

**FILOZOFICKÁ FAKULTA UNIVERZITY KARLOVY V PRAZE**

**DIPLOMOVÁ PRÁCE**

**2013**

**Lenka Šíková Hrejsemnová**



FILOZOFICKÁ FAKULTA UNIVERZITY KARLOVY V PRAZE

Ústav obecné lingvistiky

# DIPLOMOVÁ PRÁCE

Lenka Šíková Hrejsemnová

Monokolokabilní slova v češtině a dalších evropských  
jazycích

Words with Extremely Restricted Collocability  
in Czech and Other European Languages

2013

Vedoucí práce: Mgr. Václav Cvrček, Ph.D.

## **Poděkování**

Ráda bych poděkovala lidem, bez nichž by tato práce nikdy nebyla dokončena: Dr. Václavu Cvrčkovi za trpělivost a cenné rady, svému muži Janu-Markovi za vytrvalou podporu a obětavost, své rodině a přátelům za jejich shovívavost. Poděkování také patří všem, kteří v průběhu psaní práce přispěli radou či povzbudivým slovem.

## **Čestné prohlášení**

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

V ..... dne.....

podpis

## **Abstrakt**

Předkládaná diplomová práce se zabývá v české odborné literatuře doposud málo pojednávaným tématem monokolokability na lexikální rovině. Vymezuje se proti jejímu stávajícímu pojetí jako okrajové záležitosti v rámci frazeologie a idiomatiky a na datech získaných z korpusu ukazuje, že se jedná o jev podstatně širší. Navrhuje statistické nástroje, kterými je možné monokolokabilní jednotky v korpusu úspěšně vyhledávat, a na základě dat takto získaných přináší klasifikaci a popis monokolokability v češtině v jejích různých podobách, včetně některých sporných jevů. Těchto poznatků následně využívá pro mezijazykové srovnání s výsledky stejným způsobem získanými ve slovenštině, francouzštině a angličtině.

### **Klíčová slova:**

monokolokabilita, korpus, čeština, evropské jazyky, klasifikace, popis, srovnání, lexikologie

## **Abstract**

This thesis presents an insight into the phenomenon of cranberry collocations. In opposition to their prevailing conception as a marginal part of phraseology, it shows, on the basis of corpus data, that their nature is very diverse and deserves a broader view. It suggests statistical tools which can be used to search for cranberry words in corpora. Based on an analysis of the data obtained by this method, it presents a classification and description of the phenomenon in Czech, including some problematic points. These findings are subsequently used for comparison of cranberry collocations in Czech with data obtained in the same way in Slovak, French and English.

### **Key words:**

cranberry collocations, corpus, Czech, European languages, classification, description, comparison, lexicology

# Obsah

<b>1. ÚVOD</b>	<b>9</b>
<b>1.1 DEFINICE</b>	<b>10</b>
<b>1.2 DOSAVADNÍ TEORIE A VÝZKUM</b>	<b>13</b>
<b>1.3 VÝZKUMNÝ CÍL</b>	<b>19</b>
<b>2.1 ZDROJ DAT</b>	<b>20</b>
<b>2.2 MOŽNOSTI VYHLEDÁVÁNÍ MONOKOLOKABILNÍCH JEDNOTEK</b>	<b>21</b>
2.2.1 PRAVDĚPODOBNOST VÝSKYTU PRVKŮ V MONOKOLOKABILNÍM SPOJENÍ	24
2.2.2 ENTROPIE	26
2.2.3 HERFINDAHL-HIRSCHMANŮV INDEX	28
2.2.4 HRANICE MONOKOLOKABILITY	29
<b>2.3 VYHLEDÁVÁNÍ MONOKOLOKABILNÍCH SLOV PODLE HHI</b>	<b>30</b>
2.3.1 PRVNÍ KROK – VYTVOŘENÍ SEZNAMŮ	30
2.3.2 POČET TYPŮ	32
2.3.3 HHI JAKO UKAZATEL MONOKOLOKABILITY U LEMMAT	35
2.3.4 SROVNÁNÍ S ENTROPIÍ	41
2.3.5 SLOVNÍ TVARY	42
<b>2.4 SHRNU TÍ</b>	<b>44</b>
<b>3. MONOKOLOKABILITA A ČEŠTINA</b>	<b>46</b>
<b>3.1 TYPY MONOKOLOKABILITY</b>	<b>47</b>
3.1.1 KLASIFIKACE PODLE POČTU REFERENČNÍCH SLOV A KOLOKAČNÍCH PARADIGMAT	47
3.1.2 KLASIFIKACE PODLE DRUHU REFERENČNÍHO SLOVA	50
3.1.3 KLASIFIKACE PODLE DRUHU MONOKOLOKABILNÍHO SLOVA	51
<b>3.2 CHARAKTER MONOKOLOKABILNÍCH JEDNOTEK</b>	<b>54</b>
3.2.1 FORMÁLNÍ CHARAKTERISTIKA	54
3.2.2 SÉMANTICKÁ CHARAKTERISTIKA	56
3.2.3 POLYSÉMIE A HOMONYMIE	58
3.2.4 VZÁJEMNÉ VZTAHY MONOKOLOKABILNÍCH SLOV	62
<b>3.3 HRANICE MONOKOLOKABILITY A SPECIFICKÉ PŘÍPADY</b>	<b>64</b>
3.3.1 VLASTNÍ JMÉNA	64
3.3.2 TERMINOLOGIE	65
3.3.3 ZKRATKY	66
3.3.4 VLIV GRAMATIKY	67
3.3.5 INTERPUNKCE JAKO KOLOKÁT	69
3.3.6 OTÁZKA (NE)DĚLITELNOSTI	71
<b>3.4 TEORIE A PRAXE</b>	<b>72</b>
<b>3.5 SHRNU TÍ</b>	<b>74</b>
<b>4. MONOKOLOKABILITA V DALŠÍCH JAZYCÍCH</b>	<b>77</b>
<b>4.1 SLOVENŠTINA</b>	<b>78</b>
4.1.1 KLASIFIKACE	79
4.1.2 BLIŽŠÍ SPECIFIKACE	81
<b>4.2 FRANCOUZŠTINA</b>	<b>82</b>
4.2.1 KLASIFIKACE	83

4.2.2 BLIŽŠÍ SPECIFIKACE	86
<b>4.3 ANGLIČTINA</b>	<b>87</b>
4.3.1 KLASIFIKACE	88
4.3.2 BLIŽŠÍ SPECIFIKACE	90
<b>4.4 SHRUTÍ</b>	<b>92</b>
<b>5. ZÁVĚR</b>	<b>94</b>
<hr/>	
<b>PŘÍLOHY</b>	<b>100</b>
<hr/>	
<b>PŘÍLOHA 1: KRAJNÍ HODNOTY</b>	<b>100</b>
1A. LEMMATA A SLOVNÍ TVARY S NEJNIŽŠÍM POČTEM TYPŮ NA POZICI L1	100
1B. LEMMATA A SLOVNÍ TVARY S NEJNIŽŠÍM POČTEM TYPŮ NA POZICI R1	101
1C. LEMMATA A SLOVNÍ TVARY S NEJNIŽŠÍM SOUČTEM TYPŮ NA POZICÍCH L1 A R1	102
1D. LEMMATA S NEJVYŠŠÍ HODNOTOU HHI PRO POZICE L1 A R1 SAMOSTATNĚ	103
1E. LEMMATA S NEJVYŠŠÍ HODNOTOU HHI PRO POZICE L1 A R1 DOHROMADY	104
1F. LEMMATA S NEJNIŽŠÍ HODNOTOU ENTROPIE PRO POZICE L1 A R1	105
<b>PŘÍLOHA 2: PŘEHLED MONOKOLOKABILNÍCH SLOV A JEJICH KOLOKÁTŮ V ČEŠTINĚ</b>	<b>106</b>
2A. MONOKOLOKABILNÍ LEMMATA	106
2B. MONOKOLOKABILNÍ SLOVNÍ TVARY	112
<b>PŘÍLOHA 3: MONOKOLOKABILNÍ JEDNOTKY V OSTATNÍCH SLEDOVANÝCH JAZYCÍCH</b>	<b>117</b>
3A. SLOVENŠTINA	117
3B. FRANCOUZŠTINA	122
3C. ANGLIČTINA	126
<b>PŘÍLOHA 4: PŘEHLED KLASIFIKACE</b>	<b>130</b>
4A. KLASIFIKACE PODLE VELIKOSTI A STRUKTURY KOLOKAČNÍHO PARADIGMATU	130
4B. KLASIFIKACE PODLE DRUHU REFERENČNÍHO SLOVA	130
4C. KLASIFIKACE PODLE DRUHU MONOKOLOKABILNÍHO SLOVA	130



# 1. Úvod

Přirozený jazyk, ač bývá obvykle popisován jako systém, zdaleka není systémem dokonalým: obsahuje velké množství nepravidelností. Zatímco rodilý mluvčí je používá zpravidla úplně automaticky a jejich zvláštnost si ani nemusí uvědomovat, jakmile se pokoušíme jazyk popsat nebo se mu naučit, narážíme na tento fakt jako na problematičtější místo, které může naši práci značně zkomplikovat. O obtížnosti osvojování nepravidelností v cizích jazycích ostatně svědčí objem literatury a pomůcek specializujících se na objasnění a nacvičení nejčastějších problematičtějších jevů.

Nepravidelnosti přitom nejsou v jazyce nikterak vzácné a můžeme se s nimi setkat v nejrůznějších podobách na kterékoli rovině – ať už jde o střídání fonémů při ablautu, neobvyklé způsoby odvozování, frazeologická spojení, různé asymetrie ve významu nebo překvapivé pravopisné konvence. Ačkoliv většinu takových jevů lze vysvětlit diachronním náhledem, tedy skrze vývoj daného jazyka a různé faktory, které jej ovlivňují, v současném, synchronním stavu jazyka jde o prvky pravidlům se vymykající a mnohdy obtížně uchopitelné. Přesto – nebo možná právě proto – má jejich popis ve studiu jazyka své místo a zvláštní důležitost, stejně jako tyto prvky samy často najdeme nikoli na periférii jazykového úzu, ale naopak v běžně používaných výrazech. Analýzou takto výjimečných jevů můžeme ostatně dojít až k přehodnocení jimi porušovaných pravidel; nikdy totiž není jisté, zda systém, skrz nějž lingvista jazyk popisuje, v něm skutečně existuje, a nakolik jde jen o nedokonalou interpretaci jevů v jazyce nacházených.

Jedním z takových nepravidelných jevů je zvláštní kolokační charakteristika jazykových jednotek, a zejména její extrémní případ: monokolokabilita, neboli vázanost prvku, např. slova, na jedno či několik málo jiných konkrétních slov. Monokolokabilita je ve své podstatě vlastně porušením jedné ze základních vlastností jazyka, totiž jeho schopnosti kombinace jednotek. Jelikož monokolokabilní jednotky se mohou kombinovat jen s jednou nebo několika dalšími jednotkami, jsou z hlediska jazykové úspornosti značně nevýhodné. Navíc tak nevstupují do běžných kontrastů a skrze to mohou být i po stránce formální a sémantické obtížně vymežitelné či přímo neanalyzovatelné. Jejich status v rámci inventáře jednotek je tedy velmi specifický, podobně jako opačný extrém, totiž prvky s extrémně vysokou kolokabilitou, u nichž můžeme taktéž vysledovat určité deformace oproti „běžným“ či „průměrným“ jednotkám. Přes to všechno však nejde nutně

o jednotky vzácné nebo z úzu mizející, naopak, najdeme mezi nimi mnohdy i případy zcela běžné a ve své pozici stabilní.

Cílem předložené práce je navrhnout vhodný nástroj pro vyhledávání monokolokabilních jednotek lexikální roviny v korpusových datech, objasnit jejich podobu a zvláštnosti, zamyslet se nad jejich statusem v jazyce a možnostmi jejich popisu a porovnat jejich podobu v češtině a dalších jazycích.

## 1.1 Definice

V první řadě je třeba vymežit základní pojmy, kterých se naše téma dotýká a s nimiž se budeme setkávat.<sup>1</sup> Zásadní je zejména rozlišení syntagmatické a paradigmatické roviny. Zatímco **syntagmatem** rozumíme spojení (kombinaci) dvou nebo více jazykových jednotek propojených syntagmatickými vztahy, které se může objevit a běžně objevuje v textu, v němž pak funguje jako jeden celek, **paradigma** je soubor jednotek propojených paradigmatickými vztahy, tj. z nějakého hlediska vzájemně zaměnitelných, přičemž jde o soubor virtuální (systémový) a v textu na daném místě realizovaný pouze jedním svým členem. **Kolokační paradigma** je pak třída prvků, s nimiž je prvek jiného paradigmatu schopen se běžně a smysluplně kombinovat, čili s nimiž může tvořit syntagma (srov. Filipec – Čermák 1985, s. 171). Jde tedy o specifický druh paradigmatu, který je vymezen na základě vztahu ke konkrétní jednotce, která do tohoto paradigmatu nepatří, a jeho členy jsou vzájemně zaměnitelné v kombinaci s tímto vnějším prvkem. Příkladem může být kořen slova *černý*, tedy *čern-*, který se kombinuje s odpovídající třídou koncovek. Tato třída je jeho kolokačním paradigmatem, a kořen se může vždy kombinovat právě s jedním jejím členem, tedy s jednou koncovkou. Vedle toho se však může kořen *čern-* kombinovat i s jinými prvky, jako třeba s odvozovacími afixy (čímž může vzniknout např. *zčernat*) nebo s dalšími kořeny (*černočerný*), a ty pak tvoří jeho další kolokační paradigmata (je ovšem třeba rozlišovat rovinu morfologickou a lexikální; o kolokacích obvykle mluvíme u jednotek lexikonu).

Právě podoba a množství kolokačních paradigmat vymezuje tzv. **kolokabilitu** jazykové jednotky, ke které se tato paradigmata vztahují. Kolokabilita je definována jako „formální a sémantická spojitelnost prvku s jinými, jež je vytvářena jedním, častěji však více jeho kolokačními paradigmaty“ (Filipec – Čermák 1985, s. 173). Jde tedy o schopnost jednotky tvořit syntagma s jinými jednotkami, přičemž u každé jednotky je dáno,

---

<sup>1</sup> Vycházíme zde především z definic a popisů uvedených v Encyklopedickém slovníku češtiny s přihlédnutím k dalším zdrojům uvedeným v textu.

s jakými druhy jednotek, popř. s jakými konkrétními jednotkami se mohou takto kombinovat. Kolokabilitu je přitom třeba odlišovat od čistě sémantické **kompatibility**, totiž slučitelnosti významu jednotek, která sice s kolokabilitou souvisí, avšak – zvláště v idiomatických spojeních – nemusí být nutně splněna. Více k této tematice viz Filipec – Čermák (1985, s. 170-173).

Na pozadí tohoto terminologického a koncepčního materiálu již můžeme vymezit, co přesně pojmem **monokolokabilita** míníme. Jestliže jsme zde mluvili o kolokabilitě de facto jako o kombinačním potenciálu jazykové jednotky, pak **monokolokabilita** představuje její maximálně omezený případ, kdy kolokační paradigma tvoří jen jedna nebo několik (jako maximální počet se uvádí 7 či 8) jednotek.<sup>2</sup> F. Čermák (2007, s. 29) ji definuje jako „takovou kolokabilitu lexikální formy, která je jasně omezená na možnost kombinovat se s několika málo (2-7), resp. jedinou další formou v jazyce“<sup>3</sup>, přičemž právě a pouze spojení s touto další jednotkou jí zajišťuje přežití v jazyce. Z jiného úhlu pohledu lze monokolokabilní jednotku chápat jako prvek, jehož použití (výskyt) je podmíněn přítomností konkrétního jiného prvku, čili který se vyskytuje pouze v kolokaci s tímto prvkem. Pojem **kolokace**, na který zde narážíme a narážet nutně budeme, zahrnuje hned několik významů v závislosti na autorovi a kontextu; v této práci jej používáme obecně jako ustálené slovní spojení v kontrastu proti **syntagmatu** jako jakémukoliv formálně a sémanticky přijatelnému spojení. F. Čermák (Filipec – Čermák 1985, s. 173) v souvislosti s definicí monokolokability dále uvádí, že nejde pouze o vlastní spojitelnost jednoho prvku s jiným, ale zároveň o *vztah* určitého (monokolokabilního) prvku k jeho zúženému kolokačnímu paradigmatu.

Pro ilustraci jmenujme coby klasický příklad spojení *nechat na holičkách* – forma *holičkách* se vyskytuje jen v kombinaci se slovesem *nechat* (a jeho několika vzácnými kontextovými variantami jako *nechávat*, *být*, *octnout se* apod.) a předložkou *na* (která žádné alternativy nemá) a má tak dvě kolokační paradigmatá, z nichž jedno obsahuje jediný člen a druhé několik málo členů s jedním převažujícím, a ke každému z těchto paradigmat (resp. jejich členům) má specifický vztah. Vzhledem k tomu, že zde máme hned dvě extrémně omezená kolokační paradigmatá a forma *holičkách* se kombinuje

---

<sup>2</sup> Vymezení pomocí konkrétního počtu jednotek je pochopitelně čistě arbitrární a neodpovídá jazykové realitě, která má spíše kontinuální, škálovitý charakter. Není zde brána v úvahu frekvenci slova ani ohled na komunikační funkci (mluvčí může např. cíleně použít neobvyklou kombinaci slov, která ovšem v rámci daného kontextu není o nic méně smysluplná než kolokace s běžnými prvky kolokačního paradigmatu). Je tedy třeba tuto podmínku chápat jako konvenci, kompromis, který nevystihuje povahu monokolokability v její úplnosti a který je v případě aplikace na text nutně upravit či doplnit podle potřeb výzkumu.

<sup>3</sup> F. Čermák takto definuje monokolokabilitu na lexikální rovině, tutéž definici však lze použít i pro rovinu morfologickou.

s prvky obou najednou, je dokonce monokolokabilní dvojnásobně, totiž vůči oběma paradigmátům.

Jak již bylo naznačeno, v této práci se zaměřujeme na monokolokabilitu lexikálních jednotek; je však třeba připomenout, že se s ní můžeme setkat také na rovině morfologické i na rovinách vyšších. Oblíbeným příkladem tohoto jevu je anglický morfém *cran*, který se objevuje pouze ve slově *cranberry*.<sup>4</sup> Stojčev (1984, s. 26) navíc u řady zkoumaných případů konstatuje, že kořenový morfém daného monokolokabilního slova je sám monokolokabilním morfémem, obě formy se tedy mohou vyskytovat a také vyskytují společně.

Jelikož naše práce je založena na korpusové studii, je třeba zmínit také pojmové nástroje korpusové lingvistiky, které budeme používat. Jedná se zejména o tag, lemma, lexikální formu, KWIC a konkordanci.

V rámci Českého národního korpusu (dále jen ČNK) je každá slovní jednotka opatřena poziční značkou, nazývanou **tag**. Taková značka může nést informace nejrůznějšího druhu. V ČNK jde o údaje morfologické, lze však narazit také na korpusy využívající podobné značky např. pro zaznamenávání syntaktických vztahů. Přítomnost tagu, a zejména tak podrobného jako v ČNK, však není ani zdaleka samozřejmostí; značkování (tagování) je prováděno automaticky, je to však velmi náročný proces a i v těch nejméně úspěšných případech za sebou zanechává celou řadu chyb, ať už způsobených nepravděpodobností či víceznačností přítomnou v textu nebo nedostatečnou rozlišovací schopností tagovacího nástroje. Přes tato úskalí přináší přítomnost tagu při vyhledávání v korpusu obrovské možnosti.

Slovní tvar a lemma jsou další kritéria, podle nichž lze korpusová data vyhledávat. **Slovní tvar** (také forma, word) je každý odlišný tvar, jakého může nějaké slovo nabývat, např. *Karlovi*. Nejde přitom o tvar vymezený morfologicky, ale čistě jako sled grafémů, což znamená, že homonymní tvary jednoho slova a dokonce i homonymní tvary různých slov (např. *ženu*) spadají všechny bez rozdílu pod jedinou formu. Vedle toho jsou odlišnými slovními tvary i varianty s velkým a malým písmenem. Naopak **lemma** je reprezentativní tvar slova (v definicích se obvykle uvádí, že jde o tvar slovníkový) a zahrnuje všechny možné morfologické formy daného slova coby lexému. Výhody i nevýhody obojího je třeba zvážit při každém vyhledávání v korpusu. Zvláště významnou

---

<sup>4</sup> Jackendoff (1997, s. 165) zmiňuje také nově vzniklé slovo *cranapple*. Jeho existence sama o sobě ovšem monokolokabilitu morfému *cran* „neohrožuje“, jelikož jeho kolokační paradigma tak má stále jen dva členy, toto by se však mohlo změnit, kdyby podobným způsobem vznikly a rozšířily se další neologismy.

se tato otázka stává v případě jevů nepravidelných, jako je právě monokolokabilita. Výběr způsobu vyhledávání totiž má přímý dopad na to, jaké výsledky se nám podaří získat. Nad touto problematikou se blíže pozastavíme v oddílu 2.3.2.

Zadáním vyhledávacího dotazu získáme výsledky, neboli **konkordanci**. Ta sestává z úryvků textu, v nichž byl vyhledávaný výraz, nazývaný také **klíčové slovo**, nalezen. Vyhledávaný výraz sám je v konkordanci označován jako **KWIC** (key word in context). **Levým a pravým kontextem** se míní text těsně předcházející, resp. následující po KWIC. Pro více informací k ČNK a k termínům a metodám spojeným s jeho využitím viz např. Šulc (1999) nebo Koček et al. (2000).

Mluvíme-li v této práci o **sloveh**, je třeba také ujasnit, co tímto pojmem myslíme. Vzhledem k lexikologické praxi jde o termín poněkud nešťastný, nese však s sebou nespornou výhodu srozumitelnosti a intuitivnosti<sup>5</sup>. S ohledem na korpusová data této studie tedy budeme za **slovo** považovat formu oddělenou mezerou z každé strany přímo v korpusu (resp. v konkordanci). Toto užití se v zásadě shoduje s významem termínu **pozice**, který se v korpusové lingvistice používá, ten však zahrnuje i interpunkční a jiné znaky, které za slova nepovažujeme. I při tomto jednoznačném vymezení skrze uvedení v korpusu ovšem ještě na problematiku vymezení slova, jeho samostatnosti a hranic narazíme<sup>6</sup>. Jednotku, jejíž přítomnost monokolokabilní slovo vyžaduje, neboli její nezbytný **kolokát**, pak budeme nazývat **referenčním slovem**, v souladu s tím, jak byl tento termín zaveden a používán v mé předchozí práci (Šíková Hrejsemnová, 2011).

## 1.2 Dosavadní teorie a výzkum

Monokolokabilita patří v lingvistice mezi poměrně málo studovaná témata. Zpravidla o ní bývá pojednáváno pouze okrajově jakožto o součásti frazeologie a jen málokdy se jí dostane pečlivějšího studia.

V českém prostředí jí po teoretické stránce věnuje pozornost téměř výlučně František Čermák v rámci svého výkladu frazeologie a idiomatiky. Monokolokabilní slovo podle něj může existovat pouze jako nevydělitelná součást frazému (srov. Filipec – Čermák 1985, s. 173). To vyplývá už ze samotné definice frazému (z hlediska kolokačního) jako „jedinečného spojení minimálně dvou prvků, z nichž některý (popř. žádný) nefunguje stejným způsobem v jiném spojení (resp. více spojení), popř. se vyskytuje pouze ve

---

<sup>5</sup> I přirozená intuitivnost slova již byla napadena (srov. Haspelmath, 2011), v češtině jakožto jazyce využívajícím písmo, ve kterém jsou slova od sebe oddělována, s ní však lze počítat.

<sup>6</sup> Viz oddíl 3.3.6 v této práci.

výrazu jediném (resp. několika málo)“ (Čermák 2007, s. 31 nebo Filipec – Čermák 1985, s. 177), v níž poslední část prakticky odpovídá definici monokolokabilních jednotek. Při dalším popisu uvádí, že monokolokabilní slova nemají vlastní lemma ani význam a nelze u nich – v různé míře – určit slovní druh nebo jiné kategorie (Čermák 2007, s. 30). Tyto závěry vyplývají právě z omezené kolokability prvku, která znemožňuje jeho zařazení do běžných paradigmat, a také z jeho výskytu vázaného na frazém a jeho obměny, přičemž frazém je ucelená jednotka jen částečně rozložitelná a analyzovatelná, což ztěžuje či přímo vylučuje analýzu v jediném přirozeném kontextu monokolokabilního slova. Těžko si lze např. představit hodnověrnou sémantickou analýzu monokolokabilního *duhu* ve spojení *jít k duhu*. Jde tedy o jednotky natolik defektní, že je Čermák vůbec nepovažuje za slova a označuje je s rezervou jako „verboidy“ (Čermák 2007, s. 186)<sup>7</sup>. Přestože však jsou toto všechno zcela legitimní a logické úvahy, lze proti nim nalézt protipříklady, kdy je např. slovní druh nebo význam monokolokabilního slova naopak snadno určitelný, či dokonce kdy je na místě pochybovat o tom, zda se vůbec jedná o součást frazému – zvláště pokud do monokolokabilních slov zahrneme opravdu všechna slova, která odpovídají jejich výše zmíněné definici.<sup>8</sup>

Studie na monokolokabilitu přímo zaměřené jsou u nás vzácností. Jedním takovým případem je diplomová práce Stilijana Stojčeva (1984), která si klade za cíl podat systematický, ucelený a co možná nejúplnější popis monokolokabilních slov a v tomto směru je asi nejobsáhlejším českým dílem o monokolokabilitě vůbec. Drží se přesně mezi a předpokladů vytyčených F. Čermákem (právě Čermák ji ostatně vedl coby školitel), v jejich rámci však problematiku rozvíjí podrobnějším popisem na základě databáze vybraných českých a bulharských frazémů (možnost existence monokolokability mimo frazeologii se zde nezkoumá). Sledované frazémy pocházejí z výkladových a frazeologických slovníků obou jazyků, výjimečně z rozhovorů či autorova vlastního povědomí. Jedná se tedy o materiál z dnešního hlediska relativně málo využitelný – ukazuje monokolokabilní jednotky ve frazémech, jak jsou zachyceny lexikografy (přičemž Stojčev sám v tomto ohledu nešetří kritikou) a jejich skutečné využití v praxi je možné jen odhadovat na základě ve slovnících uvedených popisů a jazykového povědomí autora nebo případů, na které narazil. Ve srovnání s možnostmi současných metod korpusové lingvistiky jde tedy o vzorek nereprezentativní a poněkud nespolehlivý.

---

<sup>7</sup> Termín verboid navrhl Josef Filipec ve své recenzi publikované ve 3. čísle 67. ročníku *Naší řeči* jako označení pro takové komponenty frazémů, které nemají všechny flektivní tvary nebo nesou v rámci frazému jiný význam než v běžném užití odpovídajícího lexému.

<sup>8</sup> Více k této problematice viz Šíková Hrejsemnová (2011) a podkapitola 1.2 v této práci.

Přesto jeho práce přináší velmi cenný vhled do charakteru monokolokabilních jednotek. Podává jejich popis po stránce zvukové, gramatické, sémantické, kolokační, funkční, sleduje jejich platnost coby znaků i hodnotu v rámci teorie informace, příležitostně také porovnává češtinu s bulharštinou atd. U všech těchto hledisek poukazuje na větší či menší anomálnost monokolokabilních slov a skrze ni na jejich specifické, značně nestabilní postavení v jazyce. Na závěr práce přináší „slovník“ frazémů s monokolokabilním komponentem, tedy databázi, z níž vycházel. Na jeho poznatky budeme v předkládané práci dále narážet (viz především podkapitola 3.2).

Na tuto studii se také odkazuje diplomová práce Otmaro Káninského, která se zaměřuje právě na frazémy s monokolokabilním komponentem. Káninský taktéž vychází z pojetí Čermáka a pomocí vyhledávání v korpusu a excerptce z vybraných slovníků se snaží najít všechny případy monokolokabilních slov v českých frazémech. Při korpusovém vyhledávání používá jako kritéria frekvenční charakteristiky předpokládané u monokolokabilních slov a zaměřuje se na lexikální formy, nikoliv lemmata. Vzhledem k značnému množství a často nevalné relevanci získaných dat pak přistupuje k jejich další redukci a posléze k ručnímu třídění. Výsledkem je seznam několika stovek tvarů frazémů s monokolokabilním komponentem, které dále analyzuje.

Káninského práce je pro nás zajímavá především z toho důvodu, že uvádí do praxe přístup k monokolokabilitě podle F. Čermáka, a naráží-li na stejné problémy a jevy jako my, staví se k nim často velmi odlišně. Jelikož se snaží pojmut monokolokabilitu v celé širší její definice (v rámci frazeologie), přijímá jako monokolokabilní všechna slova, jejichž kolokační paradigma tvoří maximálně 8 slov, resp. 16 slovních tvarů<sup>9</sup>, a která zároveň splňují daná frekvenční kritéria a jsou považována za frazémy (což bylo ověřováno ve slovníku). Velká část dat přitom byla vyřazena pouze na základě posledního požadavku. Jak Káninský uvádí, „většina kolokací, které jsou výsledkem takového vyhledávání, nepatří vůbec do oblasti IF“. Tento fakt přechází s tím, že případy mimo IF (frazeologii a idiomatiku) nelze za monokolokabilní považovat, neudává ovšem žádný důvod, proč by to tak mělo být, a v závěru také o tomto rozhodnutí pochybuje. Zato však podává popis typů takto vyřazených dat, a ten pro nás není bez zajímavosti, jelikož se v něm objevují také (a nejen) výrazy, na které jsem narazila i při vlastním výzkumu a zabývá se jimi taktéž předkládaná práce.

---

<sup>9</sup> K tomuto požadavku však lze vznést námitku, že 16 lexikálních forem (zavedených kvůli variantám s velkým a malým písmenem) zdaleka neobsáhne celou variabilitu tvarů i jen jednoho nebo dvou ohebných slov. V tomto smyslu jde tedy o kritérium nevhodně zvolené. Popis Káninského přístupu a jeho srovnání se svým dále rozvádím v úvodu k podkapitole 2.2.

Právě na podoby monokolokability v češtině je zaměřena má studie z roku 2011<sup>10</sup>, na niž tato práce přímo navazuje. Vycházela z pojetí Čermáková, proti němuž se vymezovala: nespokojila se s představou monokolokabilních slov jako jednotek, které se vyskytují pouze ve frazémeh a nelze o nich prakticky nic říci, a pomocí vyhledávání v korpusu, konkrétně v nereferenčním korpusu SYN v rámci ČNK, který v té době čítal 500 milionů slov, se pokusila zjistit, co všechno definici monokolokability vlastně odpovídá, a na základě výsledků pak nabídnout jiný pohled na celou problematiku. Podobně jako ve výše zmíněném výzkumu Káninského se ukázalo, že monokolokabilní je celá řada slov, která (společně s jejich referenčním slovem) za frazémy nepovažujeme. Při pokusu o klasifikaci tak byly navrženy čtyři typy monokolokability:

1. Reflexiva tantum (např. *smát se, hroutící se, postesknout si*)
2. Slova valenčně vázaná (např. *záležet na, opírat o, vyplývající z*)
3. Frazémy (např. *křížem krážem, par excellence, na počest*)
4. Víceslovná pojmenování (např. *zoologická zahrada, nadmořská výška, krychlový metr*)

Každý z těchto typů představuje specifickou problematiku a také odlišný vztah monokolokabilního slova k jeho referenčnímu slovu či slovům. **Reflexiva tantum** se vyskytují jen se zvratným zájmenem, které však má v těchto případech ke zvratnému zájmenu daleko, a mimo tuto vázanost se chovají úplně stejně jako jiná slovesa nebo přídavná jména. **Slova valenčně vázaná** jsou snad nejzvláštnějším typem – jedná se o jednotky, které obligatorně vyžadují doplnění jiným slovem, jež se na ně váže přes konkrétní gramatické slovo, a právě tento gramatický prvek je pro ně referenčním slovem. Jde tedy svým způsobem o monokolokabilitu sekundární, jelikož referenční slovo se tu váže nikoli s monokolokabilním slovem samotným, ale teprve s jeho doplněním. A stejně jako u reflexiv tantum s sebou ani v tomto případě monokolokabilita nenese defektnost po stránce formální či významové, která je průvodním rysem frazémů (pokud se nerozhodneme považovat všechny uvedené typy kolokací také za frazémy právě na základě monokolokability jedné jejich části). **Frazémy**, přinejmenším v tom smyslu, jak si je obvykle představujeme, tvoří mezi případy monokolokability překvapivě malou část (může ovšem jít zčásti o důsledek toho, že byla zkoumána pouze spojení přesahující absolutní frekvenci 100 výskytů). Díky obvyklému přístupu k monokolokabilitě jako poddruhu frazeologie jim však je v literatuře věnováno nejvíce pozornosti. Z tohoto

---

<sup>10</sup> Klauzurní práce na Ústavu lingvistiky a ugrofinistiky FFUK.



důvodu jsem se ve své práci jejich studiem podrobně nezabývala. Poslední skupinou jsou **víceslovná pojmenování** tvořená zpravidla monokolokabilním adjektivem a substantivem. Tyto případy, často terminologické povahy, stojí na okraji frazeologie nebo mimo ni; např. podle Čermákova členění (1985, s. 205) by mohly spadat mezi kvazifrazémy. K jednotlivým typům, jejich zvláštnostem a zastoupení v datech se budu průběžně vracet v kapitolách 3 a 4.

V klauzurní práci jsem k vyhledávání v korpusu využila lemma (na rozdíl od metody Káninského, viz výše v této kapitole) a orientovala se primárně na monokolokabilitu absolutní, tedy na slova, jejichž kolokační paradigma je tvořeno jen jedním prvkem, popř. která se vážou obligatorně se dvěma kolokačními paradigmaty, z nichž alespoň jedno je pouze jednočlenné. Takovým případem je třeba *zavolanou ve spojení (jako) (na) zavolanou*<sup>11</sup> (srov. Šíková Hrejsemnová 2011, s. 6). Monokolokabilní slova, jejichž kolokační paradigma mělo více prvků, tak do mého tehdejšího seznamu pronikla pouze v případě, že jeden z těchto prvků se objevil v drtivé většině výskytů. Počet zkoumaných monokolokabilních slov byl dále (pravděpodobně ve velké míře) redukován požadavkem minimální četnosti výskytů v korpusu. Inventář monokolokabilních slov, se kterými jsem při analýze pracovala, tak i přes značnou velikost použitého korpusu nebyl ani zdaleka vyčerpávající. Vzhledem k tomuto faktu i k nevelkému rozsahu práce zůstala řada otázek nezodpovězena či pouze naznačena. Cílem této práce je tedy navázat, kde jsem přestala, a pokusit se o hlubší a zároveň ucelenější náhled.

V zahraniční literatuře (s výjimkou slovenské) na termín monokolokabilita (monocollocability) nenarazíme vůbec, resp. pouze v pracích přeložených z češtiny. Místo toho se setkáme s celou řadou nejasných a mnohoznačných pojmů; Čermák (1982, s. 208) jmenuje kupříkladu **nekrotismy** (u některých sovětských autorů), v angličtině **nonce forms** či **dummy words**, popř. i zcela nepřesné **archaismy**. V současné anglicky psané literatuře se pro spojení s monokolokabilním slovem jako nejvhodnější jeví termín **cranberry collocations** („brusinkové kolokace“). Toto označení je motivováno analogií k **cranberry morpheme** („brusinkový morfém“), totiž morfému, který se vyskytuje pouze v jednom slově. Již zmíněným klasickým příkladem tohoto jevu je právě anglické *cranberry* („brusinka“, „klikva“), jehož první morfém, *cran*, se v žádném jiném slově

---

<sup>11</sup> V příkladech uvedených přímo v textu je obvykle referenční slovo (nebo slova) uvedeno v jednoduchých závorkách. Pokud se jedná o víceprvkové kolokační paradigma, jsou jednotlivé kolokáty vypsané ve složených závorkách.

nevyskytuje a lze jej blíže specifikovat jen skrze opozici k dalším druhům bobulí (*blackberry* – „ostružina“, *strawberry* – „jahoda“ apod.).<sup>12</sup>

Podobně jako v českém prostředí, i v zahraniční literatuře je monokolokabilita zmiňována spíše okrajově jako vlastnost určitého typu frazémů. Tento relativní nezájem konstatuje v roce 1982 také F. Čermák (ibid.), když píše, že „systematická registrace monokolokabilních slov a jejich komplexní popis není dosud k dispozici v žádném více studovaném jazyce, i když se o jejich existenci mluví již déle“, a zdá se, že se od té doby v tomto směru mnoho nezměnilo.

