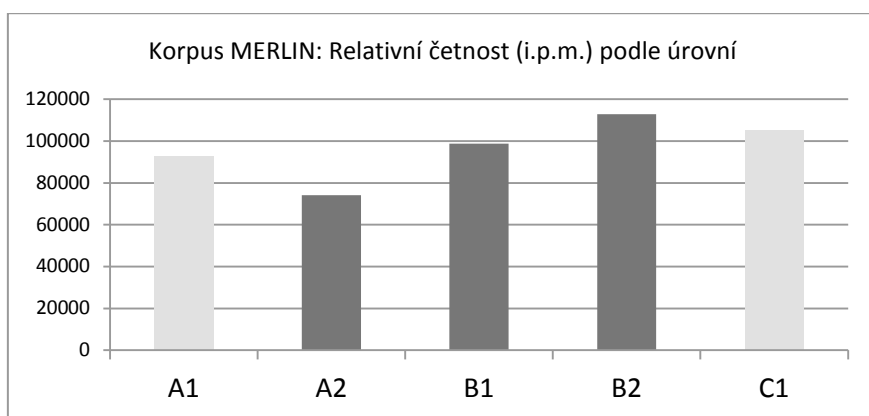
Graf 13: MERLIN: Relativní četnost – autosémantické vs. synsémantické spojovací prostředky

⁸⁰ Jako synsémantické byly započítány i příklady, kdy byl spojovací prostředek sice užit nekorektně, ale z kontextu vyplývalo, o jaké užití se jedná. Např. *Přijedu na víkend, jak (→jestli) to není problém?* (MERLIN)

4.1.2 Repertoár a četnost podle úrovně (celkové hodnocení)

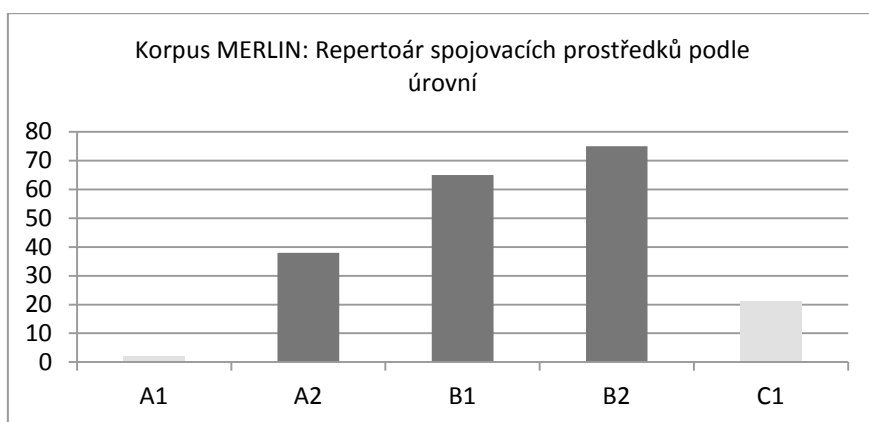
Předchozí podkapitola poskytla globální pohled na spojovací prostředky obsažené v korpusu MERLIN. Máme již představu o celkové šíři repertoáru i frekvenci užití jednotlivých výrazů, o podílu různých druhů prostředků podle formálně syntaktických vztahů, které vyjadřují, i podle jejich větněčlenské platnosti. V této podkapitole se na výše uvedené oblasti podíváme podrobněji a rozšíříme přehled o hledisko úrovně hodnocení písemné produkce podle SERRJ. Přitom si budeme všimnout jak toho, v čem jsou získaná data pro jednotlivé úrovně podobná, tak toho, čím se od sebe odlišují. Tento postup nám umožní identifikovat charakteristické rysy syntaxe mezijazyka daných úrovní.

Podíváme-li se na zastoupení spojovacích prostředků na jednotlivých úrovních, zjistíme, že jejich počet s rostoucí úrovní znalosti češtiny stoupá (graf 14).⁸¹



Graf 14: Korpus MERLIN: Relativní četnost (i.p.m.) spojovacích prostředků podle úrovní

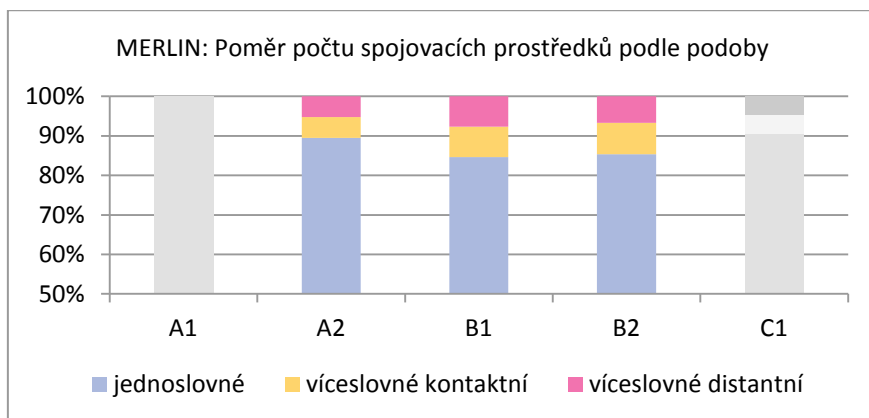
Zároveň se na vyšších úrovních rozšiřuje také repertoár použitých výrazů. Z pouhých 38 spojovacích prostředků na úrovni A2 až po 75 na úrovni B2 (graf 15).



Graf 15: Korpus MERLIN: Repertoár (počet různých) spojovacích prostředků

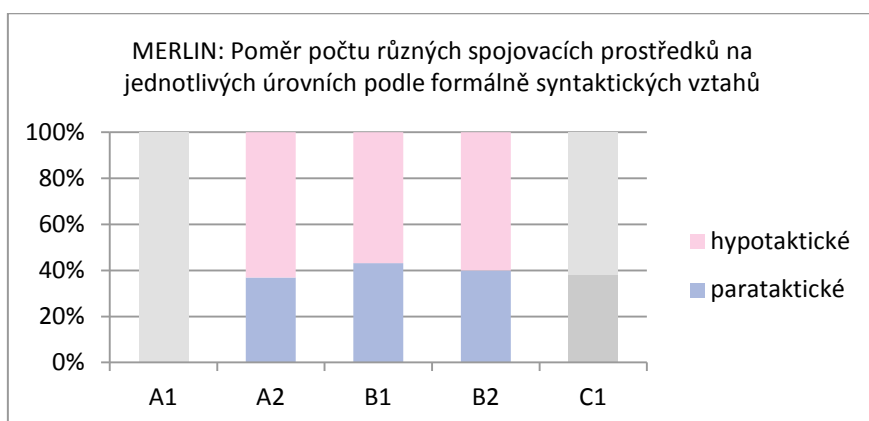
⁸¹ Upozorňujeme na tomto místě znovu na omezení v interpretacích a závěrech pro úrovně A1 a C1, které jsou v korpusu MERLIN zastoupeny pouze okrajově. Jádro korpusu tvoří úrovně A2, B1 a B2 (viz kap. 3.1.1.1).

Zastoupení spojovacích prostředků podle podoby se mění při přechodu z úrovně A2 na úroveň B1, kdy přibývají víceslovné spojovací prostředky (graf 16). Mezi úrovní B1 a B2 k výraznější proměně nedochází. Jednoslovné prostředky tvoří cca 85 % z celkového počtu a víceslovné se dělí o zbývajících 15 % téměř stejným poměrem.

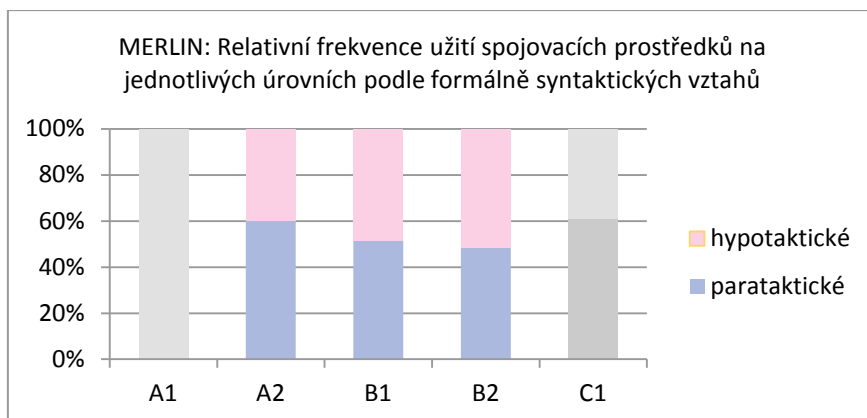


Graf 16: MERLIN: Poměr počtu spojovacích prostředků podle podoby

Poměr počtu různých spojovacích prostředků v repertoáru podle jejich funkce (parataktické vs. hypotaktické) je na všech úrovních obdobný (graf 14). Jinými slovy lze říci, že se repertoár s rostoucí úrovní rozšiřuje stejnou měrou o parataktické i hypotaktické prostředky. Co se týká frekvence užití, můžeme sledovat, že se stoupající úrovní roste počet dokladů s hypotaktickým spojením vět (graf 15). U úrovně B2 je pak poměr v zastoupení téměř vyrovnaný.

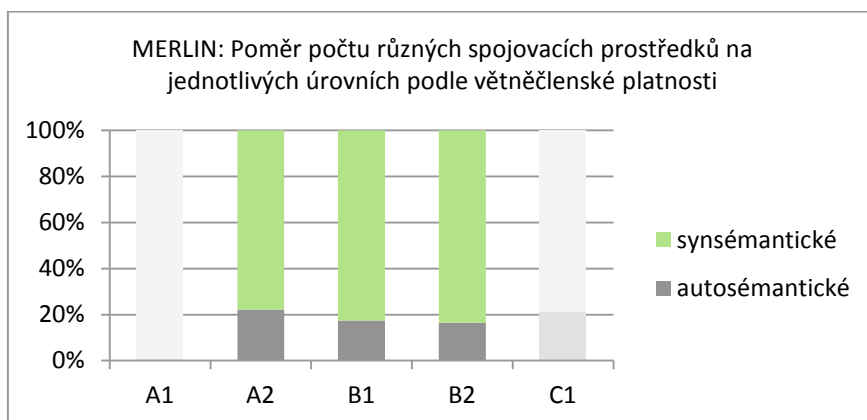


Graf 17: MERLIN: Poměr počtu různých spojovacích prostředků na jednotlivých úrovních podle formálně syntaktických vztahů (repertoár)

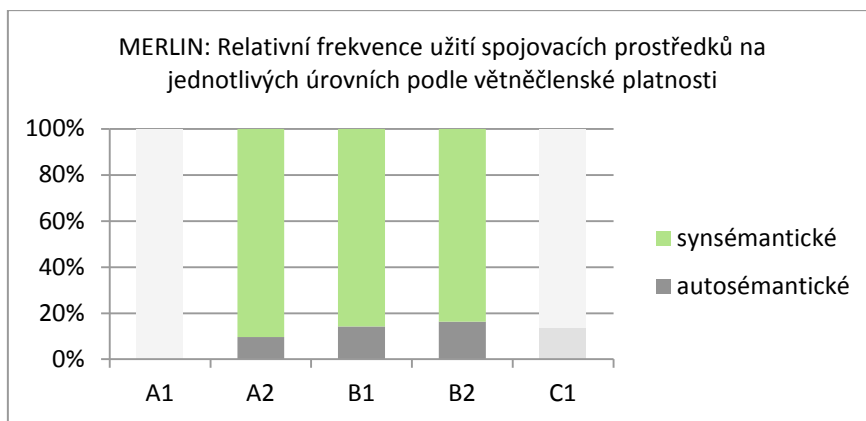


Graf 18: MERLIN: Relativní frekvence užití spojovacích prostředků na jednotlivých úrovních podle formálně syntaktických vztahů

Postupme nyní k charakteristice úrovní podle větněčlenské platnosti spojovacích prostředků, které autoři textů užívají. Již jsme zmínili výše, že v repertoáru korpusu MERLIN se celkově nachází více sysémantických než autosémantických spojovacích prostředků. Zatímco nové sysémantické spojovací prostředky obohacují repertoár každé úrovně, počet autosémantických výrazů v repertoáru se od úrovně B1 téměř nemění (graf 19). I když se repertoár rozšiřuje o sysémantické výrazy, četnost jejich užití je relativně nízká. A protože zároveň výrazněji roste také frekvence užití některých autosémantických spojovacích prostředků (*který, jaký, co*), je v celkovém počtu dokladů na úrovních B1 a B2 patrný zvyšující se podíl autosémantických (vztažných) spojení (graf 20).



Graf 19: MERLIN: Poměr počtu různých spojovacích prostředků na jednotlivých úrovních podle větněčlenské platnosti (repertoár)



Graf 20: MERLIN: Relativní frekvence užití spojovacích prostředků na jednotlivých úrovních podle větněčlenské platnosti

Od skupinových charakteristik se nyní přesuneme k detailnější analýze složení repertoáru na jednotlivých úrovních. Pohledem na nejfrekventovanější spojovací prostředky podle úrovní odhalíme, že se na prvních deseti pozicích každé z nich (kromě A1) vždy objevuje sedm výrazů (*a, ale, když, který, nebo, protože, že*) (tabulka 25). Od úrovně B1 pak přibývají ještě další dva (*co* a *aby*). Všechny uvedené spojovací prostředky (s výjimkou *protože*) obsazují první pozice (do 10. místa) i v psaných korpusech řady SYN (viz zde jako sloupec SYN2015). Z hlediska mezivětných vztahů, které navozují, se jedná o výrazy s relativně vyhraněnou funkcí (paratactická spojení: slučovací – *a*, odporovací – *ale*, vylučovací – *nebo*; hypotactická spojení: vztažná – *který*, obsahová – *že*, příčinná – *protože*, účelová – *aby*). Budeme je proto v rámci našeho dalšího výzkumu považovat za spojovací prostředky základní-frekventované.

Tabulka 26: MERLIN a SYN2015: 10 nejčtenějších spojovacích prostředků

	MERLIN					celkově	SYN 2015
	A1	A2	B1	B2	C1		
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>

Na opačném pólu, tedy na posledních místech podle frekvence se vyskytují především spojovací prostředky hypotaktické nebo některé varianty parataktických víceslovných prostředků (např. *a tedy*, *bud'to – (a)nebo*). Z velké části se jedná o konkurenční varianty frekventovanějších výrazů (*ačkoli – i když*; *nýbrž – ale* apod.), kterých přibývá s tím, jak se repertoár na vyšších úrovních rozšiřuje. V tabulce 27 uvádíme přehled spojovacích prostředků, které jsou v korpusu MERLIN zastoupeny pouze jednotlivými výskyty. Z ní můžeme usuzovat, že na úrovni B2 je repertoár prostředků již značně rozvinutý a uživatelé jazyka na této úrovni jsou schopni využívat i méně frekventované možnosti navození hlavních mezivětných vztahů.

Tabulka 27: MERLIN: Nejméně zastoupené spojovací prostředky

A1	A2	B1	B2	C1	celkově
			<i>ačkoli</i>		<i>ačkoli</i>
		<i>a tedy</i>			<i>a tedy</i>
	<i>bud'to – (a)nebo</i>				<i>bud'to – (a)nebo</i>
			<i>byť</i>		<i>byť</i>
			<i>coby</i>		<i>coby</i>
		<i>čili</i>			<i>čili</i>
			<i>jakoby</i>		<i>jakoby</i>
		<i>jakkoli</i>			<i>jakkoli</i>
			<i>jakožto</i>		<i>jakožto</i>
			<i>nýbrž</i>		<i>nýbrž</i>
			<i>pak</i>		<i>pak</i>
		<i>přece</i>			<i>přece</i>
			<i>třebaže</i>		<i>třebaže</i>
			<i>tudíž</i>		<i>tudíž</i>

Na závěr této kapitoly se pokusíme shrnout dosavadní zjištění v přehledové tabulce (tabulka 28). Z ní by měly být patrné charakteristické rysy jednotlivých úrovní v souvislosti s rozsahem a složením repertoáru spojovacích prostředků podle funkce a typu na úrovních s dostatečným počtem dokladů (tj. A2–B2). Uvádíme vždy hlavní rysy, kterými se daná úroveň odlišuje od ostatních.

Tabulka 28: Globální charakteristika úrovní podle spojovacích prostředků

	A2	B1	B2
frekvence užití	kontinuálně roste společně se zvyšující se úrovní		
rozsah repertoáru	podstatně se rozšiřuje		rozšiřuje se
typy prostředků podle podoby		zvyšuje se poměr zastoupení víceslovných prostředků	
formálně syntaktické vztahy	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
funkčně syntaktické hledisko		přibývají synsémantické a v menší míře také autosémantické prostředky	přibývají hlavně synsémantické prostředky
	roste frekvence užití autosémantických spojovacích prostředků		

4.1.3 Tendence v závislosti na úrovni (celkové hodnocení)

Již během globálního popisu úrovní v předchozí podkapitole jsme narazili na doklady toho, že se od sebe jednotlivé úrovně odlišují, a to jak složením repertoáru, tak frekvencí konkrétních spojovacích prostředků. Omezíme-li se pouze na úrovně A2 až B2, u nichž máme reprezentativnější vzorek dat, můžeme vyzorovat jisté tendence v užívání spojovacích prostředků v závislosti na úrovni osvojení češtiny. Hlavní z nich popíšeme v této podkapitole.

4.1.3.1 Složení repertoáru

V souvislosti s konstatováním, že se repertoár spojovacích prostředků kontinuálně rozšiřuje společně se zvyšující se úrovní, nás zajímá, které výrazové prostředky přibývají. Podle toho lze totiž usuzovat, kterým mezivětným významovým vztahům se na dané úrovni dostává explicitního vyjádření.⁸² Na úrovni B1 se oproti úrovni A2 objevují vztahy srovnání a přirovnání pomocí víceslovných spojovacích prostředků (69)–(70). Autoři textů na úrovni

⁸² V úvahu bereme pouze čteněji zastoupené výrazy a stranou ponecháváme ty, které se na dané úrovni vyskytují pouze jednotlivě (srov. tabulka 27).

B1 také začínají užívat nepřímé rozkazovací věty se spojkou *at'* (71). Poprvé se také objevují konstrukce se vztažným zájmenem *jehož* a zájmeno *což* vztahující se k obsahu celé řídicí věty (72). Na úrovni B2 můžeme nově najít vztažné zájmeno *jenž* a také přípustkové věty s kondicionálem – *i kdyby* (73).

- (69) *Hodně lidí dobře se oblikaji a dělaji špatne věci, nikdy hodně porušuji **jak** zakon, **tak** i dustojnost člověka.* (MERLIN)
- (70) *Všechný lidi ví, že **čím** více lidi **tím** více nazorů.* (MERLIN)
- (71) *Když přijelo na začátku k placení, řekla jsem manželovi, **at'** zaplatí jenom jeden týden, protože ještě nevíme jaké bude počasí, atd'...* (MERLIN)
- (72) *Fakt je, že opravdu realně přítele mám možna jenom jeden a nebo dva, **což** je hodně!* (MERLIN)
- (73) ***I kdybychom** chtěli někam utéct a obránit se proti informacím, nebyli bychom schopni najít nějakou divokou jeskyně.* (MERLIN)

Celkový přehled všech nově se objevivších spojovacích prostředků podává tabulka 29.

Tabulka 29: MERLIN: Změny v repertoáru podle úrovní (pořadí podle frekvence)

nově se objevivší			chybějící oproti nižší úrovni	
A2	B1	B2	B1	B2
<i>že</i>	<i>či</i>	<i>jenž</i>	<i>bud'to – (a)nebo</i>	<i>jakmile</i>
<i>který</i>	<i>i když</i>	<i>i kdyby</i>		<i>a tedy</i>
<i>když</i>	<i>však</i>	<i>vždyť</i>		<i>čili</i>
<i>protože</i>	<i>což</i>	<i>zatímco</i>		<i>jakkoli</i>
<i>nebo</i>	<i>jak – tak</i>	<i>jenomže</i>		<i>přece</i>
<i>co</i>	<i>a ne</i>	<i>nicméně</i>		
<i>aby</i>	<i>at'</i>	<i>proto, že</i>		
<i>i</i>	<i>avšak</i>	<i>ačkoli</i>		
<i>jako</i>	<i>čím – tím</i>	<i>byť</i>		
<i>proto</i>	<i>nejenom – ale (i)</i>	<i>coby</i>		
<i>tak</i>	<i>bud' – (a)nebo</i>	<i>jakoby</i>		
<i>kde</i>	<i>přesto</i>	<i>jakožto</i>		
<i>jak</i>	<i>sice</i>	<i>nýbrž</i>		
<i>kdy</i>	<i>ačkoliv</i>	<i>pak</i>		
<i>než</i>	<i>jelikož</i>	<i>třebaže</i>		
<i>jestli</i>	<i>tedy</i>	<i>tudiž</i>		
<i>kdo</i>	<i>aniž</i>			
<i>kdyby</i>	<i>ba</i>			
<i>ani</i>	<i>jakmile</i>			
<i>jaký</i>	<i>jestliže</i>			
<i>a proto</i>	<i>kdykoli(v)</i>			
<i>až</i>	<i>jednak</i>			
<i>kolik</i>	<i>jehož</i>			
<i>nejen – ale (i)</i>	<i>a tedy</i>			
<i>proč</i>	<i>čili</i>			
<i>dokonce</i>	<i>jakkoli</i>			
<i>pokud</i>	<i>přece</i>			
<i>takže</i>				

a tak
-li
totiž
anebo
zda
kam
přestože
bud'to – (a)nebo

Kromě toho, že uživatelé jazyka na vyšších úrovních dokážou explicitně vyjádřit větší množství mezivětných vztahů, disponují také alternativními prostředky k vyjadřování vztahů užívaných již na nižších úrovních. V mezijazyce uživatelů se tak postupně tvoří řady konkurenčních spojovacích prostředků, které se na vyšších úrovních rozrůstají o výrazy schopné vyjadřovat i speciální významové nebo stylistické odstíny těchto spojení. V korpusu MERLIN máme tento proces doložený na úrovni B1 i B2. Na úrovni B1 dochází k rozšiřování ve více sémantických skupinách jak u parataktických, tak hypotaktických spojení; u B2 se rozvíjejí především hypotaktické skupiny přípustkových a důsledkových prostředků (tabulka 30). Tendence v rozšiřování skupin konkurenčních prostředků s ohledem na úroveň shrneme v tabulce 31.

Tabulka 30: MERLIN: Skupiny konkurenčních prostředků

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

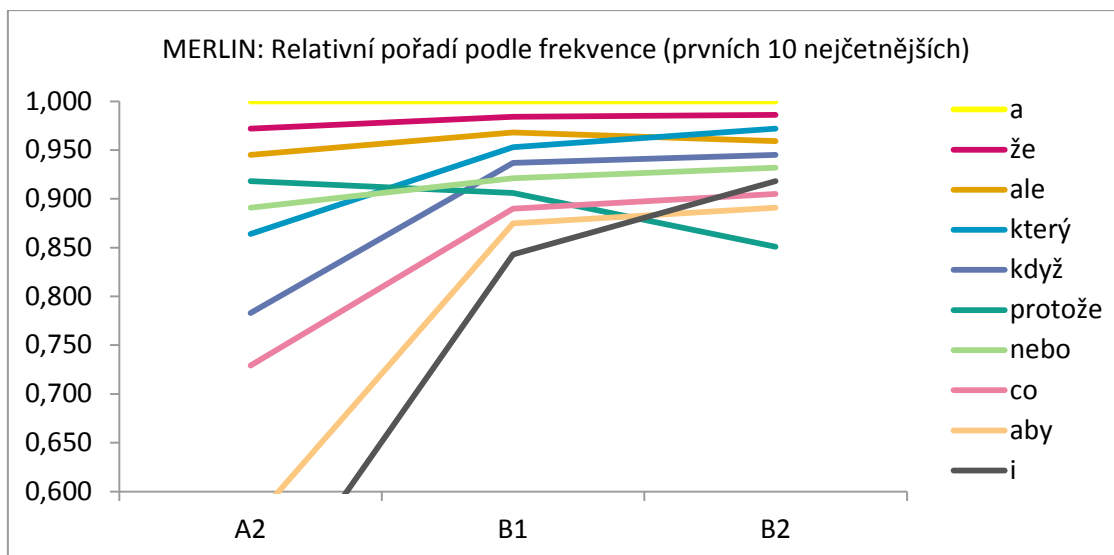
Tabulka 31: MERLIN: Tendence v rozšiřování skupin konkurenčních prostředků

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

4.1.3.2 Frekvence spojovacích prostředků

Výše jsme již narazili na to, že první pozice na všech úrovních obsazuje víceméně uzavřená skupina spojovacích prostředků (viz kap. 4.1.1). Můžeme ale ještě podrobněji sledovat, jak se do jejich frekvence, resp. pořadí podle frekvence promítá zvyšující se úroveň jazykové kompetence. Na předních pozicích pořadí podle relativní frekvence (graf 21) si víceméně setrvalou pozici bez ohledu na úroveň udržují tři spojky (*a*, *ale*, *nebo*), které slouží k vyjadřování parataktických vztahů, a také hypotaktická obsahová spojka *že*.⁸³ Toto zjištění není překvapující, spojky *a* a *že* jsou, jak už jsme zmínili, vůbec nejfrekventovanější i v korpusech řady SYN a před ostatními výrazy mají značný početní předstih. Spojky *ale* a *nebo* zase vyjadřují základní mezivětné vztahy (odporovací a vylučovací) a oproti ostatním konkurenčním výrazům jsou ve svých skupinách nejuniverzálnější a významově bez speciálních příznaků.

⁸³ Protože se počet různých spojovacích prostředků v repertoáru na jednotlivých úrovních liší, užíváme pro srovnání pořadí percentil frekvence (i.p.m.) pro jednotlivé spojovací prostředky v rámci dané úrovně. Tím se zohlední různý počet prvků v množinách, pro které se pořadí stanovuje, a je možné je mezi sebou porovnat. Percentil totiž udává relativní umístění vzhledem k ostatním posuzovaným. Např. u spojky *proto* bychom získali absolutní pořadí (rank) podle hodnot i.p.m.: A2 = 10 z 38; B1 = 14 z 65, B2 = 13 z 75, zatímco percentil podle i.p.m. činí: A2 = 0,756; B1 = 0,796 a B2 = 0,837. Tyto hodnoty převedeny na procenta (75,6 %; 79,6 %; 83,7 %) vlastně udávají, kolik procent jiných spojovacích prostředků se na příslušné úrovni nachází v pořadí za spojkou *proto*. Je tedy zřejmé, že se tato spojka se stoupající úrovní posouvá na vyšší pozici před ostatní spojovací prostředky.



Graf 21: MERLIN: Relativní pořadí podle frekvence (percentil podle i.p.m.) – prvních 10 nejčtetnějších

U jiných spojovacích prostředků se pořadí podle frekvence proměňuje v závislosti na úrovni. Patrný je předěl mezi úrovní A2 a B1, kdy se repertoár rozšiřuje téměř o 70 % a zároveň se některé spojovací prostředky výrazněji posouvají v pořadí dopředu. Na úrovni B1 se v textech začínají častěji objevovat vztažné věty připojené pomocí *který* a také časové věty s *když*. Do první desítky na této úrovni nově pronikají výrazy *co*, *aby* a *i*, které se dále posouvají výše i na úrovni B2. Nárůst výskytů spojky *aby* může souviset s tím, že se na úrovni B1/B2 ve výuce češtiny jako cizího jazyka, obvykle zařazuje kondicionál a jeho užití v účelových větách (srov. např. ČZKZK 2009, ČpCB2 2013). U výrazu *co* se na úrovni B1 projevuje nejen větší zastoupení vztažných vět, ale také dovednost užívat jej v různých typech konstrukcí. Zatímco na úrovni A2 se ve velké většině setkáváme s užitím autosémantického *co* v nepřímých otázkách (74) a vztažných větách (75), na úrovni B1 je početněji zastoupený i výskyt synsémantického *co*, a to především v některých ustálených konstrukcích typu *co se týká* (76) nebo *co + superlativ* (77).

(74) *V posledním emaile jsi se zeptala co dělám.* (MERLIN)

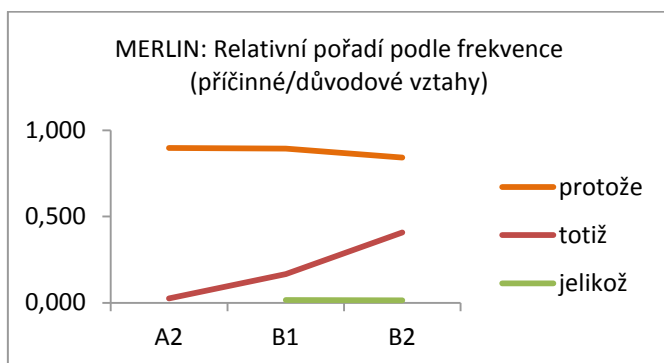
(75) *Koupím na darěk něco co potřebuje duchodci.* (MERLIN)

(76) *Co se týká diplomky, už začínat psát diplomku.* (MERLIN)

(77) *Doufám že uvidíme se co nejdříve.* (MERLIN)

U spojky *i* se do četnosti částečně promítá obohacení repertoáru o víceslovné spojovací prostředky (*i když; nejenom – ale i*).

Klesající tendence se u výrazů v první desítce podle celkové frekvence projevila pouze u výrazu *protože*. Možnou příčinou by mohla být preference parataktického vyjadřování důvodových vztahů na vyšších úrovních, čemuž by napovídala zvyšující se frekvence výrazu *totiž* (viz graf 22). Navíc ani v korpusech SYN se *protože* neobjevuje mezi první desítkou spojovacích prostředků. Můžeme se tedy domnívat, že je tento typ spojovacího prostředku na nižších úrovních nadužívaný a že se teprve na vyšších úrovních dostává do pozadí, a přibližuje se tak stavu užívání v korpusech rodilých mluvčích.



Graf 22: MERLIN: Relativní pořadí podle frekvence (příčinné/důvodové vztahy)

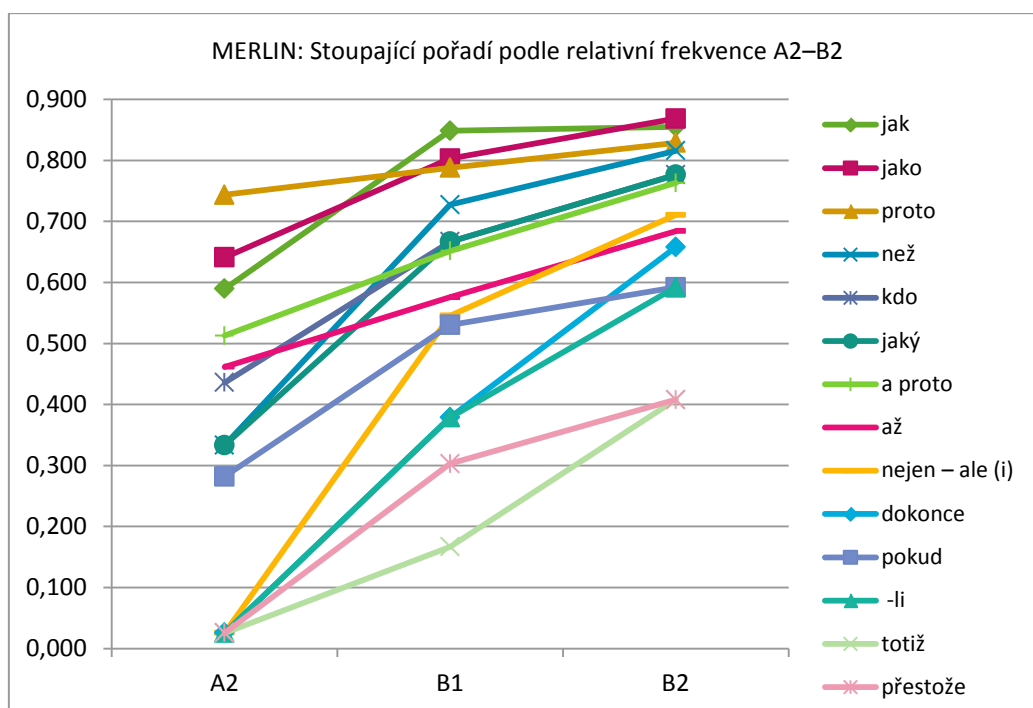
Rozšířme nyní záběr analýzy i na ostatní výrazy, které se v celkovém pořadí podle frekvence pohybují od 10. pozice níže. Pro tuto analýzu můžeme vyčlenit dvě skupiny spojovacích prostředků podle toho, zda se objevují na všech třech úrovních A2–B2, nebo zda v repertoáru přibyly až u úrovni B1.

(A) Spojovací prostředky z první skupiny nalezneme v repertoáru již od úrovně A2, můžeme tedy srovnávat tendence v jejich užívání napříč třemi úrovněmi, tj. až do B2. Na jedné straně stojí prostředky, které vykazují různě silnou stoupající tendenci, tj. jejich frekvence se zvyšuje jak mezi úrovní A2 a B1, tak mezi úrovní B1 a B2. Na druhé straně stojí výrazy, jejichž frekvence v roli spojovacích prostředků na vyšších úrovních klesá. Uprostřed pak stojí skupina prostředků, kde není interpretace trendu jednoznačná. Přehled všech tří skupin udává tabulka 32. Výrazy jsou v ní řazeny sestupně podle celkové frekvence v korpusu, tmavší podbarvení signalizuje větší rozdíl frekvence mezi úrovní A2 a B2. Vlevo jsou uvedeny výrazy se stoupající tendencí, v pravém sloupci výrazy s tendencí klesající. Prostřední sloupec zahrnuje výrazy, kde se neprojevila vyhraněná tendence.

Tabulka 32: MERLIN: Tendence v pořadí podle relativní frekvence A2–B2

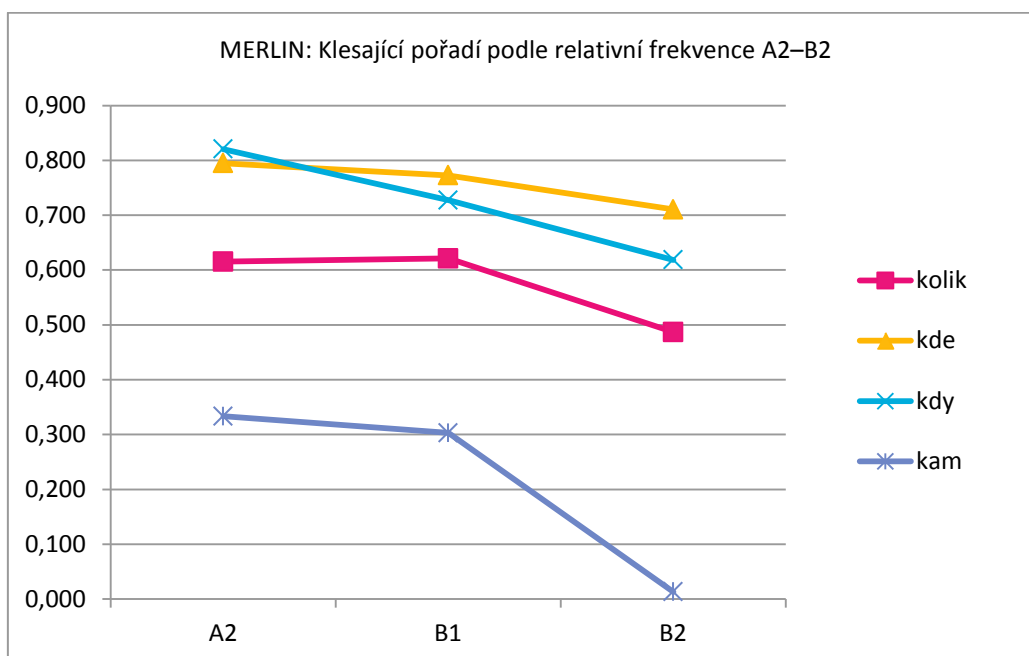
rozdíl relativní frekvence A2–B2		
stoupající	stoupavě-vyrovnaný	klesající
<i>jak</i>	<i>tak</i>	<i>kde</i>
<i>jako</i>	<i>jestli</i>	<i>kdy</i>
<i>proto</i>	<i>ani</i>	<i>kolik</i>
<i>než</i>		<i>kam</i>
<i>kdo</i>	stoupavě-klesající	
<i>jaký</i>	<i>a tak</i>	
<i>a proto</i>	<i>zda</i>	
<i>až</i>	<i>proč</i>	
<i>nejen – ale (i)</i>	<i>takže</i>	
<i>dokonce</i>	<i>kdyby</i>	
<i>pokud</i>		
<i>-li</i>		
<i>totiž</i>	klesavě-stoupající	
<i>přestože</i>	<i>anebo</i>	

Jak vidíme, stoupající užívání je charakteristické téměř pro polovinu zkoumaných spojovacích prostředků. Nejvýraznější nárůst frekvence od úrovně A2 k úrovni B2 byl zjištěn u spojovacích prostředků vyjadřujících parataktické vztahy stupňovací (*nejen – ale (i); dokonce*) a u hypotaktických podmiňovacích spojení s *-li* (graf 23).



Graf 23: MERLIN: Stoupající pořadí podle relativní frekvence A2–B2

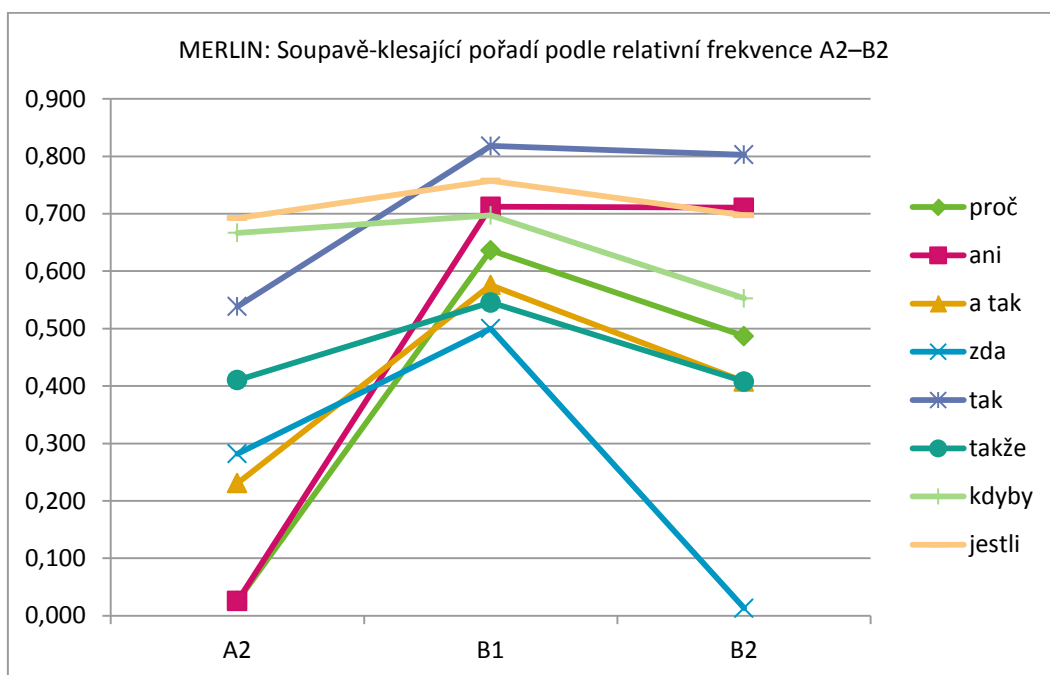
Naopak některé spojovací prostředky na vyšších úrovních ustupují. Tato klesající tendence je, jak vyplývá z analýzy, charakteristická pro autosémantické spojovací prostředky – zájmenná příslovce *kdy*, *kde*, *kam* a zájmennou číslovku *kolik* (graf 24). Nápadný je pokles především u zájmenných příslovcí *kdy* a *kde*, které na úrovni A2 zaujímají místo v první desítce, zatímco na úrovni B2 se dostávají až na 17., resp. 23. pozici. Při interpretaci těchto výsledků je však třeba vzpomenout na povahu testových úloh, na jejichž základě texty úrovně A2 a B1 obsažené v korpusu vznikaly (srov. kap. 3.1.1.2). Obrátíme-li totiž pohled k jejich zadání, zjistíme, že v nich byli kandidáti vedeni k formulování dotazů na místo, čas nebo cenu konání akce/ubytování apod., a tím potažmo k užití nepřímých otázek s uvedenými výrazy. Vzhledem k tomuto vlivu, nelze v korpusu MERLIN klesající tendenci u jmenovaných výrazů dostatečně prokázat.



Graf 24: MERLIN: Klesající pořadí podle relativní frekvence A2–B2

Vedle spojovacích prostředků, jejichž četnost má na všech třech úrovních (A2–B2) tendenci stoupavou, nebo klesavou, se objevují i takové, u nichž tomu tak není (graf 25). Jejich frekvence stoupá při přechodu z úrovně A2 na úroveň B1, mezi úrovněmi B1 a B2 však další nárůst nepokračuje. Frekvence je dále buď setrvalá, a to u těch celkově nejpočetněji zastoupených výrazů (*tak*, *jestli*, *ani*), nebo mírně klesající, což se projevuje u méně frekventovaných prostředků. Právě omezené zastoupení většiny spojovacích prostředků z této

skupiny neumožňuje dalekosáhlejší generalizaci těchto závěrů. Pokles frekvence na úrovni B2 může být důsledkem omezeného množství dokladů.



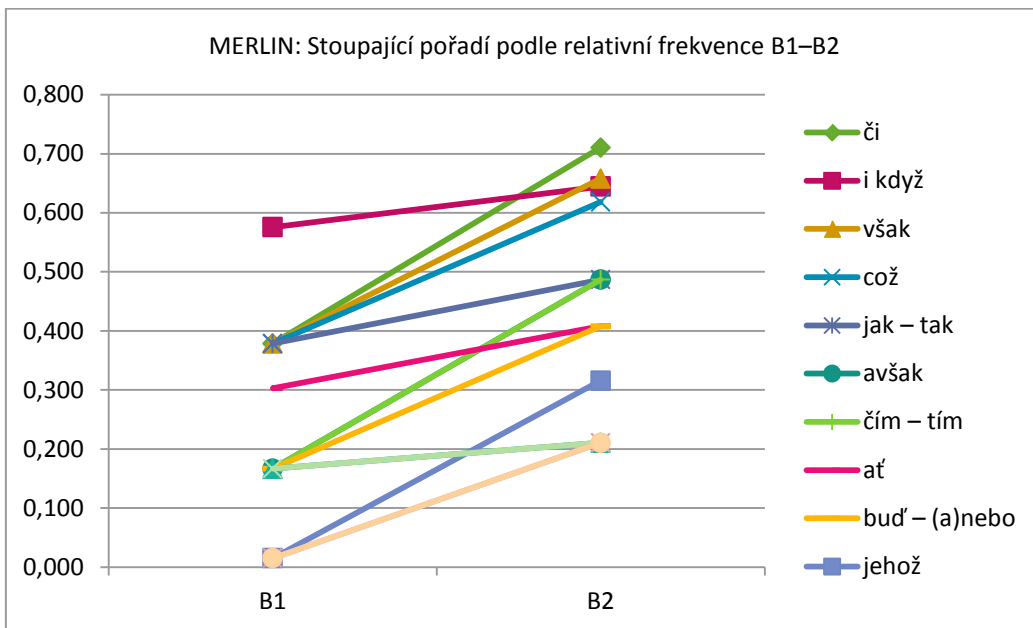
Graf 25: MERLIN: Soupavě-klesající pořadí podle relativní frekvence A2–B2

(B) Druhou zkoumanou skupinu tvoří výrazy objevující se až od úrovně B1 (jejich soupis viz výše tabulka 29). Většinu v této skupině představují prostředky se stoupající tendencí, nejmarkantnější nárůst vykazují výrazy *či*, *avšak*, *čím – tím* a *jehož* (graf 26). Naproti tomu klesající tendence se zásadnějším způsobem neprojevila, nebo není dostatečně podložena. Výrazy, u kterých jsme pokles zjistili, jsou v korpusu totiž celkově velmi málo zastoupené (v jednotkách výskytů). Stejně je tomu i u výrazů bez podstatného rozdílu frekvencí mezi úrovněmi B1–B2 (*přesto*, *jednak* a *jelikož*). Přehled všech podskupin uvádí tabulka 33. Spojovací prostředky jsou v ní řazeny sestupně podle frekvence v celém korpusu, tmavší podbarvení značí výraznější rozdíl mezi úrovní B1 a B2. Rozdíly se dají odečíst také ze znázornění v grafech 25 a 26.

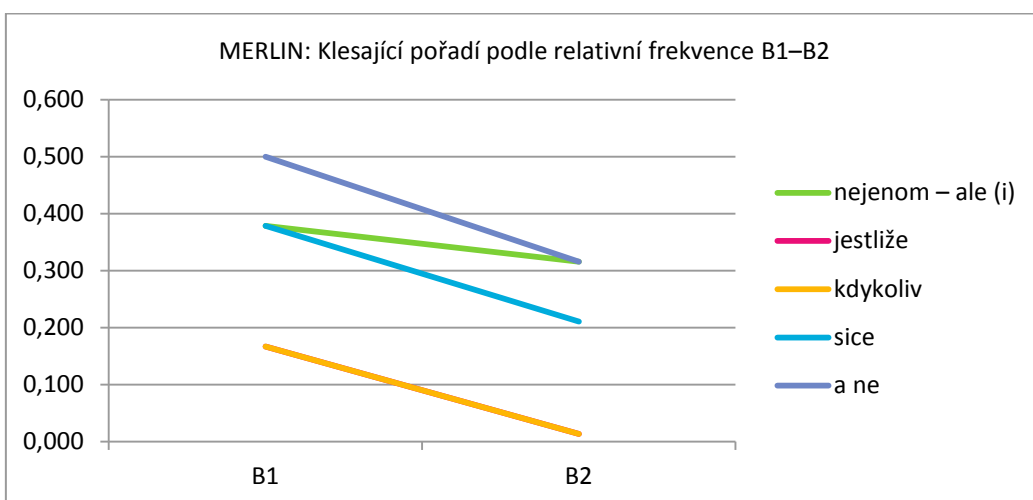
Tabulka 33: MERLIN: Tendence v pořadí podle relativní frekvence B1–B2

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>

avšak
čím – tím
ať
bud' – (a)nebo
jehož
ačkoliv
tedy
aniž
ba



Graf 26: MERLIN: Stoupající pořadí podle relativní frekvence B1–B2



Graf 27: MERLIN: Klesající pořadí podle relativní frekvence B1–B2

4.1.3.3 Shrnutí

V této kapitole (4.1.3.) jsme se věnovali sledování tendencí v užívání spojovacích prostředků v závislosti na úrovni hodnocení textů podle SERRJ. Na základě provedených analýz můžeme formulovat následující závěry:

(1) Se stoupající úrovní se repertoár spojovacích prostředků jednak rozšiřuje o spojovací prostředky vyjadřující nové mezivětné vztahy, jednak se rozšiřují řady konkurenčních prostředků pro vyjádření již zastoupených mezivětných vztahů. První případ, tj. prohloubení dovednosti explicitně vyjadřovat mezivětné vztahy na nižších úrovních doposud nerealizované, se týká na úrovni B1 nepřímých rozkazovacích vět se spojovacím *at'* a srovnávacích konstrukcí s víceslovnými spojovacími prostředky (*čím – tím, jak – tak*), na úrovni B2 se projevuje u vztažných konstrukcí se zájmenem *jenž* a u přípustkových vět s kondicionálem (*i kdyby*). Druhý případ, tj. rozšiřování výrazových možností pro vyjádření vztahů v mezijazyce nižší úrovně již realizovaných, probíhá nejrozsáhleji na úrovni B1, a to zejména u vztahů odporovacích, přípustkových a časových, na úrovni B2 pak již jen v menší míře hlavně v hypotaktických vztazích přípustkových a důsledkových.

(2) S rostoucí úrovní se zvyšuje frekvence užívání většiny spojovacích prostředků. Jako nejprogresivnější se v tomto trendu jeví parataktické spojovací prostředky *i* a *dokonce* a hypotaktické *aby*, *totiž*, *-li*. V mezijazyce se tak postupně posiluje pozice explicitně vyjádřených účelových, příčinných a podmínkových vztahů.

Zároveň platí, že spojky *a*, *že*, *ale*, *nebo*, *jež* jsou celkově v češtině nejfrekventovanější, vykazují nejvyšší četnost setrvale na všech úrovních a jejich pořadí podle frekvence se mezi úrovněmi nemění.

Pokles frekvence se projevil pouze u některých autosémantických spojovacích prostředků (zájmených příslovci *kdy*, *kde*, *kam* a číslovky *kolik*) a synsémantické spojky *protože*.

4.1.4 Korespondenční analýza spojovací prostředky – úrovně SERRJ

V předchozích podkapitolách jsme detailněji popsali hlavní tendence, které charakterizují změnu složení repertoáru a frekvence užití konkrétních spojovacích prostředků v závislosti na vývoji jazykové kompetence. Vycházeli jsme přitom z analýzy užívání jednotlivých výrazů a v některých případech skupin výrazů podle funkce. I když jsme získali řadu jedinečných poznatků a formulovali dílčí závěry, našemu zkoumání ještě chybí pohled na makrorovinu vztahů spojovacích prostředků a úrovní. Teprve na této rovině lze totiž posuzovat validitu

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 úrovně 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.

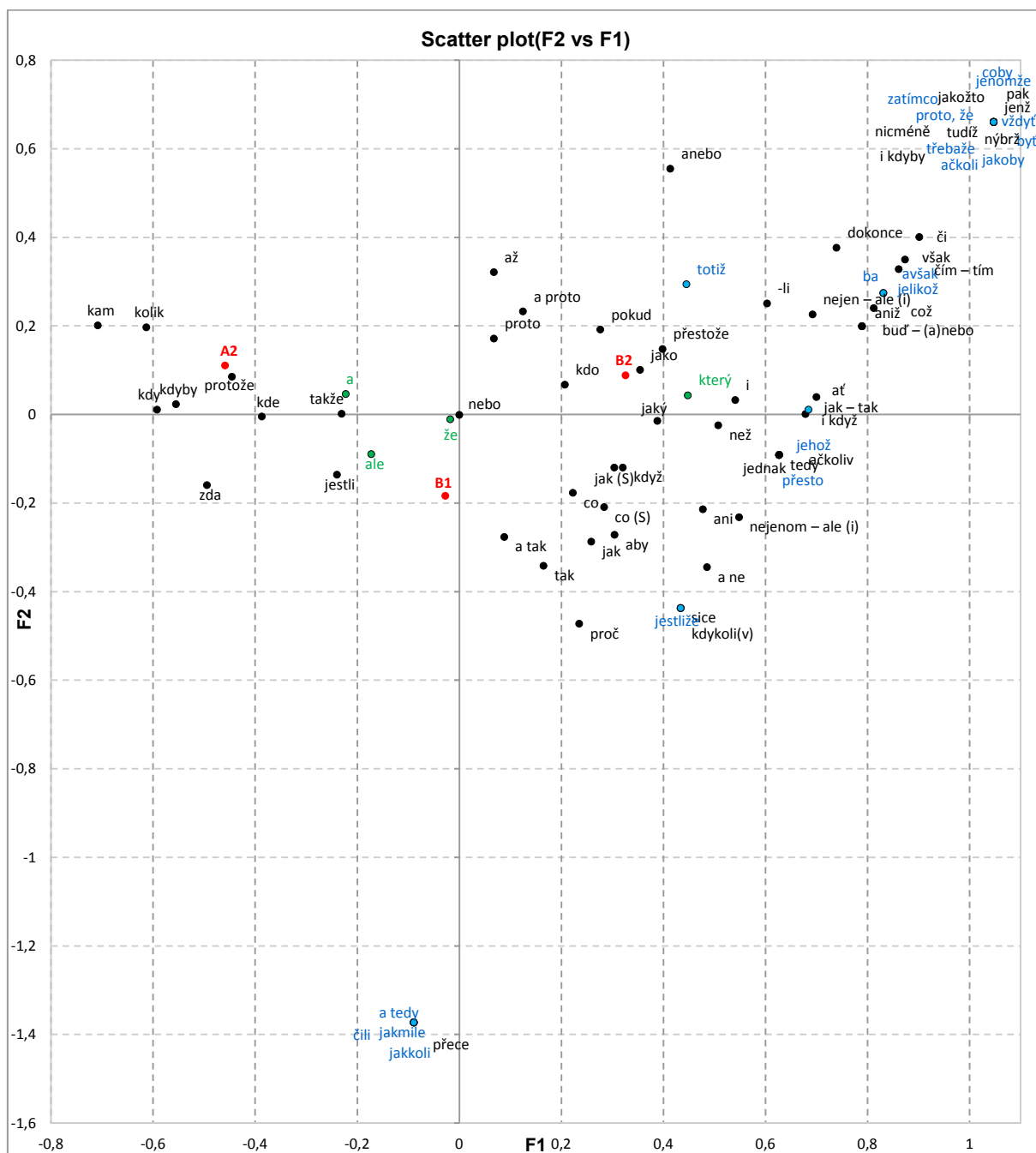
Výsledky analýzy ukázaly, že se původně komplexní data dají převést a zachytit pouze pomocí dvou dimenzí (osa x: inerce = 84,37 %, osa y: inerce = 15,6 %), což značí vysokou korelaci mezi úrovněmi podle SERRJ a užitím spojovacích prostředků. Zároveň je indikován rozdíl mezi jejich používáním na nižších a vyšších úrovních.

V korespondenční mapě je na první pohled patrné, že texty vyšší úrovně podle SERRJ obsahují širší repertoár spojovacích prostředků. Větší počet spojovacích prostředků se kumuluje vpravo kolem červeného bodu zastupujícího úroveň B2. Naproti tomu vlevo kolem úrovně A2 je shromážděna pouze menší skupina spojovacích prostředků. Lze tedy konstatovat, že se 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.

Toto zjištění samo o sobě není nijak překvapující, podle SERRJ ale užívá omezené množství kohezních prostředků ještě uživatel na úrovni B2 a teprve na úrovni B2+ dokáže používat rozmanité spojovací výrazy (srov. SERRJ 2001, s. 128). Na základě naší analýzy bychom ovšem mohli vyslovit pochybnost, zda toto odlišení skutečně platí. Kolem úrovně B2 jsou již zastoupeny spojovací prostředky důležité pro vyjádření všech hlavních mezivětných vztahů (viz kap. 2.1.2) a u některých z těchto vztahů jsou k dispozici i konkurenční prostředky (např. *proto, tak, a tak, tedy*). Zdá se spíše, že linie oddělující uživatele omezeného repertoáru od uživatele schopného užít rozmanité prostředky, leží již na rozhraní úrovní B1+ a B2.