Práce R. Moonové (1998) na téma monokolokability naráží v rámci klasifikace „fixed expressions and idioms“, tedy frazeologie a idiomatiky v širším slova smyslu, pro anglický jazyk. Slovní spojení s monokolokabilním prvkem řadí do kategorie „anomalous collocations“ („anomální kolokace“) společně s dalšími podtypy: „ill-formed collocations“ („špatně utvořené kolokace“), totiž kolokacemi, které porušují gramatická pravidla, „defective collocations“ („defektní kolokace“), kde jeden z prvků nenesé význam nebo je jeho význam jiný než v ostatních spojeních, a „phraseological collocations“ („frazelogické kolokace“), které představují velmi málo produktivní spojení v tom smyslu, že na pozici jednoho prvku může stát jen několik jiných prvků (Moon 1998, s. 19-21). Tato skupina je definována svou anomálností vůči lexikogramatické struktuře a stojí v opozici k dalším dvěma kategoriím, tzv. „formulae“ (ustálená přirovnání, přísloví atd.), vymezeným pragmaticky, a „metaphors“, čili více či méně průhledným metaforám (srov. tamtéž). Navrhuje pak i jemnější členění těchto velkých skupin. Jednotlivé případy přitom mohou patřit do více kategorií zároveň. Moonová uvádí, že kolem 25% spojení v její databázi (jednalo se o data získaná z korpusu o 18 milionech slov a dalších zdrojů) vykazovalo rysy dvou typů najednou a 1% dokonce spadalo do tří kategorií.

Opačný přístup představuje článek F. Richtera a M. Sailera (2002). Jedná se o pokus zahrnout specifické vlastnosti monokolokabilních slov do teorie generativní gramatiky. Autoři na základě analýzy příkladů monokolokabilních slov v němčině ukazují, že s výjimkou omezené kolokability se jedná o jednotky a kolokace vykazující zcela pravidelné chování a není proto vhodné je považovat za součást frazeologie a idiomatiky. Poukazují dále na nedostatky různých nástrojů generativní gramatiky při pokusu o pro uchopení omezené kolokability a navrhují pro tento účel nástroj nový, atribut COLL, který začleňují do stávající teorie.

---

<sup>12</sup> *Cranberry* už zřejmě není jediným možným použitím morfému *cran*, stále je však k němu v tomto smyslu v literatuře odkazováno. Srov. pozn. 4 v této práci.

### 1.3 Výzkumný cíl

Jak již bylo naznačeno, cílem této práce je navázat na mou studii z roku 2011, či přesněji na pojetí monokolokability, které jsem v ní přijala, a na zjištění, která jsem učinila, a skrz to podrobit tuto problematiku podrobnějšímu, kritickému a v rámci možností ucelenému pohledu, který by v ideálním případě bylo možné vztáhnout i na další jazyky.

Nejprve tedy provedu nové vyhledávání ve velkém českém korpusu, přičemž se pokusím získat i takové výsledky, které se do výsledného seznamu v předchozí studii nemohly dostat a přesto je u nich smysluplné o monokolokabilitě hovořit. To nutně obnáší zhodnocení stávajících metod vyhledávání a navržení a uplatnění co nejvhodnějšího možného postupu.

Na základě získaných výsledků pak provedu revizi své klasifikace a rozšíření popisu monokolokabilních jednotek a jejich kolokací. Součástí těchto úvah bude také zamyšlení nad některými konkrétními problémy s monokolokabilitou spojenými a nad otázkou jejich hranic.

V poslední části práce pak využiji těchto poznatků a pokusím se je aplikovat také na další jazyky: vyhledám obdobným způsobem monokolokabilní slova v korpusech vybraných evropských jazyků, konkrétně angličtiny, francouzštiny a slovenštiny, a pokusím se výsledky porovnat s češtinou jak po stránce obecné charakteristiky a vymezených typů, tak v otázce, zda se jednotlivé jazyky v konkrétních případech monokolokability podobají či přímo shodují.

## 2. Hledání monokolokability v korpusu

### 2.1 Zdroj dat

Téma jako monokolokabilita si nutně žádá velké množství dat. Jde totiž o jev, který se plně projeví teprve při pohledu na desítky, v lepším případě stovky a tisíce příkladů. Podíváme-li se např. na větu (1), můžeme z ní samotné o formě *holičkách* vyvodit maximálně to, že existuje, a o její kolokabilitě můžeme říct, že zahrnuje spojení s předložkou *na* a se slovesem *necat*. Chceme-li být ještě důslednější, můžeme toto poslední tvrzení upřesnit s ohledem na přítomný tvar *nenechá*. Tento konkrétní příklad nám však neříká nic o tom, zda se slovo reprezentované zde tvarem *holičkách* vyskytuje i v dalších tvarech a zda připouští spojení s jinými slovy, natož s kolika a kterými.

(1) Ať se stane cokoli, nenechá svého přítele na holičkách.<sup>13</sup>

Budeme-li však k tomuto příkladu přidávat další, povaha slova *holičkách* se dříve či později začne projevovat; zjistíme tak například, že přípustné jsou i jiné tvary slovesa *necat* a že toto sloveso se v příkladech objevuje nápadně často. Zároveň sotva narazíme na jinou formu našeho KWIC než *holičkách*.<sup>14</sup> Avšak teprve po nashromáždění velkého množství příkladů můžeme začít vyvozovat o vlastnostech tohoto slova v obecné rovině. Nenarazíme-li na jinou formu než *holičkách* napotřetí, neznamená to nic, pokud na ni ale nenarazíme ani poté, můžeme předpokládat, že jiná forma neexistuje nebo je velmi vzácná, příp. že existuje pouze v nějaké varietě jazyka, kterou naše data nezahrnují. Podobně pokud ze 100 příkladů řekněme 80 bude v kombinaci se slovesem *necat*, 15 s *necávat*, 5 se *zůstat* a všech 100 zároveň s předložkou *na*, můžeme z toho již vyvozovat zajímavé závěry o kolokabilitě. I v tomto případě je však možné, že dalších 100 příkladů bude obsahovat kombinace jiné nebo v jiném poměru, a proto stejně jako při jakémkoli jiném zkoumání platí, že čím více příkladů máme, tím spolehlivější budou naše zobecnění.

Z tohoto důvodu je nutné při studiu monokolokability použít zdroj dat tak rozsáhlý a zároveň bohatý, jak je jen možné (pochopitelně s ohledem na množství dat, jaké je výzkumník schopen zpracovat), a zabývat se jen jednotkami, které se v takovém zdroji

---

<sup>13</sup> Není-li uvedeno jinak, pocházejí všechny příkladové věty prezentované v této práci z korpusu SYN v rámci ČNK.

<sup>14</sup> Pochopitelně po vyřazení případů, kde jde o homonymní slovo označující kadeřnici.

vyskytují mnohokrát. Pro svou studii jsem proto zvolila SYN2010, nejaktuálnější referenční korpus v rámci ČNK, který je žánrově vyvážený a dostatečně velký (100 milionů slov). Pro rozsáhlejší výzkum by bylo jistě vhodnější použít korpus SYN, který je mnohonásobně větší, vzhledem ke zde zvolenému způsobu vyhledávání a zpracování dat by však byl jako zdroj až příliš velký a výsledky by mohla ovlivnit i jeho žánrová nevyváženost (výrazně v něm převažují publicistické texty).

## 2.2 Možnosti vyhledávání monokolokabilních jednotek

Způsobů, jak v korpusu vyhledávat ustálená slovní spojení, je vícero, a každý z nich má své výhody i nedostatky. Obecně lze říci, že nástroj, který by v daném korpusu dokázal vyhledat všechna monokolokabilní slova a právě jen ona, zatím neexistuje a je otázkou, zda je něco podobného vůbec možné. Výzkumník je proto vždy nucen uchýlit se k výběru kritérií, která mu přinesou výsledky co nejčistší (tj. obsahující co nejméně chyb a slov s širší kolokabilitou) a co nejbližší tomu, co pro své záměry potřebuje.

Více o tomto problému napovídá pohled na studie, které se s ním různým způsobem potýkaly. Nejjednodušší je samozřejmě případ, kdy sledujeme jen jeden či několik předem vybraných příkladů – zde se této problematice vlastně zcela vyhýbáme a jedinou potíží je vhodný výběr kandidátů, jejichž výskyty v korpusu vyhledáváme. Takto jsem postupovala např. ve své klauzurní práci z roku 2012, kde jsem se zaměřila na 8 francouzských slov se silně omezenou kolokabilitou a zkoumala jejich vlastnosti skrze výskyty v korpusu InterCorp. Další případovou studií je článek I. Bozděchové, v němž se autorka pozastavuje nad kolokabilitou slova *tratoliště*.

Chceme-li však vyhledáváním získat více než několik již známých případů a naopak postihnout co nejširší škálu jednotek, je prakticky jediným možným postupem využití statistických nástrojů. Řada korpusových manažerů (vyhledávacích programů pracujících s korpusovými daty) umožňuje počítání kolokací, filtrování a další operace, s jejichž pomocí se dá monokolokabilita poměrně snadno zjistit či ověřit; tyto nástroje však zpravidla přicházejí na řadu teprve tehdy, když máme před sebou konkordanci nebo alespoň seznam potenciálně monokolokabilních kandidátů, tedy když samotné vyhledávání monokolokabilních jednotek již proběhlo. Je proto třeba definovat si monokolokabilitu prostřednictvím nějaké statistické vlastnosti, kterou jsme schopni v korpusu zjistit.<sup>15</sup> Tento

---

<sup>15</sup> Tento krok již překračuje možnosti manažeru Bonito a tím i možnosti běžného uživatele ČNK.

postup s sebou nese nutnost buďto omezit množství či druh získaných dat, nebo naopak pracovat i s velkým množstvím nerelevantních výsledků.

S obojím se potýkal např. O. Káninský (2003, s. 17-27). Jeho cílem bylo získat frazémy s monokolokabilním komponentem. Tyto kolokace vyhledával podle slovních tvarů pomocí dvou kritérií: podíl absolutní frekvence celé kolokace a absolutní frekvence hledaného prvku musel být větší nebo roven 0,25 a kolokační paradigma prvku mohlo tvořit maximálně 16 lexikálních forem. První požadavek znamenal, že daná kolokace musela tvořit alespoň čtvrtinu případů výskytu hledaného prvku, čili že se toto spojení často opakovalo. Druhá podmínka de facto shrnovala požadavek monokolokability. Proti ní však lze snadno vznést námitku, že číslo 16 (určené jako dvojnásobek maximálního počtu členů kolokačního paradigmatu prvku, který by bylo možné podle Čermákovy definice považovat za monokolokabilní) samo o sobě mnoho nevypovídá. Situaci zde totiž komplikuje počítání ve slovních tvarech. Například slovo, které se kombinuje s neohebnými slovními druhy, takto může mít až 16 různých kolokátů a hranici monokolokability pak značně přesahuje, ale přesto se dostane do databáze. Naopak slovo, které se kombinuje třeba jen se dvěma slovesy a je tedy určitě monokolokabilní, se do databáze dostat nemusí, jestliže tato slovesa mohou v kolokaci nabývat různých tvarů či stát na začátku věty (čili začínat velkým písmenem). Nehledě na to, zda počítáme lemmata či slovní tvary, také pouhý součet prvků není dostačujícím kritériem, jelikož nevypovídá o tom, jak jsou spojení s jednotlivými prvky frekventovaná, zda nejde v některých případech o záměrně neobvyklou kombinaci (např. v poesii) apod. Káninský navíc stanovuje hranici minimální četnosti každého hledaného prvku v korpusu jen na 4 výskyty, což je množství pro konstatování monokolokability zcela nedostačující (srov. oddíl 2.1.1 v této práci) a rapidně zvyšující počet kritériím vyhovujících, ale nepoužitelných prvků. Důsledkem takto zvolených kritérií bylo značné množství dat, z nichž jen malá část byla pro cíle výzkumu relevantní, a autor tak byl nucen provést velmi rozsáhlé a náročné třídění. Jde tedy o postup s výhradami použitelný, avšak poněkud neefektivní.

Kritéria podobného typu, ale mnohem přísnější, byla použita pro vyhledávání monokolokability v mé práci z roku 2011. Primárním cílem v tomto případě byla slova, která mají pouze jedno referenční slovo, ačkoliv výsledky s několikačlenným kolokačním paradigmatem nebyly následně z databáze vyřazovány. Vyhledávání proběhlo na lemmatech (u hledaných i referenčních slov), takže spojení jako např. *nadmořská výška* byla počítána ve všech tvarech dohromady, a klíčovým slovem mohla být pouze substantiva, adjektiva, slovesa a adverbia. Vlastní monokolokabilita byla při vyhledávání

formulována požadavkem, aby se klíčové lemma vyskytovalo minimálně v 85% případů ve spojení s jedním konkrétním lemmatem (přičemž toto referenční slovo mohlo být tvořeno jen malými českými alfabetskými znaky). Jde v zásadě o totéž kritérium, jaké použil Káninský (viz výše), ovšem podstatně striktnější a také přesnější vzhledem k cílům výzkumu – to je dáno faktem, že na rozdíl od Káninského jsem se nesnažila zahrnout do výsledků monokolokabilní prvky s více referenčními slovy. Minimální četnost v korpusu byla stanovena na 100 výskytů, tedy počet, kdy už vliv náhody není tak velký a ve výsledcích lze sledovat obecnější trendy. Takto vysoký práh byl opodstatněn také tím, že práce probíhala na lemmatech a tudíž bylo od řady kolokací k dispozici více výskytů v různých tvarech; vyčleňování lexikálních forem při zachování tohoto limitu by pravděpodobně ve výsledném seznamu značně zredukovalo počet přítomných lexémů a frekventovanější slova by naopak mohla být zaznamenána hned několikrát v různých tvarech. Variabilita v datech byla dále umocněna tím, že referenční slovo se mohlo nacházet kdekoliv ve vzdálenosti do 3. pozice nalevo či napravo od daného monokolokabilního slova. Toto kritérium bylo zavedeno s ohledem na poměrně volný slovosled češtiny. Kombinace těchto tří kritérií se ukázala jako praktická – výsledků nebylo příliš málo a ačkoliv ruční třídění bylo nutné, vyřazených slov při něm nakonec nebylo mnoho, také v důsledku toho, že jsem se nevyhýbala žádnému typu monokolokability a naopak se snažila postihnout tento jev v co nejvíce podobách. I tento postup však měl své nevýhody, a to především právě ve své zaměřenosti na slova směřující k absolutní monokolokabilitě (tedy s jediným referenčním slovem). Je docela dobře možné, že u slov s o něco širším kolokačním paradigmatickým by se našly i další typy případů se specifickými vlastnostmi, a tak i když se tato práce snažila o co nejširší pojetí, zůstala de facto omezena jen na monokolokabilitu v její nejsilnější podobě. Podobně mohlo zakrýt některé druhy monokolokability nebo tvarové nepravidelnosti s monokolokabilitou spojené vyhledávání pomocí lemmatu. Zvolená kritéria tak navzdory zmíněným výhodám nelze jednoduše převzít a použít pro další výzkum, pokud by měl obsáhnout více, než v této klauzurní práci bylo cílem.

Z tohoto důvodu a také ve snaze najít co nejvýhodnější způsob vyhledávání monokolokabilních jednotek jsme se pro účely této práce rozhodli použít metodu odlišnou. V jejím jádru stojí postupy využívané v lingvistice zejména teorií informace, totiž výpočet pravděpodobnosti výskytu jedné jednotky v závislosti na výskytu jiné jednotky.

### 2.2.1 Pravděpodobnost výskytu prvků v monokolokabilním spojení

Obecně řečeno prvek nese v rámci kontextu tím větší míru informace, čím méně pravděpodobná je jeho přítomnost na dané pozici. V tomto směru představují monokolokabilní jednotky velmi specifický případ. Tuto zvláštnost konstatuje již Čermák (1982, s. 63), když píše, že monokolokabilní jednotka nenese informaci žádnou, protože pravděpodobnost jejího výskytu se blíží jistotě<sup>16</sup>. Domnívám se, že tato úvaha je nepřesná – ve skutečnosti se jistotě blíží pravděpodobnost výskytu referenčního slova, objeví-li se v kontextu slovo monokolokabilní. Pro ilustraci uveďme příklad (2).

- (2) a. Druhý očitý svědek se jmenoval Storm.  
b. „Podepište se jako svědek,“ řekl Sellers Dr. Wellsovi.

Ve větě (2a) vidíme monokolokabilní jednotku *očitý* a její referenční slovo *svědek*. *Očitý* se v korpusu SYN2010 vyskytuje celkem 347krát, a to vždy ve společnosti referenčního slova *svědek* nebo slov *svědectví*, *svědkyně* a *účastník*, které tvoří ostatní členy jeho kolokačního paradigmatu. Můžeme proto říci, že ve chvíli, kdy se v kontextu objeví *očitý*, pravděpodobnost výskytu jednoho ze čtyř členů jeho kolokačního paradigmatu se blíží jistotě. Oproti tomu přítomnost *svědek* v kontextu ještě zdaleka neznamena, že se objeví i *očitý*. Věta (2b) je příkladem, kdy se *svědek* vyskytuje zcela nezávisle na *očitý*, a takových příkladů bychom našli celou řadu – právě proto, že *svědek* má podstatně širší kolokační paradigma(ta) než *očitý*, může být rozvíjen nejrůznějšími jinými adjektivy nebo se vyskytovat samostatně jako v příkladu (2b). O asymetrii tohoto páru svědčí i nepoměr absolutních frekvencí obou slov – oproti 347 výskytům *očitý* stojí 7 001 *svědek*.<sup>17</sup>

Podobnou úvahu naznačuje ve své diplomové práci také Stojčev: „V tomto smyslu se doprovodný komponent MS jeví jakožto de facto nadbytečný, redundantní. MS jako by absorbuje, přivlastňuje si jeho informační hodnotu.“<sup>18</sup> Je však třeba mít na paměti, že připsání informační hodnoty spojení jako celku nebo monokolokabilnímu prvku v něm (vyjma případů, kdy je jedna z jednotek elidována) nelze volně zaměňovat – jedná se také

<sup>16</sup> „Protože informační hodnota prvku (zde slova) je nepřimo úměrná stupni jeho pravděpodobnosti, prediktability, potvrzuje se nám tak mj., že prvky monokolokabilní informaci nenesou (téměř) žádnou, jsou informačně prázdné (pravděpodobnost jejich výskytu se mění v krajních případech v jistotu).“ (Čermák 1982, s. 63)

<sup>17</sup> Tento vztah je možné kvantitativně vyjádřit pomocí obligatornosti, popř. vzájemné obligatornosti (srov. Cvrček, v tisku).

<sup>18</sup> MS je ve Stojčevově práci zkratka pro monokolokabilní slovo.



o otázku hloubky analýzy. Považujeme-li spojení s monokolokabilním slovem za nedělitelný celek, počítáme samozřejmě informační hodnotu celého spojení jako každého jiného slova v kontextu a v tomto směru pravděpodobně ani nenajdeme výrazné rozdíly oproti zbytku slovní zásoby; avšak ve chvíli, kdy monokolokabilní spojení rozdělíme na jednotlivé prvky a počítáme pravděpodobnost jejich přítomnosti vzhledem k sobě navzájem a k ostatnímu kontextu, můžeme již očekávat velmi specifické výsledky (takový krok je teoreticky napadnutelný, jelikož izoluje ve skutečnosti nesamostatné prvky, právě této neúplnosti však může být využito k automatickému rozlišení monokolokabilních slov od kolokačně neutrálních).

Naše práce je případem tohoto druhého přístupu. Z výše uvedeného lze vyvodit následující tvrzení:

- A. Pravděpodobnost výskytu monokolokabilního prvku v daném kontextu bude prakticky nulová, pokud se v tomto kontextu nevyskytuje žádné z jeho referenčních slov (a není-li referenční slovo pouze elidováno); v případě výskytu s některým referenčním slovem bude rovna pravděpodobnosti výskytu celého spojení coby jednotky vyššího řádu.
- B. Pravděpodobnost výskytu referenčního slova v daném kontextu, pokud tento kontext nezahrnuje monokolokabilní slovo, jehož kolokačního paradigmatu je součástí, se nebude lišit od jednotek slovní zásoby, které se nevyskytují v monokolokabilním spojení (není-li samo referenční slovo monokolokabilní); v případě výskytu s patřičným monokolokabilním slovem bude úměrná postavení v jeho kolokačním paradigmatu, oproti ostatním typům spojení však vysoká a blízcí se či rovná jedné, jedná-li se o monokolokabilitu absolutní.

Tyto předpoklady můžeme využít jako základ pro vlastní vyhledávání. Jsou-li správné a použijeme-li je vhodným způsobem, získáme relevantní výsledky. Metoda studií uvedených výše v tomto oddíle pracovala v zásadě s tvrzením A, jen v mírně posunutém smyslu – kritériem pro vyhledávání byl poměr počtu výskytů daného (potenciálně monokolokabilního) slova ve spojení s jiným (potenciálně referenčním) a celkově. Předkládaná práce využívá naopak důsledky tvrzení B – zaměřuje se na statistické vlastnosti referenčního slova vzhledem k monokolokabilnímu.

## 2.2.2 Entropie

Postupem, který bývá pro podobné účely využíván, je výpočet **entropie (H)**. Tato veličina byla zavedena do teorie informace a popsána pro její potřeby v práci Claude Shannona (1948). Jde o nástroj, který dokáže (mimo jiné) kvantifikovat množství informace obsažené v určitém prvku na základě pravděpodobnosti jeho výskytu na dané pozici; a to je přesně typ kritéria vhodný pro sledování monokolokability.

Máme-li například systém o  $n$  jednotkách A, B, C, ..., N a pravidla tohoto systému říkají, že po prvku A mohou následovat některé z ostatních prvků s určitou pravděpodobností, můžeme spočítat entropii pozice prvku následujícího po A a získaná hodnota nám řekne, jakou volnost máme při výběru prvku do této pozice, čili s jakou jistotou či nejistotou můžeme predikovat, který z prvků se zde objeví. Čím vyšší je hodnota entropie, tím „náhodnější“, méně předvídatelný je výběr takového prvku. Kupříkladu v systému o třech prvcích, z nichž má každý stejnou pravděpodobnost výskytu na určité pozici, bude entropie pro tuto pozici nižší než v systému s 50 prvky o stejné pravděpodobnosti výskytu.

V případě monokolokability tedy bude hodnota entropie v pozici referenčního slova velmi nízká – máme zde totiž na výběr jen z velmi omezeného počtu prvků a i z těch mnohdy ve velké míře převažuje prvek jediný. Pro absolutní monokolokabilitu by se entropie měla blížit nule.

Máme tak v rukou nástroj, který je schopen kvantifikovat kolokabilitu slov, a zdálo by se, že pro vyhledávání monokolokability by mohl být ideální. I v tomto případě však narážíme na určité obtíže. Jednou z nich je relativně složitý a nepříliš intuitivní výpočet entropie. Tento bod je ovšem snadno vyřešen využitím počítače, kdy je výpočet proveden automaticky. O něco problematičtější je fakt, že entropie de facto nepopisuje kolokabilitu daného slova, ale volnost výběru prvků na určité pozici v jeho okolí. Podle Shannona si můžeme představit produkci jako sled stavů, kdy je na základě předchozího kontextu vybírán příští prvek z určité množiny, a entropie se v tomto smyslu váže právě k oněm stavům, nikoli ke konkrétním prvkům. Chceme-li proto využít entropii jako kritérium pro vyhledávání monokolokability, musíme si vytyčit pozice, kde je nejpravděpodobnější výskyt referenčního slova, a entropii počítat pro ně.

Rozmezí  $\langle +3, -3 \rangle$  (tj. kdekoliv mezi 3. pozicí nalevo a 3. pozicí napravo od klíčového slova), které bylo použito ve studiích uvedených výše, je pro tyto účely až příliš široké a značně by zkomplikovalo práci s daty. Mnohem výhodnější by bylo počítat

entropii pro každou konkrétní pozici zvlášť a zároveň co možná nejvíce snížit počet takto sledovaných pozic. Ačkoliv referenční slovo (čili pozice s výrazně sníženou entropií) se může nacházet kdekoli v této vzdálenosti a mnohdy ještě dále – takový případ vidíme hned ve větě (1), kde se referenční slovo *nenechá* objevuje na pozici -4 vůči *holičkách*, a to ve zcela přirozeném slovosledu – lze zároveň očekávat, že ustálené spojení bude mít tendenci klást své členy do co nejtěsnější blízkosti. Platnost tohoto předpokladu pro monokolokabilní jednotky dokazuje Václav Cvrček (v tisku), když sleduje proximitu (blízkost) zvrtného zájmena vůči reflexivům tantum a referenčního slova vůči ostatním typům monokolokabilních jednotek. Pro tuto druhou skupinu se v téměř 90 % případů referenční slovo vyskytovalo na pozici +1 nebo -1, tedy hned vedle monokolokabilního slova. Pro reflexiva tantum byla průměrná vzdálenost zvrtného *se/si* o něco vyšší, avšak nejčastěji se opět vyskytovala bezprostředně u klíčového slova. Tento poznatek je o to významnější, že pozice zvrtného zájmena u sloves je podle pravidel slovosledu dána jako následující po prvním přízvučném celku ve větě, čili nezávisle na pozici slova, k němuž se váže. To umožňuje mezi oběma prvky značnou vzdálenost – a přesto je nejčastěji najdeme v intervalu  $\langle +1, -1 \rangle$ , přičemž větší vzdálenost, jak Cvrček konstatuje, je zpravidla způsobena přítomností jiného obligatorního doplnění (např. zájmena pro nepřímý předmět) mezi reflexivem a jeho zvrtným zájmenem.

S ohledem na tento trend můžeme bez velkého rizika „znečištění“ či naopak nežádané redukce dat počítat entropii pro pozice +1 (R1) a -1 (L1) vůči klíčovému slovu. Výskyty, kdy se referenční slovo nachází ve větší vzdálenosti, by se měly projevit jen jako mírné zvýšení entropie na sledovaných pozicích. Pro rozlišení případů, kdy referenční slovo preferuje či obligatorně stojí jen na jedné straně, budeme počítat entropii pro levou a pravou pozici zvlášť i pro obě dohromady.

Využití entropie pro vyhledávání však komplikuje ještě jedna nesnáz: předpokládáme, že čím více se budeme blížit absolutní monokolokabilitě, tím více se hodnota entropie pro danou pozici bude blížit nule. Nevíme však, jaké nejvyšší hodnoty může entropie u monokolokabilních slov dosáhnout. Navíc pravděpodobně zjistíme odlišné hodnoty u případů, kdy referenční slovo preferuje jen jednu stranu (a dostaneme tak pro tuto stranu entropii velmi nízkou a pro druhou stranu naopak relativně vysokou). Vzhledem k tomu, že kolokabilita je vlastnost, kde se jasné hranice hledají jen těžko, s hodnotami entropie to sotva bude jiné.

Při vyhledávání v korpusu tedy nemůžeme rovnou zadat kritérium, že hledáme taková slova, pro něž platí, že entropie pro dané pozice je nižší než daný limit – tento limit

jednoduše neznáme. Použijeme proto mírně odlišný postup a jedním z cílů této práce bude najít ve výsledcích alespoň přibližný trend v poměru monokolokabilních slov mezi nimi v závislosti na výši entropie nebo jiného kritéria. Orientace v takovém trendu by pak mohla být využita jako jednoduchá a praktická pomůcka pro další výzkumy.

### 2.2.3 Herfindahl-Hirschmanův index

Entropie ovšem není jedinou veličinou, kterou můžeme vyjádřit tento typ vztahů. Podobně lze využít také tzv. **Herfindahl-Hirschmanův index (HHI)**. Tento index se zatím v lingvistických pracích příliš často neobjevuje; jeho původním účelem je vyjadřovat diversitu trhu v rámci ekonomických věd, srov. Hirschman (1964). Stejně dobře však může posloužit k vyjádření diversity kontextu daného slova, jak jej používá také Cvrček (v tisku). Metafora situace na trhu ostatně není představě kolokability tak vzdálená, jak by se mohlo zdát. Danou pozici si můžeme představit jako odvětví, v němž operují různé firmy. Je-li těchto firem mnoho a rozdělují si trh přibližně rovnoměrně, diversita takového trhu je velká. Stejně tak pokud na dané pozici může stát velké množství různých slov a žádné z nich výrazně nepřevažuje, můžeme na této pozici mluvit o velké diversitě – to znamená, že zde buďto stojí slova, která se ke klčovému slovu nevážou, nebo je klčové slovo vzhledem k prvkům na této pozici velmi široce kolokabilní. Absolutní monokolokabilitu je zase možné připodobnit k situaci, kdy na trhu existuje jen jedna firma – dané referenční slovo má na určitou pozici „monopol“ a dochází k minimální diversitě. A konečně v případech, kdy má monokolokabilní slovo kolokační paradigma o několika prvcích, které se na dané pozici relativně rovnoměrně střídají, bude hodnota HHI podobná jako na trhu, který si rozdělilo několik firem, jde tedy svým způsobem o „kartel“ těchto referenčních slov.

Oproti entropii má HHI tu výhodu, že jeho výpočet je podstatně jednodušší a také intuitivně snáze uchopitelný. Počítá se jako součet pravděpodobností jednotlivých prvků, které se na dané pozici mohou vyskytnout, umocněných na druhou. Vyšší diversitě tak odpovídají nižší hodnoty HHI a naopak; v případě absolutní monokolokability se hodnota HHI bude blížit nebo dokonce rovnat jedné, protože pravděpodobnost obligatorního prvku bude také blízká jedné, v případě maximální diversity se bude blížit (ale nikoli rovnat) nule, protože každý z mnoha prvků bude mít jen velmi malou pravděpodobnost výskytu. Umocnění na druhou zvyšuje rozdíl mezi vysokými a nízkými pravděpodobnostmi.

Kromě tohoto však nese HHI v zásadě tytéž výhody a nesnáze pro použití k vyhledání monokolokability jako výpočet entropie. Stejně jako u entropie se potřebujeme soustředit na co nejmenší počet pozic, pro které budeme hodnoty počítat, a opět nevíme, jak vysokých (resp. nízkých) hodnot mohou monokolokabilní slova vůči pozici s referenčními slovy dosahovat. Zůstává také otázkou, zda a nakolik se výsledky získané těmito dvěma metodami budou lišit – k tomuto uvádíme více v oddílu 2.3.4 této práce.

#### 2.2.4 Hranice monokolokability

Z výše uvedeného se zdá, že vyhledávání monokolokability pomocí hodnot entropie a HHI by mělo být poměrně snadné. Situaci však komplikuje řada faktorů. Zřejmě nejvýznamnějším je fakt, že tyto metody počítají diversitu nikoliv pro slova, ale pro pozice v textu, na kterých tato slova stojí. I když referenční slovo stojí ve většině případů bezprostředně vedle monokolokabilního, nemusí zde stát vždy – stačí jednoduchá transformace a může se přesunout na opačnou stranu nebo na některou vzdálenější pozici. Ani u slov výrazně preferujících jednu stranu nevíme předem, o kterou jde. Monokolokabilní spojení se ve volnosti referenčního prvku různí; některá mají slovosled daný a neměnný, u některých se může referenční slovo vzdálit třeba na sedm i více pozic daleko – typickým a asi nejextrémnějším příkladem tohoto jsou reflexiva tantum (ani u nich k tomu ovšem nedochází často, viz výše). Roli hraje také vlastní způsob výpočtu – vyšších hodnot budou dosahovat primárně pozice u slov směřujících k absolutní monokolokabilitě, kde se pravděpodobnost výskytu jednoho referenčního slova blíží jedné – s klesajícím HHI, resp. stoupající entropií, však tento vliv nejčtenějšího prvku přetrvává a může se stát, že slovo, které není monokolokabilní, ale většinu jeho výskytů tvoří určitá kolokace, se bude dle indexů zdát „monokolokabilnější“ než jiné slovo, které má jen několik možných kolokátů, ale žádný z nich není výrazně dominantní.

Z těchto důvodů lze jen těžko očekávat, že v hodnotách obou indexů existuje něco jako „hranice monokolokability“. Mnohem pravděpodobnější je, že se stoupající entropií, resp. klesajícím HHI, se bude ve výsledcích poměr monokolokabilních slov oproti ostatním pozvolna zmenšovat takřka donekonečna. Hledáme-li tedy hranici, která by se dala použít při vyhledávání monokolokability podle HHI nebo entropie, musíme spíše než hranici monokolokability samotné hledat hranici, kdy se monokolokabilní slova mezi výsledky vyskytují relativně málo – tak málo, že ruční třídění, které by je vydělilo

z ostatních o podobné hodnotě indexu, by již zabralo příliš mnoho úsilí a času a počet takto získaných monokolokabilních slov by byl již z hlediska výzkumu zanedbatelný.

Takovou pragmatickou hranici je ovšem nutné určit pro každý výzkum zvlášť s ohledem na jeho cíle, metody a časové možnosti. Očekáváme však, že poměr monokolokabilních slov oproti ostatním, resp. jeho klesající průběh podle hodnot indexů, se nebude výrazně měnit v závislosti na typu textu ani jazyce, jelikož HHI i entropie vyjadřují čistě statistické informace o diversitě okolí dané jednotky, nehledě na ostatní její vlastnosti. Podaří-li se nám tedy zmapovat, jak tento průběh vypadá, měly by tyto výsledky být použitelné i pro další fáze našeho výzkumu.

## **2.3 Vyhledávání monokolokabilních slov podle HHI**

### **2.3.1 První krok – vytvoření seznamů**

S tímto cílem, totiž zjistit, jakým způsobem ve skutečnosti hodnoty HHI a jiných ukazatelů odpovídají monokolokabilitě, byla z korpusu nejprve vybrána a vypsána všechna plnovýznamová slova (čili slova automaticky označená jako substantivum, adjektivum, sloveso nebo adverbium) s absolutní frekvencí minimálně 200 výskytů. Paralelně přitom byly vytvořeny dva seznamy – zvlášť byla vyhledána lemmata a lexikální formy. Ke všem položkám pak bylo přiřazeno několik statistických údajů:

- (a) Absolutní frekvence sledovaného prvku
- (b) Počet typů (lemmat/lexikálních forem dle seznamu) přítomných na pozici L1
- (c) Počet typů na pozici R1
- (d) Počet typů na obou pozicích dohromady
- (e) Hodnota normalizovaného HHI pro pozici L1
- (f) Hodnota normalizovaného HHI pro pozici R1
- (g) Hodnota normalizovaného HHI pro obě pozice dohromady
- (h) Hodnota entropie (H) pro pozici L1
- (i) Hodnota H pro pozici R1
- (j) Hodnota H pro obě pozice dohromady
- (k) Podíl interpunkce na pozici L1 (vyjádřeno v procentech)
- (l) Podíl interpunkce na pozici R1
- (m) Podíl interpunkce na obou pozicích dohromady
- (n) 3 slova, která se na pozicích L1 a R1 vyskytovala nejčastěji

Počtem typů rozumíme to, kolik různých lemmat, resp. slovních tvarů, se vyskytlo na dané pozici. Např. slovo *odiv* v seznamu lemmat má na pozici L1 jen 1 typ (*na*), v seznamu slovních tvarů však má forma *odiv* na téže pozici 2 typy (*na*, *Na*). Hodnota (d) není pouhým součtem hodnot (b) a (c); jedná se o počet typů, které najdeme minimálně na jedné z pozic L1 a R1; existují-li typy, které se mohou vyskytovat na obou pozicích, jsou započítány jen jednou. Totéž platí také pro hodnoty (g), (j) a (m). Různý počet kontextových typů pro jednotlivé položky s sebou nese nutnost použít normalizovanou verzi výpočtu HHI. Jeho hodnoty, které se pohybují jen v intervalu  $(0,1)^{19}$ , byly zaokrouhleny na desetistícinu, hodnoty entropie na tisícinu.

Jelikož na pozicích L1 a R1 se mnohdy objevovaly také interpunkční znaky a ty nebyly z našich seznamů vyloučeny, uvedli jsme v rámci (k), (l) a (m) do seznamu také jejich procentuální podíl ve výskytech, zaokrouhlený na desetiny procenta. Ačkoliv přítomnost interpunkce s monokolokabilitou přímo nesouvisí, může jít z hlediska analýzy o zajímavý údaj a automatickým vyřazením těchto výsledků by také mohla být ztracena některá relevantní data. Více k roli interpunkce uvádíme v oddílu 3.3.5.