Z deskriptorů SERRJ také vyplývá, že uživatelé nižších úrovní ovládají především spojovací prostředky v cílovém jazyce frekventované, zatímco na vyšších úrovních se objevují i prostředky v cílovém jazyce méně frekventované (srov. SERRJ 2001, s. 128). Prokázání tohoto vztahu sice nebylo hlavním cílem předkládané práce, ale protože jsme měli

k dispozici data o frekvenci spojovacích prostředků z korpusu SKRIPT2012, bylo nasnadě zakomponovat je do korespondenční analýzy. Podle barevného kódování v korespondenční mapě tak můžeme sledovat, jakou frekvenci mají dané výrazy v akvizičním korpusu rodilých mluvčích češtiny (graf 29). 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 29: MERLIN: korespondenční analýza relativní frekvence a úrovně SERRJ + barevně znázorněná frekvence výrazů v L1

Pohled do takto upravené korespondenční mapy výše uvedené popisy rozdělení spojovacích prostředků podle frekvence z deskriptorů SERRJ potvrzuje. Na levé straně, tj. na úrovni A2 se objevují výhradně spojovací prostředky více nebo středně frekventované (černá a modrá barva). Větší množství v češtině rodilých mluvčích méně frekventovaných (modrých) spojovacích prostředků můžeme pozorovat až od úrovně B2. Výjimku tvoří několik spojovacích prostředků u úrovně B1 (*a tedy, jakmile, čili* atd.) na spodním okraji mapy uprostřed, podle pozice celé skupiny je ale patrné, že se v textech B1 objevily pouze okrajově v jednotkách výskytů. Naopak na úrovni B2 se méně frekventované výrazy dostávají blíže do centra využívaných prostředků (např. *totiž, jestliže, přesto* atd.). I když stejně jako u B1 i na této úrovni můžeme narazit na skupinu méně frekventovaných výrazů, které se objevují jen okrajově, viz pravý horní roh mapy (*coby, byt', i kdyby* atd.)

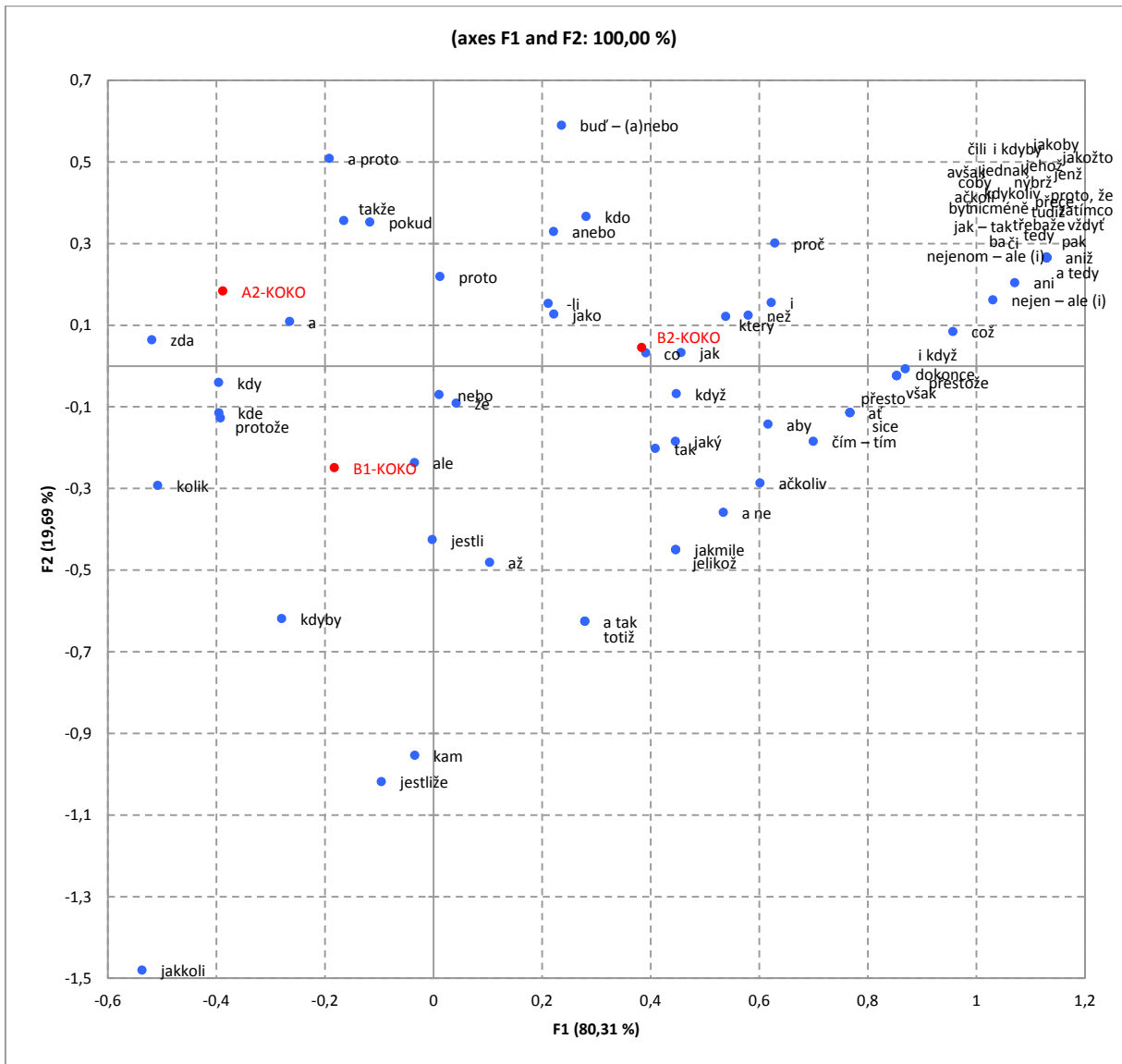
Závěrem této podkapitoly můžeme shrnout, že korespondenční analýza potvrdila závislost užití spojovacích prostředků na úrovni jazykové kompetence. Výsledky analýzy podpořily adekvátnost deskriptorů SERRJ, které uvádějí, že na vyšších úrovních lze (na rozdíl od nižších úrovní) očekávat užívání spojovacích prostředků, jež jsou v cílovém jazyce méně frekventované. Naopak co se týká rozsahu repertoáru spojovacích prostředků, výsledky nasvědčují tomu, že uživatelé jazyka ovládají širší rozsah spojovacích prostředků již na nižší úrovni, než vyplývá z deskriptorů SERRJ.

4.1.5 Korespondenční analýza spojovací prostředky – hodnocení koheze a koherence

Až doposud jsme se ve svém zkoumání zabývali pouze užíváním spojovacích prostředků v závislosti na celkovém hodnocení textu podle SERRJ. Metadata korpusu MERLIN však obsahují u každé produkce i detailnější hodnocení podle dílčích kritérií (pravopis, gramatická kompetence, koheze a koherence apod.), a proto jsme se rozhodli využít shromážděných dat a analyzovat také vztah spojovacích prostředků k hodnocení koheze a koherence. Hodnocení podle dílčích kritérií nemusí totiž nutně korespondovat s úrovní podle celkového hodnocení. Například text, který je v kohezi a koherenci hodnocen na úroveň B2, může být vlivem ostatních kritérií (např. souhrou nedostatků v lexiku a gramatice) hodnocen celkově jako odpovídající úrovni B1.

Zajímá nás, jak se zohlednění pouze tohoto dílčího hodnocení promítne do rozložení repertoáru spojovacích prostředků, a proto stejně jako v předchozí podkapitole užíváme

korespondenční analýzu (graf 30). Pracujeme s relativní frekvencí spojovacích prostředků a hodnocením textů podle koheze a koherence.

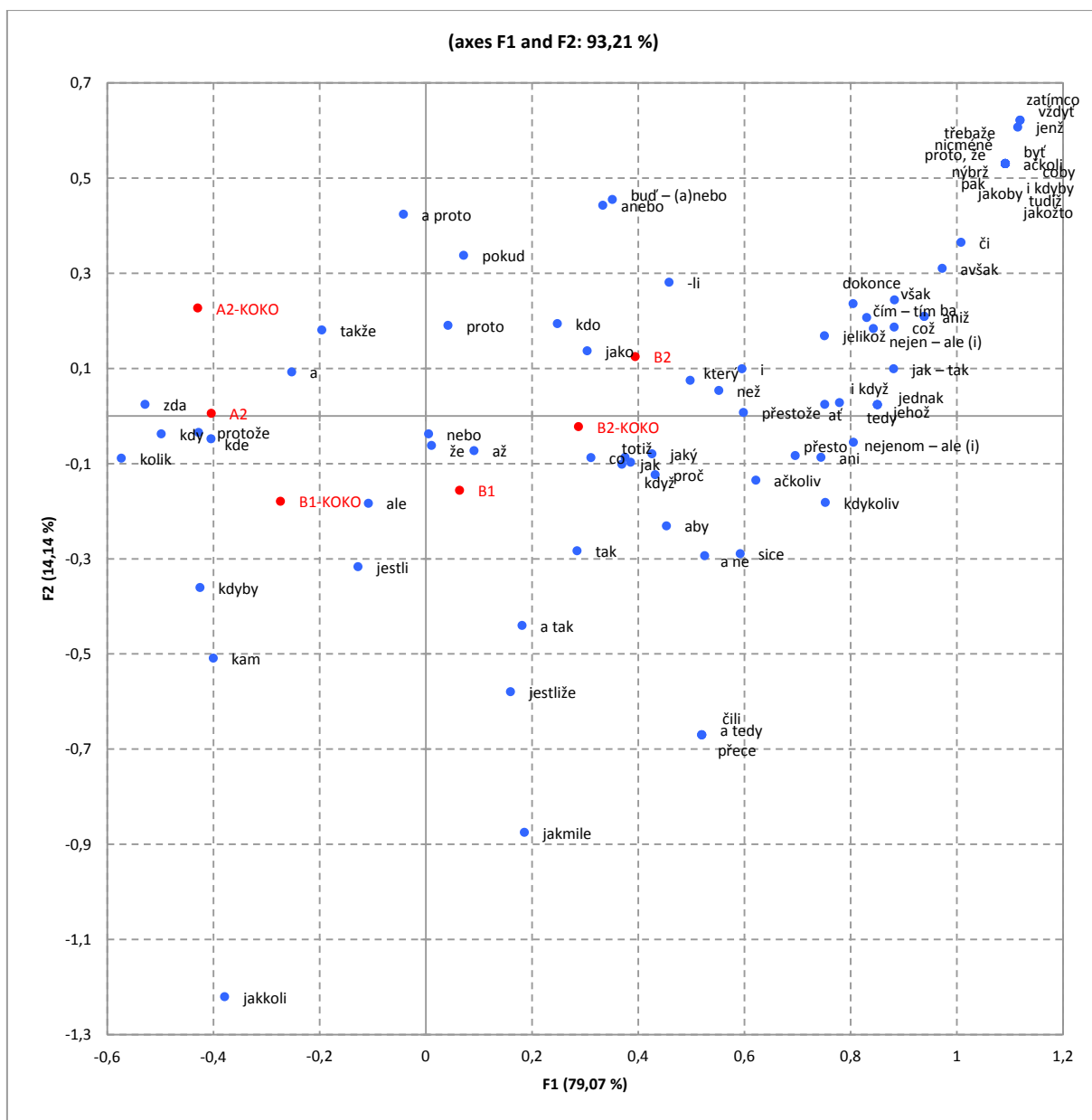


Graf 30: MERLIN: korespondenční analýza relativní frekvence a hodnocení koheze a koherence

Pohled do výsledků ukazuje, že závislost spojovacích prostředků na hodnocení koheze a koherence je silnější než závislost na celkovém hodnocení. Pole obsahující spojovací prostředky je v korespondenční mapě vztahující se k hodnocení koheze a koherence kompaktnější a sevřenější. Výrazněji se zde projevuje předěl mezi úrovněmi A2/B1 a úrovní B2, což opět potvrzuje naši hypotézu, že k užívání širšího repertoáru spojovacích prostředků dochází již na spodní hranici úrovně B2. Svědčí o tom i to, že se rozpustila skupina spojovacích prostředků, jež byly podle celkového hodnocení textů na úrovni B1 solitéry

s jednotlivými výskyty (čili apod.). Hodnocení koheze a koherence textů, ve kterých se vyskytovaly, totiž odpovídá úrovni B2.

S tímto zjištěním také vyvstává domněnka, že posun k vyšší úrovni podle hodnocení koheze a koherence, který závisí na repertoáru spojovacích prostředků, probíhá dříve, než uživatel dosáhne vyšší jazykové kompetence v celé šíři. Tomu může nasvědčovat i korespondenční analýza se zahrnutím celkového hodnocení, kterou uvádí graf 31. Ve výstupu můžeme ledovat, že body znázorňující hodnocení podle koheze a koherence (A2-KOKO, B1-KOKO a B2-KOKO) na vodorovné ose předcházejí bodům odpovídajícím hodnocení celkovému. Zvláště nápadný je tento jev u úrovně B1.



Graf 31: MERLIN: korespondenční analýza relativní frekvence, celkového hodnocení a hodnocení koheze a koherence

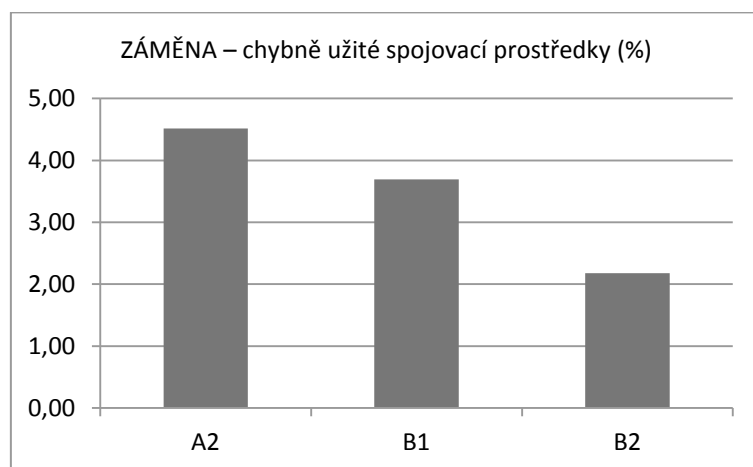
Uvedenou domněnku ovšem nelze na základě této kvantitativní analýzy potvrdit. Bylo by totiž třeba zohlednit také míru správnosti a funkčnosti užití daných spojovacích prostředků. Jisté je, že se v tomto směru otevírá široké pole pro další výzkum.

4.1.6 Chyby v užití spojovacích prostředků

Součástí hlavního záměru předkládané práce je nejen analýza složení repertoáru a frekvence užití spojovacích prostředků, již jsme prezentovali v předchozích kapitolách, ale také pohled na správnost jejich užití. Z toho důvodu jsme se v teoretické části práce věnovali otázkám chybové analýzy (kap. 2.3), kterou jsme aplikovali na dokladech z korpusu MERLIN postupem uvedeným v kapitole 3.3.4. Hlavním typům chyb, které jsme vymezili, se budeme věnovat v následujících podkapitolách.

4.1.6.1 Záměna spojovacího prostředku⁸⁴

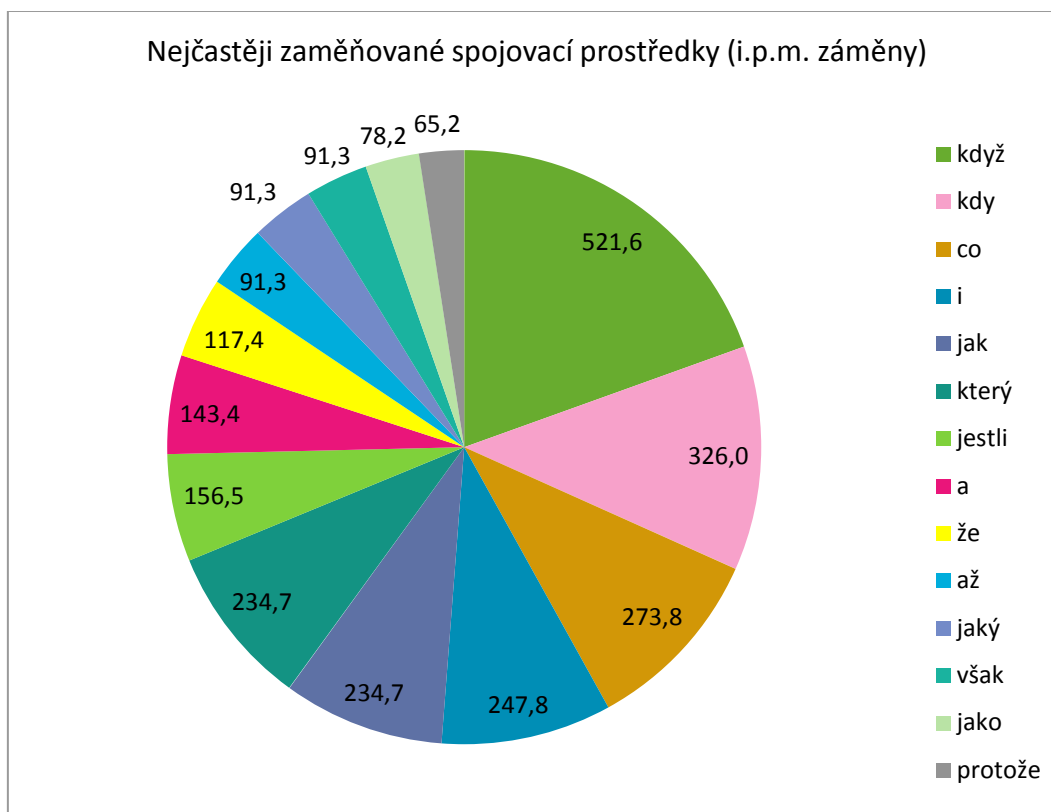
Nejčastější chybou, která se v korpusu objevuje a váže se k užívání spojovacích prostředků, je jejich záměna. Dochází k ní nesprávným užitím spojovacího prostředku na místě, kde má stát jiný spojovací prostředek. Na všech úrovních dohromady bylo nalezeno 243 dokladů tohoto typu chyby, což při celkovém počtu 7 333 spojovacích prostředků, které korpus obsahuje, představuje 3,31 %. Přitom platí, že na nižších úrovních je poměr zastoupení záměn vyšší (A2 – 4,51 %) a s rostoucí úrovní klesá (B2 – 2,18 %) – graf 32.



Graf 32: Záměna spojovacích prostředků (procentuální zastoupení záměn v celkovém počtu dokladů)

⁸⁴ V podstatě se jedná o nadbytečné užití jednoho výrazu a zároveň chybějící užití jiného.

K záměně došlo celkem u 36 různých spojovacích prostředků. Prvních 14 spojovacích výrazů, které nejčastěji stojí na místě jiných spojovacích prostředků, ukazuje graf 33.



Graf 33: Spojovací prostředky s nejčastější záměnou

Analýzou nejčastěji zaměňovaných prostředků jsme zjistili, že můžeme záměny popisovat z několika různých úhlů pohledu.

(A) Jedním z nich je popis podle toho, jak se záměna promítá do formálně syntaktických vztahů mezi větami.

(Ai) V korpusu se vyskytují záměny mezi dvěma mezivětnými významovými vztahy na stejné formálně syntaktické úrovni. Jako příklad na rovině parataktické můžeme uvést spojku *a*, která nejčastěji stojí místo *ale*, tj. dochází k posunu odporovacího nebo stupňovacího významu na slučovací (78)–(79).

(78) *Hodně lidi dobře se oblikaji a (↓ale) dělají špatne věci, nikdy hodně porušuji jak zakon, tak i dustojnost člověka.* (MERLIN – B1)

(79) *Lidé nosí oblečeny podle ne jenom počasí, a (↓ale) také podle situace.* (MERLIN – B2)

U hypotaktických vztahů se setkáme kupříkladu s užitím *když* místo *jestli/pokud*, při němž dochází k posunu podmínkového vztahu na časový (80).

(80) **Když** (\downarrow *jestli*) *chceš, pozvám Ty doma v pátku odpoledne.* (MERLIN – A2)

(Aii) Druhou podskupinu tvoří záměny působící změnu parataktického vztahu na hypotaktický a naopak (81).

(81) *Ted' se musím učit a protože* (\downarrow *proto*) *dokončím dopis.* (MERLIN – A2)

(Aiii) Některé záměny mohou mít z hlediska formálně syntaktických vztahů nejasnou interpretaci, takže není jasné, zda se jedná o záměnu vztahů stejné roviny, či nikoliv (srov. 82).

(82) *Bohužel musím ukončit psaní ale* (\downarrow *protože / neboť*) *musím pracovat na seminární práci.* (MERLIN – B1)

(B) Na základě větněčlenské platnosti lze hovořit o záměnách autosémantických spojovacích prostředků za synsémantické a naopak. V korpusu se často vyskytuje v podobě užití *když* místo *kdy* (83), res. *kdy* místo *když*.

(83) *Studium je ještě hezké, protože mám hodně čas až do té doby, když* (\downarrow *kdy*) *diplom musí být dodělané.* (MERLIN – A2)

(84) *Jezdím mnoho, a často po Evropě. divám se na zajímavé koncerty, pracuju, ale kdy* (\downarrow *když*) *se vrátím doma jsem šťastna.* (MERLIN – A2)

(C) Podle toho, zda je záměna spojovacího výrazy spojena se změnou třídy okolnostního významu (např. časové okolnosti vs. místní okolnosti) případně poměru mezi větami hlavními (např. slučovací vs. odporovací), nebo zda se pouze promítá do významového posunu v rámci jedné třídy (např. zesílení vylučovacího významu), můžeme charakterizovat záměny proměňující a modifikující.

(Ci) Záměny proměňující třídu okolnostního významu nebo poměr mezi větami jsme ukázali již v příkladech *a* místo *ale* (78) a *když* místo *jestli* (80).

(Cii) Příkladem záměny modifikující okolnostní význam může být užití *když* místo *až* (85).

(85) *Hledám dobrou gramaticu českého jazyku. Znaš jednou a mužeš koupit? Platím se když* (\downarrow *až*) *sedjeme.* (MERLIN – A2)

S využitím výše navržené kategorizace jsme se pokusili vytvořit ukázkou možné přehledové charakteristiky záměn spojovacích prostředků nejčastěji pozorovaných v korpusu MERLIN (tabulka 34). Z tabulky je patrné, že většina záměn probíhá mezi vztahy na stejné formálně syntaktické úrovni, tj. nejčastěji mezi synsémantickými spojovacími prostředky na rovině hypotaktické, méně na rovině parataktické. Pouze ojediněle se objevují případy, kde jeden spojovací prostředek vstupuje do záměn napříč těmito vztahy (např. *a*, *že*).

Dále se ukazuje, že v záměnách nehraje zásadnější roli větněčlenská platnost výrazů, synsémantické prostředky se objevují místo synsémantických i autosémantických a opačně. Výjimkou je pouze výraz *jak*, který ve sledovaných dokladech nikdy nestál na místě jiného autosémantického spojovacího prostředku.

Všechny sledované výrazy se také alespoň v několika výskytech objevily v záměnách proměňujících třídu mezivětného významu. Synsémantické prostředky se nadto vyznačují i záměnami modifikačními, čímž se liší od autosémantických spojovacích prostředků.

Tabulka 34: Charakteristika nejčastěji zaměňovaných spojovacích prostředků

výraz	užitý místo	H>P	P>H	P>P	H>H	A>S	S>A	S>S	A>A	PROM	MOD
<i>když</i>	<i>jestli, kdy, až, kdyby, aby, než, že</i>				X		X	X		X	X
<i>kdy</i>	<i>když, kde, i když</i>				X		X	X		X	
<i>co</i>	<i>což, který, jaký, kolik, že, jak</i>					X			X	X	
<i>i</i>	<i>a, ani, ale</i>			X						X	X
<i>jak</i>	<i>jako, když, kdyby, jestli, než, jakmile, jaký</i>				X	X				X	
<i>který</i>	<i>kdo, co, jaký</i>				X			X		X	
<i>jestli</i>	<i>jestliže, kdyby, když, že</i>				X			X		X	X
<i>a</i>	<i>ale, i když</i>		X	X				X		X	
<i>že</i>	<i>jestli, aby, ať, co, který, kdy, a</i>	X			X		X	X		X	
<i>až</i>	<i>když, že, kdy, než</i>				X		X	X		X	X
<i>jaký</i>	<i>který, jako</i>				X	X			X	X	X
<i>však</i>	<i>avšak, ale</i>			X				X		X	X
<i>jako</i>	<i>jak, jaký, než</i>				X		X	X		X	
<i>protože</i>	<i>proto, ale</i>	X						X		X	

vysvětlivky:

H>P: hypotaktický prostředek místo parataktického

P>H: parataktický prostředek místo hypotaktického

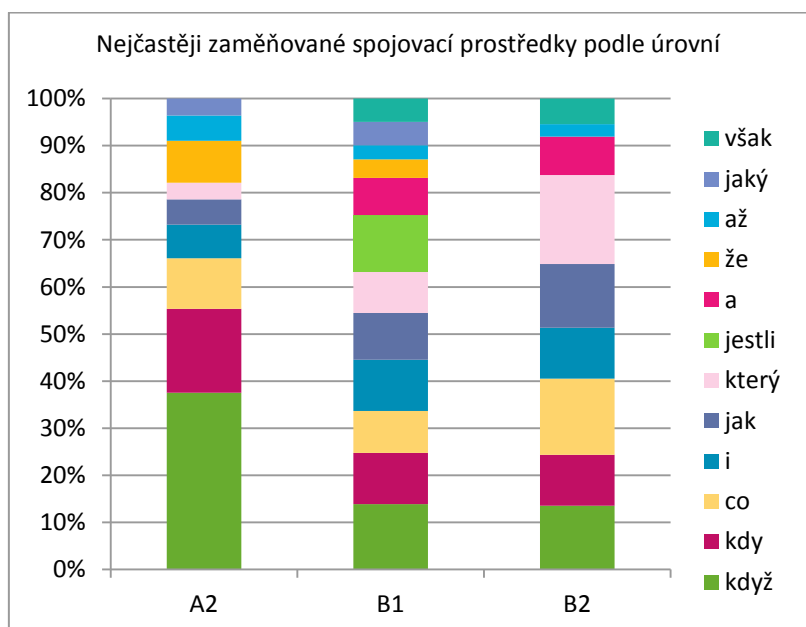
A>S: autosémantický prostředek místo synsémantického

S>A: synsémantický prostředek místo autosémantického

PROM: proměna okolnostního vztahu nebo poměru mezi hlavními větami

MOD: významový posun okolnostního vztahu nebo poměru mezi hlavními větami

Naše chybová analýza se pro účely této práce spokojí s uvedenou rovinou podrobnosti a pokročí ke zkoumání zastoupení chyb-záměn konkrétních spojovacích prostředků na jednotlivých úrovních podle SERRJ.⁸⁵ Pro tuto analýzu jsme vybrali prvních 14 výrazů, ve kterých uživatelé nejčastěji chybovali a které se podílejí na téměř 80 % všech chyb tohoto typu. V grafu 34 můžeme sledovat, jak se mění podíl spojovacích prostředků na záměnách v závislosti na úrovni.



Graf 34: Nejčastěji zaměňované spojovací prostředky podle úrovní (i.p.m.)

Na úrovni A2 si můžeme všimnout nápadné převahy záměn *kdy* a *když*, které představuje více než polovinu záměn v této skupině. Dále stojí za pozornost užití spojky *že* místo *jestli* a *co* místo *že*, k nimž na této úrovni dochází v obsahových větách (86)–(87).

⁸⁵ Nicméně i na základě našich víceméně kvantitativních popisů několika kvalitativně vytknutých kategorií, je zřejmé, že se zde otevírá široké pole pro detailnější zkoumání. Bylo by kupříkladu možné zjišťovat, které významové vztahy se vyznačují největší chybivostí a které konkrétní spojovací prostředky mají na záměnách podíl. Od toho by se následně mohla odvíjet i případná didaktická doporučení pro výuku a tvorbu učebních materiálů.

(86) *Ted', nevím, že (↓jestli) budu navštěvovat těbe v létě, protože chcem pojet do Prievdzi s letadlem na letiště a tam chcem letát od Vysokou do Vysokou tatra a nazpět.* (MERLIN – A2)

(87) *Dnes hodne lidí mysly co (↓že) oni mame hodne „pratelej” No to je nespravne.* (MERLIN – A2)

Na úrovni B1 záměn *kdy* a *když* ubývá, zato přibývá chybného užití *jestli* místo *kdyby* (88). Zdá se, že uživatelé úrovně B1 se snaží vyjadřovat podmínkové vztahy, pro něž ale ještě nemají zcela osvojené explicitní výrazové prostředky. Obdobně tomu je u záměny *a* za *i* v některých případech, které by mohli být interpretovány kromě slučovacího významu také významem důsledkovým (89).

(88) *Vyhřali jsme intelektuální hru v Berlině a vyhlali jsme taky peníze. Jestli (↓kdyby) nebyl tam někdo z nás, myslím, že jsme nemohli vyhrat.* (MERLIN – B1)

(89) *Srdečne Ti dekuju za Tve pozvání na grilovací party. bohužel budu mět dovolená v práci i (↓a/a proto) nemůžu přijet.* (MERLIN – B1)

S úrovní B2 je spojená vyšší frekvence užití zájmena *který*, jež je ovšem zejména v korelativních konstrukcích užíváno neadekvátně na místo vztažných zájmen *kdo* a *co* (90)–(91). Naopak zájmeno *co* autoři textů často používají pro vyjádření vztahu závislé věty k obsahu věty řídicí na místě, kde by se mělo objevit vztažné *což* (92). Tato záměna se v korpusu vyskytuje již od úrovně B1 a na úrovni B2 její frekvence vzrůstá.

(90) *Myslím, že ten stesk má každý, který (↓kdo) bud' často cestuje nebo je v zahraničí delší dobu.* (MERLIN – B2)

(91) *Podle mého názoru je toto přísloví pravda a když mám něco které (↓co) musím vyřešit, obrácím se na další lidi.* (MERLIN – B2)

(92) *Zaměstnavatele vybercí zaměstnancy vždycky budou ovlivňovani tím, jak kandidati vypadají, na co (↓což) má také vliv, zda přichází v dobrým obleku nebo v džinsach a tričku.* (MERLIN – B2)

Speciálním případem záměny spojovacího výrazu je několik dokladů, v nichž místo spojovacího prostředku stojí výraz, který v češtině spojovací funkci neplní (93) nebo do textu pronikl z mateřského jazyka autora textu (94). V celém korpusu MERLIN jsme narazili pouze na dvanáct dokladů tohoto typu.

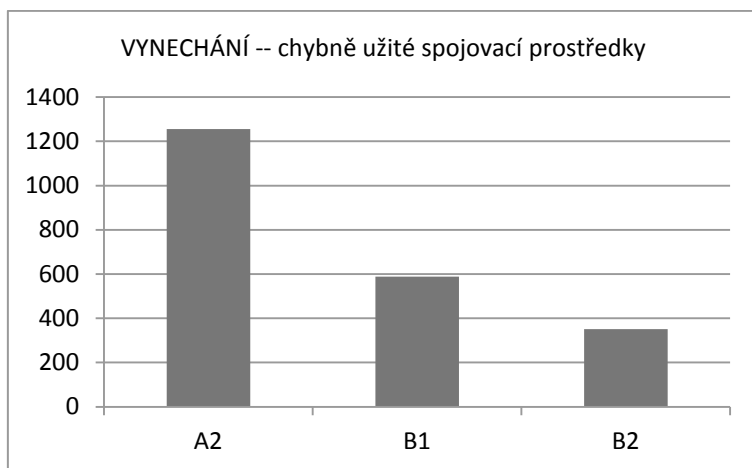
(93) *Také chcem ti posielat nejaké fotky, ked' (↓to toho, když) jsem byl na vylet s Martinem na horach.* (MERLIN – A2)

(94) *Když se podíváme na nejaké velké a úspěšné organica, to (↓tak) vždycky uvidime, že tam pracuje opravdu hodně lidí.* (MERLIN – A2)

Po prostudování poměru a povahy chyb-záměn na jednotlivých úrovních, bychom závěrem tohoto oddílu mohli vyslovit domněnku, že k záměnám dochází zpravidla tam, kde se uživatelé jazyka snaží vyjadřovat nové mezivětné vztahy, pro něž si doposud (dostatečně) neosvojili adekvátní spojovací prostředky. Tehdy totiž sahají po výrazech, které vyjadřují vztahy v jistých ohledech podobné (srov. vztažné kdo a který), nebo jsou naopak významově nepřilíš vyhraněné (srov. *že*).

4.1.6.2 Vynechání spojovacího prostředku

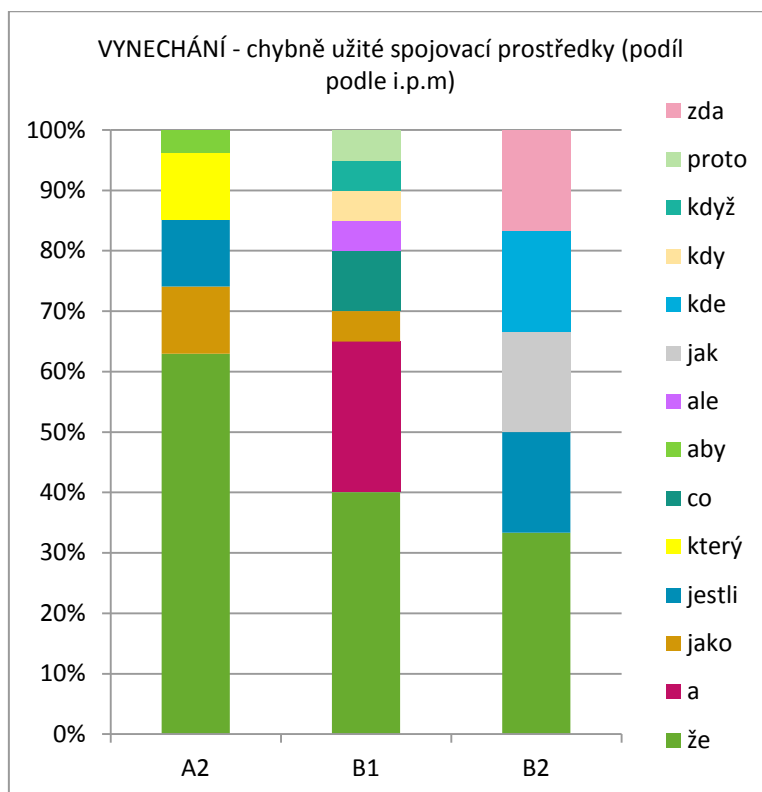
Druhým typem chyb sledovaným v korpusu MERLIN jsou případy vynechání spojovacího prostředku. Oproti záměnám (viz výše 245 dokladů) je vynechávání méně frekventované, celkem jsme shromáždili 53 dokladů. Frekvence také vykazuje klesající tendenci směrem k vyšší úrovni jazykové kompetence (graf 35).



Graf 35: MERLIN: Relativní frekvence vynechání spojovacího prostředku (i.p.m)

Zajímavá skutečnost se ukáže při bližším pohledu na to, u kterých spojovacích prostředků k vynechávání dochází (graf 36). Celkově se tato chyba týká pouze čtrnáct výrazů z celého repertoáru (81 prostředků), z nichž pouze u jednoho (spojky *že*) se projevuje na všech

úrovních. U ostatních se chybovost objevuje pouze na jedné, výjimečně na dvou úrovních. To ukazuje silnou spojitost úrovně a frekvence chyby-vynechání.



Graf 36: Chybějící spojovací prostředky (podíl podle i.p.m)

Na úrovni A2 uživatelé kromě spojky *že* nejčastěji vynechávali *jako* uvádějící doplňkové výrazy (95), uvození závislých otázek s *jestli* (96) a zájmeno *který* ve složitějších konstrukcích s korelativem (97).

(95) *Studovala vysokou školu, teď pracuje ve střední škole (↓jako /učitelka/) učitelkou.*

(MERLIN – A2)

(96) *Nevíš (↓jestli) má ta jazyková škola ještě kurs češtiny v lete?* (MERLIN – A2)

(97) *Neznam tvoje město a proto prosím napišu me budou neco lidé (/nějací lidé/ ↓které /je/) je znám?* (MERLIN – A2)

Úroveň B1 se projevila největší šíří vynechávaných výrazů (celkem 8). Převahu opět tvoří spojka *že* (98), naopak vyšší frekvenci vynechání spojky *a* je třeba interpretovat s vědomím toho, že se většina dokladů této chyby objevila pouze ve dvou textech, kde několikrát chybělo *a* při spojení několikanásobných větných členů (99). Ve více různých textech úrovně B1 bylo

zjištěno vynechání zájmena *co* v komplexnějších korelativních strukturách (srov. výše vynechání *který* na úrovni A2) (100)–(101).

(98) *Myslím, (↓že) v března všechno končí a potom mám čas.* (MERLIN – B1)

(99) *Důležité je že naše tradice (↓a) kultura mluví jak člověk musí se oblect u rozmanitých okolnostech (↓a) situacích!* (MERLIN – B1)

(100) *Ale stále má jinou funkci ukázat, kdo jsem a to záleží **na to**, (↓co) jini lidé myslí o tom.* (MERLIN – B1)

(101) *Určitě můžete ji koupit v pekárně ale ta nebude mít stejnou chuť **než tu** (↓co/kerou) jste pekli sami.* (MERLIN – B1)

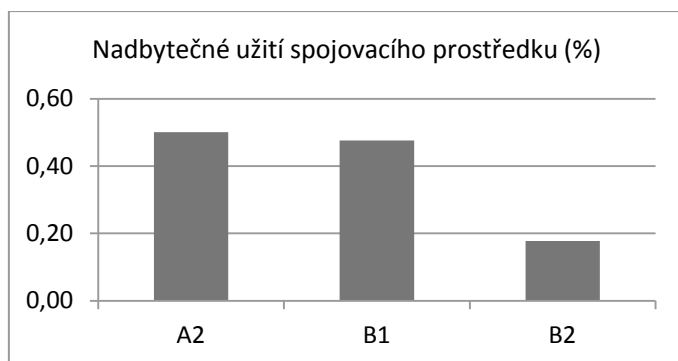
Posunem k úrovni B2 se vynechávání spojovacích prostředků stává spíše ojedinělou záležitostí a kromě obsahových vět se týká především nepřímých otázek (*jestli, zda*) (102).

(102) *Cena oblečení dá také informaci, jakou socialní pozici má ten člověk, (↓**jestli/zda**) je bohatý nebo chudý.* (MERLIN – B2)

Analýza všech dokladů s chybějícím spojovacím prostředkem, napovídá, že k vynechání dochází zejména tehdy, jedná-li se o komplexnější strukturu obsahující několik vět a/nebo zájmené korelativní spojení, kde se zájmeno uvozující větu závislou vztahuje k větnému členu věty řídící, jenž je vyjádřen zájmenem.

4.1.6.3 Nadbytečné užití spojovacího prostředku

Nadbytečné užití spojovacího prostředku se v korpusu MERLIN objevuje zcela okrajově, celkově se jedná o 0,4 % všech dokladů spojovacích výrazů. Se stoupající úrovní tento typ chyby téměř mizí, na úrovni B2 se jedná pouze o nahodilé případy (srov. graf 37).



Graf 37: Nadbytečné užití spojovacích prostředků (procentuální zastoupení v celkovém počtu dokladů)

Na více než jedné úrovni se projevilo nadbytečné užití pouze u spojek *a*, *ale*, *že* a *aby* (103)–(104). Některé případy navíc dovolují dvojí interpretaci (105).

(103) *Děkuji za tvůj pozvání, ale (↓0) abych neměl čas, protože mi měla v pondělí - 16. 3 - to je zkouška.* (MERLIN – A2)

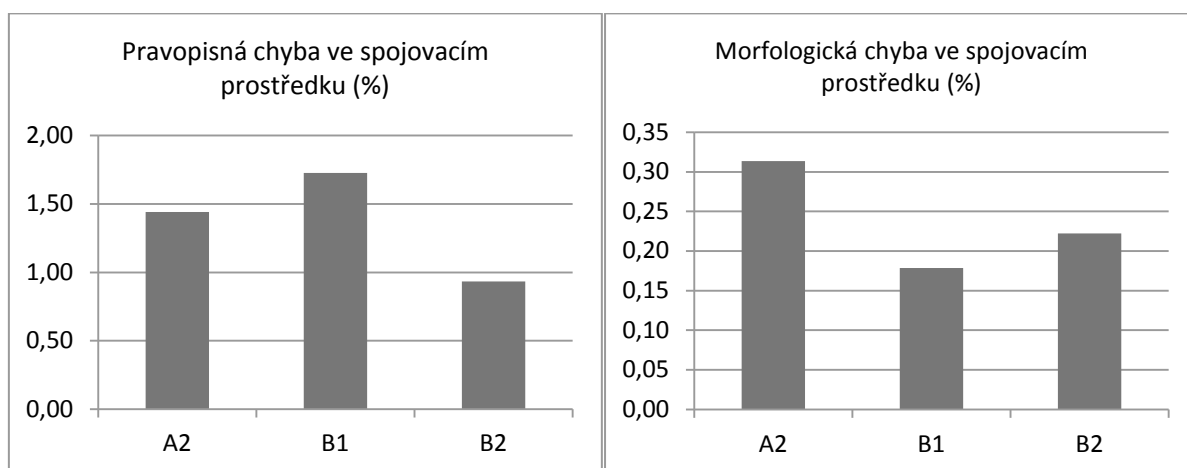
(104) *Protože nemam čas 13. srpna a (↓0) nepůjedu na grilovací party.* (MERLIN – A2)

(105) *Určitě, že (↓0/Je jisté, že) v 21. století, školní vzdělávací program má být doplněn a rozšířen podle současných potřeb našeho života.* (MERLIN – B1)

4.1.6.4 Pravopisné a morfologické chyby

V analyzovaných dokladech jsme také naráželi na chyby spojené s ortografickou realizací spojovacích prostředků a s tvořením jejich formy. Ani jedna z těchto chyb nepatří do výše zmíněných typů chyb, a proto je vyčleňujeme zvlášť. Daný výraz totiž vždy stojí na správném místě (nechybí ani nepřebývá) a je vybrán v souladu se syntaktickým vztahem, který spojení vyžaduje. Jedná se tedy pouze o formální stránku jeho realizace. Relativní frekvence výskytu těchto chyb je velmi malá – pravopisné chyby se objevují celkem u 1,39 % všech spojovacích prostředků a gramatickou chybou je zasaženo pouze 0,22 % z nich).

Procentuální zastoupení chyb (podíl správných a chybných výskytů) udávají grafy 38 a 39.



Graf 38: Pravopisné chyby ve spojovacím prostředku (procentuální zastoupení v celkovém počtu dokladů)

Graf 39: Morfologické chyby ve spojovacím prostředku (procentuální zastoupení v celkovém počtu dokladů)

Při jejich interpretaci je třeba mít na paměti složení repertoáru spojovacích prostředků a jejich frekvenci na daných úrovních. Morfologické chyby mohou nastávat pouze u spojovacích prostředků autosémantických ohebných (zájmena, číslovky), a proto není jejich

četnost tak vysoká. V korpusu se objevily u výrazů (*který, co, kdo, jaký, jenž*) (106). Ty jsou navíc na každé úrovni jinak frekventované a některé z nich se na nižších úrovních vůbec neobjevují (*jenž*).

(106) *Přestože každý člověk zná toto přísloví, nekaždý ví, že jsou vernější přátelé, které (↓kteří) zůstávají svami v radosti.* (MERLIN – B1)

U interpretace pravopisných chyb je třeba zohlednit také fakt, že jejich variabilita je vyšší – chybná diakritika, záměna písmen, chybná hranice slov atd. Například u úrovně B1 je nárůst pravopisných chyb ovlivněn rozšířením repertoáru o výrazy s diakritikou (*než, takže, proč*) (107) a o víceslovné výrazy (*a proto, a tím vs. anebo*) (108).

(107) *Lide, kteří pracují moc, vetsonou měli by mít v životě víc, než (↓než) precující min.* (MERLIN – B1)

(108) *Nouzovy situací jsou vždycky něco který nemáme rádi řešit a v tom případě se chováme jinak než normalně ikdyž (↓i když) to se týka o naše přitele.* (MERLIN – B1)

Dodejme na okraj, že se morfologické a ortografické chyby mohou kumulovat i v rámci jednoho výrazu (109).

(109) *Na štěstí (doufám!) je to výjimečna sitauce, a každý má někoho na kim (↓koho) se může spolechnout.* (MERLIN – B1)

4.1.6.5 Shrnutí

V této kapitole jsme se věnovali analýze chyb v užívání spojovacích prostředků obsažených v korpusu MERLIN. Na základě povrchové taxonomie (srov. kap. 2.3.1) jsme klasifikovali chyby spočívající v záměně, vynechání nebo nadbytečném užití spojovacích výrazů. U jednotlivých typů chyb jsme dále zkoumali, jaká je jejich frekvence a vztah k úrovním podle SERRJ. Zároveň jsme sledovali, u kterých spojovacích prostředků k chybám dochází, a pokusili jsme se o základní interpretaci s ohledem na proměny mezijazyka s rostoucí úrovní jazykové kompetence.

V produkcích nerodilých mluvčích nejčastěji dochází k prvnímu typu chyb, tj. k záměnám jednoho spojovacího výrazu za jiný. Celkový počet chybně užitých prostředků ale není nijak významný, představuje pouze 3,31 % z celkového počtu všech výskytů. I přesto, že je počet chyb na všech úrovních relativně nízký, potvrdila se pozitivní korelace stoupající úrovně

podle SERRJ a bezchybného ovládní spojovacích prostředků. Závislost počtu chyb na úrovni je signifikantní na hladině významnosti 0,001 (Chí-kvadrát test nezávislosti). Jinými slovy řečeno na vyšších úrovních se uživatelé dopouštějí menšího počtu chyb. K záměnám dochází nejčastěji mezi synsémantickými spojovacími prostředky v hypotaktických spojeních. Záměny parataktických spojovacích prostředků jsou méně časté a vyznačují se jen omezenou variabilitou. 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. Frekvenční analýza a rozbor dokladových vět nasvědčují, že k záměnám dochází tam, kde se uživatelé jazyka snaží vyjadřovat nové mezivětné vztahy, pro něž si doposud (dostatečně) neosvojili adekvátní spojovací prostředky.

Druhý typ chyb – vynechání spojovacího prostředku – je sice podstatně méně frekventovaný, ale i u něho se projevuje klesající tendence v závislosti na rostoucí úrovni podle SERRJ. Týká se pouze omezeného počtu výrazů a nejsilněji se projevuje u spojek *že* a *jestli/zda* v nepřímých otázkách a u vztažných zájmen *co* a *který* v komplexních souvětých strukturách s korelativním spojením. Ojediněle je v korpusu doloženo nadbytečné užití spojovacích prostředků *a*, *ale*, *že* a *aby*, a to zejména na úrovních A1 a B1.

Zvlášť jsme vyčlenili ortografické a morfologické chyby, které přímo nesouvisí s dovedností explicitního vyjadřování mezivětných vztahu pomocí spojovacích prostředků. Odráží se pouze ve formální stránce jejich realizace. Morfologické chyby vykazuje pouze 0,22 % ze všech spojovacích prostředků. Svou podstatou zasahují výhradně autosémantické spojovací prostředky – v korpusu zájmena *který*, *co*, *kdo*, *jaký*, *jenž*. Pravopisné chyby jsou sice variabilnější a týkají se všech druhů spojovacích prostředků, i přesto se ale objevují pouze u 1,39 % z nich.

4.2 SONDA DO KORPUSU CZESL-SGT A SKRIPT2012 (A SYN2015)

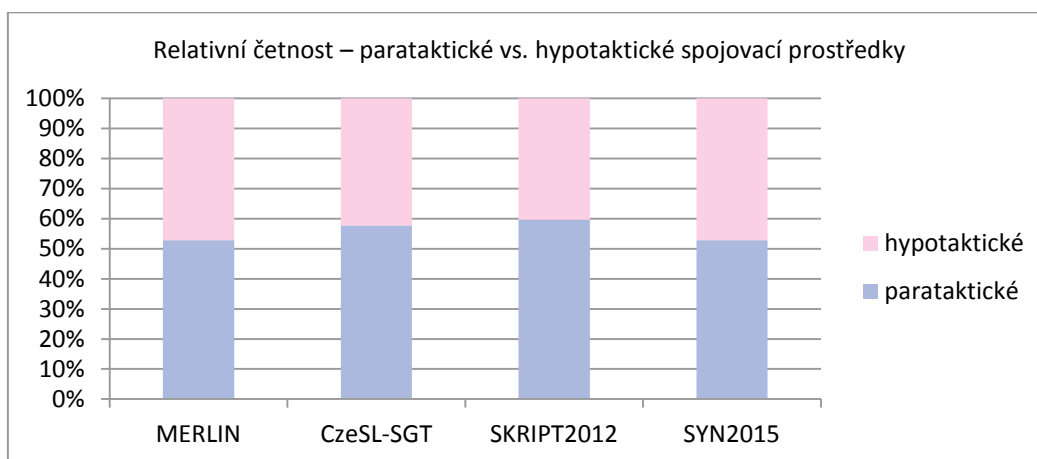
V předchozí kapitole jsme provedli podrobnou analýzu dokladů z korpusu MERLIN. Následující sondou do korpusu CzeSL-SGT bychom chtěli ověřit, zda se zjištěné tendence projevují i v korpusu většího rozsahu. Pohled do korpusu SKRPIT2012 nám umožní sledovat, zda/jak tyto tendence korespondují s užíváním spojovacích prostředků rodilými mluvčími, kteří jsou také ve fázi osvojování jazyka. U některých přehledů přidáváme na okraj i data z korpusu SYN2015, který můžeme v kontextu předkládané práce považovat za korpus „profesionálních“ rodilých mluvčích.

V teoretické části předkládané práce (3.1.2) jsme uvedli, že korpus CzeSL-SGT nedisponuje stejně ověřeným přiřazením textů k úrovním SERRJ jako MERLIN (korpus SKRIPT2012 není k SERRJ vztažen vůbec), omezíme se proto pouze na vybrané globální analýzy. Bude nás zajímat: (1) celkový poměr četnosti spojení podle syntaktické funkce; (2) pořadí spojovacích prostředků podle frekvence a (3) závislost užití spojovacích prostředků na typu korpusu.

4.2.1 Poměr četnosti spojení podle syntaktické funkce

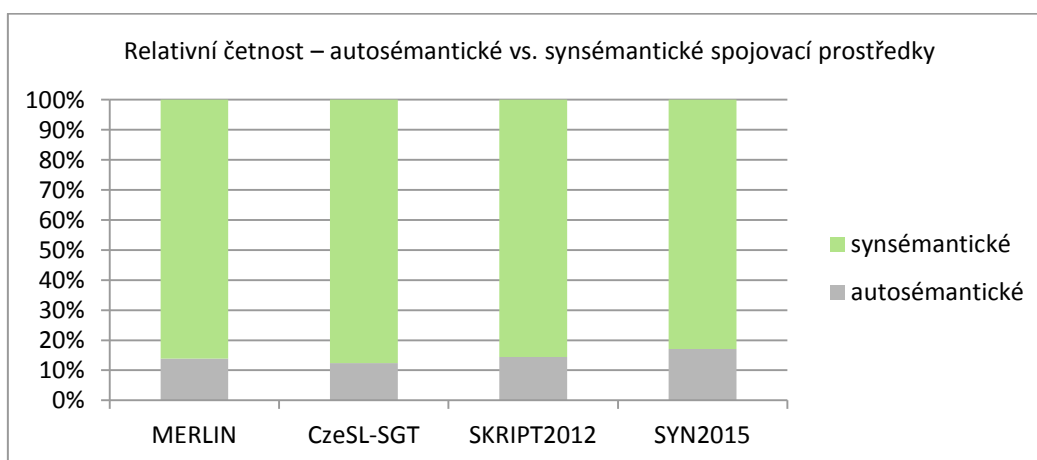
Popis spojovacích prostředků korpusu MERLIN jsme uvedli analýzou poměru, jakým jsou v dokladech zastoupeny prostředky podle jejich syntaktické funkce (parataktické vs. hypotaktické; autosémantické vs. synsémantické). K této charakteristice složení dokladového materiálu jsme se dále odvolávali například při chybové analýze. Považujeme proto za důležité alespoň orientačně ověřit, že je zjištěné rozložení pro písemné produkce nerodilých mluvčích rozložením běžným a že není zkresleno například omezenou velikostí korpusu nebo jeho složením. Zároveň nás zajímá, zde se nějak podstatně odlišuje od stavu v korpusech rodilých mluvčích.

Za tímto účelem jsme nejprve porovnali četnosti parataktických a hypotaktických spojovacích prostředků, které jsou zastoupeny v korpusu MERLIN (81 spojovacích prostředků, viz výše kap. 4.1), s jejich četností v korpusu CzeSL-SGT a SKRIPT2012 a SYN2015. Analýza sice ukázala nepatrné rozdíly mezi korpusy MERLIN a CzeSL-SGT, poměr četnosti dokladů je však v obou případech vyrovnaný s náznakem převahy parataktických spojení (graf 40). Zároveň se nijak výrazněji neliší od poměru získaného z korpusů rodilých mluvčích.



Graf 40: Relativní četnost (i.p.m.) – parataktické vs. hypotaktické spojovací prostředky

Stejným způsobem jsme porovnávali četnost spojovacích prostředků autosémantických a synsémantických. I v tomto případě všechny sledované korpusy vykazují převahu spojení synsémantických. Korpusy MERLIN a CzeSL-SGT jsou poměrem blíže ke korpusu SKRIPT2012, mezi sebou navzájem nevykazují podstatný rozdíl.