Posledním údajem zavedeným do seznamu jsou 3 nejfrekventovanější slova, s nimiž se sledované slovo na daných pozicích vyskytuje, vyjma interpunkce. Jde spíše o orientační a ilustrativní prvek. U monokolokabilních slov by se zde mělo objevit referenční slovo nebo slova, nejde však o žádné závazné kritérium.

Výpočet těchto údajů byl proveden automaticky pomocí speciálního ad hoc zkonstruovaného perlovského skriptu.<sup>20</sup>

Lemmat, která odpovídala zadaným kritériím, bylo celkem 21 384, slovních tvarů 33 686. Vzhledem k tomuto objemu a poměrně mizivému poměru pro tento výzkum zajímavých slov v něm jsem s kompletním seznamem nepracovala (ani jej zde neuvádím), ale vytvořila jsem několik dílčích seznamů pro každý sledovaný údaj (b) – (j) zvlášť. V těchto tabulkách byly slovní tvary a lemmata vždy seřazeny podle dosažené hodnoty. Tímto způsobem se automaticky shromáždila na začátku seznamů slova, která by měla být monokolokabilní, a dále v seznamu omezení kolokability prvků postupně sláblo.

---

<sup>19</sup> Dosažení hodnoty 1, přestože je teoreticky možné, by bylo zcela výjimečným případem, jako realističtější se proto jeví zápis intervalu  $(0,1)$ ; v případě monokolokability však coby ideální případ hledáme právě takový extrém.

<sup>20</sup> Za nezbytnou technickou pomoc v této fázi výzkumu děkuji Dr. Václavu Cvrčkovi.

### 2.3.2 Počet typů

U ideálního monokolokabilního slova bychom očekávali na jedné ze stran počet typů roven jedné. Takový výsledek by znamenal, že dané slovo je absolutně monokolokabilní a jeho referenční slovo se nachází vždy na pozici L1 nebo vždy na pozici R1.<sup>21</sup> V seznamu lemmat jsme se s ním setkali jen pětkrát: na levou stranu bylo orientované pouze (*na*) *odiv* a v pozici R1 stálo referenční slovo vždy v případech *de (facto)*, *tab* (.), *křížem (krážem)* a *pod* (.). Stojí za povšimnutí, že ve dvou případech byl „referenčním slovem“ interpunkční znak, konkrétně tečka. Zde jde pouze o zkratku za celé slovo (*tab.* pro *tabulka* a *pod.* pro *podobný* nebo *podobně* – nejedná se o předložku) a tečku za touto zkratkou, tedy o jev, který lze za monokolokabilitu označit jen s velkými výhradami (více k otázce interpunkce jako kolokátu viz oddíl 3.3.5). Zbývající tři příklady jsou však monokolokabilními slovy par excellence. Výsledky o počtu typů do 20 pro jednu ze stran uvádíme v přílohách č. 1a a 1b. Všechna lemmata až do hodnot 7 (resp. 8) musí nutně patřit monokolokabilním slovům nebo zvláštním případům jako výše uvedené *tab*. Pokud se totiž na této pozici střídá jen málo slov, znamená to, že tato slova tvoří kolokační paradigma daného prvku. Kdyby na pozici L1 mohlo běžně stát i slovo, které s klíčovým slovem netvoří syntagma (čili nemůže patřit do jeho kolokačního paradigmatu), dosažené hodnoty by byly mnohem vyšší. Nízká hodnota říká, že klíčové slovo na pozici L1 pustí jen slova, s nimiž je svázáno, a pokud sem pronikne slovo jiné, jde o ojedinělý případ, který ovšem jen dále redukuje předpokládané kolokační paradigma. Pokud by se navíc takto nějaký „vetřelec“ často opakoval, znamenalo by to naopak, že klíčové slovo je pravděpodobně součástí většího ustáleného slovního spojení. Stejně jako v případě Káninského práce zde také platí, že samotný počet typů neříká nic o poměru jejich frekvence v kombinaci s klíčovým slovem, a je třeba jej vztáhnout i k absolutní frekvenci klíčového slova samotného. Stejný počet typů u slova o dvou stovkách a o několika tisících výskytů nevypovídá o stejné situaci.

U hodnot 9 a více už se mohou objevovat prvky, které hranici monokolokability překračují; jejich kolokabilita je stále nízká, do definice monokolokability – přinejmenším v její stávající podobě – se však již nevejde. Vzhledem k tomu, že kromě sledování každého případu zvláště neexistuje způsob, jak od sebe odlišit slova, kde počet typů skutečně odráží šíři kolokačního paradigmatu, od slov, v jejichž bezprostřední blízkosti se vyskytují i „vetřelci“, nelze prostý součet typů pro jednu z pozic použít pro stanovení hranice monokolokability: můžeme pouze vyčlenit slova, která jsou nutně

---

<sup>21</sup> To ovšem nevylučuje možnost, aby takové slovo bylo monokolokabilní i vůči jinému kolokačnímu paradigmatu, jehož prvky by se objevovaly na jiných pozicích.



monokolokabilní (a vytrídít z nich případy jako *tab*, na což by dost možná stačil požadavek, aby do typů nebyla započítána interpunkce), většina monokolokabilních prvků však zůstane ve druhé skupině jednoduše proto, že referenční slovo nebo slova se mohou (i kdyby jen výjimečně) od klíčového vzdálit nebo se objevit na opačné straně.

Součet typů na obou stranách, tedy hodnota (d), je kritériem ještě hůře použitelným. Lze si sice představit slovo, které by bylo monokolokabilní vůči dvěma slovům zároveň a každé se vyskytovalo na jedné straně, vícenásobně monokolokabilní slova však jsou poměrně vzácná a ve všech ostatních případech již hraje roli různě proměnlivá situace na druhé, volné pozici. Při pohledu do seznamu (výsledky s nejnižšími součty uvádím v příloze 1c) zjistíme, že na vrcholu se umístily především zkratky a výrazy nezačleňující se do větné struktury a celkově jsou dosažené hodnoty podstatně vyšší. Od hodnoty 30 výše vidíme mezi výsledky již velmi hustou koncentraci monokolokabilních slov; ve všech případech jde nutně o slova se silně omezenou kolokabilitou, opět však nemáme žádný plošně použitelný nástroj, jak mezi nimi vybrat čistě monokolokabilní lemmata. Tyto výsledky by však mohly velmi dobře posloužit jako zdroj dat pro studii zaměřenou na omezenou kolokabilitu v širším měřítku.

Seznam slovních tvarů byl pochopitelně o mnoho delší, čítal celkem 33 686 položek. Stejně jako u lemmat jsem proto vytvořila dílčí seznamy podle jednotlivých kritérií.

Hned na začátku je nutno připomenout, že údaje o formách nám přinášejí jiné informace než tytéž údaje pro lexém reprezentovaný lemmatem. Zatímco pro lemma, které představuje všechny možné formy daného lexému, výsledky přinášejí vnitřně nerozlišený součet všech kolokátů vyskytujících se minimálně s jednou formou, a tak obraz o kolokabilitě lexému jako celku, jednotlivé slovní tvary umožňují sledovat, jak je tato kolokabilita vnitřně uspořádána. S jejich pomocí můžeme zjistit, že dané slovo (je-li ohebné) se v různých podobách chová velmi podobně, pokud budou jeho tvary na seznámech blízko u sebe, nebo naopak že jednotlivé tvary jsou různě zatíženy.<sup>22</sup> Pokud v blízkosti jedné formy, eventuelně v celém seznamu nenajdeme další, může to v zásadě znamenat dvě věci: další tvary buďto neexistují, popř. jsou vzácnější (nedosáhly minimálního počtu 200 výskytů), nebo mají výrazně odlišnou kolokabilitu. První případ může v extrému znamenat, že dané slovo má jen jeden nebo několik ustrnulých tvarů, jako je tomu v případě (*nechat*) (*na*) *holičkách*. Ve druhém případě může dojít i k jevu, kdy je určité slovo monokolokabilní jen v jedné nebo několika ze svých forem, jako např. tvar *řo*,

---

<sup>22</sup> Stejného výsledku je pochopitelně možné dosáhnout i sledováním jednotlivých lemmat v jejich tvarech pomocí základních statistických nástrojů v manažeru Bonito.

*poř* *víkendech*. Něco podobného lze nalézt pouze při vyhledávání na slovních tvarech, jelikož lemmata zahrnují i formy s běžnou kolokabilitou, která monokolokabilní formy „skryje“.

Kromě tohoto platí pro formy v zásadě tytéž úvahy jako v případě lemmat; v souladu se současnou definicí jsou jednoznačně monokolokabilní pouze ty, které mají počet typů na jedné straně max. 7 (resp. 8)<sup>23</sup>. Oproti lemmatům takových případů vidíme více (výsledky s nejnižšími dosaženými hodnotami uvádím v přílohách č. 1a a 1b). Tento fakt je možné vysvětlit jednak tím, že konkrétní formy ohebných slov mají obecně užší kolokabilitu než celý lexém (např. určité pády se mohou pojít pouze s některými předložkami, vokativ bývá v textu zpravidla oddělen čárkami apod.), a za druhé – a to se týká případů na vrcholu tabulky – může jít právě o případy, kdy se určitá forma vyskytuje pouze v jednom či několika ustrnulých spojeních.

Všimněme si kupříkladu, že ačkoliv slovo *zásluha* se vyskytuje v různých formách a spojeních (v našem seznamu figurují coby tvary s minimální absolutní frekvencí 200 formy *zásluhou*, *zásluhy*, *zásluhu*, *zásluha* a *zásluze*), tvar *zásluze* měl na pozici L1 pouze 8 typů. Tímto se výrazně lišil od ostatních uvedených tvarů, které na stejném místě dosáhly hodnot od 90 do 483. Při pohledu do korpusu zjistíme, že se vyskytuje téměř výlučně ve spojeních *po zásluze* a *o zásluze*, toto druhé užití je navíc jen okrajové. Jedná se tedy právě o případ, kdy je monokolokabilní jen jedna z možných forem daného slova, a jak již bylo naznačeno, pohled do tabulek s lemmaty na tento fakt nijak neupozorňuje, je možné jej objevit jen při práci přímo s jednotlivými formami.

Vedle toho se ovšem ve výsledcích setkáme i s případy, kdy je referenčním slovem interpunkční znak, ať už se jedná o zkratky (tytéž, které jsme viděli u lemmat) či o jiné specifické tvary, jako je již zmiňovaný vokativ. Ten se obzvláště nápadně prosadil na vrchol tabulky s počtem typů pro obě pozice dohromady (případy s nejnižšími hodnotami uvádím v příloze 1c), jelikož je z textu zpravidla vyčleněn interpunkcí – zejména čárkami, ale i dalšími znaky. V mnoha případech je navíc takřka obligatorně předcházen slovem *pane* nebo *paní*. Hned na první pozici s pouhými 13 typy pro obě pozice tak stojí *mylorde*, o něco dále najdeme i *mylady* (i přesto, že jde o slovo nesklonné) a s nevelkým odstupem následují další případy jako *prezidente*, *majore*, *doktoro* nebo *tatínku*.

I když si však odmyslíme tento typ nalezených tvarů, vidíme, že výsledky se mnohdy od dat získaných skrze lemma liší. Jedním z případů, které se při práci s lemmaty zcela

---

<sup>23</sup> Počtem typů se zde však myslí počet různých slovních tvarů – může přitom jít o různé lexémy nebo jen různé tvary stejného lexému. Narážíme tak na stejné obtíže jako Káninský, srov. oddíly 1.2 a 2.2 v této práci.

ztratily, je slovo *neposlední*, které formálně spadá pod lemma *poslední*. Užívá se však prakticky výlučně ve spojení *v neposlední řadě* – je tedy monokolokabilní vůči dvěma různým jednotkám (dvěma jednočlenným kolokačním paradigmatům), které vůči němu stojí téměř vždy v pozici L1 a R1.

Podobné případy ukazují, jak je důležité při vyhledávání monokolokability sledovat zvlášť lemmata i lexikální formy: díváme-li se jen na lemmata, mohou nám uniknout i zcela jednoznačné příklady a nutně přehlédneme typ „dílní“ monokolokability, totiž monokolokabilitu jen jedné z více forem lexému, jako je výše zmíněné *{po, o} zásluze*. Na druhou stranu sledujeme-li jen jednotlivé formy, nemůžeme z výsledků jedné zobecňovat na celý lexém, protože ostatní formy mohou mít zcela jiné vazby, a v případě, že pracujeme jen s výsledky o určité minimální četnosti, můžeme přehlédnout mnoho méně frekventovaných ohebných slov.

Jak je vidět, počet typů je pomůckou užitečnou, ale nikoliv dostačující pro vyčlenění monokolokabilních slov z korpusu. S její pomocí jsme schopni určit některé případy jako jednoznačné, při o něco vyšších číslech však tento nástroj nepomůže odlišit nízkou kolokabilitu od monokolokability.<sup>24</sup> Stejně tak jeho účinnost klesá úměrně k tomu, jak pevně je referenční slovo vázáno na konkrétní pozici, tj. může-li se pohybovat po levé i pravé straně nebo se od monokolokabilního slova vzdálit. Počet typů nám tak je schopen přinést několik jasných případů a nastínit, jaká další slova by monokolokabilní být mohla. K přesnějšímu vymezení však již potřebujeme jiný nástroj.

### 2.3.3 HHI jako ukazatel monokolokability u lemmat

Tabulky zarovnané podle výše indexu HHI již nabízejí poněkud komplexnější náhled na situaci. Jak již bylo řečeno, HHI může nabývat hodnot od 0 do 1, přičemž vyšší čísla odpovídají homogennějšímu kontextu. Zastoupení hodnot v datech však není rovnoměrné: pro drtivou většinu slov se pohybovaly v intervalu od 0 do 0,1. Průměrná hodnota HHI v celém seznamu byla u lemmat 0,0426 pro pozici L1 a 0,0733 pro R1. S ohledem na tyto údaje je možné za příznakově vysoké považovat již hodnoty nad 0,2 či dokonce nad 0,1. Nemusí se však nutně jednat hned o monokolokabilitu: zvýšená hodnota může značit, že se klíčové slovo kombinuje s relativně málo jinými prvky nebo že výraznou část jeho výskytů tvoří určitá kolokace, za monokolokabilitu však označujeme teprve extrémní případy takového chování.

---

<sup>24</sup> Vzhledem ke škálovité povaze kolokability je ovšem stanovení jakékoliv pevné hranice záležitostí čistě arbitrární.

Lemmata s nejvyššími hodnotami uvádím v přílohách č. 1d a 1e. Již při prvním letmém porovnání s odpovídajícími seznamy pro počet typů vidíme, že vyjma oněch pěti případů, kdy se na jedné straně objevil jediný typ, se jednotlivé položky v obou tabulkách liší – některé jen v rámci několika pozic v pořadí, jako např. *potaz* či *krážem*, jiné mnohem výrazněji: kupříkladu *lordstvo* najdeme v seznamu podle počtu typů na 23.-24. místě (stejně jako lemma *dem* má na levé straně 9 typů), kdežto podle hodnoty HHI „až“ na místě 107., a ještě markantnější rozdíl najdeme u lemmata *kdy*, které se podle HHI řadí na 60. místo, ale počet 804 typů na pozici L1 jej podle tohoto kritéria řadí až téměř na konec seznamu (18 562.-18 565. místo).

Tyto rozdíly jsou dány faktem, že HHI nepočítá pouze s množstvím typů na dané pozici či pozicích, ale taktéž s poměrem zastoupení těchto typů. To z něj také činí mnohem preciznější a tím i vhodnější kritérium pro konstatování monokolokability a nerovnoměrné kolokability vůbec. Pohled do tabulky naznačuje, že i zde se nevyhneme zkratkám typu *tab*; tečka za nimi je chápána jako typ a jelikož spojení s ní tvoří drtivou většinu výskytů, hodnota HHI je nutně vysoká. Něco podobného vidíme také u slov, která se používají k uvození vedlejší věty a vyskytují se tak v bezprostřední blízkosti čárky – to je důvodem, proč dosáhlo tak vysoké hodnoty lemma *kdy*. Podobných dat, nechceme-li je do svých výsledků zahrnout, se ovšem lze snadno zbavit přeskokováním interpunkce při výpočtech.

Následujícím krokem bylo porovnávání hodnot dosažených jednotlivými slovy s jejich kolokační charakteristikou zjistitelnou v korpusu. U každého prvku bylo zaznamenáno, s jakými kolokáty se na dané pozici vyskytuje, až do počtu 8 různých lemmat. Sledovali jsme také, zda se tyto kolokáty vyskytují vždy na jedné pozici nebo jsou pohyblivé. Pokud jsme našli jiné způsoby použití slova (např. s dalšími kolokáty nebo v úplně jiném typu konstrukce, jiném významu...), zapsali jsme jejich počet a v případě zvláštností i povahu. Na základě všech těchto údajů pak každé slovo bylo označeno za monokolokabilní nebo ne.

Zároveň s tím bylo ovšem nutné stanovit pevná kritéria, podle kterých by se dalo jednoznačně rozhodovat, zda je dané slovo monokolokabilní. Kolokabilita je totiž vlastnost škálovitého charakteru a tomu je třeba přizpůsobit i aplikaci definice monokolokability na reálná data. V řeči je možné se setkat s užitími ovlivněnými různou stylizací a hrou s jazykem, s metajazykovými užitími, s nejrůznějšími chybami nebo výrazy patřícími jen do určité části jazyka, zejm. slangů. Jednoduše řečeno, v jazyce je možné leccos, nás však v tomto případě zajímá to, co je ustálené (nebo se alespoň vyskytuje opakovaně) a patří do běžného úzu. Výzkumník pak stojí před řadou jedinečných rozhodnutí, ve kterých hraje

značnou roli jeho intuice. Z těchto důvodů jsme se rozhodli zavést plošná kritéria, díky nimž bylo sníženo riziko, že ovlivníme data nevhodnými rozhodnutími u jednotlivých případů či subjektivními názory, a která nám umožnila považovat za monokolokabilní i slova, která se skrze rozvrstvenost a tvárnost jazyka okrajově vyskytují i v jiných vazbách – takových je totiž většina.

Jako referenční slova byly přijaty kolokáty, které měly vůči klíčovému slovu syntagmatický vztah a vyskytovaly se v jeho sousedství nejčastěji, minimálně však dvakrát. Jejich počet byl omezen na 8 (tedy podle „volnější“ hranice odpovídající definici monokolokability). Zároveň jsme kladli důraz na to, aby zaznamenávané kolokáty byly pokud možno zařaditelné do téhož kolokačního paradigmatu nebo měly vůči klíčovému slovu podobný syntaktický či sémantický vztah. To bylo nutné u slov, která se běžně vážou s více slovy najednou (u různě rozvitých substantiv, sloves apod.). Kupříkladu slovo *zapotřebí* stálo téměř vždy v kombinaci se slovesem *být* či *mít*. Zároveň se vedle něj velmi často objevovalo zájmeno *ten*, to však má vůči *zapotřebí* jiný vztah než obě slovesa a v příkladech, kde se vyskytovalo, zároveň vedle stálo i jedno ze sloves. Jako referenční slova tak byla zaznamenána obě slovesa, zájmeno *ten* však nikoliv. U některých polysémních slov ovšem byly zaznamenány kolokáty různého charakteru, specifické pro jednotlivé významy. Polysémie však zpravidla souvisí s rozšířením kolokačních paradigmat slova, a tak je mezi monokolokabilními prvky poměrně vzácná. Jejím příkladem je adjektivum *uhličitý* (HHI na L1 0,7066), jehož kolokační paradigma se skládá ze dvou odlišných skupin substantiv a ty tvoří jeho různé významy:  $k_1$  {oxid, kysličník, kyselina} a  $k_2$  {koupel, voda, macerace, pramen, lázeň}.

Po vyčlenění výskytů prvku s jeho referenčním slovem či slovy pak zpravidla zbylo různě velké množství případů, kde byl použit buďto s jinými slovy spadajícími do téhož kolokačního paradigmatu, nebo ve zcela jiném typu vazby. Z důvodů uvedených výše jsem jako monokolokabilní přijala všechna slova, u kterých tato „jiná užití“ tvořila max. 3% všech výskytů. Tříprocentní hranice vznikla teprve v průběhu analýzy – při větším množství sledovaných slov se postupně vyrýsovala jako dostatečně volná, aby zahrnula jasné případy, i dostatečně přísná, aby vyloučila některé nevhodné případy, jako např. reflexiva, která mají sice málo používanou, ale nezanedbatelnou nereflexivní variantu (srov. *nacházející*, které je ze svých 448 výskytů v 17 případech, tj. 3,82%, nereflexivní a váže se pak zpravidla s akuzativem). Zpočátku jsem pracovala s volnější pětiprocentní hranicí a slova, která splňovala tuto, ale nikoliv tříprocentní hranici, byla posléze ze seznamu vyloučena.

Zvlášť byl ještě zaznamenán počet zjevných chyb – překlepů, vynechání působících negramaticky, záměn klíčového slova za jiné apod. S ostatními daty bylo nakládáno, jako by tyto chybné výskyty neexistovaly – tedy např. i tam, kde se zřejmě jednalo o překlep na zaznamenaném referenčním slově, nebyl chybný výskyt k tomuto referenčnímu slovu připočítán, a nebyly ani zahrnuty do výpočtu poměru „jiných užití“.

Výsledkem zde popsaného třídění byla tabulka zachycující kolokabilitu všech slov s vysokým HHI a vedle toho samostatný seznam slov přijatých jako monokolokabilní.

Naším cílem v této fázi výzkumu bylo zjistit, u jakých hodnot HHI se mezi výsledky vyskytují monokolokabilní slova ve výrazném poměru oproti ostatním a zda různým hodnotám HHI odpovídají určité „typické“ druhy monokolokability.

Odpověď na první otázku naznačuje tabulka 1. Zde vidíme počet slov ( $n$ ), která dosáhla hodnoty HHI v uvedeném intervalu, počet monokolokabilních slov mezi nimi ( $m$ ) a jejich poměr ( $m/n$ ). Uvedeny jsou údaje pro pozici L1 a R1 od klíčového slova. U nižších hodnot, kde již počet nalezených slov ztatečně narůstal, jsem pro větší názornost zmenšila zaznamenávané intervaly na polovinu.

Rozmezí HHI	Pozice L1			Pozice R1		
	$n$	$m$	Poměr $m/n$	$n$	$m$	Poměr $m/n$
1 - 0,9	19	18	0,947	19	13	0,684
0,8999 - 0,8	14	10	0,714	29	16	0,552
0,7999 - 0,7	25	6	0,240	38	16	0,421
0,6999 - 0,6	21	4	0,190	46	17	0,370
0,5999 - 0,55	15	1	0,067	31	9	0,290
0,5499 - 0,5	17	4	0,235	39	8	0,205
0,4999 - 0,45	33	6	0,182	44	7	0,159
0,4499 - 0,4	37	1	0,027	58	13	0,224

**Tabulka 1:** Poměr monokolokabilních slov podle hodnoty HHI na pozicích L1 a R1. Jako  $n$  označujeme celkové množství slov, která dosáhla hodnoty HHI v daném intervalu,  $m$  je počet monokolokabilních slov mezi nimi.

Již na první pohled jsou patrné dvě věci: zaprvé rychle klesající poměr monokolokabilních slov ve vysokých hodnotách a jeho „stabilizace“ o něco níže, kde však mají stále poměrně vysoké zastoupení, a zadruhé výrazný rozdíl mezi levou a pravou stranou.

Celkem pochopitelně je u hodnot 1-0,8 monokolokabilních slov většina. Jejich o něco nižší zastoupení vzhledem k pozici R1 je způsobeno tím, že nejvyšších hodnot dosahuje také několik zkratk, za nimiž následuje tečka označená v korpusu jako samostatná pozice. Od hodnot 0,7999 níže se pak již monokolokabilních slov objevuje mezi výsledky výrazně méně, klesání jejich poměrného počtu je však následně pozvolné a velmi pravděpodobně pokračuje dále tímto způsobem ještě v nižších hodnotách. Zároveň si však lze povšimnout, že toto klesání není rovnoměrné – vyskytují se v něm výkyvy, jako např. jen jediné monokolokabilní slovo u hodnot 0,5999-0,55 a 0,4499-0,4 na L1, což při výzkumu svádělo k předpokladu, že dále se už monokolokabilní slova vyskytnou jen ojediněle (nic takového se však nestalo, namátkovým průzkumem byla objevena hned 4 další monokolokabilní slova v intervalu 0,3999-0,35), a vedle toho téměř čtvrtina monokolokabilních slov u hodnot 0,5499-0,5.

Ve srovnání s na první pohled chaotickými údaji na levé straně má pozice R1 znatelně „hladší“ průběh. Možná je to způsobeno výrazně vyšším počtem prvků, čímž se podobné výkyvy mohou navzájem vyrušit, nebo jde jen o náhodu. Faktem je, že s výjimkou úplně nejvyšších hodnot HHI najdeme na pravé straně zhruba dvojnásobek příkladů a také poměr monokolokabilních slov mezi nimi je o něco vyšší – zdá se, že klesá pomaleji. Přitom analýza variability kontextu, kterou provedl ve své práci V. Cvrček (v tisku) naopak ukazuje, že v obecném měřítku je méně variabilní levá pozice. Tento nepoměr je ovšem možné vysvětlit naším prvotním zaměřením na plnovýznamová slova; na levou stranu jsou nejčastěji orientována substantiva, pokud v jejich kombinacích převažují určité předložky či přídavná jména. Na pravé straně najdeme především přídavná jména, která se vážou jen na některá substantiva, a slovesa, která vyžadují doplnění přes předložku. Roli také jistě hrají reflexiva tantum, ať už adjektivní či slovesná, u nichž se zvrtné zájmeno nejčastěji vyskytuje bezprostředně napravo. V mé klauzurní práci (Šíková Hrejsemnová, 2011) reflexiva tantum sama o sobě tvořila většinu získaných monokolokabilních slov. Za nepoměrem v počtu případů tak dost možná stojí právě odlišné slovnědruhové složení.

Záznamy o kolokabilitě jednotlivých prvků ukazují, že s hodnotou HHI se mírně mění i složení tabulky co do typu monokolokabilních slov. Mezi hodnotami 1 a 0,8 je velké zastoupení absolutně monokolokabilních prvků, tedy těch, které mají jen jedno referenční slovo. Je-li referenčních slov více, jedno z nich tvoří drtivou většinu výskytů a ostatní se opakují jen několikrát. Zároveň jde o slova, která neumožňují pohyb referenčního prvku – tedy především frazeologická spojení neproměnlivého charakteru (*na*

*odiv, do třetice, křížem krážem*), adjektivní reflexiva tantum (*týkající se, odehrávající se*), víceslovné předložky (*na úkor*) a spojení monokolokabilního relačního adjektiva se substantivem (*nadmořská výška, bobkový list*). Uvedené vlastnosti pochopitelně úzce souvisí s vysokou hodnotou HHI. Jak již bylo uvedeno, způsob výpočtu HHI má za následek, že nejvyšších hodnot dosahují pozice, kde se pravděpodobnost jednoho z prvků blíží jedné (bez ohledu na to, zda se ve zbytku případů opakuje jen několik slov nebo má každý další výskyt jiný prvek), a to také znamená, že se převažující prvek nesmí přesunovat jinam, protože sledovanou pozici by pak obsadil jiný.

Kolem hodnot 0,8 již převažují případy s více referenčními slovy a rychle (zejm. na levé straně) přibývá slov, která sice vykazují silnou preferenci kombinace s určitým prvkem, ale nelze je považovat za monokolokabilní. Samostatnou kapitolu tvoří slova vázaná na interpunkci a spojky, jako např. tázací, resp. vztažná zájmena (*kdy, kde* – obojí o hodnotách těsně pod 0,7), slova užívaná zpravidla jako oslovení (*mylady*) a slovesa uvozující přímou řeč (*odtušit, odseknout, podotknout* – od 0,65 níže, objevují se průběžně). Zatímco adjektivní reflexiva tantum se hromadí v horní části seznamu a v nižších hodnotách se s nimi setkáváme méně, u slovesných je tomu naopak – slovesa umožňují zvratnému zájmenu mnohem více pohybu, a tak se s nimi setkáváme spíše na nižších hodnotách (nejvýše se umístilo *chýlit se* s HHI 0,6711, u hodnot kolem 0,5 a níže začínají tvořit výraznou část monokolokabilních případů). Snad jediným typem, na který narážíme průběžně bez výrazných změn v zastoupení, jsou různé způsoby vázanosti na předložky (zpravidla na L1 u substantiv a na R1 u sloves) a na pravé straně adjektiva rozvíjející jen jedno několik substantiv.

Kromě levé a pravé strany byl vytvořen také seznam pro prvky s vysokým HHI počítaným pro obě pozice najednou. Dosažené hodnoty zde byly nutně o něco nižší; u oněch několika málo slov, která vykazují monokolokabilní charakter pro obě pozice, se pohybovaly kolem 0,4. Jako příklad jmenujme třeba *{s, bez, po, při} přihlédnutím {k, na}*. Jen o kousek dále (kolem hodnoty 0,3 a níže) již tabulka udržovala poměrně stálou podobu – vyskytovala se zde jednak slova, u nichž na obou stranách převažoval jeden prvek, nebyla však monokolokabilní vůči žádnému z nich (jako např. *morek*, u nějž převažuje užití v kolokaci *do morku kostí*, ale cca 17% jeho výskytů je jiného typu) a jednak slova monokolokabilní jen na jednu stranu, což se však samo o sobě podepsalo na výši HHI. Poměr obou těchto skupin se až do hodnoty 0,19 (výsledky o nižších hodnotách jsem procházela jen namátkově) pohyboval poměrně zešíroka kolem 0,5 a nevykazoval žádný jasně klesavý trend. Ukázalo se navíc, že všechna slova zde zachycená se již objevila



v seznamech pro jednotlivé strany. Krom vizuálního přiblížení výsledků tak výzkum slučující obě strany nepřinesl nic zásadně nového a v dalších fázích byl tento krok již vypuštěn.

### 2.3.4 Srovnání s entropií

Jak již bylo řečeno, na podobném principu jako u HHI lze monokolokabilní jednotky vyhledávat také s pomocí výpočtu entropie. V jejím případě výrazně homogennímu okolí odpovídají naopak co nejnižší hodnoty – tedy prvky, které mají na určité pozici HHI o výši 1, mají pro tutéž pozici nulovou entropii. Tabulky se slovy, která dosáhla nejnižších hodnot, uvádím v příloze 7a pro pozici L1 a 7b pro pozici R1. Již letmé srovnání s odpovídajícími tabulkami pro HHI (přílohy 1d a 1e) naznačují, že výsledky jsou velmi podobné: seznamy nejsou přesně stejné, obsahují však tytéž jednotky jen v mírně přeházeném pořadí. S klesajícím HHI, resp. rostoucí entropií, se ovšem rozdíly v relativním pořadí postupně zvětšují. Na pozici L1 námi zkoumaná slova o HHI mezi 1 a 0,4 dosahovala hodnot entropie 0 až 3,5. V tomto intervalu se postupně objevovala i další slova, ovšem srovnatelného typu. Monokolokabilní slova v tomto seznamu nevykazovala žádné výrazně odlišné uspořádání – stejně jako v seznamu řazeném podle hodnot HHI se hromadila na nejvyšších místech seznamu a pomalu, podobným tempem jako u HHI, řídla. U seznamů pro pozici R1 je situace obdobná. Z tohoto usuzujeme, že oba způsoby vyhledávání monokolokability jsou v zásadě rovnocenné a přinášejí srovnatelné výsledky.

Tento předpoklad můžeme ověřit pomocí korelační analýzy pořadí jednotlivých prvků v seznamech seřazených podle HHI a podle entropie. Pro tento účel je vhodným nástrojem Spearmanův korelační koeficient ( $r_{sp}$ ). Pro 20 prvků o nejnižší hodnotě entropie, kde jsou patrné jen drobné odchylky v pořadí týchž jednotek zarovnaných podle HHI, byly zjištěny následující hodnoty:

- $r_{sp} = 0,964$  pro pozici L1 ( $n = 20, \Sigma d_i^2 = 48$ )<sup>25</sup>
- $r_{sp} = 0,658$  pro pozici R1 ( $n = 20, \Sigma d_i^2 = 455$ )

Oba výsledky přesahují kritickou hodnotu Spearmanova koeficientu pro tento počet prvků, která činí 0,445 pro hladinu statistické významnosti  $p = 0,05$  a 0,568 pro  $p = 0,01$ . Můžeme proto tvrdit, že pořadí dotyčných prvků vzájemně koreluje. Nápadně nižší

---

<sup>25</sup> V tomto zjednodušeném zápisu  $n$  označuje počet sledovaných prvků a  $\Sigma d_i^2$  součet druhých mocnin rozdílů mezi pořadím každého prvku v obou seznamech.

hodnota pro pozici P1 je důsledkem relativně velkého rozdílu v pořadí u jednoho prvku (jednalo se o lemma *kutný*, rozdíl v pořadí byl 20 pozic). Při výpočtu korelačního koeficientu pro celý soubor dat docházíme k následujícím výsledkům:

- $r_{sp} = 0,724$  pro pozici L1 ( $n = 21\ 384$ ,  $\Sigma d_i^2 = 450\ 250\ 475\ 230$ )
- $r_{sp} = 0,877$  pro pozici R1 ( $n = 21\ 384$ ,  $\Sigma d_i^2 = 199\ 728\ 749\ 946$ )

Jak je vidět, hodnoty  $r_{sp}$  jsou nápadně vysoké i pro celý seznam lemmat. Předpoklad, že výsledky obou způsobů kvantifikace diversity jsou navzájem srovnatelné, se tak potvrzuje. Pro analýzu tedy můžeme použít jen jeden z nich, aniž bychom se museli obávat významnějších ztrát dat, a v tom případě se jako vhodnější jeví ten, který je snáze interpretovatelný, čili index HHI.

### 2.3.5 Slovní tvary

Předpokládáme, že monokolokabilita může být buďto vlastností lexému jako takového, nebo jen některé z jeho forem, jde-li o ohebné slovo. V tomto druhém případě však nebude v seznamu lemmat zachycena, protože formy s běžnou kolokabilitou omezení vztahující se k jedné, která je monokolokabilní, „zamaskují“. Výjimkou jsou samozřejmě ty vzácné případy, kdy je monokolokabilní forma zároveň zdaleka nejčastější či jedinou používanou – jako např. tvar *holičkách*, který spadá (v korpusu) pod lemma *holička*; jiné tvary však najdeme leda ojedinele, jako v příkladu (3), a taková použití můžeme v zásadě považovat za hru s jazykem či obecně za okrajová a potenciálně zanedbatelná. Skrze takovou zvláštní distribuci lze také pochybovat o tom, zda vůbec u podobných slov můžeme uvažovat o pravidelném skloňování či na jeho základě určit coby základní nějaký tvar, který se v úzu nevyskytuje (srov. Čermák – Blatná, 1995, s. 118).

- (3) Pokud Slovanstva se týče, tato bludička v bažinách dějin našeho národu zavedla mnoho našich duchů, když ne do zkázy, tož na holičky.

V případech, jako je zmíněné *holičkách*, přinese vyhledávání podle lemmatu a podle slovního tvaru stejné či velmi podobné výsledky. Pro odhalení slov, která běžně využívají více svých možných tvarů a jen jeden nebo část z nich je monokolokabilní, si však již s lemmatem nevystačíme. Proto jsme vytvořili zvláštní seznamy také pro vyhledávání podle jednotlivých slovních tvarů. Pro tyto tvary platila stejná pravidla jako předtím pro

lemmata – musely dosáhnout absolutní frekvence 200 nebo vyšší a byly pro ně zaznamenány stejné statistické údaje (srov. oddíl 2.3.1 v této práci). Výpočet těchto údajů byl přitom proveden ve vztahu k jednotkám téhož typu, totiž ke slovním tvarům (tedy např. u počtu typů na dané pozici nebyl udáván počet lemmat, jako tomu bylo v seznamu pro lemmata, ale počet různých slovních tvarů). To pak mělo pochopitelně vliv na zjištěné hodnoty, které pro jednotlivé tvary týchž slov zdánlivě poukazovaly na širší kolokabilitu, než jaká byla zjištěna u lemmat. Při ručním třídění však již byla referenční slova zaznamenávána jako lemmata.

Ukázalo se, že nejvyšších hodnot HHI dosáhly tvary týchž slov, se kterými jsme se setkali již v seznamu lemmat. Zároveň zde převažovala slova neohebná nebo se vyskytující z drtivé většiny jen v jednom tvaru, jako např. *odiv*, *rozmyšlenou*, *úkor* nebo již zmíněné *holičkách* (vše monokolokabilní vůči *na*). Na podobné případy ovšem průběžně narážíme i dále (např. *krážem* o HHI pro L1 0,6973 či *opírající* o HHI pro R1 0,8491), mezi nimi však postupně přibývají jiné. Nacházíme zde monokolokabilní tvary slov, která v seznamu lemmat jako monokolokabilní přijata nebyla kvůli vysokému podílu „jiných užití“ nebo vůbec nedosáhla tak vysokých hodnot HHI, aby se dostala do ručně tříděné skupiny. U slov tohoto druhého typu jde většinou o monokolokabilitu vůči gramatickému prvku (nejčastěji předložce). Jde tedy o propojení omezení gramatického, kdy lze např. určitý pád kombinovat jen s některými předložkami, sémantického, které může některé z těchto předložek vyloučit, protože by spojení s nimi nedávalo smysl, a konvenčního, kdy se ze zbývajících možných předložek v běžném úzu používá jen jedna nebo několik jednoduše na základě ustálenosti takového spojení. Může se přitom jednat o slova, u nichž by nás monokolokabilitu ani nenapadlo hledat. Kupříkladu forma *předprodeji* se kombinuje jen s předložkou *v* a okrajově ještě s *k*. Pro pozici L1 má HHI 0,8060. Odpovídající lemma, *předprodej*, však má pro tutéž pozici HHI jen 0,2580 – takové hodnota je sice nadprůměrná, vypovídá však spíše obecně o nesouměrné či nepřilíš bohaté kolokabilitě než o potenciální monokolokabilitě. Pro tvar *předprodeji* si přitom snadno lze představit i kombinace s jinými předložkami (např. *po*, *při*...), ty se však v korpusu nevyskytují nebo se ani jednou neopakují. Jako další obdobné příklady jmenujme formy *vikendech* či *záchodě*. Zvláštním typem takových „gramaticky“ orientovaných monokolokabilních forem jsou specifické tvary např. pro duál (*modrýma*) nebo vokativ (*kamaráde*).

Dalším zajímavým poznatkem je, že u ohebných monokolokabilních slov máme většinou v seznamu slovních tvarů záznam jen o jednom nebo dvou tvarech. Důvodem je především relativně malá četnost těchto slov; pokud klíčové lemma není v korpusu

zastoupeno minimálně několika stovkami, lépe však několika tisíci výskytů, limit minimální absolutní frekvence může přesáhnout jen málo jeho tvarů. K nerovnoměrnému zastoupení různých tvarů však docházelo i u některých slov frekventovaných. V tom případě již může jít o jev přímo související s monokolokabilitou. Spojuje-li se např. adjektivum jen se substantivy určitého rodu, tvary ostatních rodů se budou vyskytovat vzácně nebo vůbec. A podobně pokud ve výsledcích tvoří většinu výskytů spojení s jedním kolokátem, který má zároveň vliv na tvar klíčového slova, bude tento tvar taktéž převažovat. Srov. např. monokolokabilní formu {k, v, o} *dispozici*, která představuje 9497 případů z celkových 10282 výskytů lemmatu *dispozice*. Požadavek minimální absolutní frekvence vedle ní překračuje pouze forma *dispozice*, zastoupená 450 případy.

Podstatný je ovšem především fakt, že vyhledávání podle slovních tvarů přineslo výsledky velmi odlišné od těch, které jsme získali vyhledáváním podle lemmat. Přináší jednak podrobnější informace o slovech, jež byla již zahrnuta v seznamu lemmat – skrze přítomnost či nepřítomnost (tedy četnost) různých forem a případných odlišností v jejich kolokabilitě. Zároveň nám umožňují najít i jiný druh monokolokability – ten, který je vázán jen na určitý tvar slova, ne na lexém jako takový. V tomto směru je možné tímto způsobem získané výsledky považovat za bohatší. Na druhou stranu však nelze lemmata opomenout úplně – jen na základě jednoho nebo dvou tvarů totiž nemůžeme dělat závěry o kolokabilitě lexému, a používáme-li hranici minimální absolutní frekvence, mohli bychom také o řadu monokolokabilních slov přijít, jelikož jejich jednotlivé tvary takovou hranici nemusí překročit, i kdyby celý lexém všechny podmínky kolokability i četnosti splňoval.

## 2.4 Shrnutí

V této části jsme uvažovali nad různými způsoby, jak v korpusu vyhledávat monokolokabilní jednotky, a s cílem vyhledat jich co nejvíce a s co nejširší škálou jejich typů jsme vybrali přístup využívající nízkou informativní hodnotu referenčního slova, resp. diversitu prvků na pozicích, kde by se referenční slovo mělo nacházet. Jako nejvhodnější se jevily dva prostředky, které právě diversitu na dané pozici kvantifikují – entropie a Herfindahl-Hirschmanův index.

V korpusu SYN2010 jsme vyhledali všechna plnovýznamová slova o absolutní frekvenci min. 200 a zaznamenali k nim řadu statistických údajů včetně hodnot obou indexů dosažených na pozicích L1 a R1. Následně bylo provedeno ruční třídění výsledků, které hodnotou indexu naznačovaly velmi výrazné omezení kolokability. Jako primární byl

použit HHI. Zvláště přitom byla zkoumána lemmata a slovní tvary. V obou případech jsme konstatovali, že poměr monokolokabilních slov mezi výsledky s klesající hodnotou HHI postupně slábne od drtivé většiny v nejvyšších hodnotách po relativně malé zastoupení na hodnotách kolem 0,4 a nižších. Zjistili jsme také, že vedle tohoto trendu lze mezi výsledky vysledovat také určitou koncentraci jednotlivých typů monokolokabilních slov v různých pásmech dosažených hodnot HHI. Naopak výsledky získané výpočtem HHI pro obě pozice dohromady nepřinesly prakticky žádné zásadní obohacení a můžeme je proto v následujících fázích vypustit. Podobně srovnání s výsledky entropie ukázalo, že seznamy vzniklé zarovnáním podle hodnot obou indexů jsou si velmi podobné; obě metody tedy lze považovat za srovnatelné a použijeme-li ve výzkumu jednu z nich, můžeme předpokládat, že druhá by nepřinesla mnoho dalších relevantních výsledků.

Porovnání seznamů získaných z vyhledávání podle lemmat a slovních tvarů potvrdilo naše očekávání, že oba způsoby, byť se v některých případech překrývají, přinášejí každý jiný typ výsledků. S oběma jsou pochopitelně spojeny výhody i nevýhody, u slovních tvarů navíc častější homonymie způsobuje určitou nevyrovnanost mezi jednotlivými formami a tím i jejich kolokabilitou. Domníváme se však, že oba způsoby jsou stejně užitečné a pro každý obsáhlý výzkum je vhodné použít oba.

V neposlední řadě byla v této fázi zpřesněna kritéria pro vymezení monokolokabilních slov s ohledem na povahu skutečných korpusových dat. Zatímco výchozí definice operuje pouze s konkrétním počtem možných denotátů, úzus představuje podstatně komplexnější situaci. Pro rozhodování o monokolokabilitě proto byly zavedeny požadavky na opakování kolokátů, které mohou být označeny za referenční slova, a na maximální poměr jiných způsobů použití, které tak mohou být označeny za okrajové. Počet přijatelných kolokátů byl zachován, ačkoliv jde o čistě arbitrární rozhodnutí – škálovitost monokolokability je však v tomto přístupu částečně kompenzována podílem „jiných užití“, která mohou zahrnovat také další, méně využívané kolokáty. S ohledem na tato kritéria můžeme formulovat pracovní definici monokolokability použitou v této studii jako takové kolokability dané jednotky, která je minimálně v 97% svých výskytů omezena na možnost kombinovat se s jedním až 8 prvky, přičemž kolokace s těmito prvky se v dostupných datech vyskytují opakovaně. Přehled monokolokabilních výsledků je uveden v příloze 2.

Poznatky, ke kterým jsme dospěli, i vlastní seznamy získaných monokolokabilních slov a záznamy o jejich kolokátech, využijeme v následujících částech výzkumu, které spočívají v další analýze a popisu již získaných dat na češtině a následně v pokusu podobným způsobem alespoň zhruba zmapovat situaci v dalších jazycích.

### 3. Monokolokabilita a čeština

Jak jsem již naznačila výše, studie přímo zaměřené na monokolokabilitu v českém jazyce jsou vzácné. V pracích Stojčeva a Káninského i ve většině studií, které na monokolokabilitu narážejí jen jako na dílčí problém v rámci širšího tématu, jako je např. článek Mileny Hnátkové (2006), přitom bývá prakticky bez výhrad přijímán přístup Čermákův. Výjimku v tomto směru tvoří můj výzkum z roku 2011, který si vzal za cíl pojmout monokolokabilitu v rámci jiného, širšího pohledu a zaměřit se právě na to, co tradiční přístup přechází. V tomto přístupu pokračuje také tato práce, nejedná se však o pouhou replikaci téhož výzkumu na bohatších datech. Pokusila jsem se naopak opět posunout a dále rozšířit svůj pohled na monokolokabilitu a její projevy, a to jak skrze použitou metodu vyhledávání v korpusu, tak i při analýze a úvahách nad získanými daty.

V předchozí kapitole jsem s cílem získání co nejbohatší škály dat vyzkoušela metodu vyhledávání monokolokabilních slov v korpusu pomocí entropie a především HHI. Využití obou indexů se ukázalo, byť pochopitelně s jistými výhradami, jako efektivní i praktické. Seznam monokolokabilních slov s údaji o jejich monokolokabilitě, získaný skrze výpočet HHI a ruční třídění dat, nám poslouží také v následující fázi výzkumu. Na základě další analýzy této databáze, spočívající především v porovnávání jednotlivých položek seznamu a sledování jejich vlastností, se v této části pokusím podat podrobnější popis monokolokabilních jednotek z různých hledisek a také ve vztahu k přístupu a popisu ve své předchozí práci i v pracích jiných autorů.

Nejprve vytvořím návrh takové klasifikace monokolokabilních slov, která by sledovala několik různých kritérií a zahrnovala i druhy monokolokability v datech dřívějších prací nezastoupené nebo opomíjené. Poté s ohledem na tuto klasifikaci přednesu bližší popis monokolokabilních slov a porovnáím svá zjištění s popisy uvedenými ve starších pracích, zejména Stojčevově. Vedle těchto obecnějších vlastností se pozastavím také nad některými zvláštními případy na hranici monokolokability a jiného typu spojení a nad hranicí monokolokability jako takovou. Na závěr zmíním postavení monokolokabilních slov v praxi, a to jak ve vlastním úzu, tak při aplikaci vědeckého zkoumání a teorií.

### 3.1 Typy monokolokability

Ve chvíli, kdy jsme vyčlenili z korpusu skupinu slov, které lze označit za monokolokabilní, se nabízí otázka, zda je možné tuto skupinu dále smysluplně dělit, čili zda se mezi daty vyskytují nějaké opakující se vlastnosti či typy odlišné od jiných opakujících se vlastností. Již na první pohled je jasné – a vyplývá to také již z metody použité pro vyčlenění monokolokabilních jednotek – že ano: v naší databázi se očividně vyskytují různé druhy slov a různé typy vztahů k referenčnímu slovu. Lze tedy mluvit o různých podobách monokolokability. K základní klasifikaci jsem přitom dospěla již dříve ve své klauzurní práci, která ovšem pracovala s poněkud chudšími daty. Jak již bylo zmíněno v úvodu (oddíl 1.2), vyčlenila jsem tehdy mezi svými výsledky reflexiva tantum, slova valenčně vázaná, frazémy a víceslovná pojmenování. Taková klasifikace jistě poslouží k hrubé orientaci, nevypovídá však podrobněji o dalších vlastnostech zařazených slov, jako je jejich ohebnost či šíře a struktura kolokačního paradigmatu. Jelikož tehdy použitá data obsahovala převážně slova o jediném referenčním prvku a jiné případy se vyskytovaly jen vzácně, a také vzhledem k úplnému vynechání slov monokolokabilních jen v jednom nebo některých svých tvarech, pro tyto opomíjené typy pravděpodobně bude nutné zavést zvláštní kategorie či klasifikaci od základu přepracovat. Se současnou databází si ovšem můžeme dovolit podstatně komplexnější náhled na situaci a zhostit se tohoto úkolu. Za účelem co nejúplnějšího rozčlenění tedy vytvořím hned několik samostatných a poměrně jednoduchých třídění podle různých hledisek, týkajících se především povahy monokolokabilního prvku a jeho referenčního slova nebo slov. V průběhu analýz se ukázalo, že předchozí klasifikace dokáže obsáhnout většinu zaznamenaných případů (výjimku tvoří především substantiva monokolokabilní vůči předložce a slova monokolokabilní jen v některém tvaru, která by si zasloužila vlastní kategorii); přináší však odlišný, poněkud zjednodušující náhled na situaci a s následujícím tříděním se pouze volně prolíná.

#### 3.1.1 Klasifikace podle počtu referenčních slov a kolokačních paradigmat

První kritérium, které použijeme, je relativně jednoduché: jedná se o to, s kolika různými a popř. s jakými slovy se monokolokabilní slovo váže. Má-li jen jeden možný kolokát, jako je tomu např. u reflexiv tantum, mluvíme o absolutní monokolokabilitě (např. *polohrubá mouka, vztahující se, jakž takž*). Právě tento typ tvořil v klauzurní práci většinu

zpracovávaných dat. V naší současné databázi je těchto případů stále velké množství, něco přes jednu třetinu. Je to pravděpodobně opět způsobeno metodou vyhledávání, jelikož hodnota HHI, jak již bylo poukázáno, je nejvyšší u pozic, kde jeden prvek v četnosti výrazně převažuje nad ostatními. Do tohoto typu byla zařazena i slova, která se příležitostně vyskytla s jinými kolokáty (až do poměru 3 % celkového počtu výskytů, srov. oddíl 2.3.3 této práce), z nichž se však žádný neopakoval.

Jen o trochu méně často narážíme na případy, kdy je referenčních slov několik, a to buďto přímo synonymních, nebo s obdobným vztahem vůči monokolokabilnímu slovu. Zpravidla jedno z nich opět výrazně převažuje a ostatní se jeví jako jeho obměny – buďto přímo slova s podobným významem, nebo taková, která použijeme v případě, kdy se monokolokabilní slovo vyskytne v jiném typu konstrukce, nebo např. různé předložky. Ukažme si toto chování na konkrétních příkladech. U slova *tržný* (o absolutní frekvenci 224) najdeme 4 opakující se kolokáty: *rána*, *ranka*, *poranění* a *zranění*. Zdaleka nejčastější je přitom *rána*, která se objevuje ve 203 případech. Ostatní kolokáty mají každý jen několik výskytů v této kombinaci. Všechny jsou zároveň synonymní se slovem *rána* a tvoří spolu s ním jedno kolokační a sémantické paradigma; skrze to na ně můžeme nahlížet jako na přípustné varianty „hlavního“ referenčního slova (*rána*). Příkladem druhého typu je *zapotřebí*, které se kombinuje se slovesy *být* (2859 výskytů) a *mít* (335 výskytů). Opět zde vidíme velký rozdíl v poměru zastoupení obou kolokátů. Slovesa *být* a *mít* ovšem můžeme těžko považovat za synonyma jako substantiva v předchozím příkladu. Zároveň však jsou si svým vztahem k monokolokabilnímu slovu velmi podobná; konstrukce s nimi můžeme v zásadě označit za transformační varianty, v nichž si jednotlivé aktanty pouze vyměňují pozici a formu. Konečně u monokolokabilních prvků, jejichž referenčním slovem je předložka, můžeme mnohdy najít i další předložky, s nimiž se kombinují – tak je tomu i u slova *střeh*, které se kombinuje s předložkou *v* (488 užití), ale také s *do* (17 užití). V řadě případů však spojení s jednou z předložek (tou nejvíce zastoupenou) mívá, přinejmenším v části výskytů, jiný, posunutý význam, odlišený více než jen působením samotné předložky. Takové příklady bychom již radili spíše do následující skupiny.

Tu tvoří monokolokabilní slova, která mají více referenčních slov a mezi nimi lze navíc najít skupiny s odlišným významem nebo vztahem ke klíčovému slovu – tj. jedná se o monokolokabilní slova polysémní, se strukturovaným kolokačním paradigmatickým. V něm je možné vyčlenit odlišné části či dílčí paradigmata, jejichž prvky se od ostatních částí odlišují a vymezují tak různé významy celého spojení či přímo monokolokabilního slova. Navzdory našemu očekávání se nejedná o malé množství okrajových případů, tato skupina



je poměrně silně zastoupena. Hranice mezi ní a předchozím typem je ovšem pozvolná – u některých případech, kdy jsou rozdíly ve významu velmi malé nebo je význam nejasný, záleží do jisté míry na jemnosti analýzy. Existují pochopitelně zcela jasné příklady, jako je již zmíněné *uhličitý*, kombinující se se dvěma velmi odlišnými skupinami prvků (díličními kolokačními paradigmaty):  $k_1$  {oxid, kysličník, kyselina}, kde lze monokolokabilní adjektivum synonymně nahradit nepřímým přívlastkem *čtyřmocného uhlíku*, a  $k_2$  {koupelel, voda, macerace, pramen, lázeň}, kde se jako substitut za monokolokabilní prvek hodí *obsahující CO<sub>2</sub>*. V podobných situacích se s odlišením jednotlivých významů můžeme setkat i v rámci popisu odpovídajícího slovníkového hesla. V řadě případů, kdy kolokační paradigma tvoří odlišné prvky, je ovšem případný rozdíl ve významu monokolokabilního prvku hůře postižitelný, srov. např. *visací* {zámek, náušnice, lampa, hodiny}. Malé množství možných kolokátů také ztěžuje či přímo znemožňuje podrobnou analýzu. S ohledem na tyto obtíže se jako nejméně konfliktní řešení jeví považovat za strukturovaná kolokační paradigmatata všechna ta, jejichž prvky nejsou navzájem synonymní a nejde ani o transformační varianty jako v příkladech uvedených u předchozího typu, a kde spojení s jednotlivými referenčními slovy mají odlišný význam.

Konečně čtvrtým a nejvzácnějším typem je vícenásobná monokolokabilita, kdy je klíčové slovo monokolokabilní vůči dvěma nebo i více kolokačním paradigmatům zároveň. Kromě již dříve zmíněného (*nechat*) (*na holičkách*) můžeme jako příklad uvést (*s přihlédnutím (k)*), u něhož se referenční slova nacházejí pravidelně na pozicích L1 – paradigma  $K_1$  {s, bez, při, po} a R2 – paradigma  $K_2$  {k, na}. V naší databázi někdy druhé paradigma tvořila interpunkce či jiné nealfanumerické znaky, v podobných případech je však sporné, zda lze vůbec mluvit o monokolokabilitě – srov. oddíl 3.3.5 v této práci. Takovým příkladem může být zkratka *překl.*, která se na pozici L1 kombinuje se zkratkou *pozn.* a na R1 se znaky „)“, „:“, „]“ a tečkou.

Vedle těchto skupin můžeme také rozlišovat spojení, kde se monokolokabilní slovo může vázat jen a pouze se svým referenčním slovem (slovy) a nemá žádné jiné vazby, a oproti tomu spojení, kde má zároveň vztahy i s jinými slovy. Do první skupiny patří především adjektiva a adverbia, která stojí jako přívlastek, resp. příslovečné určení vůči referenčnímu slovu (substantivu, resp. adjektivu či adverbium), jako je tomu např. ve spojeních *nadmořská (výška)* nebo *zbrusu* {nový, nově, čerstvý}, a zařadili bychom sem pravděpodobně i ta spojení, kde vztah mezi jednotlivými prvky není zcela jasný a nelze je proto ani hypoteticky oddělit, jako např. u oboustranně monokolokabilního *křížem krážem*.

Do druhé skupiny pak patří všechna ostatní spojení, zvláště ta, kde je monokolokabilním prvkem sloveso (rozvíjené zpravidla dalšími argumenty a doplněními).

### 3.1.2 Klasifikace podle druhu referenčního slova

Kromě jejich počtu můžeme u referenčních slov také sledovat, o jaký slovní druh jde. Toto kritérium nám především rozdělí monokolokabilní spojení na ta, kde je referenční prvek gramatické povahy, a na spojení dvou plnovýznamových slov. Zvláštním případem jsou slova vážící se k interpunkci, v zásadě je ale můžeme přiřadit k referenčním slovům gramatickým.

U skupiny, kde referenční prvek patří k neplnovýznamovým slovním druhům, můžeme dále rozlišovat, zda se jedná o předložku, zájmeno či spojku. Ke spojkám lze počítat i interpunkci. V rámci naší databáze tvořila největší část monokolokabilní slova s předložkou jako referenčním prvkem. Většinou šlo o substantiva nebo slova substantivům podobná (např. *potaz, třetice, neurčito, holičkách*), v menší míře však také o valenčně vázaná slovesa (*přirovnat, zalapat*) a adjektiva (*sestávající, považovaný*). Zájmeno bylo referenčním slovem téměř výlučně pro reflexiva tantum, ať už slovesná nebo adjektivní (*postesknout, otázat, lesknoucí*); výjimku tvoří *lordstvo*, které se obligatorně pojí s přivlastňovacími zájmeny (*jeho, váš a jejich*). Třetí skupina, kde je referenčním prvkem spojka nebo interpunkce, byla malá a i tak poměrně heterogenní. Patřily sem příklady jako *(a) pod, onak* {či, nebo, ani, i, podruhé} nebo *načež* (na L1 vždy stojí interpunkční znaménko). K monokolokabilitě vůči interpunkci se ještě vrátíme v oddílu 3.3.5.

Druhá skupina, kde je referenčním prvkem plnovýznamové slovo, byla co do počtu nalezených příkladů přibližně stejně velká jako první. Drtivou většinu v ní tvoří slova monokolokabilní vůči substantivu. Jedná se převážně o přídavná jména (relačního typu), jako např. *pravdivostní, pokutový* nebo *vyměřovací*, a několik sloves (*pokývat, přimhouřit* apod.). Pro slovesa v této skupině je charakteristické, že jejich referenční slova označují části těla. Zároveň právě monokolokabilní spojení se substantivem coby referenčním prvkem tvoří v celé databázi většinu případů, kdy je referenčních slov více než tři. Referenčním slovem může být také adjektivum, sloveso nebo adverbium. Od každého jsem však našla jen několik případů. Navíc se mnohdy jedná o slova, která se mohou kombinovat s prvky patřícími k různým slovním druhům a tvoří tak transformační varianty téhož výrazu nebo různé významy. Např. *nasnadě* má jako referenční sloveso *být* a jemu příbuzné adjektivum *jsoucí*, podobně *zbrusu* se kombinuje s adjektivy *nový* a *čerstvý* a s

adverbiem *nově*. Příkladem jiného způsobu použití v souvislosti s jiným druhem referenčního slova může být *povšimnutí*, které se vyskytuje v kolokacích (*bez povšimnutí*, *(za) povšimnutí*, ale také *(hodný) povšimnutí*. Poslední podskupinu tvoří referenční slova, u nichž je z nějakého důvodu určení slovního druhu obtížné nebo nejednoznačné. Sem patří např. cizojazyčné výrazy, jako v případě {de, ipso, post} *facto*, referenční slova, jež jsou sama monokolokabilní (typickým příkladem je dvojice *křížem krážem*, týká se to ovšem také cizojazyčných slovních spojení), popř. jiné nejasné případy.

Pochopitelně není vyloučeno, že se monokolokabilní slovo může vázat s referenčními slovy různého typu. Takovými příklady jsou výše zmíněné *nasnadě*, *zbrusu* či *povšimnutí*. Podobná víceznačnost je ovšem poměrně vzácná a při výběru monokolokabilních slov z korpusu jsme se jí vyhýbali; s výjimkou případů, kdy se jedná o blízké příbuzná slova (jako *nový* – *nově*, *být* – *jsoucí*), by totiž mohlo dojít k zanesení dat nerelevantními výsledky. Za přijatelné byly proto považovány pouze situace, kdy po vyčlenění výsledků s jedním typem referenčního slova zbytek dat vykazoval výraznou preferenci jiného typu spojení (ostatní typy nesměly překročit 3% všech výsledků, srov. oddíl 2.3.3) a kdy se ukázalo, že dané klíčové slovo se nemůže kombinovat s prvky obou typů zároveň. Jako příklad, kdy se zároveň nejednalo o slova přímo příbuzná, uveďme sloveso *zavrtět*, které má několik substantivních referenčních slov {hlava, ocas, ocásek, zadek, hlavička, rameno, palec}, jež ve vztahu k němu fungují jako předmět, ale vedle toho se vyskytuje i v reflexivní podobě (*zavrtět se*, popř. *zavrtět sebou*). Zde vidíme značnou podobnost ve vztahu mezi oběma členy spojení. Zcela odlišnou roli pak hrají referenční slova u výše zmíněného *povšimnutí*. Mnohem častější jsou ovšem případy vázání s různými slovními druhy u slov vícenásobně monokolokabilních. Vzhledem k různorodosti podobných příkladů a tomu, že se jedná pouze o kombinaci již existujících typů, se zdá vhodnější je všechny zařadit do více kategorií podle toho, s jakými všemi referenčními slovy se mohou kombinovat, spíše než pro ně vytvořit speciální skupinu.

### 3.1.3 Klasifikace podle druhu monokolokabilního slova

Dosud jsme se při hledání vhodných kritérií k třídění zaměřovali na referenční slova. Také u klíčového, monokolokabilního prvku však lze podobné vlastnosti vysledovat a využít k jinému typu klasifikace. Situaci zde ovšem komplikuje deklarovaná neanalyzovatelnost monokolokabilního prvku, který skrze nemožnost použití v mnoha různých srovnatelných spojení, a tudíž i neexistenci vysledovatelných analogií

v chování, nelze blíže popsat (srov. např. Filipec – Čermák 1985, s. 174). Avšak, jak již naznačují příklady v této práci dosud uváděné, monokolokabilitu nacházíme i u prvků, které se zdají fungovat po všech ostatních stránkách jako „normální“ slova. Setkáváme se tedy jak se slovy a kolokacemi téměř pravidelnými, tak i s případy zcela obskurními a synchronně neprůhlednými. Pokoušet se třídit výsledky na základě jakési obecné „míry (ne)pravidelnosti“ se jeví jako málo vypovídající a nejisté. Místo toho však můžeme použít jednodušší a jednoznačná kritéria, která nám monokolokabilní slova roztrídí na více menších skupin s konkrétními projevy pravidelného či naopak abnormálního chování (samozřejmě s výjimkou monokolokability samotné). To, co bychom v obecnější rovině popsali zjednodušeně a nepřesně jako míru pravidelnosti, se pak projeví v přináležitosti daného slova k jedné či více skupinám podle konkrétních kritérií.

Jako základní hledisko můžeme použít ohebnost daného prvku, tj. jeho přítomnost pouze v jednom tvaru nebo ve dvou a více tvarech. Je přitom třeba rozlišovat ohebnost teoreticky odvoditelnou (představitelnou na základě analogie vůči jiným slovům) a skutečně prokázanou přítomností daných tvarů v úzu prostřednictvím korpusu. Důsledky opomíjení tohoto rozdílu jsou patrné na příkladu slova *holičkách*, které se na první pohled jeví jako ustrnulý tvar 6. pádu slova *\*holičky*, resp. *\*holička*. Při pohledu do korpusu ovšem zjistíme, že jiné tvary odpovídající takovému imaginárnímu paradigmatu (počopitelně po odstranění příkladů homonymního slova *holička* ve významu „kadeřnice“) se vyskytují zcela ojediněle leda v užití metajazykovém nebo takovém, které spíše evokuje hru s jazykem, jako v příkladu (3). V zásadě proto můžeme *holičkách* označit za neohebné a pokusy při teoretickém popisu domýšlet jeho další tvary, či dokonce je používat jako lemma apod., za neopodstatněné a nerespektující skutečnou povahu tohoto slova.

Při ověřování existence různých tvarů pak můžeme u monokolokabilních slov konstatovat, že jsou zcela neohebná, tj. vyskytují se jen v jediném tvaru (např. *být libo*, {*de*, *ipso*, *post*} *facto*). To ovšem nutně neznamená, že jde o nepravidelné chování – je třeba brát zřetel také na to, o jaké slovo jde, jakému slovnímu druhu se podobá apod. Např. u zkratk (považujeme-li je za zvláštní slovo), cizích slov či některých příslovčí není neohebnost nikterak příznaková. Vedle toho může dojít i k homonymii – kupříkladu slovo *náručí* najdeme pouze v tomto tvaru, jedná se však o homonymii 6. pádu (*v náručí*) a 2. pádu (*do náručí*, *z náručí*).

Jindy najdeme ke konkrétnímu slovu několik různých tvarů, ale ani zdaleka ne všechny teoreticky možné podle odpovídajícího paradigmatu. Takové případy můžeme

označit jako částečně ohebné (např. {z, na, v, mimo} *doslech, polohrubá mouka*). Je však třeba mít na paměti, že u všech slov, nehledě na jejich frekventovanost či volnost kolokability, existuje jistý nepoměr v zatížení jednotlivých tvarů. Je to způsobeno jak sémantikou daného slova, která může preferovat určitý způsob použití a blokovat jiný, a samozřejmě také asymetrií v gramatických kategoriích jako takových, kdy např. vokativ je sám o sobě poměrně vzácný a přítomný čas u dokonavých sloves nemožný. Z toho důvodu není nic zvláštního, pokud ve výčtu nalezených slovních tvarů některé chybí (úměrně počtu tvarů, kterých dané slovo teoreticky nabývat může). Za omezenou ohebnost lze považovat pouze případy, kdy výčet obsahuje jen velmi malou část možných tvarů. K takové situaci dochází zvláště často u podstatných jmen monokolokabilních vůči předložkám. Můžeme sem také zařadit případy slov omezených jen na tvary odpovídající určité gramatické kategorii, jako např. tvary pouze mužského rodu.

Konečně celá řada slov se vyskytuje v mnoha různých tvarech a lze je tak označit za plně ohebná. To pochopitelně nevylučuje, že některé jejich možné tvary chybí – může jít jednoduše o důsledek zmíněné asymetrie běžné i u zcela pravidelných slov, popř. mohou některé tvary chybět pouze v daném korpusu a v širším úzu být okrajově nebo i běžně zastoupené (to se ostatně týká i většiny nespisovných tvarů). Plnou ohebnost nacházíme především u monokolokabilních slov slovesné a adjektivní povahy, jako např. *zavrtět* {hlava, se, ocas, ocásek, zadek, hlavička, rameno, palec} či *závazkový* {vztah, právo, kampaň, hnutí, úprava, instrument}.

U slov, pro něž je doložen více než jeden tvar, je také třeba sledovat, zda jsou monokolokabilní všechny nebo jen jeden, popř. některé. Narážíme zde vlastně na rozlišení monokolokability lexému oproti monokolokabilitě jednotlivých tvarů, se kterým jsme počítali již při dvojitým vyhledávání v korpusu, totiž podle lemmat a slovních tvarů.

Další kategorií, kterou je vhodné u monokolokabilních slov rozlišit, je slovní druh, či přesněji podobnost některému slovnímu druhu. Monokolokabilní slova jsou často nositeli různých abnormálních vlastností, ať už formálních (jako např. snížená ohebnost, viz výše) či sémantických, avšak není tomu tak zdaleka ve všech případech. Nevidím důvod, proč slova zvláštní pouze omezenou kolokabilitou a jinak pravidelná považovat co do slovního druhu za blíže neurčitelná. Na druhou stranu však u mnoha nacházíme i další anomálie, a tehdy se zdá vhodnější mluvit spíše o podobnosti než o jasné příslušnosti, popř. přímo o neurčitelnosti.

Z důvodu omezené kolokability se jako vhodná kritéria rozlišení slovního druhu jeví jednak podobnost jeho tvarů některému flektivnímu paradigmatu a morfologická stavba

slova jako takového (je-li zjevná), jednak jeho alespoň zdánlivý syntaktický vztah k referenčnímu slovu, popř. v rámci celé věty. Na tomto základě můžeme mezi monokolokabilními slovy najít substantiva (*doslech, odkvět, stroužek*), adjektiva (*pokutový, pohybující, považovaný*), slovesa (*postesknout, zavrtět, sestávat*) i adverbia (*zrakově, památkově, zbrusu*), byť tyto třídy jsou co do počtu příslušných prvků různě zastoupeny. Jednotlivé případy se také liší v tom, jakými vlastnostmi se danému slovnímu druhu podobají. Kromě těchto čtyř kategorií nám ovšem zůstává také nezanedbatelná skupina slov, která nelze k žádnému slovnímu druhu jednoznačně přiřadit na základě žádného z uvedených kritérií. Jedná se pochopitelně o cizí slova jako *facto* nebo *loci*, ale také o některá neohebná slova, u nichž použitá kritéria nevedou k jednoznačnému výsledku, jako např. u *libo* nebo zkratky *pod.*, která může stát jak za *podobný*, tak za *podobně*.

Třídění, která jsme zde v jejich základním principu uvedli, pochopitelně nejsou jediná možná. Jako kritérium k další klasifikaci můžeme použít také syntaktické funkce nesené celým monokolokabilním spojením (jako zvláštní skupina se zde vydělují např. víceslovné předložky), sémantickou charakteristiku monokolokabilního slova a jiné vlastnosti. I takto však uvedená kritéria a způsoby třídění v souhrnu vypovídají o charakteru daného monokolokabilního slova a umožňují tak o něm podat poměrně obsáhlý obraz. Přehled početního zastoupení jednotlivých typů ve zkoumaném vzorku je uveden v tabulkách v příloze 4.

## 3.2 Charakter monokolokabilních jednotek

Již uvedená klasifikace nabízí určitou představu o tom, jak monokolokabilní jednotky a jejich kolokace vypadají a jak se chovají. V této kapitole jejich popis ještě dále rozvedu. Vyjdu přitom z komplexního a velmi obsáhlého výkladu uvedeného v práci S. Stojčeva (1984), který ovšem počítal pouze s monokolokabilitou v rámci frazeologie a idiomatiky, a také z popisu O. Káninského (2004), jenž na něj navazuje s novějšími daty i metodami. Vzhledem k tomu, že v těchto pracích už je o monokolokabilních jednotkách řečeno mnoho, zaměřím se především na to, v čem se mé výsledky a závěry – v důsledku zahrnutí dat i mimo frazeologii a idiomatiku – od jejich odlišují.

### 3.2.1 Formální charakteristika

Hlavní vlastností, kterou se monokolokabilní slova podle Stojčeva vyznačují, je jejich defektnost po různých stránkách. Z fonologického a intonačního hlediska se od

ostatní slovní zásoby patrně neliší (s výjimkou toho, že mezi nimi neexistuje žádné tvořené jen jedním fonémem); v některých případech, kdy referenční slovo má pevně danou pozici vůči monokolokabilnímu, však může dojít ke stírání hranice mezi oběma prvky a jejich splývání v jedno slovo. Stojčev jako příklady v češtině uvádí *mermo (moci) – mermomoci* a *(na) pospas – napospas*. Ze synchronního hlediska se oba případy již vyskytují převážně v jednoslovné podobě<sup>26</sup> a přechází tak vlastně na monokolokabilitu na morfologické úrovni.

Zaměříme-li se na morfologické složení monokolokabilních slov, zjistíme, že vedle synchronně neprůhledných, nemotivovaných, jako je *krážem* (kde sice můžeme vidět něco jako pádovou koncovku *-em*, ale připojenou k jinde neexistujícímu slovotvornému základu *kráž*; můžeme pak tento základ označit za monokolokabilní morfém nebo považovat celé slovo za nerozložitelné), najdeme i velké množství zřejmě pravidelně utvářených slov, jako např. *památkově*. Právě jejich morfologické složení pak u nich umožňuje určit význam jinak než rozdělením významu celé kolokace, což je ovšem postup mnohdy nevhodný (viz oddíl 3.2.2). Stojčev zvláště upozorňuje právě na to, že monokolokabilní slova mohou mít a mívají monokolokabilní kořenový morfém, jako u *pif-ka* nebo *kráž-em*. U řady slov ve výsledku i zdánlivá formální motivovanost naráží na fakt, že význam jednotlivých částí je synchronně neprůhledný nebo jej nelze použít pro složení významu celku pravidelným způsobem, jako u slova *holičkách*.

Pro řadu monokolokabilních prvků je dále typické omezení deklinačních a konjugačních paradigmat. Jedná se o vlastnost natolik rozšířenou, že jsem ji použila i jako jedno z hlavních kritérií ke klasifikaci (viz kapitola 3.1.3). Stojčev uvádí, že flektivní paradigmata jsou většinou omezena na jeden nebo dva tvary a jen ojediněle jich zahrnují více. Tento závěr však vychází především z toho, že pracoval pouze s monokolokabilními prvky v rámci frazémů. Naše databáze zahrnuje mnohem širší spektrum výsledků a plně nebo téměř plně ohebných prvků je v ní značné množství. Částečně ohebná nebo zcela neohebná slova<sup>27</sup> však stále tvoří její nezanedbatelnou část. Zároveň je třeba počítat se slovy, která mají plné flektivní paradigma a monokolokabilní je jen jeden nebo několik z jeho členů.