Graf 41: Relativní četnost – autosémantické vs. synsémantické spojovací prostředky

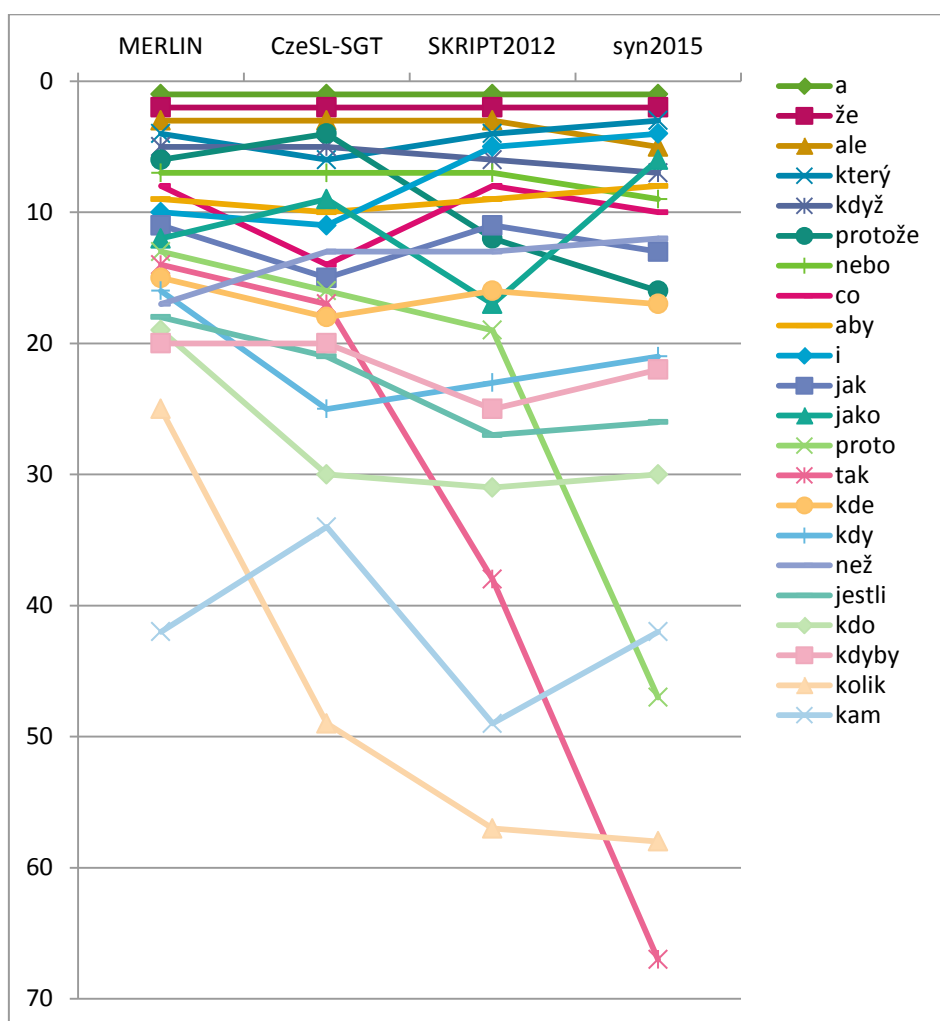
Jsme si vědomi, že se v obou případech jedná pouze o velmi hrubé srovnání, protože, jak znovu připomínáme, jsme se při porovnání omezili pouze na četnosti spojovacích prostředků obsažených v repertoáru korpusu MERLIN, ale všechny tři zbylé korpusy podle namátkového vyhledávání obsahují širší repertoár spojovacích prostředků. Jeho složení samozřejmě může ovlivnit i poměr spojení tříděných podle formálně syntaktických vztahů a větněčlenské platnosti. Přesto však tyto výsledky považujeme za důležité,

neboť se u průnikové množiny spojovacích prostředků nevyskytují žádné zásadní signály, které by nasvědčovaly tomu, že korpus MERLIN obsahuje abnormální složení dat.

4.2.2 Pořadí spojovacích prostředků podle frekvence

V kapitole 4.1.3 jsme ukázali, jak se v korpusu MERLIN mění pořadí spojovacích prostředků podle frekvence v závislosti na stoupající úrovni. U některých výrazů (zájmenná příslovce *kdy*, *kde*, *kam* a číslovka *kolik*) jsme získali podezření, že jejich celková frekvence může být ovlivněna návodnou formulací zadání testových úloh, na základě nichž texty vznikly. Abychom zjistili, jaká je situace v korpusu CzeSL-SGT, sestavili jsme i pro něj pořadí frekvence všech 81 spojovacích prostředků z korpusu MERLIN (graf 42).

Porovnáme-li v něm celkové pořadí těchto výrazů, zjistíme největší rozdíl u číslovky *kolik* a příslovce *kdy* (a také zájmena *kdo*). Zde by se tedy mohlo jednat o projev nadužívání v korpusu MERLIN. Výraz *kde* tak výrazný rozdíl nevykazuje a u příslovce *kam* je pořadí v korpusu CzeSL-SGT dokonce vyšší.



Graf 42: Srovnání korpusů: pořadí podle frekvence

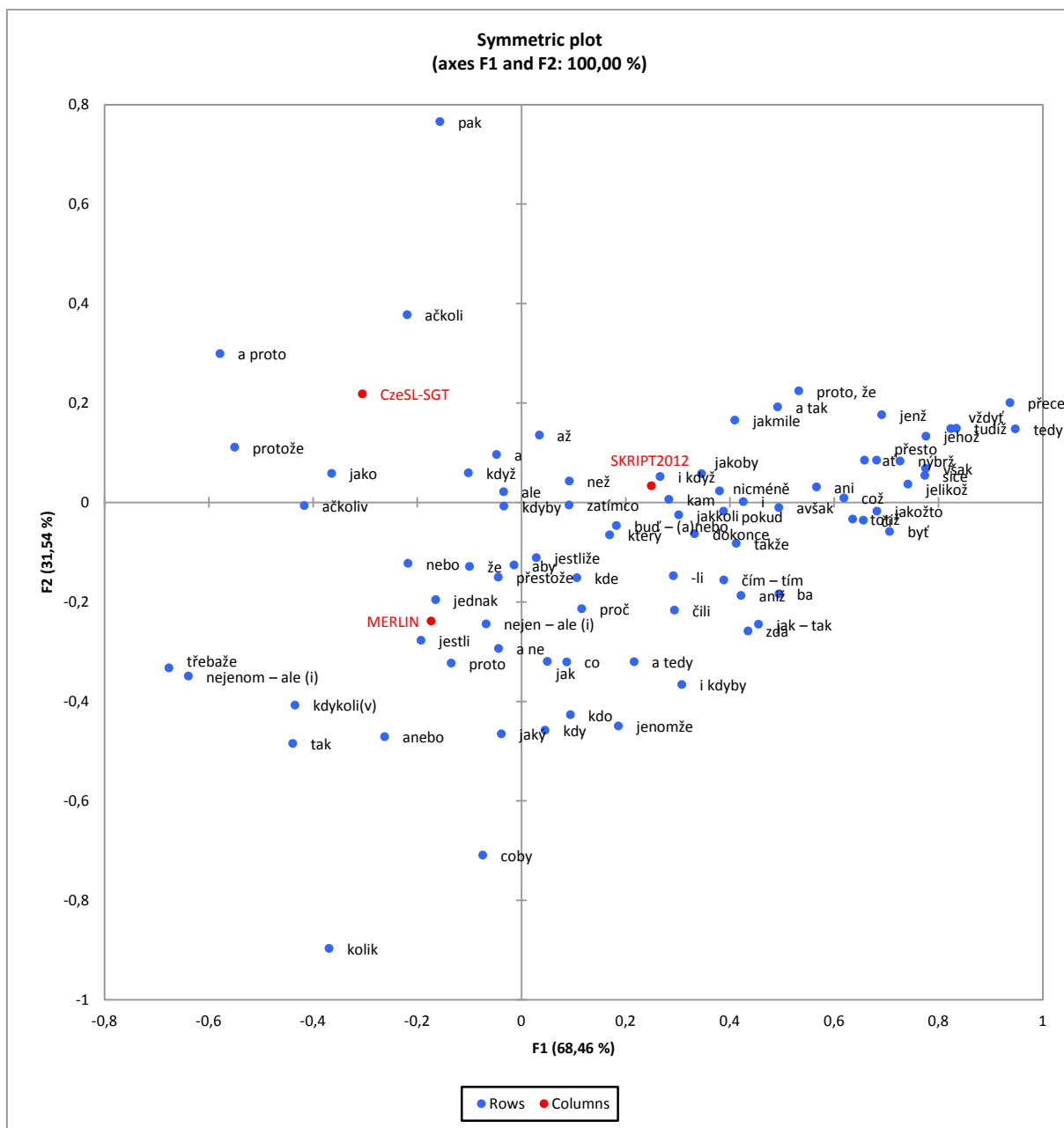
Další dva sloupce v grafu udávají pořadí podle frekvence v korpusech SKRIPT2012 a SYN2015 a nabízejí tak možnost sledovat, jak často konkrétní spojovací prostředky užívají rodilí mluvčí češtiny. Při detailnějším pohledu se ukazuje větší blízkost výsledků z obou korpusů nerodilých mluvčích s výsledky z korpusu SKRIPT2012. To je nejspíš dáno akvizičním charakterem, který je spojuje. I přesto můžeme u některých výrazů sledovat v produkci nerodilých mluvčích odlišné vzorce užívání. Jedná se o náznaky nadužívání – pokles spojnice směrem doprava (např. *protože, tak, jestli, kdo*), anebo naopak opomíjení – nárůst směrem doprava (např. *i*). Oba uvedené jevy představují běžný rys žakovských textů i v jiných jazycích, např. v angličtině se staly předmětem zkoumání řady studií (srov. Shea 2009, Muhdi – Hussein 2014 aj.). V češtině zatím podobné studie chybí a námi provedená sonda, jak doufáme, se stane jejich předzvěstí.

4.2.3 Závislost užití spojovacích prostředků na typu korpusu

V předchozí sondě jsme narazili na fakt, že se některé spojovací prostředky v textech rodilých mluvčích objevují častěji než v písmených produkcích nerodilých uživatelů češtiny a naopak. Z grafu 42, který udává pořadí spojovacích prostředků podle relativní frekvence, však nebyly zcela zřejmé některé souvislosti (např. seskupení výrazů s podobným vzorcem užívání), a navíc v něm nejsou zaneseny všechny výrazy, jimiž jsme se až doposud zabývali. I když tato problematika není předmětem našeho zkoumání, rozhodli jsme se, že se pokusíme alespoň o základní znázornění vztahů frekvence prostředků a typu korpusu pomocí korespondenční mapy, abychom tím poskytli východisko případným návazným výzkumům v budoucnosti. Relativní frekvence spojovacích prostředků z korpusu SKRIPT2012 jsme shromáždili a použili již při korespondenční analýze závislosti úrovní SERRJ a rozložení repertoáru v korpusu MERLIN (kap. 4.1.4). Nyní tato data využijeme a budeme sledovat vztah s frekvencemi ze zbylých dvou korpusů nerodilých mluvčích, který byl v podobě korespondenční mapy převeden do grafu 43.

Hlavní svislá osa tvoří pomyslnou hranici mezi texty nerodilých mluvčích (od osy doleva) a rodilých mluvčích (od osy vpravo). Zhruba polovina spojovacích prostředků, jež leží v blízkosti průsečíku hlavních os (souřadnice 0,0), je stejně využívána v obou skupinách. Mezi tyto prostředky patří nejfrekventovanější spojky *a, ale, že, když* a také spojky, které ve svém významovém okruhu nemají mnoho konkurenčních výrazů, např. *aby, kdyby, až*. Další kumulaci spojovacích prostředků můžeme pozorovat v pravé polovině grafu.

Jsou to prostředky, jež můžeme očekávat spíše v textech rodilých mluvčích a jejichž frekvence v produkcích nerodilých mluvčích obvykle stoupá až na vyšších úrovních (srov. kap. 4.1.3).



Graf 43: Vztah užití spojovacího prostředku v závislosti na typu korpusu

Naopak čím více vlevo se konkrétní spojovací prostředek nachází, tím silnější je jeho vztah k jazyku nerodilých mluvčích. Můžeme si všimnout, že na této straně grafu nedochází ke kumulaci, jako tomu je na straně opačné, vyskytují se zde spíše jednotlivé případy. Za nejspolehlivější pro formulování závěrů můžeme považovat výskyt prostředků v oblasti

kolem vodorovné osy, která představuje průnik domén korpusů CzeSL-SGT a MERLIN. Zde leží například spojka *protože*, což značí, že je preferovanější nerodilými mluvčími než rodilými. Podobnou vlastnost vykazují také výrazy *ačkoli* a *jako*.

Toto zjištění by mohlo zavdat důvod k dalšímu podrobnějšímu zkoumání, aby se ukázalo, jaká je příčina popsaného jevu. Bylo by třeba ověřit, zda se skutečně jedná o nadužívání, nebo pouze o anomální výskyty. Generalizaci na základě naší analýzy totiž omezuje skutečnost, že ani jeden z korpusů není úplně vhodný k takovému typu srovnání. Korpus MERLIN je svým rozsahem omezen pouze na úroveň A2–B2, CzeSL-SGT texty nižších úrovní obsahuje, ale celkově v něm převažují texty autorů se slovanským mateřským jazykem. SKRIPT2012 má sice akviziční charakter, ale odlišuje se věkovým složením autorů a zadáním i charakterem produkovaných textů. Nejprve by tedy bylo třeba uspokojivě vyřešit tuto skutečnost, např. vhodným sestavením srovnávacích subkorpusů a/nebo volbou náležitých statistických nástrojů, pak by teprve bylo možné vyvozovat závěry obecného charakteru.

4.3 SHRNUÍ VÝSLEDKŮ

Čtvrtá kapitola práce se věnovala analýze autentických textů s cílem zmapovat, jak nerodili 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.

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.

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*). 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. Jako nejprogresivnější se v tomto trendu jeví parataktické spojovací prostředky *i* a *dokonce*

a hypotaktické *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 (zájmenných 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.

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.

Další korespondenční analýza brala v potaz nikoliv celkové hodnocení, ale pouze hodnocení koheze a koherence podle SERRJ. Potvrdilo se očekávání, že závislost spojovacích prostředků na hodnocení koheze a koherence je silnější než závislost na celkovém hodnocení. Zároveň se objevil náznak toho, že posun k vyšší úrovni podle hodnocení koheze a koherence, která závisí na repertoáru spojovacích prostředků, probíhá dříve, než uživatel dosáhne vyšší jazykové kompetence v celé šíři.

Poslední část výzkumné práce 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. 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 *kteřý* v komplexních souvětých strukturách s korelativním spojením. Nadbytečné užití bylo doloženo pouze u *a*, *ale*, *že* a *aby*, a to zejména na úrovních A1 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. Zajímalo 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. 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.

5 ZÁVĚR

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.

Před samotnou analýzou autentického materiálu jsme v druhé kapitole nejprve identifikovali nezbytná teoretická východiska. Věnovali jsme se podrobněji otázce vymezení spojovacích prostředků v současném lingvistickém popisu češtiny a přiklonili jsme se ke klasifikaci na základě formálně výrazového kritéria. Hlavními se nám staly opozice hypotaktických vs. parataktických a autosémantických vs. synsémantických prostředků. Zvláště jsme se dotkli problematických míst, jako například polyfunkčnosti některých výrazů, u nichž jsme předpokládali, že je bude nutné zohlednit při zpracování korpusových dat. Nastínili jsme také charakteristiku úrovní jazykové kompetence s ohledem na užití spojovacích prostředků podle SERRJ a podle popisů referenčních úrovní pro češtinu. Při tom jsme zjistili, že zmíněné dokumenty uvádějí buď velmi obecné a/nebo empiricky nedostatečně validované rozlišení úrovní. Tím se mimo jiné potvrdila důležitost a aktuálnost našeho výzkumu. V neposlední řadě jsme se v teoretické části práce zabývali teorií mezijazyka a základy chybové analýzy.

Ve třetí kapitole byly charakterizovány korpusy, se kterými jsme pracovali. Nejpodrobněji jsme rozebírali složení a stavbu korpusu MERLIN, který pro nás představoval primární zdroj dat. Vysvětlili jsme, jakým způsobem je v něm provedeno hodnocení úrovní podle SERRJ, s nímž jsme operovali téměř ve všech našich analýzách. Podstatnou část této kapitoly dále zaujímá rozbor možností vyhledávání spojovacích prostředků a zjištění jejich úplného repertoáru. Vypořádat jsme se při něm museli jak se značnou rozmanitostí spojovacích prostředků, tak s jejich funkční homonymií (např. u výrazů *jak, co* apod.). Na základě tohoto rozboru jsme navrhli postup, který nám umožnil v co největší míře zapojit automatické nástroje vyhledávání a zajistit srovnatelnost vzorků získaných z různých korpusů. I přesto

jsme podstatnou část analýzy nakonec prováděli ručně. To se týkalo především chybové analýzy dokladů, která byla v předkládané práci relativně úzce zacílená, a proto se při ní nedala uspokojivě využít chybová anotace tvůrců korpusu.

Čtvrtá kapitola představuje jádro disertační práce. Věnuje se popisu a charakteristice užívání spojovacích prostředků ve shromážděných dokladech. Vycházeli jsme od globálního pohledu na rozložení repertoáru a frekvenci užití konkrétních výrazů i jejich skupin, dále se pozornost obrátila na jednotlivé úrovně podle SERRJ, u nichž bylo pomocí korespondenční analýzy zkoumáno, jak se v užívání spojovacích prostředků odlišují a které prostředky jsou pro ně typické. Provedená chybová analýza posloužila jako podklad pro stanovení závislosti četnosti chyb na dosažené úrovni jazykové kompetence.

Na základě výsledků výše popsaných dílčích kroků 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* $\langle \rangle$ *jestli < že* $\langle \rangle$ *co < který* $\langle \rangle$ *jaký* $\langle \rangle$ *jako* $\langle \rangle$ *jak > kdyby < když* $\langle \rangle$ *kdy; proto* $\langle \rangle$ *protože > ale < a* $\langle \rangle$ *i > ale < však* $\langle \rangle$ *avšak*; kde *X > Y* značí, že výraz *X* se chybně užívá místo výrazu *Y*. Všem typům chyb je společné, že se zvyšující úrovní ustupují.

Možnost využití uvedených výsledků analýz a závěrů této disertační práce spatřujeme především v jejich dalším zpracování a didaktické transformaci do učebních materiálů a metod výuky češtiny jako cizího jazyka. Protože se jedná o první takto pojatou studii syntaxe písemné produkce nerodilých mluvčích, přináší s sebou 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.

LITERATURA

- 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.
- Arras, U. (2010): Subjektive Theorien als Faktor bei der Beurteilung fremdsprachlicher Kompetenzen. In: Berndt, A. – Kleppin, K. (eds.): *Sprachlehrforschung: Theorie und Empirie – Festschrift für Rüdiger Grotjahn*. Frankfurt: Lang, s. 169–179.
- Bauer, J. (1958): Parataxe a hypotaxe při studiu souvětí. In: *O vědeckém poznání soudobých jazyků*. Praha: ČSAV, s. 270–278.
- Bauer, J. (1962): Spojky a příslovce. In: Grepl – Mrázek, R. – Večerka, R. (eds.): *Syntactica slavica: vybrané práce ze slovanské skladby*. Brno: Univerzita J. E. Purkyně, s. 29–37.
- Bauer, J. (1962/63): Relativa a spojky. In: *Acta Universitatis Carolinae. Slavica Pragensia* 4, s. 221–226.
- Bauer, J. (1964): Spojky a částice. In: *Sborník prací Filozofické fakulty brněnské univerzity*, A 12, s. 131–139.
- 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.
- Biber, D. – Xeppen, R. (1998): Comparing native and learner perspectives on English grammar: A study of complement clauses. In: Granger, S. (ed.): *Learner English on Computer*. London: Addison Wesley Longman, s. 145–158.
- Bulejčíková, P. (2014): *Problematika spisovnosti se zřetelem k výuce češtiny jako cizího jazyka*. Disertační práce. Praha: FF UK.
- 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>
- Carlsen, C. (ed.) (2013): *Norsk Profil. Det felles europeiske rammeverket spesifisert for norsk. Et første steg*. Oslo: Novus.
- Corder, S. P. (1981): *Error Analysis and Interlanguage*. Oxford: OUP.
- Corder, S. P. (1974): Idiosyncratic Dialects and Error Analysis. In: Richards, J. (ed.). *Error analysis: Perspectives on Second Language Acquisition*. Essex: Longman, s. 158–171.
- 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.

- ČKZK2 = Holá, L. – Bořilová, P. (2009): *Česky krok za krokem 2: Czech step by step 2*. Praha: Akropolis.
- ČpCB2 = Kestřánková, M. – Hlínová, K. – Pečený, P. – Štěpánková, D. (2013): *Čeština pro cizince: úroveň B2: vhodná pro všechny národnosti*. Brno: Edika, 2013.
- ČRaJ = Čechová, M. a kol. (2000): *Čeština – řeč a jazyk*. Praha: ISV.
- Daneš, F. (1977): Zpracování souvětí v nové mluvnici češtiny. *Slavica Slovaca*, 12, s. 244–252.
- Daneš, F. (1985): *Věta a text*. Praha: Academia.
- 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.
- Dulay, H. – Burt, M. – Krashen, S. (1982): *Language Two*. Oxford: OUP.
- Ellis, R. (1994): *The Study of Second Language Acquisition*. Oxford: OUP.
- Ellis, R. – Barkhuizen, G. (2009): *Analysing learner language*. Oxford: OUP.
- ESČ = *Encyklopedický slovník češtiny* (2002). Praha: NLN.
- Grepl, M. (1977–8): Souvětí v sémanticky orientovaném popisu syntaxe. In: *Sborník prací Filozofické fakulty brněnské univerzity*, A 25/26, s. 7–29.
- Grepl, M. – Karlík, P. (1986): *Skladba spisovné češtiny*. Praha: SPN.
- Grepl, M. – Karlík, P. (1998): *Skladba češtiny*. Olomouc: Votobia.
- Havránek, B. – Jedlička, A. (1960): *Česká mluvnice*. Praha: SPN.
- Havránek, B. – Jedlička, A. (1988): *Česká mluvnice*. Praha: SPN.
- Havránek, B. – Jedlička, A. (1996): *Stručná mluvnice česká*. Praha: Fortuna.
- Hebák a kol. (2007). *Vícerozměrné statistické metody (3)*. Praha: Informatorium.
- Hošnová, E. (2005): *Studie z vývoje novočeské syntaxe (konjunkce, pronominalizace)*. Praha: Karolinum.
- Hrbaček, J. (1964a): O syntaktické funkci korelativního spojení řídicí a závislé věty. *Slovo a slovesnost*, 25, s. 81–95.
- Hrbaček, J. (1967): K poměru mezi spojovacími prostředky členskými a větnými (Podřadicí spojky v jednoduché větě). *Naše řeč*, 50, s. 138–144.
- 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.
- Komárek, M. (1956): K otázce slovních druhů v češtině (Odpověď na kritiku). In: *Slovo a slovesnost*, 17, s. 160–168.
- Kopečný, F. (1962): *Základy české skladby*. Praha: SPN.
- Mašín, J. (2015): *Pragmalingvistika a osvojování češtiny jako cizího jazyka*. Praha: FF UK.
- MERLIN project (2014): *User Manual*. [cit. 15. 3. 2017] Dostupné z: http://www.merlin-platform.eu/docs/MERLIN_user-manual-CZ.pdf
- MČ2 = *Mluvnice češtiny 2* (1986). Praha: Academia.
- MČ3 = *Mluvnice češtiny 3* (1987). Praha: Academia.
- MSČ1 = Cvrček, V. a kol. (2010): *Mluvnice současné češtiny 1 – Jak se píše a jak se mluví*, Praha: Karolinum.
- Mudhhdí, S. K. – Hussein, R. F. (2014): A Corpus-based Study of Conjunctive Adjuncts in the Writings of Native and Non-native Speakers of English. *English Linguistics Research Vol. 3, No. 2*.
- NESČ = *Nový encyklopedický slovník češtiny online*. [cit. 15. 3. 2017] Dostupné z: <https://www.czechency.org>
- Pečený, P. (2010): K sémantice srovnávacích konstrukcí se spojovacím „jak“. *Jazykovědné aktuality*, roč. 47, č. 1–2, s. 23–37.
- Petkevič, V. (2006): Reliable Morphological Disambiguation of Czech: Rule-Based Approach is Necessary. In: Šimková, M. (ed.): *Insight into the Slovak and Czech Corpus Linguistics*. Veda, Bratislava, s. 26–44.
- 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.
- PMČ = *Příruční mluvnice češtiny* (1995). Praha: NLN.
- 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>

- Rešková, I. – Pintarová, M. (2004): *Communicative Czech (Intermediate Czech)*. [Brno]: I. Rešková.
- 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. – Mirovský, 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
- 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.
- Skutil, M. a kol. (2011): *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Praha: Portál.
- Spinelli, B. – Parizzi, F. (eds.) (2010): *Profilo della lingua italiana*. Firenze: La Nuova Italia.
- SSJČ⁸⁶ = *Slovník spisovného jazyka českého I–IV*. Praha: Nakladatelství Československé akademie věd / Academia, 1960–1971. [cit. 15. 3. 2017] Dostupné z: <http://ssjc.ujc.cas.cz>
- SSČ = *Slovník spisovné češtiny pro školu a veřejnost*. Praha: Academia, 2003. [cit. 15. 3. 2017] Dostupné z: <http://prirucka.ujc.cas.cz>
- 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
- Šmilauer, V. (1962): *Nauka o českém jazyce*. Praha: SPN.
- Šmilauer, V. (1966): *Novočeská skladba*. Praha: SPN.
- Štěpán, J. (2013): *Souvětí současné češtiny*. Liberec, TUL.

⁸⁶ Pokud není uvedeno jinak, citujeme z příslušných hesel elektronické verze slovníků SSČ a SSJČ.

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. (2011): *Evaluace chybové anotace v žakovském korpusu češtiny*. Disertační práce. Praha: FF UK Ústav českého jazyka a teorie komunikace.

Štindlová, B. (2012): Chybové taxonomie a možnosti chybové anotace v žakovských korpusech. In: Šebesta, K. – Škodová, S. a kol. (2012): *Čeština – cílový jazyk a korpusy*. Liberec: TUL, s. 35–60.

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

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

Trávníček, F. (1949): *Mluvnice spisovné češtiny. Část 2, Skladba*. Praha: Melantrich.

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.