---

<sup>26</sup> V korpusu SYN2010 najdeme 136 výskytů *mermomoci* oproti pouhým 7 *moci mermo* nebo jen *mermo*. Na *pospas* je dosud poměrně časté, korpus obsahuje 135 případů oproti 331 *napospas*.

<sup>27</sup> Pochopitelně je řeč o slovech, která se podobají ohebným slovním druhům a otázka ohebnosti je tak u nich relevantní.

S morfolo­gickým slo­žením sou­visí také pří­sluš­nost do vir­tuálních pa­ra­digmat<sup>28</sup> z hlediska slo­votvo­rného. Stojčev upozor­ňuje na to, že mono­koloka­bilní slo­va jsou často tvo­řena způ­so­bem, k němuž ana­logické pří­pady v pra­videlném ja­zyce ne­naj­deme. Jako pří­klad lze uvést spo­jení *do třetice*; k ostatním číslovkám ana­logické tvary češ­ti­na nemá (*dvojice*, *čtveřice* atd. jsou sou­částí jiného pa­ra­digmatu, kam ná­leží tvar *trojice*). Mono­koloka­bilní slo­vo tak samo o sobě tvo­ří jediný člen vir­tuálního pa­ra­digmatu. V ně­kterých pří­padech je ovšem stej­ně utvo­řených slo­v více, všechna jsou však mono­koloka­bilní, popř. se po­dobají i pra­videlným slo­vům, fungují však jiným způ­so­bem. To je pří­pad desubstantivních ad­jektiv typu *tma tmoucí* a *smrt smrtoucí* či de­verbativních sub­stantiv (přesněji slo­v sub­stantivům po­dobných) typu *na shledanou* a *na rozloučenou*, vedle nichž exis­tuje také skupina po­dobně utvo­řených, ale pra­videlných názvů her jako *kopaná*, *schovávaná* atd. Rozší­ření da­tabáze o prvky mimo fra­zeologii a idi­omatiku ovšem zna­mená, že tyto vlast­nosti nelze vz­táhnout na všechny mono­koloka­bilní prvky. Řada mono­koloka­bilních slo­v je sou­částí zcela pra­videlného vir­tuálního pa­ra­digmatu – např. *zrakově* {postí­žený, vní­maný, hendikepovaný, ná­ročný} můžeme vz­táhnout k *sluchově* či *čichově*, *výkupní* {cena, orga­nizace} ke *kupní* či *nákupní* apod. Objevují se navíc nové ana­logické skupiny – např. u ad­jektivních reflexiv tantum je zpra­vidla obli­gatorně zvr­atné i zá­kladové slo­veso, de­verbativní ad­jektiva typu *považovaný* taktéž pocházejí zpra­vidla z mono­koloka­bilních slo­ves. V tomto smě­ru je tedy po­vaha mono­koloka­bilních slo­v velmi roz­manitá.

### 3.2.2 Sémantická charakteristika

Při práci s formou mono­koloka­bilních slo­v, byť vy­kazuje různé dru­hy de­formace, se stále můžeme opřít o ně­co hmatatelného a ne­popiratelného a řeší se především otáz­ky in­terpre­tace nalezených jevů. Pole sémantiky je však nesro­v­natelně ná­ročnější; po­mí­neme-li fakt, že i o tom, co je vlastně vý­znam, se dosud vedou spory, pracujeme zde zpra­vidla s ab­straktními, těž­ko uchopitelnými před­stavami a kate­goriemi. A co je nesnadné v pra­videlném ja­zyce, stává se u mono­koloka­bilních slo­v ještě komplikovanějším.

V první řadě narážíme na fakt, že množství mono­koloka­bilních prvků, resp. jejich kolokací, spadá do oblasti fra­zeologie a idi­matiky a v takovém pří­padě je sporné, zda vůbec má smysl dělení fra­zému na jeho části a určování jejich vý­znamu zkoušet. Jelikož

---

<sup>28</sup> Vir­tuálním pa­ra­digmatem rozumíme třídu „(široce) ek­vi­valentních prvků spjatých společnou funk­cí a vý­znamem, jež je dána ab­straktně, v ja­zykovém sys­tému a je vyme­zitelná svým místem v jeho hierar­chii“, srov. Filipec – Čermák (1985, s. 171).



frazeologie, resp. idiomatika se vyznačuje mimo jiné právě tím, že dochází k posunu nebo i výraznější změně významu a že v doslovném vyznění mnohdy frazeologické spojení vůbec nedává smysl, domnívám se ve shodě s Čermákem a dalšími autory (srov. např. Čermák – Blatná, 1995, s. 118 a odkazy zde uvedené), že analýza významu jednotlivých částí frazémů na základě jejich domnělých vztahů v rámci z podstaty nepravdělného spojení je neprůkazná. Totéž platí i pro pokusy popsat taková slova na základě jejich morfologického složení, právě kvůli posunům a změnám významu, jimiž celé slovní spojení prochází.

Vedle toho však naše databáze obsahuje i řadu slov, která do frazeologie a idiomatiky nepatří, a zde již analýza jistě smysl má. Přestože kolokabilita slova je silně omezena a nemůžeme tak hledat analogie v syntagmatických vztazích k různým kolokátům, význam takových slov se mluvčímu zdá zřejmý; je definovatelný právě na základě paradigmatických vztahů a analogií s příbuznými slovy.

Stále však zůstává otázka, zda se skutečně jedná o význam daného slova jako takového, nebo celého monokolokabilního spojení. Jinak řečeno, i když je jeho význam odvoditelný z morfologického složení, monokolokabilní slovo samo o sobě nemůže fungovat jako celistvá jednotka, vždy tvoří kolokaci s některým referenčním slovem. I v případech, kdy stojí samostatně, se jedná jen o kontextovou elipsu. Je pak možné se ptát, zda referenční slovo není inherentní součástí monokolokabilního slova, zda netvoří po stránce významové nedělitelný celek. Například u sloves, která jsou monokolokabilní vůči předložce, se jedná de facto o vztah k danému valenčnímu doplnění, které je vázáno přes předložku. Je-li nějaké doplnění obligatorním argumentem slova, tvoří vztah k němu vlastně součást významu tohoto slova. Např. u *záviset (na)* význam slovesa předpokládá nějaký předmět, ke kterému se závislost vztahuje a jenž je na syntaktické rovině vyjádřen doplněním vázaným přes předložku *na*. Ostatně podle PMC tvoří předložková vazba sloves a z nich odvozených slov součást vícečlenné lexikální jednotky. Ve prospěch takového pohledu svědčí i fakt, že monokolokabilní slova se velmi často kombinují s prvky gramatickými či co do významu značně vyprázdněnými. Například u spojení typu *být nasnadě* nese sloveso v podstatě jen transformační roli, která z celku vytváří výraz schopný nést slovesné funkce, a lze je tak přirovnat k odvozovacím sufixům. Snad nejvýraznější je pak vyprázdňenost zvrtných zájmen u reflexiv tantum, kde tyto prvky již nenesou funkci vůbec žádnou – kromě toho, že jsou pro dané sloveso nebo adjektivum nezbytné.

Vedle toho však existují také monokolokabilní spojení dvou prvků, z nichž každý zjevně nese specifický a vydělitelný význam. Jedná se především o spojení monokolokabilního adjektiva se substantivem, jako např. *polohrubá (mouka)* nebo

*objízdny* {trasa, komunikace, kolej, úsek, cesta}. V takových případech se zdá, že je to naopak silná vyhraněnost významu monokolokabilního prvku, která způsobuje jeho omezenou kolokabilitu. Také u slov, která jsou monokolokabilní jen v jedné nebo několika ze svých forem, zřejmě není důvod tvrdit, že nemají vlastní význam. Zároveň je však mezi těmito slovy mnoho případů, kdy v monokolokabilním spojení (přinejmenším s některým, zpravidla nejfrekventovanějším z referenčních slov) dochází k posunu významu. Na takové případy je pak třeba nahlížet jako na spojení frazeologického typu.

### 3.2.3 Polysémie a homonymie

Následující oddíl se blíže pozastavuje nad dvěma jevy, na něž v datech narážíme a které se týkají poměrně velké části výsledků mimo frazeologii a idiomatiku. Polysémie a zejména homonymie potenciálně monokolokabilních slov mnohdy přinášely komplikace při ručním třídění dat. Vedle zjevné polysémie, která se projevuje v odlišnosti jednotlivých referenčních slov, navíc narážíme i na její „skrytý“ typ, který se týká jednotlivých kolokací použitých různým způsobem. Při studiu monokolokability je proto na oba jevy třeba pamatovat.

Jednotlivá monokolokabilní spojení mívají poměrně stabilní význam. Je to dáno jednak tím, že řadu z nich tvoří frazémy, u nichž na polysémii příliš nenarážíme, a pojmenování blízká svou povahou termínům, u nichž je významová jednoznačnost jednou z nejdůležitějších vlastností. K této skupině můžeme přiřadit poměrně početnou skupinu adjektiv monokolokabilních vůči substantivu, kde mívá celé spojení poměrně přesně vymezený a často konkrétní význam.

V kontextu se pochopitelně i význam monokolokabilních kolokací aktualizuje, mírně přizpůsobuje tématu apod. Můžeme však narazit také na taková spojení, u nichž existují relativně různé způsoby použití a v souvislosti s tím také větší či menší posuny ve významu. Pro ilustraci uveďme příklady (4) a (5). V první dvojici je použito spojení (*do budoucna*). Obě užití se však mírně liší. Rozdíl mezi nimi je snáze viditelný, když se je pokusíme substituovat jiným výrazem při zachování významu. U (4a) by tomu odpovídalo *v budoucnosti* nebo *pro příště*, kdežto u (4b) spíše *do budoucnosti*; substituty pro (4a) přitom nelze použít pro (4b) a naopak. Ve druhém případě je navíc přítomnost předložky *do* ovlivněna slovesem *hledět* a spojení s monokolokabilním slovem se tak více podobá volné kombinaci kolokačně neutrálních slov.

- (4) a. Ne snad z morálních důvodů, ale prostě proto, aby se do budoucna vyhnuli možnosti vydírání.  
b. Podle Britů nastal čas hledět do budoucna se vzájemnou důvěrou.

Podobně vidíme rozdíl také v příkladu (5). Zde je použita kolokace *do střehu*. V (5a) má konkrétní význam bojového postavení. V (5b) se však již jedná o širší význam stavu ostražitosti, zbystrění smyslů nebo pozornosti.

- (5) a. Felix Jaeger vytrhl meč z pochvy a postavil se do střehu.  
b. „Vážně?“ podivil se Robert mrazivým tónem, který by vnímavějšího muže už předem uvedl do střehu.

Jak je vidět, určitá variabilita ve významu a způsobu použití existuje i u konkrétního monokolokabilního spojení, přinejmenším v některých případech. Mnohem bohatší možnosti porovnávání a analýzy se však otevírají, pokud přistoupíme k rozdělení monokolokabilního spojení na jednotlivé části a považujeme je za do jisté míry autonomní prvky se specifickými vlastnostmi. To nám umožňuje sledovat rozdíly nejen mezi jednotlivými užitími jednoho spojení, ale také mezi různými kombinacemi monokolokabilního slova s jeho kolokáty.

Tento postup se v zásadě podobá analýze použité pro klasifikaci monokolokabilních slov podle počtu jejich referenčních slov a kolokačních paradigmat, a týká se případů, kdy má monokolokabilní prvek více referenčních slov. Můžeme pak sledovat, jak je množina těchto referenčních slov strukturována; zda jsou v ní prvky stejného typu a stejného nebo podobného významu, zda se k monokolokabilnímu slovu vážou stejným způsobem, nebo naopak zda jde o velmi odlišné prvky, které k němu mají úplně jiný vztah. Pohybujeme se pak na stupnici od synonymních kolokací přes spojení různého významu, kde je však vztah mezi kolokáty obdobný a také význam monokolokabilního slova stabilní, až po jasně polysémní případy s více oddělenými kolokačními paradigmaty. Nejedná se přitom pochopitelně o kategorie aristotelského typu; spíše se dá hovořit o škále s jednoznačnými i přechodnými případy.

Například u slova *lordstvo* vidíme velmi homogenní skupinu referenčních slov {jeho, váš, jejich}. Jedná se vždy o přivlastňovací zájmeno, které tvoří součást používání titulu, rozdíly ve významu se v podstatě týkají jen různých gramatických kategorií. Zato je třeba upozornit, že kolokační paradigma tohoto slova nelze zjednodušit na tvrzení, že se

kombinuje k přivlastňovacím zájmenem; charakteristické totiž pro ně je, že se kombinuje jen se třemi. Příklady jako *\*moje lordstvo* či *\*její lordstvo* se v korpusu nevyskytují buď vůbec, nebo jen ojediněle.

Podobně i *činžovní* má poměrně velké, ale jednoduché kolokační paradigma, jehož prvky jsou substantiva velmi podobného významu, označující různé druhy obydlí nebo jejich komplexů {dům, vila, blok, domek, byt, zástavba}. Referenční slova se s monokolokabilním pojí stejným způsobem (*činžovní* tvoří jejich přivlastek) a význam celku se liší jen významem referenčního slova – vždy se jedná o druh stavby s nájemním bydlením.

O něco složitější již je situace u slovesa *pokrčit*, které se kombinuje se substantivy {rameno, koleno, noha, nos, paže, obočí, čelo, nosík}. Na první pohled se celá skupina zdá poměrně konzistentní – nejde sice přímo o synonyma, všechno jsou to však slova označující části těla, čili kohyponyma. Při bližší analýze ovšem zjistíme, že se liší způsobem, jakým se s monokolokabilním slovem kombinují, svým vztahem k němu. Zcela jasně se tu vyděluje kolokace *pokrčit (rameny)*, kde je substantivum v 7. pádě a význam kolokace má kromě konkrétního denotátu pohybu také význam vyjádření pochybnosti, neznalosti apod., který je tímto pohybem obvykle dáván najevo. Oproti tomu ostatní referenční slova bývají ve čtvrtém pádě. Význam *pokrčit* pak odpovídá „ohnout“ (*koleno, nohu, paži*) nebo „svraštit“ (*obočí, čelo, nos, nosík*). Pro kolokace s *nos* (nad někým nebo něčím) SSSJČ uvádí význam „pohrdnout“ a stejně jako *pokrčit (rameny)* je toto spojení označováno jako frazém. V korpusu se případy s tímto významem vyskytují, vedle nich se objevuje také význam „projevit odpor, znechucení“ a v malé míře další, dokonce i pozitivní („Mírně spiklenecky pokrčila nos.“). Ve třech případech je navíc substantivum v 7. pádě, zřejmě analogicky ke spojení *pokrčit (rameny)*. Kolokační paradigma pro *pokrčit* tedy můžeme v zásadě rozdělit na tři části (popř. dílčí paradigmata):  $k_1$  {rameno}, která má blízko k frazeologii,  $k_2$  {koleno, noha, paže}, kde se jedná o prosté ohnutí končetiny, a  $k_3$  {obočí, čelo, nos, nosík}, kde jednotlivé kolokáty podobně jako u  $k_1$  nesou i význam vyjádření emocí skrz to, že se jedná o výrazy mimiky, přičemž *pokrčit nos* je v některých případech také frazém. V malé míře se ovšem objevují i jiné kolokáty a reflexivní forma.

Jasným příkladem polysémie je pak již zmíněné adjektivum *uhličítý* nebo např. *klíčn*. V korpusu k němu nacházíme 7 referenčních slov: {kost, kůstka, rostlina, hyfa, mycelium, rostlinka, vlákno}, z nichž zdaleka nejčastější je kolokace s *kost*. Již na první pohled se zde vydělují dvě skupiny slov – na jedné straně {kost, kůstka} pro označení části

lidské nebo zvířecí kostry, na straně druhé {rostlina, hyfa, mycelium, rostlinka, vlákno} tvořící s adjektivem kolokace patřící do botanického oboru.

*Klíční* je ovšem specifický příklad; vzniklo totiž dvěma odlišnými způsoby, jimž odpovídají uvedená kolokační paradigmatata, a jedná se tedy vlastně o homonymii dvou různých adjektiv. Přinejmenším pro *klíční* {kost, kůstka} bychom i v případě oddělené analýzy mohli konstatovat monokolokabilitu. Na synchronní rovině však již motivace obou odvozenin není zcela průhledná a rozdíl v původu slov se stírá; můžeme je proto, byť s rezervou, analyzovat i jako jediné slovo.

Při pohledu z opačné strany pak je možné narazit na polysémantická slova, která jsou monokolokabilní jen v jednom ze svých významů. Tento typ monokolokability je velmi obtížně zjistitelný a do svých dat jsem jej nezahrnula, s výjimkou případů, kdy byly jiné významy velmi vzácné a jako monokolokabilní se tak jevílo celé slovo (např. *vínek*). Jeho přijatelnost navíc není jednoznačná; jelikož význam slovní jednotky se projevuje také v tom, s jakými slovy se kombinuje, hrozí zde cykličnost analýzy. V konečném důsledku by také musela být považována za monokolokabilní všechna slova obsažená ve frazémeh, jelikož ve frazému je význam jeho částí neanalyzovatelný, čili jiný než v běžných kolokacích a volných spojeníh.

Na příkladu *klíční* je vidět, že stejně jako všude v jazyce i v rámci monokolokabilních slov narážíme kromě polysémie také na homonymii. Homonymie celých lexémů je ovšem poměrně řídká a v některých případech dochází více či méně ke splývání obou slov, pokud se jejich význam příliš neliší nebo není zcela průhledný. Naopak např. u (*nechat*) (*na*) *holičkách* je záměna monokolokabilního prvku za ženskou variantu slova *holič* velmi málo pravděpodobná.

Mnohem častěji narážíme na homonymii u jednotlivých slovních tvarů. Stejně jako v případě lexémů i nalezené homonymní tvary různých slov byly z dat vyřazeny jako nerelevantní. Mnohdy však došlo k homonymii různých tvarů téhož slova. Jednalo se především o shodné tvary různých pádů substantiv a adjektiv. Při analýzách výsledků vyhledávání podle slovních tvarů je třeba na toto pamatovat; jednotlivé výsledky neodpovídají přímo různým kombinacím gramatických kategorií pro dané slovo, ale slovním tvarům ve smyslu prostého sledu grafémů, nejsou proto ve všech případech zcela srovnatelné.

### 3.2.4 Vzájemné vztahy monokolokabilních slov

Z toho, co bylo dosud řečeno, je jasné, že monokolokabilní slova tvoří zcela náhodnou a neuspořádanou množinu. Je možné mezi nimi najít podmnožiny se společnými vlastnostmi, analogie a podobnosti po stránce významové i formální. V této kapitole se zaměřím na některé takové vztahy a podobnosti.

Synonymie se mezi monokolokabilními slovy či jejich spojeními vyskytuje jen ojediněle a spíše náhodně (např. *bez ustání* – *bez přestání*). Setkáváme se však se skupinami slov, jejichž kolokabilita má podobnou povahu a patří do podobného významového okruhu. Jsou to např. slovesa vyjadřující specifický pohyb některé části nebo částí těla, jako výše zmíněné *pokřčit*, ale také *vytřeštit* {oko, zrak}, *přimhouřit* {oko, víčko}, *pokývat* {hlava, hlavička} či *zavrtět* {hlava, se, ocas, ocásek, zadek, hlavička, rameno, palec}; podobně můžeme vymezit skupinu kuchyňských surovin, kde jde zpravidla (ale ne výlučně) o spojení monokolokabilního adjektiva se substantivem: *bobkový (list)*, *polohrubá (mouka)*, *stroužek (česneku)*, *zakysaný* {smetana, výrobek, mléko, nápoj, krém} atd.

Mezi monokolokabilními slovy je také mnoho slov spřízněných, zpravidla přímo odvozených jedno z druhého. Jedná se o adjektivní reflexiva tantum odvozená od sloves, např. *usmívat (se)* – *usmívající (se)*, *týkat (se)* – *týkající (se)* atd., adjektiva odvozená od sloves monokolokabilních vůči předložce, jako např. *sestávat {z, s}*<sup>29</sup> – *sestávající (z)*, nebo od sloves monokolokabilních vůči substantivu, jako u dvojice *přimhouřit* {oko, víčko} – *přimhouřený* {oko, víčko, pohled, řasa, okno}. Kolokační paradigmata obou slov v podobných dvojicích bývají shodná nebo se částečně překrývají; sdílí především ten prvek nebo prvky, které jsou v daných kolokacích nejfrekventovanější. Monokolokabilní slova jsou však běžně spřízněna i se slovy kolokačně neutrálními a je-li odvozené slovo monokolokabilní, neznamena to ještě, že musí být monokolokabilní i jeho základové slovo (srov. např. monokolokabilní *pohybující se* a *pohybovat se/něčím*).

Dále narážíme na skupiny slov, které tvoří virtuální paradigma z hlediska svého morfologického utváření. Sem patří např. již zmíněná substantivům podobná slova končící na *-ná*, resp. *-nou* (vzhledem k monokolokabilitě vůči předložce *na* a neexistenci jiných tvarů je forma „prvního pádu“ sice logicky utvořeným, ale zcela umělým konstruktem), jako *(na) shledanou*, *(na) rozloučenou* či *(na) rozmyšlenou*. Toto paradigma je navíc

---

<sup>29</sup> Předložka *s* je zde vypsána jako samostatné referenční slovo, přestože jde jen o (z hlediska normativní mluvnice nesprávnou) pravopisnou alternativu *z*, protože se objevuje opakovaně; lze ji však jednoduše přičíst k výskytům *s z* nebo označit zvlášť spolu s překlipy a jinými chybami.

specifické tím, že jeho členy tvoří pouze monokolokabilní slova. Většina monokolokabilních slov, která nespádají podle tradičního pojetí do oblasti frazeologie, je odvozena zcela pravidelným způsobem a můžeme je proto zařadit do virtuálních paradigmat spolu s kolokačně neutrálními slovy. V některých případech je monokolokabilní odvozené i základové slovo, není to však ani zdaleka pravidlem. Můžeme např. srovnat dvojice *přimhouřit – přimhouřený*, *pokrčit – pokrčený* a *sklonit – skloněný*. V prvním případě jsou oba členy monokolokabilní, ve druhém je monokolokabilní sloveso, ale adjektivum má možných kolokátů podstatně více, a ve třetím případě, který je odvozen stejným způsobem, není monokolokabilní ani jeden člen. Monokolokabilita tedy až na výjimky není vlastnost, která by se dala jednoznačně vyčíst z morfologického složení slova. Další příklady a poznámky k tematice virtuálních paradigmat jsou uvedeny v oddílu 3.2.1.

Při vyhledávání podle slovních tvarů lze navíc narazit na případy, kde monokolokabilita souvisí s gramatickými kategoriemi, je gramaticky podmíněná. Specifickou skupinu tak tvoří např. duálové tvary adjektiv, které se mohou spojovat pouze s označeními částí těla. Jako příklad můžeme uvést *holýma* {rukama, nohama, zádama}, *modrýma (očima)* či *zavřenýma (očima)*. V případě *holýma* si navíc můžeme povšimnout referenčního slova *záda*, které neoznačuje párový orgán; jedná se vlastně o homonymii neexistujícího duálového a nespisovného tvaru.

Krajním případem takto gramaticky podmíněné monokolokability je 6. pád substantiv a adjektiv. Jelikož tento pád se vždy kombinuje s předložkou a zároveň předložek vázících se s 6. pádem je pouze 5 – *na, o, po, při* a *v* (srov. PMČ, s. 241 a 344) – všechna slova v 6. pádě jsou vlastně monokolokabilní (u adjektiv se de facto jedná o monokolokabilitu převzatou od substantiva, které rozvíjejí, i při samostatném použití, např. *elipse* nebo *nominalizaci*, se však vždy s jednou z těchto předložek musí kombinovat). K problematice 6. pádu se ještě vrátíme v oddílu 3.3.4.

Je zřejmé, že se monokolokabilní slova vyznačují řadou vlastností, která je od ostatní slovní zásoby odlišují. Některá mají nejasný význam, jsou utvořena neobvyklým způsobem nebo přímo obsahují monokolokabilní kořenový morfém, jiná mají tendenci splývat se svým referenčním prvkem apod. Zároveň však lze mezi nimi najít řadu specifických skupin, jejichž členy sledují určitou analogii, podobně jako to vidíme v pravidelném jazyce, a jindy konstatujeme, že se dané slovo kromě omezené kolokability v ničem neliší od kolokačně neutrální slovní zásoby. Zjištěné vzájemné podobnosti a analogie nám umožňují nejen konstatovat, že se nejedná o zcela nahodilou skupinu archaismů, jak by se

mohlo na první pohled zdát, ale zčásti také formulovat pro takové analogické skupiny pravidla a tím i predikce. Na monokolokabilní slova tak lze nahlížet jako na zvláštní součást jazyka, která se částečně řídí vlastními zákonitostmi, zároveň se ale volně prolíná s pravidelnou slovní zásobou. Vyznačení hranice monokolokability je pak pochopitelně záležitostí pouhé konvence. Definice skrz maximum 7-8 kolokátů je sama o sobě velmi zjednodušující a neříká takřka nic o povaze monokolokabilních slov. Ve chvíli, kdy žádná přirozená hranice neexistuje a výzkumník je konfrontován pouze se škálou více či méně jednoznačných případů, je však i takové vodítko užitečným nástrojem.

### 3.3 Hranice monokolokability a specifické případy

Právě hraničními případy, kdy není jednoznačné, zda je vhodnější dané slovo považovat za monokolokabilní nebo celé spojení popsat úplně jiným způsobem, se zabývá následující podkapitola. Načrtnu zde několik skupin slov, resp. jejich kolokací, pro něž je charakteristická kombinace s jedním nebo několika prvky, což by odpovídalo definici, z různých důvodů se však označení monokolokability nezdá být vhodným či přesným.

#### 3.3.1 Vlastní jména

Mezi proprii je řada slov, která se kombinují jen s jedním nebo několika jinými. Jedná-li se o proprium víceslovné, jeví se pak takový prvek jako monokolokabilní. Jeho samostatná použití (bez referenčního slova) lze navíc snáze než u jiných monokolokabilních slov označit za kontextovou elipsu.

Potíž vězí ve zvláštním vyprázdněném charakteru vlastních jmen. Jelikož jejich význam<sup>30</sup> spočívá jen v odkazu na konkrétní referent nebo referenty ve skutečnosti, nedá se u nich hovořit o formálně-sémantickém vztahu k referenčnímu slovu ve stejném smyslu jako u jiných monokolokabilních prvků. Takový „význam“, zvláště u vlastních jmen osob, je navíc velice nestálý – k rozšíření počtu denotátů, a tím často i referenčních slov, stačí, aby se narodilo dítě a dostalo nové jméno nebo aby spisovatel vymyslel novou postavu či místo. Jinými slovy, kolokabilita proprii není záležitostí jazykového systému, ale vypovídá spíše o počtu denotátů, čili o momentální situaci ve světě.

Dalším problémem je u proprii nedostatečné, resp. zcela nereprezentativní zastoupení ve vzorku. Do korpusu se nutně dostává jen malá část proprii a ještě méně se jich v něm objevuje opakovaně. Tato situace je nejmarkantnější opět u vlastních jmen osob: je

---

<sup>30</sup> Odhlédneme-li od faktu, že některá propria mohou přecházet v apelativa, srov. např. „Ty jseš ale Matěj!“



jednoduše nemyslitelné, aby se do korpusu dostala všechna nebo alespoň podstatná část. Vzorek v korpusu přítomný tak zdaleka nevypovídá ani o oné momentální situaci ve světě; jeho obsah a rozsah je dán tím, o jakých lidech, místech apod. se v zahrnutých textech nejvíce píše (popř. mluví, pokud jde o mluvený korpus).

Vzhledem k těmto vlastnostem nebyla *propria* do naší databáze přijímána. Při vyhledávání v korpusu je možné se jim z velké části vyhnout automatickým vyřazením lemmat a slovních tvarů začínajících velkým písmenem. Automatická lemmatizace pochopitelně není zcela přesná, ve výsledcích se tak i přesto ojedinělé případy objevily a byly následně vyřazeny při ručním třídění. Výjimku tvořily řídké případy, kdy se jednalo o monokolokabilní slovo použité např. jako součást názvu knihy nebo organizace apod., přičemž se jeho kolokabilita v rámci takového *propria* výrazně nelišila od ostatních výskytů. Jako příklad můžeme uvést název knihy *Výprava do deštného pralesa*, který se v korpusu objevuje dvakrát a obsahuje monokolokabilní adjektivum *deštný*. Jeho kolokační paradigma přitom tvoří slova *prales*, *les* a *sezona*. Kolokace s *prales*, kterou vidíme v uvedeném názvu, tedy odpovídá kolokační charakteristice tohoto adjektiva v jiných kontextech, nízký počet výskytů v tomto konkrétním názvu ani nemá zásadní vliv na celkový poměr zastoupení jednotlivých kolokátů. Takové příklady byly zpravidla do dat započítávány.

### 3.3.2 Terminologie

Zvláštní skupinu monokolokabilních slov tvoří také termíny a spojení terminologické povahy. Podle F. Čermáka (1985, s. 205) výrazy tohoto typu patří do periferie frazeologie a idiomatiky a lze je označit za kvazifrazémy. Jejich základní vlastností bývá (resp. měl by být) jednoznačný, pevný význam. Mohou být tvořeny slovy s běžnou kolokabilitou, je však mezi nimi i řada monokolokabilních, jako např. *nadmořská (výška)*, *čtvereční* {metr, kilometr, centimetr, míle, milimetr, stopa, km, palec}, *pozměňovací* {návrh, zákon} atd. Jedná se zpravidla o spojení monokolokabilního adjektiva se substantivem.

Takováto spojení podle výskytů v korpusu splňují podmínky monokolokability a nezdá se nutné je z databáze vyčleňovat. Velmi volně navíc přecházejí do oblasti frazeologie i běžných kolokací. Je však nutno poznamenat, že v korpusu je zastoupena především běžná slovní zásoba a odborné texty, v nichž lze očekávat největší zastoupení terminologických výrazů, tvoří sice jako celek také jeho významnou část, avšak jednotlivé obory již mají zastoupení relativně malé. Každý vědní obor přitom používá především

termíny z vlastní oblasti a s ostatními se prolíná jen částečně. V důsledku toho terminologické a jim blízké výrazy, na které v korpusu typu SYN2010 narážíme (přínejmenším v míře min. 200 výskytů), jsou především z těch, jaké se používají i v každodenním životě a tvoří centrální součást jazyka. Takové jsou i výše zmíněné příklady *nadmořská (výška)* a plošné míry, které jsou zpravidla známé i mluvčím nespécializovaným v daném oboru. Terminologie používaná především vědci daného oboru je poměrně málo zastoupena a tudíž se do databáze většinou nedostala. Důsledkem takového malého zastoupení (zejm. co do počtu textů) však může být i to, že k určitému málo kolokabilnímu slovu nejsou zaznamenány zdaleka všechny možné kolokáty a dané slovo se tak jeví jako monokolokabilní, i když takové ve skutečnosti není.

Nabízí se také otázka, zda případné terminologické výrazy zahrnovat do databáze zaměřené primárně na základní, běžnou slovní zásobu. Na stejný problém narážíme u výskytů výrazně stylisticky zabarvených, kdy se např. spisovatel snaží prací s jazykem nabudit určitý dojem a používá zastaralé nebo nářeční výrazy a kolokace. Případy, kdy takové stylistické zabarvení znamenalo jiný způsob použití monokolokabilního slova, byly zpravidla označeny jako netypické a tvořily součást „zbytkové“ skupiny podobně jako neobvyklé kombinace slov v poesii apod. Do stejné skupiny ovšem příležitostně spadaly také kolokace a termíny běžné pouze v rámci určitého oboru. Dochází tak vlastně k dvojímu přístupu k jednotkám stejného typu pouze na základě toho, že jedna skupina se užívá i v běžném jazyce a druhá je záležitostí určitého oboru vědy nebo jiné činnosti.

### 3.3.3 Zkratky

Prakticky jakékoliv slovo, někdy i syntagma, může být příležitostně zkráceno; u některých však k tomu dochází opakovaně s takovou pravidelností, že pak vedle plných tvarů daného slova stojí jeden nebo dva ustálené tvary zkrácené. Takové ustálené zkratky mohou být používány odlišným způsobem než celé slovo, popř. se v úzu stát mnohem frekventovanějšími a původní slovo tak de facto nahradit, jako v případě *cca*, které v psaném projevu nad plným tvarem převažuje<sup>31</sup> a proniká i do mluveného.

V naší databázi se objevilo několik zkratk, které vykazovaly monokolokabilní chování. Jmenujme např. *soc.* {dem., demokrat, demokracie, zařízení, věc}, *(a) pod.* či *(pozn.) překl.* Lze si povšimnout, že v uvedených příkladech má základové slovo vždy

---

<sup>31</sup> V korpusu SYN2010 najdeme 2139 výskytů *cca* a pouze 3 *circa* (resp. 4, ve čtvrtém se ovšem jedná o sloveso v latinském textu).

podstatně širší kolokabilitu než příslušná zkratka a i když sdílí její kolokáty, chová se jako kolokačně neutrální. Narážíme tedy na zásadní rozdíl v používání obou forem.

Je pak třeba rozhodnout, jak s takovými případy naložit. Na jednu stranu, zvláště když je monokolokabilní zkratka, ale plné tvary slova ne (jako je tomu v uvedených příkladech), se nabízí možnost zaznamenávat zkrácené formy jako samostatné prvky. Ty nesou stejný, popř. podobný význam jako základové slovo, avšak používají se jiným způsobem, v jiných kontextech (resp. jen v části možných kontextů základového slova), tedy spíše jako synonymní nebo hyponymní výrazy. Na druhou stranu považovat zkratku za jiné slovo než základový lexém se nezdá vhodné. Spíše se jako řešení jeví zaznamenání zkratky jako specifického tvaru základového slova, podobně jako tvary pro různé pády, rod apod. V takovém případě je ovšem třeba započítat případy výskytů zkratky také do záznamů o kolokabilitě lexému reprezentovaného lemmatem (jinými slovy, neoddělovat výsledky lemmatu plného tvaru a lemmatu zkratky). V případech, kdy může zkratka stát za různá slova (např. *překl.* může být zkratkou za *překladař*, ale také za *překlad*), je také nutné vymezit, ve kterých výskytech zkratka zastupuje to které slovo, a to i tehdy, kdy takové homonymní zkratky mají stejná kolokační paradigmatata. V praxi ovšem může podobné rozlišování činit problémy, pokud kontext nenabízí dostatečné informace.

### 3.3.4 Vliv gramatiky

Na způsob, jakým se slova mezi sebou kombinují, mají pochopitelně vliv také gramatická pravidla. U slov, která se jeví jako monokolokabilní vůči prvku nebo prvkům gramatického charakteru, je pak třeba věnovat pozornost tomu, jaké povahy je vztah mezi nimi a nakolik se u jednotek téhož typu opakuje (zda se jedná o jednotlivé relativně izolované případy nebo o všechna slova v dané gramatické kategorii). Domnívám se, že o monokolokabilitě lze mluvit pouze tehdy, není-li širší kolokačního paradigmatu dána pouze gramatickými pravidly a hraje-li roli také další vlastnosti daného slova (sémantická stránka, analogie k příbuzným slovům apod.).

Tyto podmínky jsou splněny u skupiny slov, která jsem ve své klauzurní práci (2011) nazvala valenčně vázanými. Jedná se o poměrně heterogenní skupinu – patří sem substantiva, adjektiva i slovesa, která mají nějaké obligatorní (valenční) doplnění, jež se k nim váže přes konkrétní předložku. O zvláštnosti této vazby jsem se již zmínila výše (viz oddíly 1.2 a 3.2.2). Kolokační paradigmatata takto valenčně vázaných slov bývá nejčastěji jednočlenné, jako v případě *kontrastovat (s)* či *sestávající (z)*. Za zvláštní případy tohoto

typu můžeme považovat i víceslovné předložky, v nichž původní předložka stojí za plnovýznamovým prvkem (např. *s ohledem* {na, k}). Někdy má ovšem takové monokolokabilní slovo referenčních prvků více. U víceslovných předložek narážíme nejčastěji na dvojí valenci, v naší databázi se jednalo o střídání *k* a *na* jako v případě zmíněného (*s*) *ohledem* {na, k}, tedy o valenci dativní a akuzativní. R. Blatná (2006, s. 123-128) zaznamenává i další typy a uvádí, že v případě {na, k} „jde o záležitost vývojovou, neboť progresivní předložka *na* pojící se s akuzativem nahrazuje v koncové pozici těchto VSP původní předložku *k* pojící se s dativem“. Také u původních předložek předcházejících plnovýznamovému prvku narážíme na víceprvková kolokační paradigmatata, jako u {s, bez, při, po} *přihlédnutí* {k, na}, v takových případech ale bývají kombinace s jednotlivými předložkami rozlišovány jako různé víceslovné předložky, popř. jiný typ kolokace. S širším kolokačním paradigmatem se setkáváme i u jiných typů valenčně vázaných slov. Nemusí přitom jít nutně jen o předložky, ale i o slova z nich odvozená, jako např. u *pocházet* {z, od, odkud, odtud, odtamtud, odněkud, odsud, odjinud}. Jak je vidět, většinu takových spřízněných slov tvoří zájmenná příslovce, ojediněle se ale najdou i jiná, např. *zevnitř* u monokolokabilního adjektiva *pocházející*. Podobné příklady se již pohybují na hranici mezi monokolokabilitou gramatického a lexikálního typu.

Existují však také případy, kdy se jedno slovo bez druhého neobejde čistě na základě gramatických pravidel a jejich omezení. S tímto jevem se v češtině setkáváme jen u jednotlivých slovních tvarů (přinejmenším nebyl u lemmat v naší databázi žádný takový případ zaznamenán). Příkladem může být již zmíněný 6. pád substantiv a adjektiv (viz oddíl 3.2.4), který se vždy pojí s předložkou, přičemž přeložek pojících se v češtině s 6. pádem je jen pět (o, na, v, po, při). V tomto smyslu by bylo nutné automaticky označit za monokolokabilní všechna slova v 6. pádě. To se ovšem neslučuje s pojetím monokolokability jako zvláštní vlastnosti lexému či konkrétního slovního tvaru (nebo obdobně prvku jiné roviny), jíž se taková jednotka liší od ostatních, kolokačně neutrálních. Označení těchto jednotek za monokolokabilní navíc nepřináší žádnou jinou informaci, než jakou získáme z toho, že jde o 6. pád, jelikož taková monokolokabilita je 6. pádem automaticky implikována. Z těchto důvodů je podle mého názoru zbytečné, ne-li přímo nevhodné tyto případy plošně označovat za monokolokabilní.

I v případě tvarů 6. pádu ovšem může dojít k monokolokabilitě v plném slova smyslu, pokud se tvar daného slova kombinuje např. pouze s jednou z možných předložek. S cílem nevynechat takové případné prvky jsem při třídění výsledků uznávala jako

monokolokabilní ty tvary 6. pádu, které byly homonymní s jiným pádem (zpravidla třetím) a přesto se jejich použití jeví jako monokolokabilní, a také příslušné slovní tvary, které se z uvedených pěti předložek kombinovaly pouze s jednou nebo dvěma.

Ke stejnému závěru jako u tvarů 6. pádu je možné dojít také v případě trpných příčestí, která se zpravidla – vyjma eliptických konstrukcí – kombinují se slovesem *být*. Zde se jedná de facto jen o víceslovný tvar slovesa vyjadřující trpný rod, tedy o podobnou situaci jako v 1. a 2. osobě minulého času. Na takovou zdánlivou monokolokabilitu můžeme pravděpodobně narazit i v jiných jazycích – taktéž u složených slovesných tvarů, ale také u obligatorních členů, popř. jiných prvků vázaných na určitou gramatickou kategorii.

### 3.3.5 Interpunkce jako kolokát

Za zvláštní případ monokolokability vázané na gramatiku, či přesněji pravopis, můžeme považovat souvšyt s interpunkčními, popř. i jinými nealfanumerickými znaky. Zde se ovšem už nejedná o kolokace (a tudíž i monokolokabilitu) v pravém slova smyslu, jelikož nejde o spojení dvou či více slov, a nelze také mluvit přímo o vztahu monokolokabilního slova k takovému referenčnímu prvku. Případy, kdy narážíme na „monokolokabilitu“ vůči interpunkci, spojkám a podobným prvkům, tak přinášejí poněkud jiný typ informace o daném slově.

Vysokých hodnot HHI dosáhlo např. poměrně velké množství slov uvozujících přímou řeč (*vyhrknout, přitakat, zamumlat* apod.), kterým zpravidla předcházela interpunkce. Tyto případy se však vzhledem k přísným kritériím do konečného seznamu nedostaly – zpravidla totiž překračovaly tříprocentní hranici poměru jiných způsobů použití. Jindy, což je podobný případ, se jako monokolokabilní jeví prvky, jejichž obligatorní doplnění bývá vyjádřeno vedlejší větou, jako v případě slovních tvarů (*nebýt divu a vězte*). Ve výsledcích se objevuje také několik zkratk; tečka za zkratkou bývá v korpusu automaticky součástí daného slovního tvaru, v některých případech však je od něj oddělena a pokud k tomu dochází ve většině výskytů, vypadá taková zkratka (i v případě vyhledávání podle lemmatu) jako monokolokabilní. Vzhledem k takto nesystematickému zpracování a také k tomu, že tečka tvoří podle mého názoru spíše součást zkratky samé než samostatný referenční prvek, nebyly tyto výsledky do konečného seznamu zahrnuty. U slovních tvarů je pak typickým příkladem vokativ, jako např. *mylorde*. Často však narážíme současně na různé způsoby použití: vedle přímého

oslovování, které je podle pravidel pravopisu od ostatního textu vždy odděleno interpunkcí, se objevují různá rozvití oslovení (což má pochopitelně vliv na výsledky), popř. jiné konstrukce, kdy je slovo ve vokativu začleněno do syntaktické struktury. S obojím se setkáváme např. u tvaru *kamaráde*, které má na pozici L1 nejčastěji čárku (v malé míře další interpunkční znaky), ale také adjektiva *starý*, *milý*, *drahý* a *dobrý* a vyskytuje se i v konstrukci *říkat (někomu) kamaráde*. Zahrnutí tvarů začínajících velkým písmenem mezi klíčová slova by pravděpodobně vedlo také k jejich zastoupení mezi prvky s vysokým HHI, a to v souvislosti s psaním velkého písmene na začátku věty, čili bezprostředně za interpunkcí.

V některých případech však je daná spojka či interpunkce součástí širšího slovního spojení a skrze souvšyk s ní pak může být odhalena monokolokabilita v rámci celé kolokace. Příkladem může být spojení (*tak*) {či, nebo, ani, i, podruhé} *onak*, které je poměrně tvárné (objevilo se např. i v podobě *fyzickým osobám tak*, *církvím onak*), vždy však obsahuje prvky *tak* a *onak* spojené spojkou, interpunkcí nebo obojím.

Jak již naznačuje předchozí popis, bylo se spojkami, interpunkcí a podobnými kolokáty vzhledem k jejich zvláštnímu charakteru při analýze nakládáno poněkud jinak než s ostatními slovy. Všechny byly zahrnuty do jediné společné kategorie, jelikož bývají často navzájem substituovatelné a shodně vyznačují buďto koordinaci, nebo hranici nějakého syntaktického celku. Zároveň však z důvodu jejich odlišného vztahu k analyzovaným slovům lze jejich vzájemnou příslušnost ve velkém měřítku jen těžko posuzovat jinak než na základě vzdálenosti v textu. Jde o prvky natolik frekventované, že vyhledávání vzdálenějších výskytů než na sousední pozici, jaké jsem prováděla u ostatních kolokací, by vždy vedlo k pozitivnímu výsledku, ve většině případů by se však nejspíše jednalo o výsledky zcela bezvýznamné (dovedeno ad absurdum, nelze např. konstatovat monokolokabilitu vůči tečce jen proto, že dříve či později věta obsahující klíčové slovo tečkou končí). Jednotlivé prvky této společné kategorie se navíc často navzájem kombinují, což by pro zpracování znamenalo další značné komplikace. Z těchto důvodů byly za relevantní považovány pouze výsledky, v nichž se interpunkce nacházela přímo na sledované pozici vůči klíčovému slovu a nebyl ani brán ohled na případné prohřešky proti pravopisu (zejm. chybějící čárky).

### 3.3.6 Otázka (ne)dělitelnosti

O rozhodování, zda je možné a vhodné monokolokabilní spojení dělit na jednotlivé prvky, a o některých aspektech tohoto problému již byla řeč (viz např. oddíl 3.2.2). V krátkosti můžeme úvahy v tomto směru shrnout následovně:

Jelikož monokolokabilní slova poskytují jen málo možností, pokud vůbec nějaké, ke srovnávání různých výskytů a hledání analogií s jinými prvky, nelze jejich spojení automaticky popisovat jako jakékoli jiné kolokace. Rozdělení monokolokabilního spojení za účelem analýzy a popisu jednotlivých prvků (a skrz to zpětně popisu celého spojení) je zcela nevhodné u frazémů a u řady jiných kolokací může vést jen k opatrným závěrům. V mnoha případech je však oddělená analýza jednotlivých částí možná a taktéž způsob jejich spojení, vyjma své kvantitativní výlučnosti, se jeví jako pravidelný.

Jinou otázkou je status jednotlivých prvků v takové kolokaci. Je nasnadě, že monokolokabilní spojení jako celek vždy tvoří zvláštní (víceslovnou) lexikální jednotku. Je však na místě se ptát, zda jde skutečně o jednotku víceslovnou. Čermák (Filipec – Čermák 1985, s. 173-174) v tomto smyslu naráží na nesamostatnost a izolovanost monokolokabilních „verboidů“ v systému. Viděli jsme však, že řada monokolokabilních slov sdílí s výjimkou relativně volné kolokability většinu vlastností ostatních jednoduchých lexikálních prvků. Roli hrají také pravopisné zvyklosti a s nimi související vnímání slov coby jednotek. Situace je tedy opět poměrně členitá a závěry lze formulovat jen pro jednotlivé skupiny. V případě frazémů, kde není možné monokolokabilnímu slovu jednoznačně připsat některé formální a sémantické vlastnosti, se neochota označit takové prvky přímo za slova zdá opodstatněná. U reflexiv tantum se jako nesamostatný a vyprázdněný paradoxně jeví naopak referenční prvek, zvrtné zájmeno, které je navzdory formální oddělitelnosti možná vhodnější považovat za morfém (srov. PMČ, s. 69). V této souvislosti by monokolokabilita u reflexiv tantum spadala spíše na rovinu morfologickou. U ostatních typů monokolokabilních spojení, včetně monokolokability vázané na konkrétní formu, ovšem klíčové slovo můžeme považovat za zvláštní, byť nesamostatnou, lexikální jednotku.

Jak je vidět, rozšíření pojmu monokolokability i za hranice tradiční frazeologie a idiomatky má za následek značnou různorodost získaných monokolokabilních spojení a jejich prolínání s jinými typy kolokací. Při určování, které prvky je možné označit za monokolokabilní a které ne, tak nemluvíme jen o hranici ve smyslu maximálního počtu kolokátů, ale také o dalších vlastnostech jednotlivých prvků i vztahů mezi nimi. Ukazuje se

pak, že i u prvků, které splňují podmínky monokolokability co do počtu a frekvenčního poměru souvýskytu s různými kolokáty, je někdy vhodnější popsat taková spojení jiným způsobem.

### 3.4 Teorie a praxe

Teoretický popis monokolokabilních jednotek sám o sobě nepřináší nic podstatného, není-li propojen s praxí (v širokém slova smyslu). Jedná se samozřejmě o propojení s územ, z něž takový popis musí vycházet, což v našem případě znamenalo analýzu dat zpřístupněných v podobě korpusu. Neméně důležitá je však také otázka aplikace teoretických poznatků. U monokolokability coby lexikálního jevu to znamená v prvé řadě lexikografickou praxi.

Je nasnadě, že silně omezená kolokabilita slov by se měla projevit také v jejich popisu ve slovnících. Slovo, které připouští pouze jednu nebo několik různých kolokací, by mělo být ve slovníku popsáno způsobem, který na takovou vlastnost explicitně poukazuje. Jednoduchý přístup aplikovatelný u ostatních slov, kde je popsán význam a forma a uveden jeden nebo několik příkladů, zde není vhodný, jelikož neukazuje, se kterými synonymními výrazy se monokolokabilní slovo kombinovat může a se kterými ne. Na nedostatečné a mnohdy zavádějící údaje ve slovnících narážejí snad všichni autoři, kteří se touto problematikou blíže zabývají. Autoři jako F. Čermák či O. Káninský přitom pracují pouze s monokolokabilním materiálem frazeologické povahy, u něž se setkáváme i s dalšími anomáliemi a zvláštní popis je proto v případě takových slov obvyklejší. Stejnou pozornost je ovšem třeba věnovat i ostatním typům monokolokabilních spojení, jelikož nerespektování jejich kolokability vede k nesprávnému či stylisticky příznakovému používání stejně jako u frazémů.

U reflexiv tantum je situace poměrně jednoduchá; heslo tvoří celé spojení a to je pak popisováno jako celistvá jednotka. V SSJČ ovšem u některých nacházíme také nereflexivní (obvykle akuzativní) varianty, ty jsou však označeny jako sekundární a řídké či zastaralé. Např. pro *setmíti (se)* je zmíněna samostatná nereflexivní varianta ve stejném významu, doložená u jedné autorky, a vedle toho i nereflexivní varianta s akuzativní vazbou ve významu „učinit temným, tmavým“. V současné době se jedná zřejmě již o zaniklé a tudíž synchronně nerelevantní způsoby použití, které vypovídají o diachronním vývoji daného slova. U slovníků založených na textech „dobrých autorů“ ovšem může dojít i k zaznamenání použití, která vypovídají spíše o kreativě konkrétních spisovatelů než o



skutečné soudobé normě. Adjektivní reflexiva tantum nejsou zaznamenávána samostatně a jejich monokolokabilitu je tak třeba odvodit z toho, že základové sloveso je také reflexivum tantum.

Jiná je situace u slov valenčně vázaných. Typ valenční vazby bývá uveden buďto explicitně v rámci formálního popisu jako v SSJČ, nebo implicitně v příkladech jako v SSČ. Valenční vazby však má velké množství slov a pouze některá z nich jsou monokolokabilní. Rozdíl mezi nimi je přitom patrný jen skrze to, že některá slova se vážou se svými doplněními přinejmenším v některých významech bez předložky (a nebudou tedy monokolokabilní vůči předložce) a jiná mají jen předložkovou vazbu (což však nemusí znamenat, že se nemohou vyskytovat bez daného doplnění). Monokolokabilita valenčně vázaných slov tak z popisu ve slovníku není zřejmá. Ještě hůře je pak poznatelná pro deverbativní adjektiva, která nemají samostatná hesla, avšak, jak jsme viděli (srov. oddíl 3.2.4), mohou se svou valencí od základového slova lišit. Zvláštním případem jsou dále slova, jejichž kolokabilita je omezena i vůči jednotlivým doplněním a dochází tak k vícenásobné monokolokabilitě. Tak např. u *kontrastovat*, které se váže přes předložku s různými substantivy, je valence uvedena v SSČ v rámci příkladu („současnost k. s historií“) a v SSJČ zvláště v rámci formálního popisu. Vedle toho u *zalapat*, které se ve většině případů (nikoli však vždy) kombinuje se substantivy *dech* a *vzduch*, SSČ uvádí obě tyto kolokace a vykládá jejich význam (význam slovesa samého nikoliv), což napovídá, že v jiných spojeních se slovo nevyskytuje. V SSJČ najdeme o něco bohatší popis; valenční vazba je popsána stejným způsobem jako u *kontrastovat* (společně s řídkým použitím bez doplnění) a obě zmíněné kolokace jsou uvedeny v dalším popisu společně s jinými, řidšími příklady.

Pro skupinu víceslovných pojmenování se pak situace liší případ od případu. Absolutně monokolokabilní adjektivum *polohrubá (mouka)* má v obou slovnících samostatné heslo (stojí za zmínku, že jako základní je uveden tvar mužského rodu, přestože není doložen), resp. odrážku (v rámci slov začínajících morfémem *polo-*). Popis však tvoří přímo celá kolokace (podobně jako u *zalapat* v SSČ). SSJČ pak ještě vykládá její význam. Naproti tomu u adjektiva *náprsní* {kapsa, taška, kapsička} je uveden výklad významu stejně jako u kolokačně neutrálních slov a jednotlivé kolokace jsou zmíněny pouze jako příklady. SSJČ opět uvádí poněkud širší (a pravděpodobně neaktuální) možnosti použití.

Jak je vidět, monokolokabilita je ve slovnících zpracována velmi nesystematicky: v některých případech je na ni poukázáno (jako u reflexiv tantum či zmíněného spojení

*polohrubá mouka*), jindy se dá z uvedených údajů částečně odvodit, mnohdy však slovníkový popis bližší informace v tomto směru jednoduše neposkytuje. Lexikografická praxe zcela nereflktuje ani teorii formulovanou a již zavedenou F. Čermákem pro monokolokabilní spojení spadající do frazeologie a idiomatiky. Kupříkladu ke slovu *odiv*, které je přinejmenším v korpusu SYN2010 doloženo pouze ve spojení (*na*) *odiv* (a úzkou kolokabilitu vykazuje i vůči slovesům) najdeme i v nejnovějším vydání SSČ informace o gramatických kategoriích (uvedena je koncovka 2. pádu a jmenný rod, stejně jako u všech substantiv), což lze z hlediska Čermákovy teorie označit za nepodložené či zcela zcestné, a SČFI uvádí a popisuje význam pro běžné způsoby použití („stavět/vystavovat/vystavit něco na odiv“), ani zde však nenajdeme zmínku o monokolokabilitě.

Tato situace je o to více zarážející, že slovníky vznikající v současné době bývají založeny na korpusových datech a právě v korpusu je šíře kolokability jednotlivých slov, přinejmenším v rámci běžného úzu, snadno zjištělná. Dopad nedostačujícího popisu monokolokability ve slovnících je přitom zřejmý. Uživatel na jeho základě může získat falešné představy o možnostech použití konkrétních slov a vyjadřovat se na tomto základě nevhodným či zcela chybným způsobem. S reprezentací ve slovníku souvisí také začlenění monokolokabilních slov do výuky jazyka. Jelikož monokolokabilní spojení tvoří vždy celistvou lexikální jednotku, měla by být tímto způsobem také osvojována – holisticky, s upozorněním nemožnost kombinování monokolokabilního prvku s jinými slovy (vyjma případných dalších referenčních slov).

Specifickou problematikou je také vlastní povědomí mluvčích o monokolokabilních slovech, která znají a používají. U spojení frazeologického charakteru se můžeme např. setkat s otázkami typu „co jsou to ty holičky?“. Bylo by zajímavé sledovat, jakým způsobem jsou monokolokabilní slova uspořádána v mentálním slovníku a zda se jejich zpracování od ostatních slov nějak liší. K tomu by ovšem bylo třeba samostatného průzkumu a rozsáhlých analýz, které zdaleka přesahují rámec této práce. Výsledky případné studie takového typu by ovšem mohly ovlivnit také pohled na monokolokabilitu přijímaný v této a jiných pracích.

### **3.5 Shrnutí**

Tato kapitola představuje výsledky analýz provedených na datech, která byla získána způsobem popsáním v kapitole předcházející. Vzhledem k bohaté škále těchto dat a absenci či nedostatečnosti předchozích teoretických popisů, které vždy pracovaly pouze

s některými podtypy v ní obsaženými, bylo třeba formulovat základní klasifikační prostředky a popis vystihující charakter monokolokability v celé její šíři.

Monokolokabilní slova a jejich spojení tak byla nejprve roztříděna do několika kategorií podle různých hledisek. Zde formulovaná klasifikace tvoří podrobnější a univerzálněji použitelnou alternativu k třídění provedenému v mé klauzurní práci (2011). Jako kritéria klasifikace byly použity počet kolokátů, struktura vícečlenného kolokačního paradigmatu či paradigmat, ohebnost monokolokabilního prvku a slovnědruhovú příslušnost obou (příp. všech) prvků monokolokabilního spojení. Tato kritéria je samozřejmě možné doplnit dalšími podle potřeb teoretického popisu či konkrétního výzkumu.

V následující části pak byla monokolokabilní slova blíže popsána co do své formy i významu, přičemž bylo vždy poukázáno na různorodost monokolokabilního materiálu a vlastnosti typické pro jeho jednotlivé typy. Ukázalo se, že dříve formulované závěry o jeho neanalyzovatelnosti při zahrnutí případů mimo frazeologii a idiomatiku již situaci příliš neodpovídají. Zmíněny byly také některé problematické jevy, které práci s monokolokabilními slovy provázejí, a vztahy, jež můžeme najít mezi jednotlivými monokolokabilními prvky navzájem i mezi nimi a ostatní slovní zásobou.

Zvlášť byly vytčeny některé případy, kdy se prvek jako monokolokabilní čistě z kvantitativního hlediska jeví, je však sporné, zda mu tuto vlastnost můžeme bez výhrad přiřknout. Potvrzuje se tak, že hranice monokolokability není jen otázkou počtu kolokátů (přičemž i zde již byl konstatován škálovitý průběh), ale také vztahu mezi jednotlivými prvky v takovém spojení a konečně i povaha spojení jako celku. Na závěr pak bylo poukázáno na to, že monokolokabilita není jevem vymezeným jen v rámci teorie, ale existujícím v realitě jazykových projevů; teorie by přitom měla úzus respektovat, nesnažit se do něj zavést modely, které v něm reálně neexistují nebo přinejmenším nejsou doloženy, a naopak jej předkládat co nejuplněji a nejvěrněji. To platí jak pro teoretický popis, tak i pro lexikografickou praxi, která za teorií v některých případech poněkud pokulhává, a v neposlední řadě pro výuku jazyka.

Monokolokabilita se může zdát jako okrajový a v rámci jazyka nepodstatný jev. Naše data však ukazují, že se nejedná jen o blíže nespécifikovatelné prvky v několika frazémeh. Objevují se mezi nimi i na první pohled běžná slova, u kterých bychom možná monokolokabilitu ani neočekávali. Zároveň jde v mnoha případech o prvky poměrně frekventované. Naše databáze zaznamenávala jen kandidáty o minimální frekvenci 200 výskytů (v korpusu o velikosti 100 milionů slov) a i tak nebyla o monokolokabilní

výsledky nouze. Výjimkou nebyly případy o jednom i několika tisících výskytů. To ovšem nevylučuje, že velké množství monokolokabilních slov je podstatně vzácnější. Vzhledem k výše popsaným vlastnostem monokolokabilita spíše ukazuje, že jazyk je schopen pojmut i jevy, které se na první pohled přičí jeho základním principům.

Teoretický popis a úvahy, které tvořily obsah předcházejících částí, vycházely ze skutečného úzu, jak jej zprostředkovává korpus, a přijímaly monokolokabilitu ve všech jejích podobách, pokud tomu nestály v cestě opodstatněné argumenty. Jedná se však pouze o jeden z možných náhledů na problematiku monokolokability. Na tytéž jevy, které zde byly popsány, by bylo možné nahlížet jinými způsoby. Stejně tak by bylo možné náš popis dále rozvést a zahrnout body, na které zde nezbyl prostor. To už je úkolem pro případné příští studie a výzkumy.

## 4. Monokolokabilita v dalších jazycích

Předcházející kapitola formulovala vlastnosti vysledovatelné u monokolokabilních jednotek v češtině. Nejedná se však o jev typický jen pro ni; objevuje se i v řadě jiných jazyků a není vyloučeno, že ve všech.<sup>32</sup> Slovy F. Čermáka (1982, s. 209) se zdá, „že existence monokolokabilních slov bude jev přirozený ve všech živých jazycích“, což následně dokládá namátkovými příklady z devíti jazyků (nutno dodat, že se jedná s výjimkou finštiny o jazyky indoevropské). Upozorňuje však také, že „jejich rozpoznání je často velmi obtížné“. Na monokolokabilní prvky lze ovšem narazit i v jazycích geograficky i typologicky vzdálených, jako jsou např. tibetobarmské – o neprůhledných prvcích, které se objevují jen v jednom nebo dvou dvouslabičných kompozitech těchto jazyků, se zmiňuje J. A. Matisoff (1991).

V takové situaci se nabízí otázka, zda se monokolokabilní jednotky v různých jazycích podobají, popř. nakolik. Na ni se pokusí odpovědět následující kapitola. Není ovšem v silách jednoho člověka provést analýzu skutečně rozsáhlého a reprezentativního vzorku jazyků a domnívám se také, že pro účely ručního třídění a interpretace výsledků je nezbytné, aby výzkumník zkoumané jazyky ovládal. Po Stojčevovi, který pracoval současně s materiálem českým a bulharským, proto tato práce představuje studii monokolokability přítomné ve slovenštině, francouzštině a angličtině. Tento průzkum ovšem představuje spíše hrubý vhled do situace v těchto jazycích a nepokouší se ji vyčerpávajícím způsobem popsat, ale spíše porovnat s výsledky získanými pro češtinu.

Pro účely výzkumu byl použit paralelní korpus Intercorp verze 6 (z dubna 2013). Jedná se o korpus nereferenční a žánrově nevyvážený; jeho výhodou je však dostupnost, lemmatizace a gramatické značkování u zkoumaných jazyků a také přiměřená velikost – v době našeho zpracování měla francouzská i anglická část přes 60 a slovenská necelých 50 milionů slov. Vyhledávání monokolokabilních prvků proběhlo prakticky stejným způsobem jako předtím na češtině: zaznamenány byly všechny pozice (na rozdíl od češtiny nebyl brán ohled na slovnědruhovou příslušnost kvůli méně přesnému tagování) o absolutní frekvenci minimálně 200 výskytů a jejich frekvenční charakteristiky, jak jsou vypsány v oddílu 2.3.1. Poté byly vytvořeny odpovídající seznamy a prvky s nejvyššími hodnotami HHI byly ručně analyzovány a tříděny. Pro tento účel bylo použito rozhraní NoSke a jednojazyčná verze korpusu.

---

<sup>32</sup> Pochopitelně je řeč pouze o přirozených jazycích.

Po vyhodnocení pilotního průzkumu na části slovenských dat, který ukázal, že žádná sledovaná slova začínající velkým písmenem nebo velkými písmeny psaná celá nepřináší z našeho hlediska relevantní výsledky, byly všechny takové prvky v seznamech lemmat i slovních tvarů vyřazeny (ještě před ručním tříděním). Tímto způsobem se poměrně značně snížil poměr nerelevantních případů ve výsledcích. Jelikož se tentokrát jednalo o studie menšího rozsahu a spíše orientačního charakteru, analyzováno bylo jen 400 jednotek pro každý jazyk: po stovce pro lemmata a slovní tvary na pozicích L1 a R1. Počet nalezených monokolokabilních prvků se pochopitelně pro každý jazyk lišil. V případě nejasností byly nalezené výsledky porovnávány s popisem ve výkladových slovnících (Petit Robert pro francouzštinu a Cambridge Advanced Learner's Dictionary pro angličtinu). Další podrobnosti související s průběhem výzkumu jsou v případě potřeby uvedeny v jednotlivých studiích.

## 4.1 Slovenština

Slovenština je jazyk češtině vůbec nejbližší, jak z hlediska historického vývoje, tak značnou podobností obou jazyků. Dá se proto očekávat, že ve slovenské slovní zásobě narazíme na stejné typy monokolokability jako v češtině a není ani vyloučeno, že se oba jazyky budou shodovat i v konkrétních monokolokabilních slovech.

Výsledky získané v českém a slovenském průzkumu ovšem nejsou zcela srovnatelné, jelikož použité korpusy obsahují odlišné typy textů. Zatímco SYN2010 je žánrově vyvážený, slovenská část Intercorpu je tvořena z největší části právními texty z korpusu Acquis Communautaire a zápisy z jednání Evropského Parlamentu (korpus Europarl). Jádro, čili beletristické texty, má v současné době jen něco přes 8 milionů slov. Tato žánrová nevyváženost má pochopitelně vliv na výsledky. Více zastoupená jsou tak na jednu stranu slova charakteristická pro texty s právní a politickou tematikou a na druhou stranou slova natolik běžná, že se vyskytují i v těchto oborech.

Ze čtyř set zkoumaných prvků bylo při analýze za monokolokabilní označeno 139. Absolutní frekvence některých z nich byla velmi vysoká – sahala až do desítek tisíc případů. V souvislosti s tím byly vyřazeny některé vysoce frekventované prvky, které sice co do počtu kolokátů a nízkého poměru jiných způsobů použití splňovaly podmínky monokolokability, avšak počet takových jiných použití byl v absolutním měřítku velmi vysoký (stovky, výjimečně i tisíce výskytů). Tyto případy by pochopitelně bylo možné ve

výsledcích zanechat, představa stovek a tisíců výskytů jako „výjimek“ je však poněkud nepřesvědčivá.

#### 4.1.1 Klasifikace

Porovnáme-li získaná data s klasifikací formulovanou pro češtinu, zjišťujeme, že je na slovenštinu aplikovatelná stejně dobře, zastoupeny jsou prakticky všechny vyčleněné typy a prvky v nich mají podobné vlastnosti, které jsme viděli i v češtině.

Slovenština má slova absolutně monokolokabilní, mezi něž patří např. reflexiva tantum (*vzťahujúci sa, nachádzajúci sa*), slova valenčně vázaná (*odhliadnuť od, odchylne od*), spojení substantiva s předložkou (*na zreteli, v zahraničí*), frazémy (*mutatis mutandis, zoči – voči*) a další kolokace, jako např. (*mim.*) *vyd., valné (zhromaždenie), (daňový) poplatník* aj. Vícečlenné kolokační paradigma tvořené podobnými či přímo synonymními slovy mají např. slova *pitný* {voda, destilát, mlieko} či {sám, samý} *osebe*. Do této skupiny patřila především adjektiva monokolokabilní vůči substantivům a prvky, jež se mohou kombinovat s více předložkami. Objevovaly se ovšem i případy (zpravidla typu víceslovných pojmenování) polysémií, s jasně strukturovaným kolokačním paradigmatickým, jako např. *makrofinančná* {pomoc, stabilita} a *orný* {pôda, plocha, plodina, poľnohospodárstvo, región, zem} a několik slov vícenásobně monokolokabilních. Z nich uveďme spojení {v, na} *neposledný* {rad, miera, miesto}, {samostatne, samostatný} *zárobkovo* (činný) a víceslovné předložky tvořené kombinacemi {s, bez, na, o} *prihliadnutie* {na, k}.

Uvedené příklady již ukazují, že monokolokabilita ve slovenštině se té, kterou jsme viděli na českých datech, velmi podobá, a naznačují také podobnost při třídění podle druhého hlediska, totiž druhu referenčního slova. S výjimkou slovesa byly nalezeny referenční prvky všech druhů zaznamenaných pro češtinu. Odlišné však bylo jejich početní zastoupení. Zdaleka nejvíce byla zastoupena skupina slov monokolokabilních vůči substantivu, z nichž většinu tvořila adjektiva, výjimečně substantiva. Monokolokabilita gramatického typu, která byla v češtině velice silně zastoupena reflexivy tantum a různými spojeními s předložkou, byla v našich slovenských datech výrazně méně početná – obě jmenované skupiny měly po devíti zaznamenaných případech. Monokolokailita vůči interpunkci byla opět vzácná. Setkáváme se však se silnějším zastoupením dalších lexikálních referenčních slov – adjektiv a prvků nejasného slovnědruhového zařazení či taktéž monokolokabilních. Je možné, že tyto rozdíly jsou přinejmenším zčásti způsobeny

odlišností textů v obou korpusech. Pro slovenštinu zaznamenaná víceslovná pojmenování (tj. spojení, kde je referenčním prvkem substantivum) se pohybují v poněkud jiných tematických okruzích než české případy – velkou část tvoří taková, která patří k tématům diskutovaným v politické sféře či k politice samé (srov. např. skleníkový, ratingový, mimovládny, hazardný, predsedajúca apod.), kdežto v češtině jsou výsledky velmi různorodé. Právnické a politické prostředí se projevuje také ve vyšším počtu cizojazyčných frazémů, jako např. *mutatis mutandis* či *inter alia*. V této souvislosti se zdá, že uvedené rozdíly nevypovídají ani tak o odlišnosti v obou jazycích, ale spíše o odlišnosti zaměření použitých korpusů.

Také u posledního způsobu třídění – podle druhu monokolokabilního slova samého – se situace ve slovenštině zdá srovnatelná s češtinou. Setkáváme se s týmiž slovními druhy, ovšem opět v jiném poměru. Mezi lemmaty (u slovních tvarů početní poměry komplikuje vícečetná přítomnost slov monokolokabilních v různých svých tvarech) byla zdaleka nejpočetnější skupina adjektiv, jako např. *blížiaci (sa)* nebo *odstredené (mlieko)*. O mnoho méně a navzájem přibližně srovnatelně jsou zastoupena substantiva ({v, na, pre, kvôli} *krátkosť, daňový poplatník*), adverbia (*životne* {dôležitý, nevyhnutný, potrebný, závislý}, *vedecko-technický*) a slova slovnědruhově neurčitelná, jako např. *zoči (voči)* či {de, ipso, post} *facto*. U této poslední skupiny je počet případů zvýšen také několika cizojazyčnými frazémy, v nichž jsou monokolokabilní oba členy, srov. např. *inter alia* nebo *know how*. Mezi lemmaty se objevilo jen jedno sloveso, a to *odhliadnuť (od)*. To je situace velmi odlišná od češtiny, kde slovesa patřila k nejvíce zastoupenému slovního druhu a kombinovala se s různými typy referenčních slov.

Z typů vytčených v dřívější klasifikaci jsou přítomny frazémy, a to zejména cizojazyčné, pocházející z latiny (*mutatis mutandis, inter alia*) a angličtiny (*know how, opt* {out, in, outs}), ale i původní, jako např. *zoči (voči)*, (*na*) *zreteli* či *vzhľadom* {na, k}. Relativně slabé zastoupení domácích frazémů oproti přejatým je pravděpodobně způsobeno malým objemem beletristických textů a absencí textů publicistických. Reflexiva tantum jsou v datech zastoupena, narazila jsem však pouze na adjektivní. V češtině přítom tento podtyp tvořil téměř polovinu případů. Ve slovenštině mají základová slovesa v seznamu přítomných adjektiv většinou podstatně nižší hodnoty HHI a zajímavým faktem také je, že vyšších hodnot dosahují na levé straně, ne na pravé, jako je to běžné v češtině. To ovšem automaticky neznamena, že nejsou monokolokabilní. Zvratné zájmeno je v tomto případě zřejmě jen častěji vzdáleno od svého slovesa.



Velkou část dat tvořila víceslovná pojmenování, jako např. *limitný* {hodnota, kritérium, suma, koncentrácia, množstvo} či *katarální* (*horúčka*). V této skupině nebyly, s výjimkou koncentrovanějšího tematického zaměření, konstatovány výrazné rozdíly oproti češtině. Slova valenčně vázaných se objevilo pouze několik (např. *odchylne od, pochádzajúci* {z, od, s}).

#### 4.1.2 Bližší specifikace

Je zřejmé, že monokolokabilita v češtině a slovenštině má víceméně stejnou povahu. Klasifikace výsledků se od češtiny prakticky neliší, a to nejen v případě třídění formulovaného v této práci co nejobecněji použitelným způsobem, ale i u klasifikace do konkrétních typů kolokací, která byla vytvořena speciálně pro češtinu. Není pak překvapením, že formální a sémantické vlastnosti slovenských monokolokabilních prvků a spojení také nevykazují výraznější rozdíly.

Mezi prvky ve výsledném seznamu převažují částečně či zcela ohebné a sémanticky průhledné. Je to dáno i velkou převahou monokolokabilních adjektiv, která jsou zpravidla odvozena od substantiv či sloves (odtud postižitelnost významu) a co do zastoupení tvarů jsou někdy omezena charakterem svých referenčních slov, např. pokud se kombinují se substantivy pouze jednoho rodu (srov. {oxid, kysličník, chlorid} *uhličité* či *psychotropná látka*). Seznam monokolokabilních slovních tvarů obsahuje především jednotlivé tvary ohebných i neohebných monokolokabilních lexémů. Nejvíce tvarů, jedenáct, je zaznamenáno pro adjektivum *týkajúci* (*sa*). Nižší či přímo nulové zastoupení konkrétních tvarů jiných slov lze zřejmě vysvětlit převážně jejich nižší absolutní frekvencí. V případě velmi frekventovaného adjektiva *členský* došlo naopak k vyřazení lemmatu z konečného seznamu kvůli vysokému absolutnímu počtu jiných použití (214), zatímco v něm figuruje hned několik jeho tvarů. V případech, kdy bylo jako monokolokabilní zaznamenáno více tvarů jednoho lexému, se jejich kolokační paradigmata někdy mírně liší. Srov. např. adjektivum *zemný* {plyn, práca, val, násyp}: s tvary *zemný* a *zemného* se coby kolokát opakovaně objevuje pouze *plyn*, s tvarem *zemným* také *val*. Objevila se ovšem také řada slovních tvarů osamocených, kde se zřejmě jedná o monokolokabilitu pouze a právě uvedeného tvaru. Mezi ně patří kupříkladu *vzhl'adom* {na, k}, (*na*) *zreteli* či *pridanú* (*hodnotu*).