Seznam tabulek

Tabulka 1:	Společné referenční úrovně: globálně pojatá stupnice (SERRJ 2001, s. 24)	33
Tabulka 2:	Společné referenční úrovně: sebehodnocení (SERRJ 2001, s. 26n.)	34
Tabulka 3:	SERRJ: Písemný projev – všeobecná stupnice (SERRJ 2001, s. 63).....	35
Tabulka 4:	SERRJ: Koherence (SERRJ 2001, s. 128).....	35
Tabulka 5:	Syntaktická kompetence podle SERRJ	36
Tabulka 6:	Spojovací prostředky očekávané na úrovni A1 (srov. PopisA1 2005, s. 262n.)	38
Tabulka 7:	Spojovací prostředky očekávané na úrovni A2 (srov. PopisA2 2005, s. 214–229)	39
Tabulka 8:	Spojovací prostředky očekávané na úrovni B1 (PopisB1 2001, s. 315–318).....	40
Tabulka 9:	Spojovací prostředky očekávané na úrovni B2 (PopisB2 2005, s.314–317).....	41
Tabulka 10:	Složení a rozsah korpusu MERLIN	49
Tabulka 11:	MERLIN: složení korpusu podle typu a tématu zadání textů	51
Tabulka 12:	Složení a rozsah korpusu CzeSL-SGT	57
Tabulka 13:	SKRIPT2012: Složení korpusu podle stupně výuky	60
Tabulka 14:	Korpusy SYN.....	61
Tabulka 15:	Tagy pro vtažná zájmena	63
Tabulka 16:	Tagy pro vzažná nebo tázací zájmena.....	64
Tabulka 17:	Tagy pro spojky jednoslovné	65
Tabulka 18:	SYN2015: Nejfrekventovanější víceslovné kombinace spojek	66
Tabulka 19:	MERLIN: Nejfrekventovanější víceslovné kombinace spojek	66
Tabulka 20:	SYN2015: Nejfrekventovanější kombinace částice a spojky.....	67
Tabulka 21:	MERLIN: Nejfrekventovanější kombinace částice a spojky	67
Tabulka 22:	SYN2015: Nejfrekventovanější kombinace spojky a příslovce.....	67
Tabulka 23:	MERLIN: Totožné výrazy se spojovací funkcí a bez ní	72
Tabulka 24:	MERLIN: Výpočet i.p.m pro spojku a.....	73
Tabulka 25:	MERLIN: Pořadí spojovacích prostředků podle frekvence	77
Tabulka 26:	MERLIN a SYN2015: 10 nejčtenějších spojovacích prostředků	83
Tabulka 27:	MERLIN: Nejméně zastoupené spojovací prostředky	84
Tabulka 28:	Globální charakteristika úrovní podle spojovacích prostředků	85
Tabulka 29:	MERLIN: Změny v repertoáru podle úrovní (pořadí podle frekvence).....	86
Tabulka 30:	MERLIN: Skupiny konkurenčních prostředků	87
Tabulka 31:	MERLIN: Tendence v rozšiřování skupin konkurenčních prostředků.....	88
Tabulka 32:	MERLIN: Tendence v pořadí podle relativní frekvence A2–B2	91
Tabulka 33:	MERLIN: Tendence v pořadí podle relativní frekvence B1–B2.....	93
Tabulka 34:	Charakteristika nejčastěji zaměňovaných spojovacích prostředků	105

Seznam grafů

Graf 1:	Ukázka vizuálního odlišení úrovní A1 a C1 (i.p.m. podle úrovní).....	50
Graf 2:	MERLIN: Rozložení podle L1 (% podle počtu tokenů).....	50
Graf 3:	MERLIN: Podíl L1 na složení korpusu (% podle počtu tokenů).....	51
Graf 4:	CzeSL-SGT vs. MERLIN: Rozložení podle úrovní (% podíl v celém korpusu na základě počtu tokenů).....	57
Graf 5:	CzeSL-SGT: Podíl L1 na složení korpusu (% podle počtu tokenů).....	58
Graf 6:	CzeSL-SGT: Podíl L1 na složení korpusu (% podle počtu tokenů).....	58
Graf 7:	CzeSL-SGT vs. MERLIN: Rozložení podle L1 (% podle počtu tokenů).....	58
Graf 8:	Korpus MERLIN: 10 nejfrekventovanějších spojovacích výrazů (podle počtu).....	76
Graf 9:	Korpus MERLIN: Počet spojovacích prostředků podle podoby.....	76
Graf 10:	MERLIN: Repertoár spojovacích prostředků – počet parataktické vs. hypotaktické.....	78
Graf 11:	MERLIN: Relativní četnost – parataktické vs. hypotaktické spojovací prostředky.....	78
Graf 12:	MERLIN: Repertoár spojovacích prostředků – autosémantické vs. synsémantické.....	79
Graf 13:	MERLIN: Relativní četnost – autosémantické vs. synsémantické spojovací prostředky.....	79
Graf 14:	Korpus MERLIN: Relativní četnost (i.p.m.) spojovacích prostředků podle úrovní.....	80
Graf 15:	Korpus MERLIN: Repertoár (počet různých) spojovacích prostředků.....	80
Graf 16:	MERLIN: Poměr počtu spojovacích prostředků podle podoby.....	81
Graf 17:	MERLIN: Poměr počtu různých spojovacích prostředků na jednotlivých úrovních podle formálně syntaktických vztahů (repertoár).....	81
Graf 18:	MERLIN: Relativní frekvence užití spojovacích prostředků na jednotlivých úrovních podle formálně syntaktických vztahů.....	82
Graf 19:	MERLIN: Poměr počtu různých spojovacích prostředků na jednotlivých úrovních podle větněčlenské platnosti (repertoár).....	82
Graf 20:	MERLIN: Relativní frekvence užití spojovacích prostředků na jednotlivých úrovních podle větněčlenské platnosti.....	83
Graf 21:	MERLIN: Relativní pořadí podle frekvence (percentil podle i.p.m.) – prvních 10 nejčetnějších.....	89
Graf 22:	MERLIN: Relativní pořadí podle frekvence (příčinné/důvodové vztahy).....	90
Graf 23:	MERLIN: Stoupající pořadí podle relativní frekvence A2–B2.....	91
Graf 24:	MERLIN: Klesající pořadí podle relativní frekvence A2–B2.....	92
Graf 25:	MERLIN: Soupavě-klesající pořadí podle relativní frekvence A2–B2.....	93
Graf 26:	MERLIN: Stoupající pořadí podle relativní frekvence B1–B2.....	94
Graf 27:	MERLIN: Klesající pořadí podle relativní frekvence B1–B2.....	94
Graf 28:	MERLIN: korespondenční analýza relativní frekvence a úrovní SERRJ.....	96
Graf 29:	MERLIN: korespondenční analýza relativní frekvence a úrovní SERRJ + barevně znázorněná frekvence výrazů v L1.....	98
Graf 30:	MERLIN: korespondenční analýza relativní frekvence a hodnocení koheze a koherence.....	100
Graf 31:	MERLIN: korespondenční analýza relativní frekvence, celkového hodnocení a hodnocení koheze a koherence.....	101
Graf 32:	Záměna spojovacích prostředků (procentuální zastoupení záměn v celkovém počtu dokladů).....	102
Graf 33:	Spojovací prostředky s nejčastější záměnou.....	103
Graf 34:	Nejčastěji zaměňované spojovací prostředky podle úrovní (i.p.m.).....	106
Graf 35:	MERLIN: Relativní frekvence vynechání spojovacího prostředku (i.p.m.).....	108
Graf 36:	Chybějící spojovací prostředky (podíl podle i.p.m.).....	109
Graf 37:	Nadbytečné užití spojovacích prostředků (procentuální zastoupení v celkovém počtu dokladů).....	110
Graf 38:	Pravopisné chyby ve spojovacím prostředku (procentuální zastoupení v celkovém počtu dokladů).....	111
Graf 39:	Morfologické chyby ve spojovacím prostředku (procentuální zastoupení v celkovém počtu dokladů).....	111
111		
Graf 40:	Relativní četnost (i.p.m.) – parataktické vs. hypotaktické spojovací prostředky.....	115
Graf 41:	Relativní četnost – autosémantické vs. synsémantické spojovací prostředky.....	115
Graf 42:	Srovnání korpusů: pořadí podle frekvence.....	116
Graf 43:	Vztah užití spojovacího prostředku v závislosti na typu korpusu.....	118

Seznam obrázků

Ukázka 1: Zadání testu A2 – popis fotografie	52
Ukázka 2: Zadání testu A2 – neformální e-mail	52
Ukázka 3: Zadání testu B1 – neformální e-mail	53
Ukázka 4: Zadání testu B2 – vyprávění / úvaha (zamyšlení).....	53

PŘÍLOHY

Příloha 1:	MERLIN: Počet výskytů v žákovském textu podle celkového hodnocení úrovně	135
Příloha 2:	MERLIN: Relativní frekvence (i.p.m.) podle celkového hodnocení úrovně	137
Příloha 3:	MERLIN: Počet různých spojovacích prostředků podle úrovně (repertoár)	138
Příloha 4:	MERLIN: Počet různých spojovacích prostředků podle funkce (repertoár).....	138
Příloha 5:	MERLIN: Počet dokladů spojovacích prostředků podle funkce.....	138
Příloha 6:	MERLIN: Frekvence spojovacích prostředků podle funkce (i.p.m.)	138
Příloha 7:	MERLIN: Počet spojovacích prostředků podle větněčlenské platnosti	139
Příloha 8:	MERLIN: Frekvence spojovacích prostředků podle větněčlenské platnosti (i.p.m.).....	139
Příloha 9:	MERLIN: Relativní pořadí podle frekvence (percentil podle i.p.m.)	139
Příloha 10:	Pořadí podle relativní četnosti.....	140
Příloha 11:	Výsledek Chí-kvadrát testu nezávislosti	143
Příloha 12:	Korespondenční analýza: spojovací prostředky vs. úrovně	144
Příloha 13:	Korespondenční analýza: spojovací prostředky vs. hodnocení koheze a koherence.....	154
Příloha 14:	MERLIN: Počet chyb – záměn	166
Příloha 15:	MERLIN: Relativní frekvence chyb – záměn.....	167

Příloha 1: MERLIN: Počet výskytů v žákovském textu podle celkového hodnocení úrovně

	A1	A2	B1	B2	C1	celkem
počet tokenů	54	21 505	34 007	19 953	1 167	76 686
počet slov	50	17 828	28 589	16 887	987	64 341
počet vět	5	2 515	2 808	1 227	73	6 628
počet textů	1	188	167	81	4	441
<i>a</i>	3	680	1 004	561	44	2 292
<i>aby</i>	0	12	104	59	5	180
<i>ačkoli</i>	0	0	0	1	0	1
<i>ačkoliv</i>	0	0	2	2	0	4
<i>ale</i>	2	154	313	140	7	616
<i>a ne</i>	0	0	5	3	1	9
<i>anebo</i>	0	2	1	6	0	9
<i>ani</i>	0	1	22	16	6	45
<i>aniž</i>	0	0	1	2	0	3
<i>a proto</i>	0	10	14	18	0	42
<i>a tedy</i>	0	0	1	0	0	1
<i>a tak</i>	0	2	9	4	0	15
<i>ať</i>	0	0	3	4	0	7
<i>avšak</i>	0	0	2	5	0	7
<i>až</i>	0	9	9	14	1	33
<i>ba</i>	0	0	1	2	0	3
<i>bud' – (a)nebo</i>	0	0	2	4	0	6
<i>bud'to – (a)nebo</i>	0	1	0	0	0	1
<i>byť</i>	0	0	0	1	0	1
<i>co</i>	0	23	116	71	4	214
<i>coby</i>	0	0	0	1	0	1
<i>což</i>	0	0	4	9	0	13
<i>či</i>	0	0	4	16	0	20
<i>čili</i>	0	0	1	0	0	1
<i>čím – tím</i>	0	0	2	5	0	7
<i>dokonce</i>	0	1	4	13	0	18
<i>i</i>	0	9	75	88	6	178
<i>i kdyby</i>	0	0	0	3	0	3
<i>i když</i>	0	0	9	11	0	20
<i>jak</i>	0	14	90	55	3	162
<i>jakmile</i>	0	0	3	0	0	3
<i>jako</i>	0	16	50	58	0	124
<i>jakoby</i>	0	0	0	1	0	1
<i>jakkoli</i>	0	0	1	0	0	1
<i>jakožto</i>	0	0	0	1	0	1
<i>jaký</i>	0	4	20	19	1	44
<i>jak – tak</i>	0	0	4	5	1	10
<i>jednak</i>	0	0	1	1	0	2
<i>jehož</i>	0	0	1	1	0	2
<i>jelikož</i>	0	0	1	3	0	4
<i>jenomže</i>	0	0	0	2	0	2
<i>jenž</i>	0	0	0	7	0	7
<i>jestli</i>	0	20	41	15	0	76
<i>jestliže</i>	0	0	2	1	0	3
<i>kam</i>	0	4	3	1	0	8
<i>kde</i>	0	30	42	16	0	88
<i>kdo</i>	0	8	20	19	2	49
<i>kdy</i>	0	31	36	9	3	79
<i>kdyby</i>	0	18	21	6	1	46
<i>kdykoli(v)</i>	0	0	2	1	0	3
<i>když</i>	0	29	162	120	4	315

<i>kolik</i>	0	15	12	5	0	32
<i>který</i>	0	33	167	186	8	394
<i>-li</i>	0	1	4	8	0	13
<i>nebo</i>	0	63	133	89	0	285
<i>nejen – ale (i)</i>	0	1	8	16	0	25
<i>nejenom – ale (i)</i>	0	0	4	3	0	7
<i>než</i>	0	4	36	37	0	77
<i>nicméně</i>	0	0	0	2	0	2
<i>nýbrž</i>	0	0	0	1	0	1
<i>pak</i>	0	0	0	1	0	1
<i>pokud</i>	0	3	6	8	0	17
<i>proč</i>	0	1	13	5	0	19
<i>proto</i>	0	28	43	45	3	119
<i>protože</i>	0	112	126	54	2	294
<i>proto, že</i>	0	0	0	2	0	2
<i>přece</i>	0	0	1	0	0	1
<i>přesto</i>	0	0	3	3	0	6
<i>přestože</i>	0	1	3	4	0	8
<i>sice</i>	0	0	4	2	0	6
<i>tak</i>	0	11	71	31	3	116
<i>takže</i>	0	5	8	4	0	17
<i>tedy</i>	0	0	2	2	0	4
<i>totiž</i>	0	1	2	4	5	12
<i>třebaže</i>	0	0	0	1	0	1
<i>tudíž</i>	0	0	0	1	0	1
<i>však</i>	0	0	4	13	0	17
<i>vždyť</i>	0	0	0	3	0	3
<i>zatímco</i>	0	0	0	3	0	3
<i>zda</i>	0	3	5	1	0	9
<i>že</i>	0	235	495	318	13	1 061
CELKEM	5	1 595	3 358	2 252	123	7 333

Príloha 2: MERLIN: Relativní frekvence (i.p.m.) podle celkového hodnocení úrovně

	A1	A2	B1	B2	C1	celkem
<i>a</i>	55 555,56	31 620,55	29 523,33	28 116,07	37 703,51	29 888,12
<i>aby</i>	0,00	558,01	3 058,19	2 956,95	4 284,49	2 347,23
<i>ačkoli</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>ačkoliv</i>	0,00	0,00	58,81	100,24	0,00	52,16
<i>ale</i>	37 037,04	7 161,13	9 203,99	7 016,49	5 998,29	8 032,76
<i>a ne</i>	0,00	0,00	147,03	150,35	856,90	117,36
<i>anebo</i>	0,00	93,00	29,41	300,71	0,00	117,36
<i>ani</i>	0,00	46,50	646,93	801,88	5 141,39	586,81
<i>aniž</i>	0,00	0,00	29,41	100,24	0,00	39,12
<i>a proto</i>	0,00	465,01	411,68	902,12	0,00	547,69
<i>a tedy</i>	0,00	0,00	29,41	0,00	0,00	13,04
<i>a tak</i>	0,00	93,00	264,65	200,47	0,00	195,60
<i>ať</i>	0,00	0,00	88,22	200,47	0,00	91,28
<i>avšak</i>	0,00	0,00	58,81	250,59	0,00	91,28
<i>až</i>	0,00	418,51	264,65	701,65	856,90	430,33
<i>ba</i>	0,00	0,00	29,41	100,24	0,00	39,12
<i>bud' – (a)nebo</i>	0,00	0,00	58,81	200,47	0,00	78,24
<i>bud'to – (a)nebo</i>	0,00	46,50	0,00	0,00	0,00	13,04
<i>byť</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>co</i>	0,00	1 069,52	3 411,06	3 558,36	3 427,59	2 790,60
<i>coby</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>což</i>	0,00	0,00	117,62	451,06	0,00	169,52
<i>či</i>	0,00	0,00	117,62	801,88	0,00	260,80
<i>čili</i>	0,00	0,00	29,41	0,00	0,00	13,04
<i>čím – tím</i>	0,00	0,00	58,81	250,59	0,00	91,28
<i>dokonce</i>	0,00	46,50	117,62	651,53	0,00	234,72
<i>i</i>	0,00	418,51	2 205,43	4 410,36	5 141,39	2 321,15
<i>i kdyby</i>	0,00	0,00	0,00	150,35	0,00	39,12
<i>i když</i>	0,00	0,00	264,65	551,30	0,00	260,80
<i>jak</i>	0,00	651,01	2 646,51	2 756,48	2 570,69	2 112,51
<i>jakmile</i>	0,00	0,00	88,22	0,00	0,00	39,12
<i>jako</i>	0,00	744,01	1 470,29	2 906,83	0,00	1 616,98
<i>jakoby</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>jakkoli</i>	0,00	0,00	29,41	0,00	0,00	13,04
<i>jakožto</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>jaký</i>	0,00	186,00	588,11	952,24	856,90	573,77
<i>jak – tak</i>	0,00	0,00	117,62	250,59	856,90	130,40
<i>jednak</i>	0,00	0,00	29,41	50,12	0,00	26,08
<i>jehož</i>	0,00	0,00	29,41	50,12	0,00	26,08
<i>jelikož</i>	0,00	0,00	29,41	150,35	0,00	52,16
<i>jenomže</i>	0,00	0,00	0,00	100,24	0,00	26,08
<i>jenž</i>	0,00	0,00	0,00	350,82	0,00	91,28
<i>jestli</i>	0,00	930,02	1 205,63	751,77	0,00	991,05
<i>jestliže</i>	0,00	0,00	58,81	50,12	0,00	39,12
<i>kam</i>	0,00	186,00	88,22	50,12	0,00	104,32
<i>kde</i>	0,00	1 395,02	1 235,04	801,88	0,00	1 147,54
<i>kdo</i>	0,00	372,01	588,11	952,24	1 713,80	638,97
<i>kdy</i>	0,00	1 441,53	1 058,61	451,06	2 570,69	1 030,17
<i>kdyby</i>	0,00	837,01	617,52	300,71	856,90	599,85
<i>kdykoli(v)</i>	0,00	0,00	58,81	50,12	0,00	39,12
<i>když</i>	0,00	1 348,52	4 763,73	6 014,13	3 427,59	4 107,66
<i>kolik</i>	0,00	697,51	352,87	250,59	0,00	417,29
<i>kteřý</i>	0,00	1 534,53	4 910,75	9 321,91	6 855,18	5 137,83
<i>-li</i>	0,00	46,50	117,62	400,94	0,00	169,52
<i>nebo</i>	0,00	2 929,55	3 910,96	4 460,48	0,00	3 716,45

<i>nejen – ale (i)</i>	0,00	46,50	235,25	801,88	0,00	326,00
<i>nejenom – ale (i)</i>	0,00	0,00	117,62	150,35	0,00	91,28
<i>než</i>	0,00	186,00	1 058,61	1 854,36	0,00	1 004,09
<i>nicméně</i>	0,00	0,00	0,00	100,24	0,00	26,08
<i>nýbrž</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>pak</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>pokud</i>	0,00	139,50	176,43	400,94	0,00	221,68
<i>proč</i>	0,00	46,50	382,27	250,59	0,00	247,76
<i>proto</i>	0,00	1 302,02	1 264,45	2 255,30	2 570,69	1 551,78
<i>protože</i>	0,00	5 208,09	3 705,12	2 706,36	1 713,80	3 833,82
<i>proto, že</i>	0,00	0,00	0,00	100,24	0,00	26,08
<i>přece</i>	0,00	0,00	29,41	0,00	0,00	13,04
<i>přesto</i>	0,00	0,00	88,22	150,35	0,00	78,24
<i>přestože</i>	0,00	46,50	88,22	200,47	0,00	104,32
<i>sice</i>	0,00	0,00	117,62	100,24	0,00	78,24
<i>tak</i>	0,00	511,51	2 087,81	1 553,65	2 570,69	1 512,66
<i>takže</i>	0,00	232,50	235,25	200,47	0,00	221,68
<i>tedy</i>	0,00	0,00	58,81	100,24	0,00	52,16
<i>totiž</i>	0,00	46,50	58,81	200,47	4 284,49	156,48
<i>třebaže</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>tudíž</i>	0,00	0,00	0,00	50,12	0,00	13,04
<i>však</i>	0,00	0,00	117,62	651,53	0,00	221,68
<i>vždyť</i>	0,00	0,00	0,00	150,35	0,00	39,12
<i>zatímco</i>	0,00	0,00	0,00	150,35	0,00	39,12
<i>zda</i>	0,00	139,50	147,03	50,12	0,00	117,36
<i>že</i>	0,00	10 927,69	14 555,83	15 937,45	11 139,67	13 835,64
CELKEM	92 592,59	74 168,80	98 744,38	112 865,23	105 398,46	95 623,71

Příloha 3: MERLIN: Počet různých spojovacích prostředků podle úrovní (repertoár)

	A1	A2	B1	B2	C1	celkem
počet	2	38	64	75	21	81

Příloha 4: MERLIN: Počet různých spojovacích prostředků podle funkce (repertoár)

	A1	A2	B1	B2	C1	celkem
parataktické	2	14	28	30	8	34
hypotaktické	0	24	36	45	13	47

Příloha 5: MERLIN: Počet dokladů spojovacích prostředků podle funkce

	A1	A2	B1	B2	C1	celkem
parataktické	5	964	1 735	1 095	75	3 874
hypotaktické	0	631	1 623	1 157	48	3 459
celkem	5	1 595	3 358	2 252	123	7 333

Příloha 6: MERLIN: Frekvence spojovacích prostředků podle funkce (i.p.m.)

	A1	A2	B1	B2	C1	celkem
parataktické	92 592,59	44 826,78	51 018,91	54 878,97	64 267,35	50 517,70
hypotaktické	0,00	29 342,01	47 725,47	57 986,27	41 131,11	45 106,02

Priloha 7: MERLIN: Počet spojovacích prostředků podle větněčlenské platnosti

	A1	A2	B1	B2	C1	celkem
autosémantické	0	154	476	369	17	1 016
syntémantické	5	1 441	2 882	1 883	106	6 317
	5	1 595	3 358	2 252	123	7 333

Priloha 8: MERLIN: Frekvence spojovacích prostředků podle větněčlenské platnosti (i.p.m.)

	A1	A2	B1	B2	C1	celkem
autosémantické	0,00	7 161,13	13 997,12	18 493,46	14 567,27	13 248,83
syntémantické	92 592,59	67 007,67	84 747,26	94 371,77	90 831,19	82 374,88

Priloha 9: MERLIN: Relativní pořadí podle frekvence (percentil podle i.p.m.)

	A1	A2	B1	B2	C1	celkem
<i>a</i>	0,667	0,974	0,985	0,987	0,955	0,988
<i>že</i>		0,949	0,970	0,974	0,909	0,976
<i>ale</i>	0,333	0,923	0,955	0,947	0,818	0,964
<i>který</i>		0,846	0,939	0,961	0,864	0,952
<i>když</i>		0,769	0,924	0,934	0,545	0,940
<i>protože</i>		0,897	0,894	0,842	0,273	0,928
<i>nebo</i>		0,872	0,909	0,921		0,916
<i>co</i>		0,718	0,879	0,895	0,545	0,904
<i>aby</i>		0,564	0,864	0,882	0,636	0,892
<i>i</i>		0,462	0,833	0,908	0,727	0,880
<i>jak</i>		0,590	0,848	0,855	0,364	0,867
<i>jako</i>		0,641	0,803	0,868		0,855
<i>proto</i>		0,744	0,788	0,829	0,364	0,843
<i>tak</i>		0,538	0,818	0,803	0,364	0,831
<i>kde</i>		0,795	0,773	0,711		0,819
<i>kdy</i>		0,821	0,727	0,618	0,364	0,807
<i>než</i>		0,333	0,727	0,816		0,795
<i>jestli</i>		0,692	0,758	0,697		0,783
<i>kdo</i>		0,436	0,667	0,776	0,273	0,771
<i>kdyby</i>		0,667	0,697	0,553	0,045	0,759
<i>ani</i>		0,026	0,712	0,711	0,727	0,747
<i>jaký</i>		0,333	0,667	0,776	0,045	0,735
<i>a proto</i>		0,513	0,652	0,763		0,723
<i>až</i>		0,462	0,576	0,684	0,045	0,711
<i>kolik</i>		0,615	0,621	0,487		0,699
<i>nejen – ale (i)</i>		0,026	0,545	0,711		0,687
<i>či</i>			0,379	0,711		0,663
<i>i když</i>			0,576	0,645		0,663
<i>proč</i>		0,026	0,636	0,487		0,651
<i>dokonce</i>		0,026	0,379	0,658		0,639
<i>pokud</i>		0,282	0,530	0,592		0,602
<i>takže</i>		0,410	0,545	0,408		0,602
<i>však</i>			0,379	0,658		0,602
<i>a tak</i>		0,231	0,576	0,408		0,590
<i>což</i>			0,379	0,618		0,566
<i>-li</i>		0,026	0,379	0,592		0,566
<i>totiž</i>		0,026	0,167	0,408	0,636	0,554
<i>jak – tak</i>			0,379	0,487	0,045	0,542
<i>a ne</i>			0,500	0,316	0,045	0,506
<i>anebo</i>		0,231	0,015	0,553		0,506
<i>zda</i>		0,282	0,500	0,013		0,506

<i>kam</i>		0,333	0,303	0,013		0,482
<i>přestože</i>		0,026	0,303	0,408		0,482
<i>ať</i>			0,303	0,408		0,422
<i>avšak</i>			0,167	0,487		0,422
<i>čím – tím</i>			0,167	0,487		0,422
<i>jenž</i>				0,579		0,422
<i>nejenom – ale (i)</i>			0,379	0,316		0,422
<i>bud' – (a)nebo</i>			0,167	0,408		0,386
<i>přesto</i>			0,303	0,316		0,386
<i>sice</i>			0,379	0,211		0,386
<i>ačkoliv</i>			0,167	0,211		0,349
<i>jehož</i>			0,015	0,316		0,349
<i>tedy</i>			0,167	0,211		0,349
<i>aniž</i>			0,015	0,211		0,253
<i>ba</i>			0,015	0,211		0,253
<i>i kdyby</i>				0,316		0,253
<i>jakmile</i>			0,303			0,253
<i>jestliže</i>			0,167	0,013		0,253
<i>kdykoli(v)</i>			0,167	0,013		0,253
<i>vždyť</i>				0,316		0,253
<i>zatímco</i>				0,316		0,253
<i>jednak</i>			0,015	0,013		0,181
<i>jelikož</i>			0,015	0,013		0,181
<i>jenomže</i>				0,211		0,181
<i>nicméně</i>				0,211		0,181
<i>proto, že</i>				0,211		0,181
<i>ačkoli</i>				0,013		0,012
<i>a tedy</i>			0,015			0,012
<i>bud'to – (a)nebo</i>		0,026				0,012
<i>byť</i>				0,013		0,012
<i>coby</i>				0,013		0,012
<i>čili</i>			0,015			0,012
<i>jakoby</i>				0,013		0,012
<i>jakkoli</i>			0,015			0,012
<i>jakožto</i>				0,013		0,012
<i>nýbrž</i>				0,013		0,012
<i>pak</i>				0,013		0,012
<i>přece</i>			0,015			0,012
<i>třebaže</i>				0,013		0,012
<i>tudíž</i>				0,013		0,012

Příloha 10: Pořadí podle relativní četnosti

	MERLIN	CzeSL-SGT	SKRIPT2012	SYN2015
<i>a</i>	1	1	1	1
<i>že</i>	2	2	2	2
<i>ale</i>	3	3	3	5
<i>který</i>	4	6	4	3
<i>když</i>	5	5	6	7
<i>protože</i>	6	4	12	16
<i>nebo</i>	7	7	7	9
<i>co</i>	8	14	8	10
<i>aby</i>	9	10	9	8
<i>i</i>	10	11	5	4
<i>jak</i>	11	15	11	13
<i>jako</i>	12	9	17	6
<i>proto</i>	13	16	19	47