Jako monokolokabilní se jeví také některá slova, která tvoří součást složenin (resp. obecně výrazů obsahujících spojovník) a nejsou doložena v jiné podobě. Jedná se o

adverbia jako *vedecko* (-) {technický, výskumný, technologický, fantastický, výchovný, technicky} či *verejno* (-) {súkromný, právny, prospešný, propagačný}. Na podobné případy jsme přitom v češtině nenarazili. V korpusu SYN2010 jsou totiž zpravidla zapsány bez spojovníku a ani výskyty se spojovníkem nejsou rozděleny na jednotlivé části. Nejedná se tedy o rozdíl v uspořádání lexika obou jazyků, ale spíše o odlišnost jejich pravopisných konvencí či pouze tokenizace (rozdělení textu na jednotlivé pozice) v rámci jednotlivých korpusů. Přítomnost těchto příkladů ve slovenštině tedy nevypovídá o specifičnosti tohoto jazyka ve vztahu k monokolokabilitě; poukazuje však na další (byť pravděpodobně poměrně vzácný) typ monokolokability, který se týká jednotlivých částí kompozit a stojí tak vlastně na hranici mezi lexikální a morfologickou rovinou. Velmi zvláštním případem tohoto typu je pak slovo *e-mail*, rozdělené v korpusu na *(e) (-) mail*, a jeho odvozenina *(e) (-) mailový*. Analýza těchto výrazů jako kompozit není irelevantní (jde o zkratku z *electronic mail*), do češtiny a slovenštiny však byly přejaty již ve zkrácené podobě a je otázkou, nakolik je *e* mluvčími vnímáno jako slovo.

Při porovnávání jazyků navzájem si tak blízkých, jako je čeština a slovenština, může být zajímavé sledovat také konkrétní monokolokabilní případy a jejich protějšky v druhém jazyce. I v našich seznamech, byť postavených na poměrně různých textech, se našly příklady monokolokabilní v obou jazycích. Např. u *(de) facto* coby latinské přejímky se nejedná o nic překvapivého; v obou jazycích se však setkáme s týmiž domácími frazémy, jako např. *(na) úkor*. Odpovídají si i některá reflexiva tantum (*týkající se – týkající sa*), víceslovná pojmenování (*pozměňovací* {návrh, zákon} – *pozmeňujúci* {návrh, rozpočet, smernica, nariadenie, dohoda, opatrenie, okolnosť'})<sup>33</sup> a slova valenčně vázaná (*pocházející* {z, od, zevnitř, odjinud} – *pochádzajúci* {z, od, s}). Kolokační paradigma takových prvků, pokud je jednočlenné, bývá shodné (resp. je tvořeno slovy v obou jazycích si odpovídajícími), vícečlenná kolokační paradigmata se však obvykle překrývají jen částečně; shodný bývá nejfrekventovanější prvek, často i další, málokdy však všechny. To ovšem může být způsobeno i odlišnými typy textů, v nichž se taková slova v obou korpusech nacházejí.

## 4.2 Francouzština

Oproti slovenštině, která je s češtinou velmi blízce spřízněná, francouzština představuje jazyk již poněkud odlišný, stále však vzdáleně příbuzný a náležící do stejné

---

<sup>33</sup> V tomto případě ani nejde o slova utvořená přesně stejným způsobem.

kulturní oblasti. I zde proto můžeme očekávat relativně podobné výsledky, zároveň však i jevy v předchozích jazycích neexistující.

Francouzská část Intercorpu je tvořena z velké části automaticky zarovnanými texty. Stejně jako v případě slovenštiny se jedná o korpusy Acquis Communautaire a Europarl, navíc však i o publicistické a zpravodajské texty převzaté z projektů Presseurop a Project Syndicate. Beletrie má sama o sobě jen necelé 4 miliony slov, tedy ještě méně než slovenská. Opět se tak jedná o vzorek vzhledem k běžné normě nereprezentativní a výsledky tím pravděpodobně budou ovlivněny.

Jako monokolokabilní byly při analýzách označeny 133 slovní tvary a lemmata, což je množství srovnatelné se slovenštinou.

#### 4.2.1 Klasifikace

Teprve při pokusu o aplikaci na relativně odlišný jazyk, jako je francouzština, se ukazuje, nakolik je klasifikace formulovaná na základě češtiny skutečně použitelná i pro další jazyky.

Kritérium počtu referenčních slov ovšem nečiní žádné větší problémy. Ve francouzštině se setkáváme se všemi vytyčenými kategoriemi. Absolutní monokolokabilita se týká víceslovných pojmenování (*livre sterling, jour férié*), frazémů (*statu quo, parce que*), slov valenčně vázaných (*susceptible de, reposant sur*) i jiných typů spojení (*de recharge, c'est, hocher la tête*). Existuje zde i skupina reflexiv tantum, která byla v předchozích jazycích také případem absolutní monokolokability. Ve francouzštině ovšem narážíme na fakt, že reflexivní slovesa včetně reflexiv tantum se nekombinují pouze s jedním neutrálním zájmenem jako v češtině a slovenštině; zvrtné zájmeno se zde mění podle kategorie osoby. Je pak třeba se rozhodnout, zda tyto tvary (*me, te, se, nous, vous*) budeme považovat pouze za flektivní varianty jednoho lexému, nebo zda každé označíme za zájmeno odlišné od ostatních. Situaci dále komplikuje i homonymie tvaru 3. osoby jednotného a množného čísla (*se*). V případě, že se rozhodneme označit jednotlivé tvary za oddělené jednotky, jako je tomu u ostatních zájmen, pravděpodobně všechna (či téměř všechna) reflexiva tantum připadnou kategorii slov s víceprvkovým kolokačním paradigmatickým.

Do ní patří i řada jiných slov, převážně víceslovných pojmenování (*{entités, autorités} adjuducatrices, {voie, réseau, canal} navigable*) a slov valenčně vázaných (*opter {pour, en faveur}, passible {de, dont, du}*). Polysémické případy se objevují o něco

méně než ve dříve analyzovaných jazycích, srov. např. {pierre, déplacement} *angulaire*, {région, zone, situation, statut, territoire} *ultrapériphérique*. Vícenásobná mnokolokabilita je podobně jako v předchozích jazycích přítomná v malé míře, a to např. u *laps (de) (temps)*, *(d') ores (et) (déjà)* či {fibre, filament, fil} *(de) polyesters*.

Druhé kritérium, totiž druh referenčního slova, je také dobře aplikovatelné a přináší podobné výsledky jako v češtině a slovenštině. Nejpočetnější je skupina substantiv a předložek. Přítomnost reflexiv tantum implikuje také kategorii zájmen; do ní lze ovšem zařadit i tázací a vztažná zájmena vzniklá spojením s předložkou, jako např. *dont* či *auquel*, která bývají alternativním výrazem pro předložková spojení v určitých konstrukcích. Analogičnost použití předložky a patřičného zájmena ilustruje příklad (6). Slovo *destiné* je valenčně vázané přes předložku *à*, ta má ovšem ve francouzštině různé podoby podle toho, s jakým slovem splývá. V příkladu (6a) vidíme formu *aux* složenou z předložky *à* a určitého členu v množném čísle (*les*). Předmět *enfants* je přes ni připojen přímo za adjektivum. Oproti tomu (6b) představuje vedlejší větu, v níž předmět adjektivu předchází a je na něj odkázáno složeným vztažným zájmenem. To je tvořeno předložkou *à*, určitým členem *les* a vlastním zájmenem *quel*.

- (6) a. Nous avons également des inquiétudes concernant les jouets destinés aux enfants de moins de trois ans.

Existuje také důvod k obavám týkajícím se hraček určených pro děti do tří let.<sup>34</sup>

- b. (...) que cette aide et les financements octroyés parviennent vraiment aux enfants auxquels ils sont destinés et ne soient pas éparpillés.

(...) aby se přiměřená pomoc a potřebné prostředky skutečně dostaly k dětem, pro které jsou určeny<sup>35</sup>, a aby se s nimi neplýtvalo.

Podobně lze přistupovat i k příslovečným zájmenům *y* a *en* (či zájmenným příslovcím, terminologie zde není ustálená; ve francouzštině *pronom adverbial* či *adverbe pronominal*), jejichž slovnědruhové určení je však nejasné a v češtině jim odpovídají střídavě zájmena a příslovce. Adjektiva jako referenční slova jsou vzácná. Naopak v souvislosti se zvýšeným zastoupením přejatých latinských spojení je třída neurčitých a

<sup>34</sup> Jako překlad je použita odpovídající pasáž v české části Intercorpu.

<sup>35</sup> V české verzi této pasáže v korpusu je vedlejší věta vynechána, proto připojuji její překlad.

monokolokabilních referenčních slov poměrně početná. Ze skupiny spojek a interpunkce se zvlášť vyděluje *que*, které tvoří součást spojkových výrazů (fr. *locutions conjonctives*), v nichž obvykle stojí jako jediný kolokát, jako např. v *parce (que)*, nebo alternuje s předložkou jako v *afin {de, que}*.

Slovnědruhovú distribuce monokolokabilních prvků je na rozdíl od předchozích jazyků poměrně vyrovnaná. Mezi lemmaty jsou nejpočetnější substantiva (*{de, en} inuline, influenza {aviaire, A, d'origine aviaire}*) a adjektiva (*désireux de, partie intégrante*), jen o něco méně slova neurčitého zařazení (*tandis que, ad valorem*). Zhruba poloviční zastoupení oproti substantivům a adjektivům mají slovesa (*opter {pour, en faveur}, s'exclamer*). Adverbium nebylo zaznamenáno žádné, daly by se tak ovšem analyzovat některé výrazy tvořící součást víceslovné spojky, jako např. zmíněné *tandis (que)*.

Porovnáme-li výsledky s dřívější klasifikací, zjišťujeme, že reflexiva tantum jsou zastoupena, zvrátané zájmeno si však u nich ponechává schopnost vyjadřovat kategorii osoby (viz výše) a např. pro sloveso *écrier* tak máme zaznamenána tři různá zájmena *{se, me, nous}*, resp. čtyři, pokud *se* chápeme jako dvě homonymní zájmena pro 3. osobu jednotného a množného čísla. Víceslovná pojmenování jsou tvořena různými způsoby: nejčastěji kombinací monokolokabilního adjektiva a substantiva, jako např. u (*fièvre catarrhale*, jindy spojením dvou substantiv, z nichž jedno může být přejaté (*livre sterling*), nepřímým přívlastkem (*dioxyde de {carbone, soufre, azote, titane, silicium}*) a objevilo se i spojení monokolokabilního substantiva a adjektiva (*encéphalopathie spongiforme*). Mezi slovy valenčně vázanými nacházíme adjektiva (*désireux de*) a slovesa (*dénuer {de, du}*). Rozlišování mezi slovesným přičestím a od něj odvozeným adjektivem ovšem není vždy snadné; na rozdíl od českých a slovenských mají ve většině případů stejnou formu a kontext nemusí vždy umožňovat rozlišení pomocí dalších (morfologických a syntaktických) faktorů. V našich datech proto mohlo dojít k částečnému překrývání výsledků adjektivních a slovesných. Konečně skupina frazémů je zastoupena různými typy, především úslovími přejatými z latiny (*ad valorem, de minimis*) a víceslovnými předložkami a spojkami (např. *afin {de, que}* může sloužit jako obojí podle toho, které referenční slovo je použito), v menší míře jinými druhy (*au fur et à mesure, d'ores et déjà*), tento nepoměr však je zřejmě dán žánrovým složením korpusu. Stejně jako v předchozích jazycích i zde narážíme i na další typy spojení – opakovaně jde o kombinace předložky a substantiva (*en suspens, de recharge*) a vzácněji o jiné (*hocher la tête, et demie, cas échéant* či *c'est*).

## 4.2.2 Bližší specifikace

Jak je vidět, monokolokabilita ve francouzštině má velice podobný charakter jako v češtině a slovenštině. Má však také svá specifika. Jedním z nich je odlišná povaha zvratného zájmena u reflexiv tantum (viz oddíl 4.2.1). Dalším je vzájemná pozice prvků ve spojeních s adjektivem: adjektiva, a především relační adjektiva, totiž ve francouzštině stojí za substantivem, kdežto v češtině a slovenštině před ním. To se pochopitelně projevuje i na poměru slov monokolokabilních vůči prvkům na levé a pravé pozici. Na levou stranu je zde monokolokabilní většina sledovaných prvků (a na této straně také ve výsledcích výrazně pomaleji klesá hodnota HHI), kdežto v češtině a slovenštině byl poměr opačný: převažovala slova orientovaná napravo.

Pro francouzštinu je také typická podstatně chudší jmenná flexe. Substantiva a adjektiva ve francouzštině nerozlišují pád (pádům odpovídající vztahy jsou vyjadřovány předložkami) a nesou pouze koncovky rodu (u adjektiv) a čísla. Vyloučen tak je zde typ monokolokability slovního tvaru specifického pádu vůči odpovídající předložce nebo předložkám.

Naopak navíc oproti oběma předchozím jazykům má francouzština determinanty. Jedná se o poměrně širokou a různorodou skupinu prvků, od členů přes čísla až po přivlastňovací zájmena. Vyjma některých typů konstrukcí a stylistických efektů je přitom použití determinantu obligatorní. Poměrně velký počet substantiv se tak jeví jako monokolokabilní vůči determinantům, zvláště pokud mezi nimi převažuje jeden (extrémním případem je slovo *plupart*, jež se ze svých 5710 výskytů až na jedinou výjimku kombinovalo vždy s určitým členem *la*). Jedná se však o vázanost způsobenou gramatickými pravidly, která z lexikologického hlediska není zvláště zajímavá a lze ji z výsledků vyřadit stejným způsobem jako případy 6. pádu v češtině či víceslovných tvarů sloves. Determinanty se ovšem objevují i v rámci monokolokabilních spojení, pokud je referenčním slovem substantivum, a koneckonců jsou stejně obligatorní i pro monokolokabilní substantiva. V případech, kdy vysoká hodnota HHI odpovídá přítomnosti determinantů, je proto třeba zkontrolovat, zda se monokolokabilita netýká následujícího slova, ke kterému determinant patří, nebo předcházejícího slova, pokud determinant patří přímo ke klíčovému prvku. Jinými slovy, je třeba determinant jednoduše přeskočit. Zanedbáním tohoto postupu by byly přehlédnuty případy jako *hocher (la tête)* či *(par) le biais*. Podobné kolokace jsou mnohdy natolik zafixované, že neumožňují vysokou variabilitu použitelných determinantů běžnou u kolokačně neutrálních slov, a vysoké HHI

odpovídající determinantům tak v mnoha případech skutečně vypovídá o ustálenosti a potenciální monokolokabilitě v dané kolokaci, jíž je determinant součástí. Nelze ani zapomínat na složené výrazy, které determinanty, konkrétně určitý člen, obsahují, jako výše zmíněné *aux* a *auxquels*. Členy tak pronikají mezi referenční slova buďto společně s jinými determinanty jako samostatné, gramatikou vyžadované prvky, nebo jako součást alternativních výrazů k základnímu referenčnímu slovu (předložce).

Při pohledu na konkrétní příklady zaujme dvojice antonym {jour, semaine} *ouvrable* a (*jour*) *férié*. Stejně jako v předchozích jazycích však jde o ojedinělý případ podobného vztahu. S češtinou nebo slovenštinou, popř. oběma jazyky, francouzské výsledky sdílí některé latinské frazémy (*de facto*, *mutatis mutandis*). Ve všech třech jazycích najdeme víceslovné pojmenování označující CO<sub>2</sub>, monokolokabilní je však ve francouzštině substantivum, kdežto v obou druhých jazycích přívlastek, shodují se tak *de facto* jen v konkrétním názvu a ostatní související kolokace se liší pro každý jazyk, srov.: {oxid, kysličník, kyselina, koupel, voda, macerace, pramen, lázeň} *uhličitý* – {oxid, kysličník, chlorid} *uhličitý* – *dioxyde de* {carbone, soufre, azote, titane, silicium}. Další výrazy bývají málokdy přímo srovnatelné, jak to bylo možné v češtině a slovenštině. Přibližně si odpovídající výrazy však najdeme např. mezi víceslovnými předložkami (*na úkor* – *na úkor* – {au, à} *détriment*) a paralela mezi *hocher (la tête)* a *pokývat* {hlava, hlavička} a *zavrtět* {hlava, se, ocas, ocásek, zadek, hlavička, rameno, palec} (význam *hocher la tête* může podle kontextu odpovídat oběma českým výrazům) naznačuje, že ve francouzštině se také může vydělovat skupina sloves kombinujících se pouze s částmi těla.

### 4.3 Angličtina

Angličtina představuje jazyk se sledovaných tří vůči češtině nejvzdálenější a z hlediska monokolokability také nejobtížněji zpracovatelný. Vyhledávání a analýzu monokolokabilních případů komplikuje vysoká polysémie a homonymie prvků. V tomto případě příliš nepomáhá ani porovnání výsledků se slovníkem: Cambridge Advanced Learner's Dictionary uvádí jako samostatná hesla i relativně málo odlišné významy jednoho lexému a homonymie je tak těžko odlišitelná od polysémie. Angličtina má dále velmi slabou flexi a kategorie jako např. slovní druh jsou určovány především skrz vzájemnou pozici; hrozí tak cykličnost analýzy.

Při třídění jednotlivých prvků se však ukázalo, že tyto vlastnosti nepůsobí až tak velké problémy. Vysokých hodnot dosahovala především slova, která se s daným prvkem

nebo prvky kombinovala ve více svých významech a homonyma pak byla poměrně snadno identifikovatelná v rámci malého počtu zbývajících případů. Vysoká homonymie slov patřících k různým slovním druhům způsobuje, že příklady, kde je monokolokabilní jen jedno z homonym a to mezi ostatními výrazně nepřevažuje, dosahují nižších hodnot HHI a nebyly tak zahrnuty mezi sledovanými prvky.

Anglická část korpusu obsahuje automaticky zarovnané texty z databází Acquis Communautaire, Europarl, Project Syndicate a Presseurop. V tomto směru má podobné složení jako část francouzská. Beletristické texty v angličtině mají objem přes 10 milionů slov; to je sice nejvíce ze tří sledovaných jazyků, v poměru k automaticky zarovnaným textům s tematikou právníkou a politickou je to však stále málo. Publicistické a zpravodajské články z korpusů Project Syndicate a Presseurop také tento nepoměr příliš nevyvažují (dohromady mají necelých 5 milionů slov). Opět se tedy musíme spokojit s ne zcela reprezentativními daty.

Vzhledem k uvedeným vlastnostem anglických dat není překvapivé, že výsledný seznam monokolokabilních slov je podstatně kratší než u předchozích jazyků: obsahuje pouze 85 prvků. Celkový součet snižuje především malé množství slovních tvarů, kterých mohou monokolokabilní lexémy nabývat. I počet samotných lemmat je však nižší než v předchozích jazycích.

#### 4.3.1 Klasifikace

Vzhledem k odlišné povaze angličtiny lze očekávat, že klasifikace podle jednotlivých hledisek přinese poněkud odlišné výsledky.

Mezi absolutně monokolokabilními případy se stejně jako v ostatních jazycích objevily frazémy cizího, zejm. latinského původu (*per annum*, *inter alia*) a víceslovná spojka *insofar (as)*, dále slova valenčně vázaná (*adhere to*, *attributable to*) a víceslovná pojmenování (*bluefin tuna*, *life expectancy*). Případy jiného typu byly zaznamenány jen výjimečně; jednalo se o spojení (*two*) *thirds* a (*in*) *the meantime*. Mezi lemmaty tvořila absolutně monokolokabilní více než polovinu případů. O něco méně početná byla slova s víceprvkovým kolokačním paradigmatickým a objevily se zde podobné typy spojení, ale ve větší míře i jiné. Z frazémů uvedme latinské {ad, post} *hoc* a původní {under, through} *the auspices*, jako víceslovné pojmenování lze označit *outermost* {region, location, area, town, part, component}, které se však již blíží polysémii, či {asylum, job, fortune, Grail} *seekers*; valenčně vázané je adjektivum *pursuant* {to, thereto} nebo slovesný tvar *tends*



{to, towards}. Vedle těchto typů se objevila také adverbia monokolokabilní buďto vůči adjektivům (*vitally* {important, necessary, dependent, urgent}) nebo interpunkci (*fourthly*, *namely*). Zcela ojedinělý je případ *eh*, které se jeví jako monokolokabilní vůči interpunkci na levé straně. Jde o jediné zaznamenané citoslovce ze všech jazyků; odpovídající výraz se objevil také v seznamu francouzském, jeho kolokabilita však byla příliš široká. Zbývající dva vytyčené typy byly zastoupeny každý jen několika příklady. Nápadně členité kolokační paradigma vidíme u slova *conducive* {to, environment} či např. u slovesného tvaru *dealt* {with, in, a blow}. Vícenásobná monokolokabilita se pak týkala zejména víceslovných předložek (*in spite of*, {in, into} *accordance* {with, to, therewith}), z jiných typů se objevilo *pave the* {way, path} {for, to} a (*primus*) *inter (pares)*, což je však jen jedna ze tří možných kolokací slova *inter* {alia, partes, pares}.

Použijeme-li druhé kritérium pro třídění, totiž druh referenčního slova, vidíme, že oproti ostatním jazykům se spektrum referenčních slov podstatně užší. Zdaleka nejčastější jsou předložky, které mohou stát před i za monokolokabilním prvkem ({in, into, through, from} *the midst*, *irrespective of*). Druhou nejpočetnější skupinu tvoří slova nejasného zařazení; v případě angličtiny jde vždy o výrazy cizojazyčné (*per capita*, *force majeure*). Frazémy obsahující neanalyzovatelné prvky, jaké jsme viděli u předchozích jazyků, zde nebyly zaznamenány, nejspíše v důsledku své nízké frekvence v korpusu. Zastoupení substantiv coby referenčních prvků je ve srovnání s ostatními jazyky velmi slabé a srovnatelné se skupinou spojek a interpunkce. Jako příklady jmenujme *bluefin (tuna)* či (*life*) *expectancy*, vůči spojce bylo monokolokabilní *insofar (as)* a vůči spojkám a interpunkci *eh* či např. *namely*. Adjektivum se coby referenční prvek objevilo pouze v případě *vitally* {important, necessary, dependent, urgent}, ojedinělým případem je také číslovka v (*two*) *thirds*.

Mezi monokolokabilními slovy samotnými se objevují hojně substantiva (*in pursuance* {of, thereof}, {at, to, in} *the forefront*) a slovesa (*adhere to*, *complied with*), v menší míře adjektiva (*bluefin tuna*, *attributable to*) a výrazy slovnědruhově neurčitelné (*de facto*, *mutatis mutandis*, *wine psr*<sup>36</sup>). Další slovní druhy se objevují vzácně. Jedná se např. o adverbia *vitally* {important, necessary, dependent, urgent} či *fourthly* nebo o zcela ojedinělé citoslovce *eh*.

Vzhledem k obrazu, který nabízí uvedené klasifikace, není překvapivé, že v datech převažují spojení typu frazeologického (ať už přejaté cizojazyčné výrazy nebo domácí

---

<sup>36</sup> Psr je zkratka za „produced in specific regions“, jedná se o označení kvality.

víceslovné spojky a předložky) a slova valenčně vázaná. Víceslovných pojmenování je velmi málo a reflexiva tantum nenajdeme vůbec. Jejich absence je dána tím, že angličtina reflexivitu používá mnohem méně a jiným způsobem než předchozí jazyky. Většinou jevů, které jsou v češtině, slovenštině a francouzštině vyjadřovány reflexivem, popř. přímo reflexivem tantum, v angličtině odpovídají běžné aktivní či pasivní konstrukce. Srov. např. *sprchovat se – sprchovat' sa – se doucher – take a shower*. V prvních třech jazycích může být tato činnost vyjádřena jak uvedeným zvrtným slovesem, tak verbonominální konstrukcí (*dát si sprchu, dat' si sprchu, prendre une douche*), kdežto v angličtině se používá pouze tato konstrukce (stojí za zmínku, že je ve všech jazycích utvořena prakticky stejným způsobem). Podobně pouze reflexivní výraz *blížící se* odpovídá slovenskému taktéž pouze reflexivnímu *blížiaci sa* a francouzskému *s'approchant*, které ovšem už existuje i v nereflexivní podobě, anglické ekvivalenty (*approaching, coming* apod.) jsou však nereflexivní. Jiné typy než tři zmíněné se objevují poměrně málo. Většina substantiv spojujících se s předložkou na levé pozici jsou víceslovné předložky, které se na pravou stranu vážou často přes *of*, jindy však přes posesivní determinant (*{in, into, through, from} the midst of...* či *{in, into, through, from} their/our/its midst*) nebo mohou nést i plný význam: např. *{on, to, upon} the verge* může znamenat „na pokraji (něčeho)“, ale také „na břehu (řeky apod.)“. Zvláštní skupinu tvoří adverbia spojující se s interpunkcí (*fourthly, namely*) a adjektivy (*vitally* *{important, necessary, dependent, urgent}*), ojedinele se objevují i další druhy spojení. Mimo jakýkoli dříve zaznamenaný typ se vydělují kolokace (*two*) *thirds* a spojení *eh* s interpunkčními znaky.

#### 4.3.2 Bližší specifikace

Z výše uvedeného je patrné, že monokolokabilní materiál získaný pro angličtinu je o něco chudší nejen co do kvantity, ale také rozmanitosti. Jelikož objem i charakter výchozích textů byl s předchozími dvěma jazyky srovnatelný, jedná se zřejmě přinejmenším zčásti skutečně o vlastnost angličtiny jako takové a ne jen o důsledek nedostatečného vzorku.

O absenci skupiny reflexiv tantum již byla řeč; angličtina je dále charakteristická používáním determinantů. Jejich typy a způsob používání jsou v hrubých obrysech srovnatelné s francouzskými. Na rozdíl od francouzských však nesplývají s jinými slovy a situace je tak o něco přehlednější. Determinanty mohou být vynechány pouze v určitých konstrukcích a stylech a také v množném čísle (angličtina nemá neurčitý člen pro množné

číslo). Řada slov tak dosahuje vysokých hodnot HHI na pozici, kde stojí determinant, a to zvláště v případech, kdy mezi výskyty převažuje určitý člen. Stejně jako v případě francouzštiny se však jedná o jev čistě gramatický a za monokolokabilitu jej proto nepovažujeme. Přítomnost determinantů měla pro analýzu stejné důsledky jako ve francouzštině: tam, kde vysoké HHI odpovídalo přítomnosti determinantu, bylo třeba bezprostředně sousedící pozici přeskočit a zaměřit se na následující, jelikož se mohlo jednat o monokolokabilitu vůči dalšímu členu kolokace. Tímto způsobem byly zaznamenány především některé výrazy s adverbialní funkcí a jim odpovídající víceslovné předložky, jako např. {in, into, through, from} *the midst* a {from, at, since} *the outset*, příp. jiné výrazy obsahující substantivum, jako např. *pave* {the way, the path}. Posesivní determinanty také často tvořily alternativu k předložkové vazbě přes *of*. V takových případech však nebyly zaznamenány jako další referenční slovo, ale jako jiný způsob použití.

Jestliže ve francouzštině nacházíme podstatně méně flexe než v češtině a slovenštině, angličtina jde v tomto směru ještě dál. Ohebná jsou pouze substantiva a slovesa (adjektiva, která ve francouzštině vyjadřují rod a číslo, jsou zde neohebná) a i pro ně se flexe omezuje jen na několik málo koncovek. V důsledku toho vyhledávání podle slovních tvarů přineslo téměř duplicitní výsledky k vyhledávání podle lemmat – pokud má monokolokabilní slovo jen jeden tvar, bude tento tvar taktéž monokolokabilní, a pokud je ohebné, nemívá více než pět různých tvarů, které zpravidla mohou vyjadřovat více různých gramatických kategorií a nést více funkcí, čímž se opět blíží výsledkům odpovídajícího lemmatu. Zároveň nízká variabilita tvarů způsobuje, že jednotlivé tvary častěji přesahují minimální frekvenci 200 výskytů. V seznamu se však objevily i slovní tvary, jejichž lemma se jako monokolokabilní nejevilo nebo dosáhlo (třeba jen z důvodu homonymie s jiným slovním druhem) nižší hodnoty HHI a nebylo proto analyzováno. Jedná se např. o sloveso *comply*, které je v seznamu zastoupeno tvary *complies*, *complied* a *complying* (všechny monokolokabilní vůči předložce *with*), tedy všemi možnými s výjimkou základního, který byl vyřazen kvůli nereprezentativním výsledkům v korpusu (k tomu více viz podkapitola 4.4) a jeho zdaleka nejvyšší absolutní frekvence přenesla tuto nereprezentativnost na celý lexém reprezentovaný lemmatem. Jindy konkrétní slovní tvar nese poněkud jiný význam či funkci než ostatní; to je případ *according*, které bylo automaticky označeno jako tvar slovesa *accord*, drtivou většinu jeho výskytů však tvoří víceslovná předložka *according (to)*.

Vedle determinantů byly zaznamenány i další druhy „monokolokability“ způsobené čistě gramatickými či pravopisnými pravidly. Jednalo se o nespisovné zkrácené tvary slov, po nichž vždy následuje apostrof (*jus'* vedle *just*, *goin'* místo *going*), a o tvary některých sloves, která při záporu splývají se záporkou. Jejich segmentace v korpusu přitom byla různá, srov. *isn + 't* či *hadn + 't* proti *ca + n't* a *wo + n't*. Tyto případy můžeme považovat čistě za chybu segmentace v korpusu nebo je analyzovat jako spojení dvou jednotek morfologické roviny, v každém případě však nejde o monokolokabilitu v tom smyslu, jak o ní uvažujeme v této práci.

Při srovnání konkrétních příkladů s předchozími jazyky docházíme k podobnému závěru jako ve francouzštině. Shodují se některé cizojazyčné výrazy (*ad valorem*, *mutatis mutandis*, *de facto*) a přibližně si odpovídají některé víceslovné předložky, např. české {na, při, k} *pokraji* a {on, to, upon} *the verge*. Jindy přítomnost analogických příkladů souvisí s tím, že texty obsažené v Intercorpu se pro jednotlivé jazyky zčásti překrývají: nacházíme např. dvojice *modroplutvý (tuniak) – bluefin (tuna)* a {région, zone, situation, statut, territoire} *ultrapériphérique – outermost* {region, location, area, town, part, component}.

## 4.4 Shrnutí

V této kapitole jsem použila metody a popisné prostředky formulované pro češtinu na tři další jazyky z evropského prostředí. Jelikož pro potřeby analýzy bylo nutné zkoumaným jazykům rozumět, výběr jazyků se řídil především jejich znalostí z mé strany a také dostupností dostatečně velkého a lemmatizovaného korpusu. Lze však předpokládat, že stejným způsobem by bylo možné pracovat s mnoha dalšími jazyky. Univerzálnost použité metody by ovšem byla podrobena opravdové zkoušce teprve při aplikaci na jazyky zcela odlišné stavby.

Analýzy ukázaly, že klasifikace vytvořená v této práci je vhodná pro materiál všech zkoumaných jazyků; ve všech jazycích jsme také narazili víceméně na stejné typy známé již z češtiny. V případě slovenštiny byly výsledky velmi podobné českým co do typů, povahy i konkrétních příkladů. Ve francouzštině jsme již narazili na určité odlišnosti související především s odlišnou morfologií. Jako nejodlišnější a zároveň nejobtížněji analyzovatelná se jeví angličtina. Není bez zajímavosti, že jednotlivé jazyky se shodují i v některých konkrétních příkladech, a to nejen v přejatých cizojazyčných výrazech. Zčásti může jít o vzájemný vliv dotyčných jazyků (např. o kalk víceslovné předložky existující

v jiném jazyce; zvláště angličtina je pro mnohé evropské jazyky zdrojem přejímání slov i frází). Přehled výsledků vykazujících monokolokabilní chování je uveden v příloze 3.

Při zpracování dat se také ukázalo, že výsledky jsou do velké míry ovlivněny složením korpusu. Výskyty v právnických textech z databáze Acquis Communautaire a v prepisech projevů v Evropském Parlamentu (Europarl) u většiny slov početně převažovaly nad ostatními typy textů. V běžném úzu přitom jde spíše o minoritní záležitost. Téhož nepoměr mohl ovlivnit četnost jednotlivých referenčních slov ve víceprvkových kolokačních paradigmatech. Poměrně slabé zastoupení beletrie a publicistických textů také souvisí s nízkým počtem domácích frazémů ve výsledcích (s výjimkou víceslovných předložek a spojek, které jsou v odborných textech běžné), jelikož neumožnilo jejich dostatečně časté opakování. U řady slov navíc došlo k tomu, že se širší kontext některých výskytů nápadně podobal či přímo shodoval. Jednalo se pravděpodobně o vícenásobnou přítomnost těchto textů nebo pasáží. V případech, kdy takové duplicitní výskyty tvořily podstatnou část dat, byl dotyčný prvek označen a vyřazen z konečných výsledků, jelikož údaje o něm vycházely i v rámci Intercorpu z nereprezentativních dat. Extrémním případem bylo např. francouzské slovo *échinodermes*, které se jevilo jako monokolokabilní vůči čárce a spojce *et*, jednalo se však pouze o důsledek drtivé převahy jediného kontextu (výčtu mořských živočichů, dokonce ve stejném pořadí, v textech pojednávajících o jejich lovu a dovážení) ve výsledcích.

Tyto obtíže ukazují, že Intercorp, přinejmenším ve své současné podobě, není ideálním zdrojem pro analýzu jevu, jako je monokolokabilita. Pro předběžný průzkum mapující základní rysy monokolokabilních slov a srovnávající je navzájem a s rysy nalezenými v češtině, což bylo cílem této části, stačil. Podrobnější a spolehlivější analýzy by však již vyžadovaly větší a především žánrově vyváženější korpus.

## 5. Závěr

Tato práce představuje průběh a výsledky zatím asi nejrozsáhlejší české korpusové studie na téma monokolokability. Přijímá poměrně široký pohled na monokolokabilitu a její projevy, zároveň však přináší i poměrně podrobný popis vybraných jevů s ní spojených.

Zabývá se nejprve otázkou, jakým způsobem můžeme monokolokabilní jednotky a kolokace v korpusu vyhledávat. Navzdory relativně jednoduchému požadavku formulovanému podle statistických vlastností monokolokabilních slov pro vyhledávání zatím neexistuje jednoduchý a vysoce efektivní nástroj, který by přinesl (téměř) všechna relevantní data a zároveň do nich zanesl jen minimum nežádoucích výsledků. Byla zde proto vyzkoušena poměrně málo používaná metoda vyhledávání podle diversity bezprostředního kontextu formou Herfindahl-Hirschmanova indexu (HHI) a entropie, které ovšem přinesly srovnatelné výsledky a pro účely studie tak byl následně využit HHI. Data touto metodou získaná bylo nutno opět ručně roztřídit, podíl monokolokabilních slov v nich však byl poměrně vysoký a zároveň se jednalo o výsledky různého charakteru.

Na základě analýzy, která proběhla současně s tříděním, pak byla formulována nová klasifikace, která měla za cíl zahrnout pokud možno všechny nalezené výsledky a zároveň být použitelná i pro jiné jazyky. S ohledem na tyto cíle byla vystavěna na velmi obecných kritériích, jimiž byla velikost a struktura kolokačního paradigmatu, druh referenčního slova či slov a příslušnost či podobnost monokolokabilního prvku jednotlivým slovním druhům. Toto třídění je možné použít současně s klasifikací uvedenou v mé klauzurní práci, která z uvedených vlastností vždy koncentruje vícero najednou a s novým tříděním se tak částečně překrývá, nezahrnuje však na rozdíl od něj všechny nalezené typy výsledků.