<i>tak</i>	14	17	38	67
<i>kde</i>	15	18	16	17
<i>kdy</i>	16	25	23	21
<i>než</i>	17	13	13	12
<i>jestli</i>	18	21	27	26
<i>kdo</i>	19	30	31	30
<i>kdyby</i>	20	20	25	22
<i>ani</i>	21	22	10	15
<i>jaký</i>	22	27	40	28
<i>a proto</i>	23	12	41	56
<i>až</i>	24	19	24	24
<i>kolik</i>	25	49	57	58
<i>nejen – ale (i)</i>	26	26	45	38
<i>či</i>	27	41	21	18
<i>i když</i>	27	24	29	31
<i>proč</i>	29	32	42	37
<i>dokonce</i>	30	29	36	55
<i>pokud</i>	31	28	32	19
<i>takže</i>	31	33	33	23
<i>však</i>	31	35	14	43
<i>a tak</i>	34	23	20	32
<i>což</i>	35	36	26	25
<i>-li</i>	35	36	44	81
<i>totiž</i>	37	51	30	27
<i>jak – tak</i>	38	64	48	40
<i>a ne</i>	39	45	58	63
<i>anebo</i>	39	50	64	50
<i>zda</i>	39	65	50	35
<i>kam</i>	42	34	49	42
<i>přestože</i>	42	38	56	46
<i>ať</i>	44	39	35	41
<i>avšak</i>	44	48	46	51
<i>čím – tím</i>	44	55	52	53
<i>jenž</i>	44	31	22	11
<i>nejenom – ale (i)</i>	44	40	76	76
<i>bud' – (a)nebo</i>	49	43	55	52
<i>přesto</i>	49	47	39	36
<i>sice</i>	49	58	34	33
<i>ačkoliv</i>	52	42	70	65
<i>jelikož</i>	52	63	43	62
<i>tedy</i>	52	71	18	20
<i>aniž</i>	55	68	61	49
<i>ba</i>	55	73	59	70
<i>i kdyby</i>	55	74	67	66
<i>jakmile</i>	55	44	54	48
<i>jestliže</i>	55	56	68	54
<i>kdykoli(v)</i>	55	59	78	72
<i>vždyť</i>	55	54	37	43
<i>zatímco</i>	55	53	63	34
<i>jednak</i>	63	62	74	61
<i>jehož</i>	63	59	47	29
<i>jenomže</i>	63	76	72	69
<i>nicméně</i>	63	59	62	45
<i>proto, že</i>	63	45	51	39
<i>ačkoli</i>	69	52	73	59
<i>a tedy</i>	69	76	77	73
<i>bud'to – (a)nebo</i>	69	79	80	79
<i>byť</i>	69	80	66	60

<i>coby</i>	69	80	79	68
<i>čili</i>	69	74	74	71
<i>jakoby</i>	69	66	69	78
<i>jakkoli(v)</i>	69	68	71	80
<i>jakožto</i>	69	76	64	75
<i>nýbrž</i>	69	71	60	57
<i>pak</i>	69	8	15	14
<i>přece</i>	69	57	27	77
<i>třebaže</i>	69	68	80	74
<i>tudíž</i>	69	67	53	64

Příloha 11: Výsledek Chi-kvadrát testu nezávislosti

Počet skupin **znaku 1**: 3

Počet skupin **znaku 2**: 2

Hladina významnosti α : 0.1%

Skutečné četnosti

	znak1 - 1. sk.	znak1 - 2. sk.	znak1 - 3. sk.	n_i
znak2 - 1. sk.	74168	98803	112865	285836
znak2 - 2. sk.	3348	3646	2455	9449
n_j	77516	102449	115320	295285

Očekávané četnosti

	znak1 - 1. sk.	znak1 - 2. sk.	znak1 - 3. sk.	n_i
znak2 - 1. sk.	75035.52	99170.67	111629.81	285836
znak2 - 2. sk.	2480.48	3278.33	3690.19	9449
n_j	77516	102449	115320	295285

G = 783.147

Kritická hodnota: $\chi_{(1-\alpha); df} = \mathbf{13.816}$

Rozhodnutí:

Na hladině významnosti 0.1 % nulovou hypotézu (H_0) o nezávislosti jednotlivých znaků zamítáme a přijímáme hypotézu H_1 , která nám říká, že zde určitá závislost existuje.

Příloha 12: Korespondenční analýza: spojovací prostředky vs. úrovně

XLSTAT 2017.02.42992 - Correspondence Analysis (CA)

Distance: Chi-square

Contingency table:

	A2	B1	B2
a	31620,553	29523,333	28116,073
že	10927,691	14555,827	15937,453
ale	7161,125	9203,987	7016,489
který	1534,527	4910,754	9321,906
když	1348,524	4763,725	6014,133
protože	5208,091	3705,120	2706,360
nebo	2929,551	3910,960	4460,482
co	976,517	3028,788	3157,420
aby	558,010	3058,194	2956,949
i	418,507	2205,428	4410,364
jak	325,506	1558,503	1403,298
jak (S)	325,506	1088,011	1353,180
jako	744,013	1470,286	2906,831
proto	1302,023	1264,446	2255,300
kde	1395,024	1235,040	801,884
kdy	1441,525	1058,606	451,060
než	186,003	1058,606	1854,358
ani	46,501	646,926	801,884
až	418,507	264,651	701,649
či	0,000	117,623	801,884
však	0,000	117,623	651,531
a tak	93,002	264,651	200,471
jenž	0,000	0,000	350,824
tedy	0,000	58,811	100,236
pak	0,000	0,000	50,118
co (S)	93,002	382,274	400,942
tak	511,509	2087,805	1553,651
jestli	930,016	1205,634	751,767
kdo	372,007	588,114	952,238
kdyby	837,015	617,520	300,707
jaký	186,003	588,114	952,238
a proto	465,008	411,680	902,120
kolik	697,512	352,869	250,589
nejen – ale (i)	46,501	235,246	801,884
i když	0,000	264,651	551,296
proč	46,501	382,274	250,589
dokonce	46,501	117,623	651,531
pokud	139,502	176,434	400,942
takže	232,504	235,246	200,471
což	0,000	117,623	451,060
-li	46,501	117,623	400,942
totiž	46,501	58,811	200,471
jak – tak	0,000	117,623	250,589
a ne	0,000	147,029	150,353
anebo	93,002	29,406	300,707
zda	139,502	147,029	50,118
kam	186,003	88,217	50,118
přestože	46,501	88,217	200,471
ať	0,000	88,217	200,471
avšak	0,000	58,811	250,589
čím – tím	0,000	58,811	250,589

nejenom – ale (i)	0,000	117,623	150,353
bud' – (a)nebo	0,000	58,811	200,471
přesto	0,000	88,217	150,353
sice	0,000	117,623	100,236
ačkoliv	0,000	58,811	100,236
jelikož	0,000	29,406	150,353
aniž	0,000	29,406	100,236
ba	0,000	29,406	100,236
i kdyby	0,000	0,000	150,353
jakmile	0,000	88,217	0,000
jestliže	0,000	58,811	50,118
kdykoli(v)	0,000	58,811	50,118
vždyť	0,000	0,000	150,353
zatímco	0,000	0,000	150,353
jednak	0,000	29,406	50,118
jehož	0,000	29,406	50,118
jenomže	0,000	0,000	100,236
nicméně	0,000	0,000	100,236
proto, že	0,000	0,000	100,236
ačkoli	0,000	0,000	50,118
a tedy	0,000	29,406	0,000
byť	0,000	0,000	50,118
coby	0,000	0,000	50,118
čili	0,000	29,406	0,000
jakoby	0,000	0,000	50,118
jakkoli	0,000	29,406	0,000
jakožto	0,000	0,000	50,118
nýbrž	0,000	0,000	50,118
přece	0,000	29,406	0,000
třebaže	0,000	0,000	50,118
tudíž	0,000	0,000	50,118

Total inertia: 0,115

Eigenvalues and percentages of inertia:

	F1	F2
Eigenvalue	0,097	0,018
Inertia (%)	84,364	15,636
Cumulative %	84,364	100,000

Results for the rows:

Weights, distances and squared distances to the origin, inertias and relative inertias (rows):

	Weight (relative)	Distance	Sq-Distance	Inertia	Relative inertia
a	0,312	0,227	0,052	0,01610	0,140
že	0,145	0,021	0,000	0,00006	0,001
ale	0,082	0,195	0,038	0,00310	0,027
který	0,055	0,450	0,202	0,01116	0,097
když	0,042	0,342	0,117	0,00496	0,043
protože	0,041	0,454	0,206	0,00838	0,073
nebo	0,040	0,001	0,000	0,00000	0,000
co	0,025	0,284	0,081	0,00203	0,018
aby	0,023	0,408	0,166	0,00383	0,033
i	0,025	0,542	0,293	0,00722	0,063
jak	0,012	0,387	0,149	0,00172	0,015
jak (S)	0,010	0,326	0,106	0,00103	0,009

jako	0,018	0,368	0,136	0,00243	0,021
proto	0,017	0,184	0,034	0,00057	0,005
kde	0,012	0,387	0,150	0,00180	0,016
kdy	0,010	0,593	0,351	0,00363	0,032
než	0,011	0,508	0,258	0,00280	0,024
ani	0,005	0,523	0,273	0,00143	0,012
až	0,005	0,328	0,108	0,00052	0,005
či	0,003	0,986	0,973	0,00313	0,027
však	0,003	0,941	0,885	0,00238	0,021
a tak	0,002	0,290	0,084	0,00016	0,001
jenž	0,001	1,238	1,532	0,00188	0,016
tedy	0,001	0,633	0,401	0,00022	0,002
pak	0,000	1,238	1,532	0,00027	0,002
co (S)	0,003	0,352	0,124	0,00038	0,003
tak	0,015	0,379	0,144	0,00209	0,018
jestli	0,010	0,276	0,076	0,00077	0,007
kdo	0,007	0,217	0,047	0,00032	0,003
kdyby	0,006	0,556	0,309	0,00190	0,017
jaký	0,006	0,388	0,151	0,00091	0,008
a proto	0,006	0,264	0,070	0,00043	0,004
kolik	0,005	0,644	0,415	0,00189	0,016
nejen – ale (i)	0,004	0,728	0,530	0,00201	0,018
i když	0,003	0,678	0,460	0,00131	0,011
proč	0,002	0,528	0,278	0,00066	0,006
dokonce	0,003	0,830	0,688	0,00196	0,017
pokud	0,003	0,337	0,113	0,00028	0,002
takže	0,002	0,231	0,053	0,00012	0,001
což	0,002	0,847	0,717	0,00143	0,012
-li	0,002	0,653	0,426	0,00084	0,007
totiž	0,001	0,533	0,285	0,00030	0,003
jak – tak	0,001	0,684	0,468	0,00060	0,005
a ne	0,001	0,595	0,354	0,00037	0,003
anebo	0,001	0,692	0,479	0,00071	0,006
zda	0,001	0,520	0,270	0,00032	0,003
kam	0,001	0,737	0,543	0,00062	0,005
přestože	0,001	0,425	0,180	0,00021	0,002
ať	0,001	0,701	0,491	0,00050	0,004
avšak	0,001	0,875	0,766	0,00083	0,007
čím – tím	0,001	0,875	0,766	0,00083	0,007
nejenom – ale (i)	0,001	0,595	0,354	0,00033	0,003
buď – (a)nebo	0,001	0,814	0,663	0,00060	0,005
přesto	0,001	0,633	0,401	0,00033	0,003
sice	0,001	0,616	0,379	0,00029	0,003
ačkoliv	0,001	0,633	0,401	0,00022	0,002
jelikož	0,001	0,921	0,849	0,00053	0,005
aniž	0,000	0,814	0,663	0,00030	0,003
ba	0,000	0,814	0,663	0,00030	0,003
i kdyby	0,001	1,238	1,532	0,00081	0,007
jakmile	0,000	1,376	1,893	0,00058	0,005
jestliže	0,000	0,616	0,379	0,00014	0,001
kdykoli(v)	0,000	0,616	0,379	0,00014	0,001
vždyť	0,001	1,238	1,532	0,00081	0,007
zatímco	0,001	1,238	1,532	0,00081	0,007
jednak	0,000	0,633	0,401	0,00011	0,001
jehož	0,000	0,633	0,401	0,00011	0,001
jenomže	0,000	1,238	1,532	0,00054	0,005
nicméně	0,000	1,238	1,532	0,00054	0,005

proto, že	0,000	1,238	1,532	0,00054	0,005
ačkoli	0,000	1,238	1,532	0,00027	0,002
a tedy	0,000	1,376	1,893	0,00019	0,002
byť	0,000	1,238	1,532	0,00027	0,002
coby	0,000	1,238	1,532	0,00027	0,002
čili	0,000	1,376	1,893	0,00019	0,002
jakoby	0,000	1,238	1,532	0,00027	0,002
jakkoli	0,000	1,376	1,893	0,00019	0,002
jakožto	0,000	1,238	1,532	0,00027	0,002
nýbrž	0,000	1,238	1,532	0,00027	0,002
přece	0,000	1,376	1,893	0,00019	0,002
třebaže	0,000	1,238	1,532	0,00027	0,002
tudíž	0,000	1,238	1,532	0,00027	0,002

Principal coordinates (rows):

	F1	F2
a	-0,222	0,046
že	-0,018	-0,011
ale	-0,173	-0,090
který	0,448	0,043
když	0,320	-0,120
protože	-0,446	0,085
nebo	0,000	-0,001
co	0,223	-0,177
aby	0,304	-0,272
i	0,541	0,033
jak	0,259	-0,287
jak (S)	0,303	-0,120
jako	0,354	0,101
proto	0,068	0,171
kde	-0,387	-0,005
kdy	-0,593	0,011
než	0,507	-0,024
ani	0,477	-0,214
až	0,068	0,321
či	0,902	0,400
však	0,873	0,350
a tak	0,088	-0,276
jenž	1,047	0,661
tedy	0,627	-0,091
pak	1,047	0,661
co (S)	0,284	-0,209
tak	0,165	-0,342
jestli	-0,240	-0,136
kdo	0,207	0,067
kdyby	-0,556	0,023
jaký	0,388	-0,015
a proto	0,125	0,233
kolik	-0,613	0,197
nejen – ale (i)	0,692	0,226
i když	0,678	0,001
proč	0,235	-0,472
dokonce	0,739	0,377
pokud	0,276	0,192
takže	-0,231	0,002
což	0,812	0,240
-li	0,603	0,251

totiž	0,445	0,294
jak – tak	0,684	0,011
a ne	0,485	-0,345
anebo	0,413	0,555
zda	-0,494	-0,160
kam	-0,709	0,201
přestože	0,398	0,148
ať	0,700	0,039
avšak	0,831	0,274
čím – tím	0,831	0,274
nejenom – ale (i)	0,548	-0,232
bud' – (a)nebo	0,789	0,199
přesto	0,627	-0,091
sice	0,434	-0,437
ačkoliv	0,627	-0,091
jelikož	0,861	0,328
aniž	0,789	0,199
ba	0,789	0,199
i kdyby	1,047	0,661
jakmile	-0,089	-1,373
jestliže	0,434	-0,437
kdykoli(v)	0,434	-0,437
vždyť	1,047	0,661
zatímco	1,047	0,661
jednak	0,627	-0,091
jehož	0,627	-0,091
jenomže	1,047	0,661
nicméně	1,047	0,661
proto, že	1,047	0,661
ačkoli	1,047	0,661
a tedy	-0,089	-1,373
byť	1,047	0,661
coby	1,047	0,661
čili	-0,089	-1,373
jakoby	1,047	0,661
jakkoli	-0,089	-1,373
jakožto	1,047	0,661
nýbrž	1,047	0,661
přece	-0,089	-1,373
třebaže	1,047	0,661
tudíž	1,047	0,661

Standard coordinates (rows):

	F1	F2
a	-0,715	0,343
že	-0,057	-0,081
ale	-0,555	-0,670
který	1,438	0,322
když	1,029	-0,896
protože	-1,433	0,638
nebo	0,000	-0,006
co	0,716	-1,321
aby	0,978	-2,027
i	1,737	0,245
jak	0,831	-2,144
jak (S)	0,975	-0,895
jako	1,139	0,750

proto	0,218	1,280
kde	-1,244	-0,035
kdy	-1,904	0,083
než	1,631	-0,181
ani	1,533	-1,598
až	0,217	2,399
či	2,897	2,989
však	2,806	2,610
a tak	0,283	-2,063
jenž	3,364	4,931
tedy	2,014	-0,682
pak	3,364	4,931
co (S)	0,911	-1,562
tak	0,531	-2,549
jestli	-0,771	-1,014
kdo	0,665	0,500
kdyby	-1,785	0,173
jaký	1,247	-0,108
a proto	0,400	1,737
kolik	-1,972	1,468
nejen – ale (i)	2,224	1,688
i když	2,180	0,008
proč	0,755	-3,527
dokonce	2,376	2,812
pokud	0,888	1,433
takže	-0,741	0,012
což	2,609	1,792
-li	1,937	1,872
totiž	1,429	2,197
jak – tak	2,198	0,082
a ne	1,559	-2,574
anebo	1,329	4,144
zda	-1,589	-1,192
kam	-2,277	1,503
přestože	1,279	1,106
ať	2,249	0,293
avšak	2,670	2,046
čím – tím	2,670	2,046
nejenom – ale (i)	1,762	-1,731
bud' – (a)nebo	2,536	1,488
přesto	2,014	-0,682
sice	1,393	-3,264
ačkoliv	2,014	-0,682
jelikož	2,767	2,448
aniž	2,536	1,488
ba	2,536	1,488
i kdyby	3,364	4,931
jakmile	-0,286	-10,248
jestliže	1,393	-3,264
kdykoli(v)	1,393	-3,264
vždyť	3,364	4,931
zatímco	3,364	4,931
jednak	2,014	-0,682
jehož	2,014	-0,682
jenomže	3,364	4,931
nicméně	3,364	4,931
proto, že	3,364	4,931

ačkoli	3,364	4,931
a tedy	-0,286	-10,248
byť	3,364	4,931
coby	3,364	4,931
čili	-0,286	-10,248
jakoby	3,364	4,931
jakkoli	-0,286	-10,248
jakožto	3,364	4,931
nýbrž	3,364	4,931
přece	-0,286	-10,248
třebaže	3,364	4,931
tudíž	3,364	4,931

Contributions (rows):

	Weight (relative)	F1	F2
a	0,312	0,159	0,037
že	0,145	0,000	0,001
ale	0,082	0,025	0,037
který	0,055	0,114	0,006
když	0,042	0,045	0,034
protože	0,041	0,083	0,017
nebo	0,040	0,000	0,000
co	0,025	0,013	0,044
aby	0,023	0,022	0,095
i	0,025	0,074	0,001
jak	0,012	0,008	0,053
jak (S)	0,010	0,009	0,008
jako	0,018	0,023	0,010
proto	0,017	0,001	0,028
kde	0,012	0,019	0,000
kdy	0,010	0,037	0,000
než	0,011	0,029	0,000
ani	0,005	0,012	0,013
až	0,005	0,000	0,028
či	0,003	0,027	0,029
však	0,003	0,021	0,018
a tak	0,002	0,000	0,008
jenž	0,001	0,014	0,030
tedy	0,001	0,002	0,000
pak	0,000	0,002	0,004
co (S)	0,003	0,003	0,007
tak	0,015	0,004	0,094
jestli	0,010	0,006	0,010
kdo	0,007	0,003	0,002
kdyby	0,006	0,020	0,000
jaký	0,006	0,009	0,000
a proto	0,006	0,001	0,019
kolik	0,005	0,018	0,010
nejen – ale (i)	0,004	0,019	0,011
i když	0,003	0,014	0,000
proč	0,002	0,001	0,030
dokonce	0,003	0,016	0,023
pokud	0,003	0,002	0,005
takže	0,002	0,001	0,000
což	0,002	0,014	0,006
-li	0,002	0,007	0,007
totiž	0,001	0,002	0,005

jak – tak	0,001	0,006	0,000
a ne	0,001	0,003	0,007
anebo	0,001	0,003	0,025
zda	0,001	0,003	0,002
kam	0,001	0,006	0,003
přestože	0,001	0,002	0,001
ať	0,001	0,005	0,000
avšak	0,001	0,008	0,005
čím – tím	0,001	0,008	0,005
nejenom – ale (i)	0,001	0,003	0,003
bud' – (a)nebo	0,001	0,006	0,002
přesto	0,001	0,003	0,000
sice	0,001	0,001	0,008
ačkoliv	0,001	0,002	0,000
jelikož	0,001	0,005	0,004
aniž	0,000	0,003	0,001
ba	0,000	0,003	0,001
i kdyby	0,001	0,006	0,013
jakmile	0,000	0,000	0,032
jestliže	0,000	0,001	0,004
kdykoli(v)	0,000	0,001	0,004
vždyť	0,001	0,006	0,013
zatímco	0,001	0,006	0,013
jednak	0,000	0,001	0,000
jehož	0,000	0,001	0,000
jenomže	0,000	0,004	0,009
nicméně	0,000	0,004	0,009
proto, že	0,000	0,004	0,009
ačkoli	0,000	0,002	0,004
a tedy	0,000	0,000	0,011
byť	0,000	0,002	0,004
coby	0,000	0,002	0,004
čili	0,000	0,000	0,011
jakoby	0,000	0,002	0,004
jakkoli	0,000	0,000	0,011
jakožto	0,000	0,002	0,004
nýbrž	0,000	0,002	0,004
přece	0,000	0,000	0,011
třebaže	0,000	0,002	0,004
tudíž	0,000	0,002	0,004

Squared cosines (rows):

	F1	F2
a	0,959	0,041
že	0,727	0,273
ale	0,787	0,213
který	0,991	0,009
když	0,877	0,123
protože	0,965	0,035
nebo	0,010	0,990
co	0,613	0,387
aby	0,557	0,443
i	0,996	0,004
jak	0,448	0,552
jak (S)	0,865	0,135
jako	0,925	0,075
proto	0,136	0,864

kde	1,000	0,000
kdy	1,000	0,000
než	0,998	0,002
ani	0,832	0,168
až	0,042	0,958
či	0,835	0,165
však	0,862	0,138
a tak	0,092	0,908
jenž	0,715	0,285
tedy	0,979	0,021
pak	0,715	0,285
co (S)	0,648	0,352
tak	0,190	0,810
jestli	0,757	0,243
kdo	0,905	0,095
kdyby	0,998	0,002
jaký	0,999	0,001
a proto	0,223	0,777
kolik	0,907	0,093
nejen – ale (i)	0,904	0,096
i když	1,000	0,000
proč	0,198	0,802
dokonce	0,794	0,206
pokud	0,675	0,325
takže	1,000	0,000
což	0,920	0,080
-li	0,852	0,148
totiž	0,695	0,305
jak – tak	1,000	0,000
a ne	0,665	0,335
anebo	0,357	0,643
zda	0,905	0,095
kam	0,925	0,075
přestože	0,878	0,122
ať	0,997	0,003
avšak	0,902	0,098
čím – tím	0,902	0,098
nejenom – ale (i)	0,848	0,152
bud' – (a)nebo	0,940	0,060
přesto	0,979	0,021
sice	0,496	0,504
ačkoliv	0,979	0,021
jelikož	0,873	0,127
aniž	0,940	0,060
ba	0,940	0,060
i kdyby	0,715	0,285
jakmile	0,004	0,996
jestliže	0,496	0,504
kdykoli(v)	0,496	0,504
vždyť	0,715	0,285
zatímco	0,715	0,285
jednak	0,979	0,021
jehož	0,979	0,021
jenomže	0,715	0,285
nicméně	0,715	0,285
proto, že	0,715	0,285
ačkoli	0,715	0,285

a tedy	0,004	0,996
byť	0,715	0,285
coby	0,715	0,285
čili	0,004	0,996
jakoby	0,715	0,285
jakkoli	0,004	0,996
jakožto	0,715	0,285
nýbrž	0,715	0,285
přece	0,004	0,996
třebaže	0,715	0,285
tudíž	0,715	0,285

Results for the columns:

Weights, distances and squared distances to the origin, inertias and relative inertias (columns):

	Weight (relative)	Distance	Sq-Distance	Inertia	Relative inertia
A2	0,259	0,472	0,223	0,058	0,504
B1	0,346	0,186	0,035	0,012	0,104
B2	0,395	0,338	0,114	0,045	0,392

Principal coordinates (columns):

	F1	F2
A2	-0,459	0,110
B1	-0,028	-0,184
B2	0,326	0,088

Standard coordinates (columns):

	F1	F2
A2	-1,475	0,824
B1	-0,089	-1,373
B2	1,047	0,661

Contributions (columns):

	Weight (relative)	F1	F2
A2	0,259	0,565	0,176
B1	0,346	0,003	0,652
B2	0,395	0,433	0,172

Squared cosines (columns):

	F1	F2
A2	0,945	0,055
B1	0,022	0,978
B2	0,931	0,069

Příloha 13: Korespondenční analýza: spojovací prostředky vs. hodnocení koheze a koherence

XLSTAT 2017.02.42992 - Correspondence Analysis (CA)

Distance: Chi-square

Contingency table:

	A2-KOKO	B1-KOKO	B2-KOKO
a	38792,656	26853,841	29613,060
aby	207,447	1296,392	3621,741
ačkoli	0,000	0,000	26,630
ačkoliv	0,000	37,040	79,891
ale	4978,737	8741,388	8495,113
a ne	0,000	74,080	133,152
anebo	103,724	37,040	186,413
ani	0,000	37,040	1011,957
aniž	0,000	0,000	79,891
a proto	933,513	185,199	719,022
a tedy	0,000	0,000	26,630
a tak	0,000	222,239	213,044
ať	0,000	37,040	133,152
avšak	0,000	0,000	133,152
až	103,724	481,517	426,087
ba	0,000	0,000	79,891
buď – (a)nebo	103,724	0,000	159,783
byť	0,000	0,000	26,630
co	1140,960	1444,551	4180,981
coby	0,000	0,000	26,630
což	0,000	37,040	319,565
či	0,000	0,000	479,348
čili	0,000	0,000	26,630
čím – tím	0,000	37,040	106,522
dokonce	0,000	74,080	372,826
i	622,342	592,636	3515,219
i kdyby	0,000	0,000	79,891
i když	0,000	74,080	399,457
jak	726,066	1000,074	3195,654
jakmile	0,000	37,040	53,261
jako	1037,237	814,875	2236,958
jakoby	0,000	0,000	26,630
jakkoli	0,000	37,040	0,000
jakožto	0,000	0,000	26,630
jaký	103,724	370,398	772,283
jak – tak	0,000	0,000	213,044
jednak	0,000	0,000	53,261
jehož	0,000	0,000	53,261
jelikož	0,000	37,040	53,261

jenž	0,000	0,000	133,152
jestli	414,895	1259,353	1011,957
jestliže	0,000	74,080	26,630
kam	0,000	185,199	79,891
kde	1452,132	1518,631	852,174
kdo	518,618	148,159	1011,957
kdy	1348,408	1222,313	772,283
kdyby	311,171	1000,074	372,826
kdykoliv	0,000	0,000	79,891
když	1140,960	2370,546	6071,742
kolik	518,618	703,756	213,044
který	1763,303	1814,949	8308,700
-li	103,724	74,080	213,044
nebo	2696,816	3259,501	4367,394
nejen – ale (i)	0,000	37,040	585,870
nejenom – ale (i)	0,000	0,000	159,783
než	311,171	333,358	1651,088
nicméně	0,000	0,000	53,261
nýbrž	0,000	0,000	26,630
pak	0,000	0,000	26,630
pokud	311,171	111,119	292,935
proč	103,724	37,040	452,718
proto	1555,855	851,915	2023,914
protože	4667,566	5000,370	2769,567
proto, že	0,000	0,000	53,261
přece	0,000	0,000	26,630
přesto	0,000	37,040	133,152
přestože	0,000	37,040	186,413
sice	0,000	37,040	133,152
tak	311,171	1074,154	2077,175
takže	311,171	111,119	266,304
tedy	0,000	0,000	106,522
totiž	0,000	111,119	106,522
třebaže	0,000	0,000	26,630
tudíž	0,000	0,000	26,630
však	0,000	74,080	372,826
vždyť	0,000	0,000	53,261
zatímco	0,000	0,000	53,261
zda	207,447	148,159	79,891
že	9231,408	12074,969	16457,618

Total inertia: 0,144

Eigenvalues and percentages of inertia:

	F1	F2
Eigenvalue	0,116	0,028
Inertia (%)	80,309	19,691
Cumulative %	80,309	100,000

Results for the rows:

Weights, distances and squared distances to the origin, inertias and relative inertias (rows):

	Weight (relative)	Distance	Sq-Distance	Inertia	Relative inertia
a	0,359	0,286	0,082	0,02936	0,204
aby	0,019	0,633	0,401	0,00773	0,054
ačkoli	0,000	1,161	1,348	0,00014	0,001
ačkoliv	0,000	0,667	0,445	0,00020	0,001
ale	0,084	0,239	0,057	0,00480	0,033
a ne	0,001	0,643	0,414	0,00032	0,002
anebo	0,001	0,397	0,157	0,00019	0,001
ani	0,004	1,091	1,189	0,00470	0,033
aniž	0,000	1,161	1,348	0,00041	0,003
a proto	0,007	0,543	0,295	0,00204	0,014
a tedy	0,000	1,161	1,348	0,00014	0,001
a tak	0,002	0,685	0,470	0,00077	0,005
ať	0,001	0,776	0,602	0,00039	0,003
avšak	0,001	1,161	1,348	0,00068	0,005
až	0,004	0,492	0,242	0,00092	0,006
ba	0,000	1,161	1,348	0,00041	0,003
bud' – (a)nebo	0,001	0,635	0,404	0,00040	0,003
byť	0,000	1,161	1,348	0,00014	0,001
co	0,025	0,393	0,154	0,00393	0,027
coby	0,000	1,161	1,348	0,00014	0,001
což	0,001	0,961	0,923	0,00124	0,009
či	0,002	1,161	1,348	0,00243	0,017
čili	0,000	1,161	1,348	0,00014	0,001
čím – tím	0,001	0,724	0,524	0,00028	0,002
dokonce	0,002	0,854	0,730	0,00123	0,009
i	0,018	0,642	0,412	0,00733	0,051
i kdyby	0,000	1,161	1,348	0,00041	0,003
i když	0,002	0,869	0,756	0,00135	0,009
jak	0,019	0,458	0,209	0,00388	0,027
jakmile	0,000	0,634	0,402	0,00014	0,001
jako	0,015	0,255	0,065	0,00100	0,007
jakoby	0,000	1,161	1,348	0,00014	0,001
jakkoli	0,000	1,575	2,480	0,00035	0,002

jakožto	0,000	1,161	1,348	0,00014	0,001
jaký	0,005	0,483	0,233	0,00109	0,008
jak – tak	0,001	1,161	1,348	0,00108	0,008
jednak	0,000	1,161	1,348	0,00027	0,002
jehož	0,000	1,161	1,348	0,00027	0,002
jelikož	0,000	0,634	0,402	0,00014	0,001
jenž	0,001	1,161	1,348	0,00068	0,005
jestli	0,010	0,426	0,181	0,00183	0,013
jestliže	0,000	1,023	1,047	0,00040	0,003
kam	0,001	0,955	0,912	0,00091	0,006
kde	0,014	0,411	0,169	0,00243	0,017
kdo	0,006	0,462	0,213	0,00135	0,009
kdy	0,013	0,397	0,158	0,00199	0,014
kdyby	0,006	0,679	0,461	0,00293	0,020
kdykoliv	0,000	1,161	1,348	0,00041	0,003
když	0,036	0,453	0,205	0,00739	0,051
kolik	0,005	0,586	0,344	0,00186	0,013
který	0,045	0,552	0,305	0,01365	0,095
-li	0,001	0,261	0,068	0,00010	0,001
nebo	0,039	0,071	0,005	0,00020	0,001
nejen – ale (i)	0,002	1,044	1,089	0,00256	0,018
nejenom – ale (i)	0,001	1,161	1,348	0,00081	0,006
než	0,009	0,593	0,352	0,00304	0,021
nicméně	0,000	1,161	1,348	0,00027	0,002
nýbrž	0,000	1,161	1,348	0,00014	0,001
pak	0,000	1,161	1,348	0,00014	0,001
pokud	0,003	0,372	0,138	0,00037	0,003
proč	0,002	0,697	0,486	0,00109	0,008
proto	0,017	0,219	0,048	0,00080	0,006
protože	0,047	0,412	0,170	0,00797	0,055
proto, že	0,000	1,161	1,348	0,00027	0,002
přece	0,000	1,161	1,348	0,00014	0,001
přesto	0,001	0,776	0,602	0,00039	0,003
přestože	0,001	0,854	0,730	0,00061	0,004
sice	0,001	0,776	0,602	0,00039	0,003
tak	0,013	0,456	0,208	0,00271	0,019
takže	0,003	0,392	0,154	0,00040	0,003
tedy	0,000	1,161	1,348	0,00054	0,004
totiž	0,001	0,685	0,470	0,00038	0,003
třebaže	0,000	1,161	1,348	0,00014	0,001
tudíž	0,000	1,161	1,348	0,00014	0,001
však	0,002	0,854	0,730	0,00123	0,009
vždyť	0,000	1,161	1,348	0,00027	0,002
zatímco	0,000	1,161	1,348	0,00027	0,002
zda	0,002	0,523	0,273	0,00045	0,003
že	0,142	0,101	0,010	0,00144	0,010

Principal coordinates (rows):

	F1	F2
a	-0,265	0,109
aby	0,617	-0,143
ačkoli	1,130	0,266
ačkoliv	0,602	-0,287
ale	-0,035	-0,237
a ne	0,534	-0,358
anebo	0,221	0,329
ani	1,071	0,204
aniž	1,130	0,266
a proto	-0,191	0,508
a tedy	1,130	0,266
a tak	0,279	-0,626
ať	0,767	-0,114
avšak	1,130	0,266
až	0,104	-0,481
ba	1,130	0,266
bud' – (a)nebo	0,236	0,590
byť	1,130	0,266
co	0,391	0,032
coby	1,130	0,266
což	0,957	0,084
či	1,130	0,266
čili	1,130	0,266
čím – tím	0,700	-0,185
dokonce	0,854	-0,024
i	0,622	0,155
i kdyby	1,130	0,266
i když	0,869	-0,007
jak	0,456	0,032
jakmile	0,446	-0,451
jako	0,222	0,127
jakoby	1,130	0,266
jakkoli	-0,537	-1,481
jakožto	1,130	0,266
jaký	0,446	-0,185
jak – tak	1,130	0,266
jednak	1,130	0,266
jehož	1,130	0,266
jelikož	0,446	-0,451
jenž	1,130	0,266
jestli	-0,002	-0,426
jestliže	-0,096	-1,019

kam	-0,034	-0,954
kde	-0,395	-0,115
kdo	0,281	0,366
kdy	-0,395	-0,041
kdyby	-0,279	-0,619
kdykoliv	1,130	0,266
když	0,447	-0,068
kolik	-0,508	-0,293
který	0,539	0,121
-li	0,211	0,153
nebo	0,011	-0,071
nejen – ale (i)	1,031	0,162
nejenom – ale (i)	1,130	0,266
než	0,580	0,124
nicméně	1,130	0,266
nýbrž	1,130	0,266
pak	1,130	0,266
pokud	-0,117	0,353
proč	0,629	0,301
proto	0,012	0,219
protože	-0,392	-0,127
proto, že	1,130	0,266
přece	1,130	0,266
přesto	0,767	-0,114
přestože	0,854	-0,024
sice	0,767	-0,114
tak	0,409	-0,202
takže	-0,165	0,356
tedy	1,130	0,266
totiž	0,279	-0,626
třebaže	1,130	0,266
tudíž	1,130	0,266
však	0,854	-0,024
vždyť	1,130	0,266
zatímco	1,130	0,266
zda	-0,519	0,064
že	0,042	-0,091

Standard coordinates (rows):

	F1	F2
a	-0,779	0,646
aby	1,814	-0,847
ačkoli	3,325	1,579
ačkoliv	1,771	-1,708
ale	-0,102	-1,407

a ne	1,572	-2,130
anebo	0,651	1,955
ani	3,152	1,213
aniž	3,325	1,579
a proto	-0,563	3,018
a tedy	3,325	1,579
a tak	0,821	-3,718
ať	2,258	-0,679
avšak	3,325	1,579
až	0,305	-2,859
ba	3,325	1,579
bud' – (a)nebo	0,695	3,504
byť	3,325	1,579
co	1,151	0,189
coby	3,325	1,579
což	2,815	0,501
či	3,325	1,579
čili	3,325	1,579
čím – tím	2,060	-1,098
dokonce	2,512	-0,141
i	1,831	0,923
i kdyby	3,325	1,579
i když	2,558	-0,044
jak	1,343	0,192
jakmile	1,313	-2,677
jako	0,653	0,752
jakoby	3,325	1,579
jakkoli	-1,579	-8,797
jakožto	3,325	1,579
jaký	1,312	-1,097
jak – tak	3,325	1,579
jednak	3,325	1,579
jehož	3,325	1,579
jelikož	1,313	-2,677
jenž	3,325	1,579
jestli	-0,006	-2,530
jestliže	-0,282	-6,053
kam	-0,101	-5,670
kde	-1,161	-0,685
kdo	0,828	2,174
kdy	-1,163	-0,242
kdyby	-0,822	-3,679
kdykoliv	3,325	1,579
když	1,316	-0,405
kolik	-1,493	-1,741
který	1,585	0,720

-li	0,622	0,911
nebo	0,031	-0,419
nejen – ale (i)	3,033	0,962
nejenom – ale (i)	3,325	1,579
než	1,707	0,735
nicméně	3,325	1,579
nýbrž	3,325	1,579
pak	3,325	1,579
pokud	-0,344	2,095
proč	1,851	1,786
proto	0,036	1,302
protože	-1,154	-0,757
proto, že	3,325	1,579
přece	3,325	1,579
přesto	2,258	-0,679
přestože	2,512	-0,141
sice	2,258	-0,679
tak	1,203	-1,200
takže	-0,486	2,115
tedy	3,325	1,579
totiž	0,821	-3,718
třebaže	3,325	1,579
tudíž	3,325	1,579
však	2,512	-0,141
vždyť	3,325	1,579
zatímco	3,325	1,579
zda	-1,526	0,379
že	0,124	-0,543

Contributions (rows):

	Weight (relative)	F1	F2
a	0,359	0,217	0,150
aby	0,019	0,064	0,014
ačkoli	0,000	0,001	0,000
ačkoliv	0,000	0,001	0,001
ale	0,084	0,001	0,166
a ne	0,001	0,002	0,004
anebo	0,001	0,001	0,005
ani	0,004	0,039	0,006
aniž	0,000	0,003	0,001
a proto	0,007	0,002	0,063
a tedy	0,000	0,001	0,000
a tak	0,002	0,001	0,023
ať	0,001	0,003	0,000
avšak	0,001	0,006	0,001

až	0,004	0,000	0,031
ba	0,000	0,003	0,001
bud' – (a)nebo	0,001	0,000	0,012
byť	0,000	0,001	0,000
co	0,025	0,034	0,001
coby	0,000	0,001	0,000
což	0,001	0,011	0,000
či	0,002	0,020	0,005
čili	0,000	0,001	0,000
čím – tím	0,001	0,002	0,001
dokonce	0,002	0,011	0,000
i	0,018	0,060	0,015
i kdyby	0,000	0,003	0,001
i když	0,002	0,012	0,000
jak	0,019	0,033	0,001
jakmile	0,000	0,001	0,002
jako	0,015	0,007	0,009
jakoby	0,000	0,001	0,000
jakkoli	0,000	0,000	0,011
jakožto	0,000	0,001	0,000
jaký	0,005	0,008	0,006
jak – tak	0,001	0,009	0,002
jednak	0,000	0,002	0,001
jehož	0,000	0,002	0,001
jelikož	0,000	0,001	0,002
jenž	0,001	0,006	0,001
jestli	0,010	0,000	0,065
jestliže	0,000	0,000	0,014
kam	0,001	0,000	0,032
kde	0,014	0,019	0,007
kdo	0,006	0,004	0,030
kdy	0,013	0,017	0,001
kdyby	0,006	0,004	0,086
kdykoliv	0,000	0,003	0,001
když	0,036	0,063	0,006
kolik	0,005	0,012	0,016
který	0,045	0,112	0,023
-li	0,001	0,001	0,001
nebo	0,039	0,000	0,007
nejen – ale (i)	0,002	0,022	0,002
nejenom – ale (i)	0,001	0,007	0,002
než	0,009	0,025	0,005
nicméně	0,000	0,002	0,001
nýbrž	0,000	0,001	0,000
pak	0,000	0,001	0,000
pokud	0,003	0,000	0,012

proč	0,002	0,008	0,007
proto	0,017	0,000	0,028
protože	0,047	0,062	0,027
proto, že	0,000	0,002	0,001
přece	0,000	0,001	0,000
přesto	0,001	0,003	0,000
přestože	0,001	0,005	0,000
sice	0,001	0,003	0,000
tak	0,013	0,019	0,019
takže	0,003	0,001	0,012
tedy	0,000	0,004	0,001
totiž	0,001	0,001	0,011
třebaže	0,000	0,001	0,000
tudíž	0,000	0,001	0,000
však	0,002	0,011	0,000
vždyť	0,000	0,002	0,001
zatímco	0,000	0,002	0,001
zda	0,002	0,004	0,000
že	0,142	0,002	0,042

Squared cosines (rows):

	F1	F2
a	0,856	0,144
aby	0,949	0,051
ačkoli	0,948	0,052
ačkoliv	0,814	0,186
ale	0,021	0,979
a ne	0,690	0,310
anebo	0,312	0,688
ani	0,965	0,035
aniž	0,948	0,052
a proto	0,124	0,876
a tedy	0,948	0,052
a tak	0,166	0,834
ať	0,978	0,022
avšak	0,948	0,052
až	0,044	0,956
ba	0,948	0,052
bud' – (a)nebo	0,138	0,862
byť	0,948	0,052
co	0,993	0,007
coby	0,948	0,052
což	0,992	0,008
či	0,948	0,052
čili	0,948	0,052

čím – tím	0,935	0,065
dokonce	0,999	0,001
i	0,941	0,059
i kdyby	0,948	0,052
i když	1,000	0,000
jak	0,995	0,005
jakmile	0,495	0,505
jako	0,754	0,246
jakoby	0,948	0,052
jakkoli	0,116	0,884
jakožto	0,948	0,052
jaký	0,854	0,146
jak – tak	0,948	0,052
jednak	0,948	0,052
jehož	0,948	0,052
jelikož	0,495	0,505
jenž	0,948	0,052
jestli	0,000	1,000
jestliže	0,009	0,991
kam	0,001	0,999
kde	0,921	0,079
kdo	0,372	0,628
kdy	0,990	0,010
kdyby	0,169	0,831
kdykoliv	0,948	0,052
když	0,977	0,023
kolik	0,750	0,250
který	0,952	0,048
-li	0,656	0,344
nebo	0,022	0,978
nejen – ale (i)	0,976	0,024
nejenom – ale (i)	0,948	0,052
než	0,956	0,044
nicméně	0,948	0,052
nýbrž	0,948	0,052
pak	0,948	0,052
pokud	0,099	0,901
proč	0,814	0,186
proto	0,003	0,997
protože	0,905	0,095
proto, že	0,948	0,052
přece	0,948	0,052
přesto	0,978	0,022
přestože	0,999	0,001
sice	0,978	0,022
tak	0,804	0,196

takže	0,177	0,823
tedy	0,948	0,052
totiž	0,166	0,834
třebaže	0,948	0,052
tudíž	0,948	0,052
však	0,999	0,001
vždyť	0,948	0,052
zatímco	0,948	0,052
zda	0,985	0,015
že	0,174	0,826

Results for the columns:

Weights, distances and squared distances to the origin, inertias and relative inertias (columns):

	Weight (relative)	Distance	Sq-Distance	Inertia	Relative inertia
A2-KOKO	0,287	0,429	0,184	0,053	0,367
B1-KOKO	0,287	0,309	0,095	0,027	0,190
B2-KOKO	0,426	0,387	0,150	0,064	0,443

Principal coordinates (columns):

	F1	F2
A2-KOKO	-0,388	0,183
B1-KOKO	-0,182	-0,249
B2-KOKO	0,384	0,045

Standard coordinates (columns):

	F1	F2
A2-KOKO	-1,141	1,089
B1-KOKO	-0,537	-1,481
B2-KOKO	1,130	0,266

Contributions (columns):

	Weight (relative)	F1	F2
A2-KOKO	0,287	0,373	0,340
B1-KOKO	0,287	0,083	0,630
B2-KOKO	0,426	0,544	0,030

Squared cosines (columns):

	F1	F2
A2-KOKO	0,817	0,183
B1-KOKO	0,349	0,651
B2-KOKO	0,987	0,013

Příloha 14: MERLIN: Počet chyb – záměn

	A1	A2	B1	B2	C1	celkem
<i>když</i>	0	21	14	5	0	40
<i>kdy</i>	0	10	11	4	0	25
<i>co</i>	0	6	9	6	0	21
<i>i</i>	0	4	11	4	0	19
<i>jak</i>	0	3	10	5	0	18
<i>kteřý</i>	0	2	9	7	0	18
<i>jestli</i>	0	0	12	0	0	12
<i>a</i>	0	0	8	3	0	11
<i>že</i>	0	5	4	0	0	9
<i>až</i>	0	3	3	1	0	7
<i>jaký</i>	0	2	5	0	0	7
<i>však</i>	0	0	5	2	0	7
<i>jako</i>	0	2	3	1	0	6
<i>protože</i>	0	5	0	0	0	5
<i>ale</i>	0	0	4	0	0	4
<i>i když</i>	0	0	2	2	0	4
<i>kde</i>	0	2	2	0	0	4
<i>kdyby</i>	0	1	2	0	0	3
<i>pokud</i>	0	3	0	0	0	3
<i>aby</i>	0	0	1	1	0	2
<i>jenž</i>	0	0	2	0	0	2
<i>než</i>	0	0	2	0	0	2
<i>proto</i>	0	1	0	1	0	2
<i>ať</i>	0	0	1	0	0	1
<i>avšak</i>	0	0	0	1	0	1
<i>bud' – (a)nebo</i>	0	0	0	1	0	1
<i>coby</i>	0	0	1	0	0	1
<i>což</i>	0	0	0	1	0	1
<i>i kdyby</i>	0	0	0	1	0	1
<i>jakmile</i>	0	1	0	0	0	1
<i>jestliže</i>	0	0	0	1	0	1
<i>kdo</i>	0	0	0	1	0	1
<i>nýbrž</i>	0	0	0	1	0	1
<i>přestože</i>	0	1	0	0	0	1
<i>tak</i>	0	0	1	0	0	1
<i>ačkoli</i>	0	0	0	0	0	0
<i>ačkoliv</i>	0	0	0	0	0	0
<i>a ne</i>	0	0	0	0	0	0
<i>anebo</i>	0	0	0	0	0	0
<i>ani</i>	0	0	0	0	0	0
<i>aniž</i>	0	0	0	0	0	0
<i>a proto</i>	0	0	0	0	0	0
<i>a tedy</i>	0	0	0	0	0	0
<i>a tak</i>	0	0	0	0	0	0
<i>ba</i>	0	0	0	0	0	0
<i>bud'to – (a)nebo</i>	0	0	0	0	0	0
<i>byť</i>	0	0	0	0	0	0
<i>či</i>	0	0	0	0	0	0
<i>čili</i>	0	0	0	0	0	0
<i>čím – tím</i>	0	0	0	0	0	0
<i>dokonce</i>	0	0	0	0	0	0
<i>jakoby</i>	0	0	0	0	0	0
<i>jakkoli</i>	0	0	0	0	0	0
<i>jakožto</i>	0	0	0	0	0	0
<i>jak – tak</i>	0	0	0	0	0	0

<i>jednak</i>	0	0	0	0	0	0
<i>jehož</i>	0	0	0	0	0	0
<i>jelikož</i>	0	0	0	0	0	0
<i>jenomže</i>	0	0	0	0	0	0
<i>kam</i>	0	0	0	0	0	0
<i>kdykoli(v)</i>	0	0	0	0	0	0
<i>kolik</i>	0	0	0	0	0	0
<i>-li</i>	0	0	0	0	0	0
<i>nebo</i>	0	0	0	0	0	0
<i>nejen – ale (i)</i>	0	0	0	0	0	0
<i>nejenom – ale (i)</i>	0	0	0	0	0	0
<i>nicméně</i>	0	0	0	0	0	0
<i>pak</i>	0	0	0	0	0	0
<i>proč</i>	0	0	0	0	0	0
<i>proto, že</i>	0	0	0	0	0	0
<i>přece</i>	0	0	0	0	0	0
<i>přesto</i>	0	0	0	0	0	0
<i>sice</i>	0	0	0	0	0	0
<i>takže</i>	0	0	0	0	0	0
<i>tedy</i>	0	0	0	0	0	0
<i>totiž</i>	0	0	0	0	0	0
<i>třebaže</i>	0	0	0	0	0	0
<i>tudíž</i>	0	0	0	0	0	0
<i>vždyť</i>	0	0	0	0	0	0
<i>zatímco</i>	0	0	0	0	0	0
<i>zda</i>	0	0	0	0	0	0
CELKEM	0	72	124	49	0	243

Příloha 15: MERLIN: Relativní frekvence chyb – záměn

	A1	A2	B1	B2	C1	celkem
<i>coby</i>				0,00		100,00
<i>nýbrž</i>				100,00		100,00
<i>však</i>			125,00	15,38		41,18
<i>i kdyby</i>				33,33		33,33
<i>jakmile</i>			0,00			33,33
<i>jestliže</i>			0,00	100,00		33,33
<i>kdy</i>		32,26	30,56	44,44	0,00	31,65
<i>jenž</i>				0,00		28,57
<i>až</i>		33,33	33,33	7,14	0,00	21,21
<i>i když</i>			22,22	18,18		20,00
<i>pokud</i>		100,00	0,00	0,00		17,65
<i>bud' – (a)nebo</i>			0,00	25,00		16,67
<i>jaký</i>		50,00	25,00	0,00	0,00	15,91
<i>jestli</i>		0,00	29,27	0,00		15,79
<i>ať</i>			33,33	0,00		14,29
<i>avšak</i>			0,00	20,00		14,29
<i>když</i>		72,41	8,64	4,17	0,00	12,70
<i>přestože</i>		100,00	0,00	0,00		12,50
<i>jak</i>		21,43	11,11	9,09	0,00	11,11
<i>i</i>		44,44	14,67	4,55	0,00	10,67
<i>co</i>		26,09	7,76	8,45	0,00	9,81
<i>což</i>			0,00	11,11		7,69
<i>kdyby</i>		5,56	9,52	0,00	0,00	6,52
<i>jako</i>		12,50	6,00	1,72		4,84
<i>kteřý</i>		6,06	5,39	3,76	0,00	4,57
<i>kde</i>		6,67	4,76	0,00		4,55

<i>než</i>		0,00	5,56	0,00		2,60
<i>kdo</i>		0,00	0,00	5,26	0,00	2,04
<i>protože</i>		4,46	0,00	0,00	0,00	1,70
<i>proto</i>		3,57	0,00	2,22	0,00	1,68
<i>aby</i>		0,00	0,96	1,69	0,00	1,11
<i>tak</i>		0,00	1,41	0,00	0,00	0,86
<i>že</i>		2,13	0,81	0,00	0,00	0,85
<i>ale</i>	0,00	0,00	1,28	0,00	0,00	0,65
<i>a</i>	0,00	0,00	0,80	0,53	0,00	0,48
<i>ačkoli</i>				0,00		0,00
<i>ačkoliv</i>			0,00	0,00		0,00
<i>a ne</i>			0,00	0,00	0,00	0,00
<i>anebo</i>		0,00	0,00	0,00		0,00
<i>ani</i>		0,00	0,00	0,00	0,00	0,00
<i>aniž</i>			0,00	0,00		0,00
<i>a proto</i>		0,00	0,00	0,00		0,00
<i>a tedy</i>			0,00			0,00
<i>a tak</i>		0,00	0,00	0,00		0,00
<i>ba</i>			0,00	0,00		0,00
<i>bud'to – (a)nebo</i>		0,00				0,00
<i>byť</i>				0,00		0,00
<i>či</i>			0,00	0,00		0,00
<i>čili</i>			0,00			0,00
<i>čím – tím</i>			0,00	0,00		0,00
<i>dokonce</i>		0,00	0,00	0,00		0,00
<i>jakoby</i>				0,00		0,00
<i>jakkoli</i>			0,00			0,00
<i>jakožto</i>				0,00		0,00
<i>jak – tak</i>			0,00	0,00	0,00	0,00
<i>jednak</i>			0,00	0,00		0,00
<i>jehož</i>			0,00	0,00		0,00
<i>jelikož</i>			0,00	0,00		0,00
<i>jenomže</i>				0,00		0,00
<i>kam</i>		0,00	0,00	0,00		0,00
<i>kdykoli(v)</i>			0,00	0,00		0,00
<i>kolik</i>		0,00	0,00	0,00		0,00
<i>-li</i>		0,00	0,00	0,00		0,00
<i>nebo</i>		0,00	0,00	0,00		0,00
<i>nejen – ale (i)</i>		0,00	0,00	0,00		0,00
<i>nejenom – ale (i)</i>			0,00	0,00		0,00
<i>nicméně</i>				0,00		0,00
<i>pak</i>				0,00		0,00
<i>proč</i>		0,00	0,00	0,00		0,00
<i>proto, že</i>				0,00		0,00
<i>přece</i>			0,00			0,00
<i>přesto</i>			0,00	0,00		0,00
<i>sice</i>			0,00	0,00		0,00
<i>takže</i>		0,00	0,00	0,00		0,00
<i>tedy</i>			0,00	0,00		0,00
<i>totiž</i>		0,00	0,00	0,00	0,00	0,00
<i>třebaže</i>				0,00		0,00
<i>tudíž</i>				0,00		0,00
<i>vždyť</i>				0,00		0,00
<i>zatímco</i>				0,00		0,00
<i>zda</i>		0,00	0,00	0,00		0,00
CELKEM	0,00	4,51	3,69	2,18	0,00	3,31