S ohledem na tuto klasifikaci pak byl podán bližší formální a sémantický popis monokolokabilních jednotek a kolokací. Ukázalo se, že monokolokabilní jednotky zahrnují poměrně širokou škálu vlastností. Našly se zde prvky očividně defektní a těžko popsitelné, ale také slova zcela pravidelná po všech stránkách s výjimkou kolokační charakteristiky. Nejednalo se ovšem o popis zcela systematický a vyčerpávající, jelikož v tomto směru může dobře posloužit diplomová práce S. Stojčeva, ale spíše o pohled doplňující Stojčevovy poznatky vztahující se ke skupině monokolokabilních komponentů frazémů o další druhy dat, které byly v této práci zahrnuty. Vedle obecného popisu jsem se zaměřila

také na vybrané specifické jevy a na některé typy kolokací a spojení, které se monokolokabilitě podobají, z nějakého důvodu však lze pochybovat o vhodnosti takového označení. Závěrem jsem pak připomněla také důsledky, které by teorie popisující monokolokabilní jednotky měla mít pro praxi, a to zejména v oblasti lexikografie a výuky jazyků.

Třetí kapitola pak doplnila předchozí poznatky o hledisko srovnávací. Stejným způsobem jako v češtině byly vyhledány monokolokabilní případy ve slovenštině, francouzštině a angličtině. Pro tyto účely byl použit paralelní korpus Intercorp; jeho nevyváženost co do typu a tematického zaměření textů měla vliv na získané výsledky, pro orientační přehled o škále monokolokabilních spojení však vystačil. Ukázalo se, že všechny sledované jazyky obsahují víceméně stejné typy monokolokability, odlišené především morfologickými vlastnostmi jednotlivých jazyků. Obecná klasifikace formulovaná v předchozí kapitole byla na všechny bez obtíží aplikovatelná. Uplatnila se i klasifikace na konkrétní typy, ačkoliv ani pro tyto jazyky nedokázala pokrýt celou škálu monokolokabilních spojení. Jednotlivé jazyky se lišily především ve skupině reflexiv tantum; ta byla ve slovenštině velmi podobná českým výsledkům, francouzština se však lišila ohebností zvrtného zájmena a v angličtině tento typ úplně chyběl. V angličtině a francouzštině dále do monokolokabilních spojení často vstupovaly determinanty a jejich obligatorní přítomnost u substantiv zanášela nový typ „šumu“ do výsledků získaných pomocí indexu HHI. Výsledky se v jednotlivých jazycích také lišily v důsledku různé míry flexe jednotlivých slovních druhů. Mnohdy se ovšem naopak stalo, že se dva nebo více jazyků shodovaly v konkrétním monokolokabilním prvku, byť jeho kolokační paradigma se v těchto jazycích obvykle překrývalo jen částečně. V obecné rovině je možné říci, že přinejmenším ve čtyřech sledovaných jazycích má monokolokabilita velmi podobnou povahu a pokud to umožňuje stavba jazyka, týká se podobných či stejných typů kolokací.

Poznatky a úvahy zde uvedené ovšem zdaleka přesahují představu monokolokability, jak ji prezentuje teorie F. Čermáka a dalších autorů, kteří jeho pohled přijímají. V tomto směru představovala poměrně razantní krok už má klauzurní práce, výsledky současné studie však poukazují i na její nedostatečnost. Monokolokabilita je jev podstatně širší a rozšířenější, než jaký zahrnuje standardní frazeologie a idiomatika, popř. její okrajové oblasti (např. terminologie). Paradoxně některé často zmiňované monokolokabilní příklady se v našich datech naopak vůbec neobjevily. V případech řídkých frazémů se pravděpodobně jednalo pouze o jejich nízkou absolutní frekvenci. Kupříkladu adjektivum *širý*, pro které F. Čermák a další autoři uvádějí kolokáty {nebe, moře, svět}, by však podle

našich měřítek jako monokolokabilní vůbec označeno nebylo.<sup>37</sup> Zdá se tedy, že teorie popisující monokolokabilitu si žádá důkladnou revizi, a dost možná by si pozornost zasloužila i definice monokolokability samotná, z níž jsem ve své studii bez větších výhrad vycházela, která však sama o sobě představuje kompromis mezi realitou a její zobecnitelností.

Popis a úvahy zde uvedené pochopitelně téma monokolokability ani zdaleka nevyčerpávají. I nadále by bylo možné je rozvádět či naopak zpochybňovat a přinášet jiná vysvětlení. Systematický a vyčerpávající popis monokolokability ve všech jejích podobách a omezeních by nejspíše vydal na celou monografii. Tato práce tedy přináší poznatky a popis odpovídající jejím možnostem, zároveň však ponechává značný prostor k dalším hypotézám a analýzám, ať už na větším či jinak zaměřeném psaném korpusu nebo na zcela jiném typu dat.

---

<sup>37</sup> V korpusu SYN2010 nacházíme 1296 výskytů. Zmíněné tři kolokáty jsou v nich nejčastější, i při započtení dalších pěti nejčetnějších (osmičlenné kolokační paradigma by tak obsahovalo prvky {nebe, moře, svět, okolí, pláň, oceán, pole, kraj}) však zbývá 231 jiné použití, což je téměř 18% všech výsledků.



## Literatura

- Bozděchová, I. (2004). Nad tratolištěm krve. *Naše řeč, roč. 87, č. 2.* 107-110.
- Cvrček, V. (v tisku). *Kvantitativní analýza kontextu*. Praha: Nakladatelství Lidové noviny.
- Čermák, F. (1982). *Idiomatika a frazeologie češtiny*. Praha: Univerzita Karlova.
- Čermák, F. (2007). *Frazeologie a idiomatika česká a obecná. Czech and General Phraseology*. Praha: Karolinum.
- Čermák, F., & Blatná, R. (1995). *Manuál lexikografie*. Jinočany: H & H.
- Filipec, J., & Čermák, F. (1985). *Česká lexikologie*. Praha: Academia.
- Filipec, J. (1984). Česká frazeologie. *Naše řeč, roč. 67, č. 3.* 139-142.
- Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica 45/1.* 31–80.
- Dostupné z:  
<http://www.eva.mpg.de/lingua/staff/haspelmath/pdf/WordSegmentation.pdf>
- Hirschman, A. O. (1964). The Paternity of an Index. *The American Economic Review, Vol. 54, No. 5.* 761.
- Dostupné z: <http://www.jstor.org/stable/1818582>
- Hnátková, M. (2006). Typy a povaha komponentů neslovesných frazémů z hlediska lexikálního obsazení. In: F. Čermák, & M. Šulc (Eds.), *Kolokace* (142-167). Praha: Nakladatelství Lidové Noviny.
- Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge: MIT Press.
- Káninský, O. (2003). *Frazémy s monokolokabilním komponentem v češtině (zpracováno na materiálu Českého národního korpusu)*. Nепublikovaná diplomová práce. Praha: Univerzita Karlova.
- Karlík, P., Nekula, M., & Pleskalová, J. (2002). *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny.
- Koček, J., Kopřivová, M., & Kučera, K. (2000). *Český národní korpus: úvod a příručka uživatele*. Praha: Filozofická fakulta UK.
- Matisoff, J. A. (1991). Lexicography of Other Tibeto-Burman Languages. In: Hausmann, F. J. et al. (Eds.), *Dictionaries. An International Encyclopedia of Lexicography, Vol. 3.* 2555-2560.

Dostupné z:

<http://books.google.cz/books?id=Kwqc7xso22wC&printsec=frontcover&hl=cs#v=onepage&q&f=false>

Moon, R. (1998). *Fixed Expressions and Idioms in English. A Corpus-Based Approach*. Oxford: Clarendon Press.

Richter, F., & Sailer, M. (2003). Cranberry Words in Formal Grammar. In: Beyssade, C., Bonami, O., Hofherr, P., C., & Corblin, F. (Eds.), *Empirical Issues in Syntax and Semantics 4*. Paris: Presses Universitaires de Paris Sorbonne.

Dostupné z:

[http://books.google.cz/books?id=qAsw56kBpKwC&dq=Empirical+Issues+in+Formal+Syntax+and+Semantics+4&hl=cs&source=gbs\\_navlinks\\_s](http://books.google.cz/books?id=qAsw56kBpKwC&dq=Empirical+Issues+in+Formal+Syntax+and+Semantics+4&hl=cs&source=gbs_navlinks_s)

Shannon, E. C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, Vol. 27. 379–423, 623–656.

Dostupné z: <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>

Stojčev, S. (1984). *Monokolokabilní slova v idiomatice a frazeologii na materiálu češtiny a bulharštiny*. Nepublikovaná diplomová práce. Praha: Univerzita Karlova.

Šíková Hrejsemnová, L. (2011). *Monokolokabilní a extrémně málo kolokabilní slova v češtině*. Nepublikovaná klauzurní práce. Praha: Univerzita Karlova.

Šíková Hrejsemnová, L. (2012). *Vybrané případy monokolokabilních slov ve francouzštině*. Nepublikovaná klauzurní práce. Praha: Univerzita Karlova.

Šulc, M. (1999). *Korpusová lingvistika: první vstup*. Praha: Karolinum.

## **Slovníky a jazykové příručky:**

Čermák, František – Hronek, Jiří – Machač, Jaroslav et al. (1988). *Slovník české frazeologie a idiomatiky. Výrazy neslovesné*. Praha: Academia.

Čermák, František – Hronek, Jiří – Machač, Jaroslav et al. (1994). *Slovník české frazeologie a idiomatiky. Výrazy slovesné. A-P, R-Ž*. Praha: Academia.

Červená, V., & Mejstřík, V. (2003). *Slovník spisovné češtiny pro školu a veřejnost: s Dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky*. Praha: Academia.

- Havránek, B., & Doležel, L. (1971). *Slovník spisovného jazyka českého. A-M, N-Q,, R-U, V-Ž*. Praha: Academia.
- Karlík, P., Nekula, M., & Rusínová, Z. (Eds.). (1995). *Průruční mluvnice češtiny*. Praha: Lidové noviny.
- Robert, P., Rey-Debove, J., & Rey, A. (Eds.). (2010). *Le Petit Robert: dictionnaire alphabétique et analogique de la langue française*. Paris: Le Robert Editions.
- Woodford, K. (Ed.). (2003). *Cambridge advanced learner's dictionary*. Cambridge: Cambridge Univ. Press.

## Použité korpusy:

- Český národní korpus - SYN2010. Ústav Českého národního korpusu FF UK, Praha 2010.  
Dostupný z: <http://www.korpus.cz>
- Český národní korpus - InterCorp. Ústav Českého národního korpusu FF UK, Praha.  
Cit.15.08.2013 , dostupný z: <http://www.korpus.cz/intercorp>
- Čermák, F., & ROSEN, A. (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics. Vol. 13, no. 3.* 411–427.
- Hajič, J. (2004). *Disambiguation of Rich Inflection (Computational Morphology of Czech)*. Vol. 1. Praha: Karolinum Charles University Press.
- Jelínek, T. (2008). Nové značkování v Českém národním korpusu. *Naše řeč, roč. 91, č. 1.* 13-20.
- 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*. Bratislava: Veda. 26-44.
- Spoustová, D., Hajič, J., Votrubec, J., Krbeč, P., Květoň, P. (2007). The Best of Two Worlds: Cooperation of Statistical and Rule-Based Taggers for Czech. In: *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing*. ACL 2007, Praha. 67-74.

# Přílohy

## Příloha 1: Krajní hodnoty

V této příloze jsou uvedeny tabulky s vybranými částmi seznamů kandidátů na monokolokabilní slova, odpovídající různým sledovaným kritériím.

### 1a. Lemmata a slovní tvary s nejnižším počtem typů na pozici L1

Počet typů	Lemma
1	odiv
2	takž
3	potaz
3	krážem
4	rozloučená
4	mání
4	ustání
4	rozmyšlená
4	přestání
4	překl
5	třetice
5	nepoznání
6	dostání
6	facto
6	protisměr
6	shledaná
6	nynějšek
6	dřívějšek
6	vysvětlená
7	hodiny

Počet typů	Forma
1	lubem
2	odiv
3	takž
3	krážem
4	nedávnem
4	potaz
4	třetice
4	nepoznání
4	přestání
4	holičkách
5	rozloučenou
5	uhličitého
5	mání
5	ustání
5	rozmyšlenou
5	mylorde
6	tuzemsku
6	nedohlednu
6	nynějška
6	nepaměti

1b. Lemmata a slovní tvary s nejnižším počtem typů na pozici R1

Počet typů	Lemma
1	de
1	tab
1	křížem
1	pod
3	pozměňovací
3	přesilový
4	nadmořský
4	kutný
4	zdaňovací
5	bobkový
5	horšovský
6	odehrávající
6	idnes
6	mezipohlavní
6	překl
7	zbrusu
7	polohrubý
7	zkrátka
7	štitný
8	výkladní

Počet typů	Forma
1	aj
1	tab
1	ap
1	orig
1	pod
2	zabývajících
3	vězte
4	nadmořské
4	zorného
4	odehrávající
5	psacímu
6	apod
6	obývacího
6	laskav
6	pokutového
6	idnes
6	rozhodně
7	pokutový
7	polohrubé
8	opepříme

1c. Lemmata a slovní tvary s nejnižším součtem typů na pozicích L1 a R1

Počet typů	Lemma
9	překl
16	pod
19	idnes
26	red
27	operka
30	mylady
30	rozmyšlená
30	zatajený
32	přihlédnutí
32	zakřenit
36	nejinak
37	polohrubý
38	zkrátka
41	berný
45	slitovný
46	domnění
49	redakčně
49	proslýchat
50	hydrometeorologický
50	zapékací
51	shledaná
52	shýbnout
52	morek
52	neurčito
53	potucha
53	ucechtnout
53	alpa
55	tab
55	řapíkatý
55	zlíbit

Počet typů	Forma
13	mylorde
17	pod
18	neposlední
18	dohledné
18	laskav
21	přihlédnutím
21	idnes
22	mylady
22	vězte
23	opepříme
26	prezidente
26	nemýlím
28	majore
31	zřetelem
31	doktorko
31	polohrubé
32	otázala
32	řapíkaté
32	nastaveném
33	tatínku
34	profesore
35	lubem
35	dobírku
36	drahoušku
36	holčičko
36	pečovatelskou
36	pamatuj
37	šéfe
37	rozmyšlenou
37	narážce

1d. Lemmata s nejvyšší hodnotou HHI pro pozice L1 a R1 samostatně

Uvedeny jsou výsledky o HHI min. 0,8500.

HHI (L1)	Lemma
1,0000	odiv
0,9908	rozloučená
0,9868	dostání
0,9809	mání
0,9770	ustání
0,9749	potaz
0,9738	třetice
0,9728	krážem
0,9716	nepoznání
0,9685	rozmyšlená
0,9660	přestání
0,9628	facto
0,9617	dem
0,9558	shledaná
0,9539	stol
0,9309	takž
0,9305	všudy
0,9126	pokraj
0,9025	úkor
0,8853	red
0,8805	nepaměť
0,8767	spadnutí
0,8744	vysvětlená
0,8632	odkvět
0,8530	naskrz
0,8517	skryt

HHI (R1)	Lemma
1,0000	de
1,0000	tab
1,0000	křížem
1,0000	pod
0,9931	nadmořský
0,9810	týkající
0,9785	zdaňovací
0,9706	zabývající
0,9660	zbrusu
0,9654	pozměňovací
0,9606	bobkový
0,9580	výkladní
0,9458	idnes
0,9453	přesilový
0,9403	odehrávající
0,9401	zkratka
0,9391	horšovský
0,9039	mezipohlavní
0,9009	blížící
0,8966	domnění
0,8874	berný
0,8826	jakž
0,8802	štítný
0,8759	výkupní
0,8742	obývací
0,8722	red
0,8709	vztahující
0,8689	orig
0,8680	zatajený
0,8646	polohrubý
0,8635	ublížení
0,8603	operka
0,8577	přihlédnutí
0,8550	soc

1e. Lemmata s nejvyšší hodnotou HHI pro pozice L1 a R1 dohromady

HHI	Lemma
0,4281	pod
0,4233	red
0,4093	domnění
0,4005	zatajený
0,3957	idnes
0,3941	přihlédnutí
0,3558	soulad
0,3487	ala
0,3461	dohledný
0,3445	hydrometeorologický
0,3194	nepoznání
0,3148	operka
0,3067	postesknout
0,3053	zkrátka
0,3024	rozmyšlená
0,2959	morek
0,2922	činění
0,2864	nadmořský
0,2853	tab
0,2849	sternum
0,2847	shledaná
0,2769	slitovný
0,2766	otázat
0,2761	berný
0,2750	skryt
0,2745	rozloučená
0,2733	bobkový
0,2694	odiv
0,2693	polohrubý
0,2690	dostání
0,2683	václavský
0,2660	udržovací
0,2660	nedohledno
0,2653	odehrávající
0,2646	všudy



## 1f. Lemmata s nejnižší hodnotou entropie pro pozice L1 a R1

Uvedeny jsou výsledky o hodnotě entropie max. 0,700.

Entropie (L1)	Lemma
0,000	odiv
0,039	rozloučená
0,062	dostání
0,072	mání
0,077	potaz
0,084	krážem
0,086	ustání
0,106	třetice
0,112	rozmyšlená
0,113	nepoznání
0,120	přestání
0,128	takž
0,143	facto
0,157	dem
0,169	shledaná
0,226	stol
0,340	všudy
0,352	pokraj
0,419	vysvětlená
0,424	red
0,427	nepaměť
0,438	úkor
0,498	spadnutí
0,550	odkvět
0,599	naskrz
0,618	skryt
0,639	nynějšek
0,665	pod

Entropie (R1)	Lemma
0,000	de
0,000	tab
0,000	křížem
0,000	pod
0,030	nadmořský
0,080	zdaňovací
0,102	pozměňovací
0,114	týkající
0,145	zbrusu
0,148	přesilový
0,150	bobkový
0,156	zabývající
0,184	výkladní
0,209	idnes
0,215	horšovský
0,227	odehrávající
0,242	zkratka
0,334	mezipohlavní
0,361	kutný
0,415	berný
0,418	štitný
0,453	polohrubý
0,481	domnění
0,491	red
0,495	jakž
0,503	výkupní
0,530	přihlédnutí
0,547	vztahující
0,563	zatajený
0,574	orig
0,575	soc
0,591	obývací
0,597	blížící
0,600	hydrometeorologický
0,602	operka
0,624	hákový
0,625	ublížení
0,654	opírající
0,674	stroužek
0,678	objízdny

## Příloha 2: Přehled monokolokabilních slov a jejich kolokátů v češtině

V této příloze je uveden seznam slov, která byla v průběhu analýzy dat označena jako monokolokabilní, s jejich absolutní frekvencí v korpusu. Zaznamenána jsou také jednotlivá referenční slova a v závorce počet jejich opakování v datech. V posledním sloupci je uveden podíl užití jiného typu, počítaný v procentech. V případě, že je slovo vícenásobně monokolokabilní, jsou referenční slova objevující se na pravé pozici barevně označena. Některá zde uvedená slova mohou být monokolokabilní i vůči vzdálenějším pozicím, ty však až na výjimky nebyly sledovány.

### 2a. Monokolokabilní lemmata

Lemma	Abs.f.	kolokáty	j. u. (%)
ala	2 705	et (2233)	
berný	236	mince (224)	0,444
blížící	1 853	se (1833)	
bobkový	251	list (249)	0,400
budoucnost	5 456	v (3194) do (2212) z (11) pro (5)	0,623
činžovní	408	dům (384) vila (9) blok (3) domek (2) byt (2) zástavba (2)	1,471
čtvereční	2 353	metr (1871) kilometr (346) centimetr (52) míle (17) milimetr (13) stopa (13) km (12) palec (7)	0,892
dem	1 284	soc (1256) nez (12)	0,393
deštný	449	prales (289) les (120) sezona (2)	1,675
diplomový	267	práce (256) zkouška (5) studium (2)	1,498
dlažební	275	kostka (207) kámen (26) materiál (11) deska (9) cihla (5) prvek (5) kostička (2) mozaika (2)	2,909
domnívat	9 346	se (9338)	0,032
doslech	333	z (201) na (57) v (40) mimo (22)	1,235
dostání	904	k (899)	0,333
dřívější	210	z (151) od (23) oproti (21) proti (12)	0,957
dušovat	228	se (228)	
facto	829	de (817) ipso (8) post (2)	

<b>Lemma</b>	<b>Abs.f.</b>	<b>kolokáty</b>								<b>j. u. (%)</b>
hákový	248	kříž (235)	úchyt (3)							1,245
hlávkový	251	salát (195)	zelí (22)	chmel (3)						0,797
holička	232	na (212)								
hraničit	318	s (311)								2,201
chýlit	399	se (385)								2,532
klíčn	279	kost (247)	kůstka (10)	rostlina (7)	hyfa (7)	mycelium (3)	rostlinka (2)	vlákno (2)		0,358
koncentračn	1 075	tábor (977)	gradient (19)	článek (18)	spád (12)	proces (5)	rozdíl (4)	změna (4)	lágr (4)	2,886
kontrastovat	577	s (567)								1,733
krážem	219	křížem (218)								0,457
křížem	217	krážem (217)								
lesknoucí	270	se (269)								
libo	271	být (242)								
lícni	451	kost (382)	strana (2)	sval (16)	nerv (8)	torba (3)	trn (3)	vak (3)	plocha (3)	2,882
loci	249	genius (185)	lex (30)	génus (11)	metoda (6)	communes (6)	stabilitas (2)			2,439
lordstvo	312	jeho (218)	váš (88)	jejich (2)						1,282
mání	557	k (554)								0,360
mezipohlavní	219	rozdíl (210)	znak (3)	poměr (3)						1,370
moučkový	343	cukr (280)								0,292
muškátový	390	oříšek (272)	květ (86)	víno (11)	ořech (8)	šalvěj (4)	réva (2)			1,795
načase	477	být (475)	mít (2)							
načež	1 541	, (1193)	. (207)	" (93)	- (15)	' (7)	? (6)	; (4)	! (3)	0,844
nadmořský	1 157	výška (1154)								0,259
náprsní	508	kapsa (276)	taška (205)	kapsička (16)						1,775
náručí	1 890	v (1479)	do (347)	z (44)						1,058
naskrz	214	skrz (199)								1,869
nasnadě	289	být (281)	jsoucí (2)							1,042
nedohledno	388	v (337)	do (41)	k (2)						1,299
nepaměť	277	od (263)	do (6)							2,527
nepoznání	349	k (345)								1,146
neurčito	213	na (173)	do (32)	v (5)						1,408
nynějšek	298	od (260)	do (34)							1,010

<b>Lemma</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
obeznámit	361	s (350)	2,507
objízdny	207	trasa (192) komunikace (4) kolej (2) úsek (2) cesta (2)	2,415
obývací	2 246	pokoj (2218) prostor (31) část (26) místnost (17) stěna (9) prostora (6) kuchyně (6) hala (5)	0,672
očitý	347	svědek (307) svědectví (28) svědkyně (8) účastník (2)	0,576
odehrávající	277	se (277)	
odiv	358	na (358)	
odkvět	307	po (290) při (6) do (3) před (2)	1,954
okvětní	395	lístek (297) plátek (73) trubka (8) cíp (5) štětinka (2) list (2)	1,023
onak	583	či (497) nebo (55) ani (11) i (7) podruhé (2)	1,379
opírající	267	se (263)	1,498
otázat	1 387	se (1376)	0,793
památkově	275	chráněný (236) chránit (22) významný (9) cenný (3) citlivý (2)	1,091
pitný	1 301	voda (1066) režim (169) kúra (24) stojan (3) účel (3) zdroj (2)	1,460
počest	860	na (858)	
pod	212	a (198)	1,980
podílový	606	fond (379) list (113) spoluvlastnictví (75) spoluvlastník (12) jednotka (4) mzda (4) odměna (3) ukazatel (2)	2,310
podít	532	se (521)	
podivit	1 464	se (1456)	0,137
pohybující	878	se (868)	0,911
pocházející	1 238	z (1162) od (65) zevnitř (4) odjinud (2)	0,323
pocházet	9 106	z (7515) od (713) odkud (593) odtud (129) odtamtud (28) odněkud (26) odsud (24) odjinud (15)	0,539
pokraj	771	na (761) při (2) k (2)	0,649
pokrčit	4 082	rameno (3848) koleno (50+) noha (50+) nos (46) paže (11) obočí (6) čelo (6) nosík (5)	0,735
pokutový	873	kop (449) území (371) blok (24) bloček (8) lístek (6) řád (3) socha (2) hod (2)	0,575
pokývat	758	hlava (714) hlavička (3)	1,058
polohrubý	248	mouka (248)	
popř	1 488	, (1157) ( (165) a (17) - (2)	0,134
pořizovací	517	cena (357) náklad (122) hodnota (25) lhůta (2) proces (2)	1,354
postesknout	860	se (857)	0,233

<b>Lemma</b>	<b>Abs.f.</b>	<b>kolokáty</b>								<b>j. u. (%)</b>
potaz	475	v (472)	na (3)							
považovaný	537	za (536)								0,186
povšimnutí	583	bez (448)	za (121)	hodný (5)						1,201
pozměňovací	258	návrh (255)	zákon (2)							0,388
poznámkový	241	blok (170)	aparát (36)	sešit (13)	záznamník (6)	vydání (4)	sešitek (3)	papír (3)	bloček (2)	1,660
pramenící	206	z (179)	v (17)	na (5)						2,427
pravdivostní	301	hodnota (238)	hodnocení (16)	podmínka (13)	ohodnocení (8)	tabulka (5)	obsah (5)	funkce (5)	nárok (3)	1,678
překl	206	pozn (201)	poz (1)	) (165)	: (19)	] (9)	. (7)			0,990
přesilový	216	hra (212)	situace (3)							
přestání	233	bez (231)								0,431
převážný	740	část (363)	většina (317)	míra (26)	bezmocnost (4)	množství (2)	zapojení (2)			2,725
přihlédnutí	410	s (369)	bez (22)	při (4)	po (3)	k (382)	na (15)			2,689 // 1,956
příjezdový	1 135	cesta (961)	kommunikace (74)	silnice (36)	turistika (9)	rampa (8)	trasa (8)	ruch (6)	brána (4)	2,291
příklon	219	k (213)	na (2)							1,826
příkročit	387	k (378)								2,326
přímhouřený	274	oko (248)	víčko (12)	pohled (6)	řasa (2)	okno (2)				1,460
přímhouřit	540	oko (532)	víčko (5)							0,556
přirovnat	771	k (753)								2,335
příspěvkový	766	organizace (713)	org (24)	péče (7)	nemocnice (5)	program (3)	společnost (2)			1,567
přivalový	295	děšť (213)	srážka (28)	vlna (23)	liják (13)	voda (11)				1,027
radno	252	být (247)	zdat se (2)							
red	1 953	pozn (1036)	( (91)	poz (7)	poznámka (3)	) (1081)	: (33)	. (17)	- (4)	1,642 // 1,901
rozloučená	871	na (870)								
rozmyšlená	252	na (249)								
rozpadající	282	se (279)								1,064
různit	278	se (278)								
řapíkatý	337	list (272)	celer (48)	lístek (3)	smil (3)	hortenzie (3)	květ (2)			1,780
sestávající	326	z (324)								0,307

<b>Lemma</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
sestávat	838	z (834) s (3)	0,119
setmít	462	se (459)	0,216
shledaná	482	na (480)	
skryt	305	v (293) z (7) do (4)	
snažící	261	se (260)	
snušní	292	prsten (190) prstýnek (91) chování (4) praktika (3) svazek (3) komůrka (2) kroužek (2) protokol (2)	1,027
soc	1 355	dem (1256) demokrat (47) (20) demokracie zařízení (5) věc (5)	1,478
spadnutí	276	na (260) před (4) hrozit (4) proti (2)	1,818
spočívající	694	v (534) na (142) u (3) vedle (3) pod (3) mimo (2)	1,009
stroužek	612	česnek (591)	0,980
sťah	519	v (488) do (17)	2,132
svědčící	246	o (190) pro (26) v (12) proti (4)	2,110
štítňý	207	žláza (196) chrupavka (5)	
takž	341	jakž (336)	
tržňý	224	rána (203) ranka (6) poranění (4) zranění (3)	0,893
třetice	379	do (376)	0,265
tyčící	202	se (202)	
týkající	3 713	se (3707)	
uhličítý	1 017	oxid (857) kysličník (104) kyselina (19) koupel (11) voda (9) macerace (6) pramen (3) lázeň (2)	0,590
úkor	1 353	na (1349)	0,296
usmívající	246	se (243)	1,220
ustání	346	bez (343)	0,867
valící	241	se (237)	0,830
vchodový	305	dveře (277) část (6) val (3) aritmetika (2) vrata (2) otvor (2) komora (2)	2,970
visací	202	zámek (177) náušnice (11) lampa (4) hodina (4)	2,970
vozový	421	park (337) cesta (33) hradba (14) plachta (6) třída (5) kolo (5) zásilka (4) skříň (3)	2,850
výkladňí	431	skříň (423) sklo (2)	1,392
výkupňí	240	cena (232) organizace (3)	2,083
vyměřovací	243	základ (215) řízení (15) oddělení (6) kolík (2) věž (2)	0,823
vyplývající	1 300	z (1281)	1,462

<b>Lemma</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
vytřeštit	451	oko (424) zrak (22)	1,109
vývěsní	204	štít (157) tabule (18) skříňka (15) poutač (3) skříň (3) plocha (3) cedule (2)	0,985
vznášející	246	se (241)	2,033
vztahující	571	se (564)	1,226
zabývající	1 157	se (1151)	0,433
zakysaný	223	smetana (185) výrobek (13) mléko (11) nápoj (8) krém (3)	1,345
zalapat	674	po (670)	0,297
zapotřebí	3 202	být (2859) mít (335)	0,062
závazkový	462	vztah (403) právo (42) kampaň (2) hnutí (2) úprava (2) instrument (2)	1,515
záviset	4 479	na (4424) od (49)	0,022
zavrtět	5 040	hlava (4607) se (348) ocas (18) ocásek (8) zadek (6) hlavička (4) rameno (3) palec (3)	0,893
zbrusu	680	nový (671) nově (4) čerstvý (2)	0,295
zdaňovací	494	období (490) pásmo (2)	0,405
zorný	650	pole (428) úhel (198) paprsek (19)	0,769
zrakově	201	postižený (192) vnímaný (2) hendikepovaný (2) náročný (2)	1,493

## 2b. Monokolokabilní slovní tvary

Slovní tvar	Abs.f.	kolokáty	j. u. (%)
al	2626	et (2233)	
autobusového	265	nádraží (226) terminál (9) stanoviště (6) spojení (6) park (5) dopravce (3) zájezd (3)	1,901
blížící	1029	se (1022)	
blížícím	334	se (326)	
budoucná	2230	do (2205) z (11)	0,539
budoucnu	3203	v (3191) o (5)	0,094
cizině	1219	v (1189) po (11) na (4) vůči (3) k (3) o (3)	0,492
členského	265	stát (218) vklad (17) příspěvek (12) podíl (5) poměr (3) charakter (3) vztah (2)	1,887
čtvrtě	971	tři (917)	
dispozici	9497	k (9308) v (33) o (12)	1,495
divu	2278	, (1933) . (203) - (47) : (28) ! (5) ; (4) ) (3) být (2223)	2,371 // 1,332
dobírku	227	na (220) za (2)	2,203
dohledné	502	v (497) doba (468) budoucnost (25) minulost (4)	0,996
domnívají	913	se (913)	
dostání	902	k (898)	0,111
eurozóně	222	v (205) k (11) vůči (2)	1,802
facto	829	de (817) ipso (8) post (2)	
havarijním	212	stav (190) pojištění (10) plán (3) výměr (2) plánování (2)	1,896
holčičko	349	, (316) můj (15) říkat (4) ale (2) milý (2)	2,865
holičkách	212	na (211)	
holýma	204	ruka (186) noha (13) záda (2)	1,471
chtě	303	chtít (295)	1,007
jaře	5012	na (4997) v (5)	0,060
jmenuješ	254	se (252)	0,395
jmenujete	270	se (263)	2,593
kamaráde	845	, (797) starý (9) milý (5) " (4) drahý (3) ' (2) říkat (2) dobrý (2)	2,485
komunitárního	252	právo (237) předpis (4) akt (2) obrat (2)	2,381



Slovní tvar	Abs.f.	kolokáty	j. u. (%)
koncentračního	243	tábor (219) gradient (8) spád (6) článek (6)	1,646
krážem	212	křížem (211)	0,472
křestním	214	jméno (191) list (15)	2,830
lubem	249	za (249)	
mání	556	k (554)	
modrýma	296	oko (289)	1,689
mykorhizních	413	houba (350) kořen (20) rostlina (18) symbióza (12) hyfa (3) druh (2)	1,937
mylorde	248	, (243) nebo (2)	1,210
načež	1237	, (1193) " (21) - (14) ; (4)	0,404
nedohlednu	338	v (334) k (2)	
nepaměti	270	od (260) do (7)	1,111
neposlední	1201	řada (1194) míra (3)	0,167
nepoznání	345	k (343)	0,580
nevíře	212	k (197) v (7) o (3)	2,358
občanští	318	demokrat (306) usedlík (4) zaměstnanec (4) aktivista (3)	0,314
oběžnou	283	dráha (269) doba (8) rychlost (4)	0,707
oblíbou	827	s (810)	2,056
obnovitelných	510	zdroj (487) energie (10) materiál (5) surovina (4)	0,784
obývacího	890	pokoj (879) prostor (10)	0,112
obývacím	804	pokoj (795) prostor (7)	0,249
odehrávající	210	se (210)	
odiv	358	na (358)	
ohledem	2935	na (2906) k (18) s (2921)	0,307 // 0,409
oknu	1740	k (1638) proti (58) naproti (12) v (3) po (2)	1,494
olivového	345	olej (326) dřevo (3) chléb (2)	2,609
onak	580	či (497) nebo (55) ani (11) i (7)	1,724
opírající	211	se (209) o (205) do (5)	0,474 // 0,948
oplátku	816	na (798) za (9) o (2)	0,614
otázala	365	se (363)	
památkového	322	ústav (274) odbor (8) fond (8) úřad (8) zákon (7) hledisko (3) objekt (3) dohled (2)	2,795

Slovní tvar	Abs.f.	kolokáty	j. u. (%)
pečovatelskou	307	služba (299) péče (4)	1,303
pitné	599	voda (583) kúra (6) stojan (2) účel (2)	1,002
play	1628	off (1324) fair (138) power (88) out (27) triple (5) station (3) down (2)	1,488
pod	212	a (198)	1,980
podivem	417	s (413)	0,959
pohonných	776	hmota (713) jednotka (31) látka (10) systém (5) směs (3) soustava (2) motor (2) technologie (2)	0,903
pokládán	267	za (260) zač (2)	1,504
pokraji	653	na (647) při (2) k (2)	0,153
pokrčí	208	rameno (193) noha (5) obočí (3) čelo (2) koleno (2)	1,442
pokrčil	2076	rameno (2046) nos (10) koleno (6) obočí (2)	0,578
pokutový	209	kop (196) lístek (5) blok (4)	1,914
polednem	512	před (479) mezi (18) s (9)	1,172
poledni	725	po (651) k (43) v (16) o (13)	0,138
polohrubé	208	mouka (207)	
posteskl	575	se (573)	0,348
potaz	474	v (471) na (3)	
považován	1639	za (1628)	0,306
považována	1092	za (1080) jako (4)	0,184
považování	292	za (291)	0,342
považováno	753	za (746) jako (4)	
procentního	243	bod (229) podíl (7) zastoupení (2)	2,058
proto	17688	, (17376) . (146) ) (20) a (5) ! (3) - (3) ; (2) ? (2)	0,741
předprodeji	271	v (265) k (2)	1,476
přepočtu	1009	v (911) při (27) po (26) podle (10) k (9)	2,480
přestání	232	bez (231)	0,431
přihlédnutím	372	s (368)	1,075
příjezdovou	291	cesta (269) komunikace (8) silnice (7) turistika (2) plocha (2)	1,031
přimhouřil	213	oko (208) víčko (2) očko (2)	0,469
příspěvkové	275	organizace (266) péče (4)	1,818

Slovní tvar	Abs.f.	kolokáty	j. u. (%)
příspěvkových	212	organizace (204) nemocnice (4)	1,887
puse	282	v (246) k (16) na (9) po (4) při (2) kvůli (2)	1,064
rámci	16698	v (16628)	0,437
roka	966	do (955) od (4)	0,208
rozloučenou	879	na (878)	
rozmyšlenou	253	na (250)	
řidičský	516	průkaz (506) trénink (2)	1,550
severovýchodě	209	na (206) po (2)	0,478
sire	201	, (192) oslovovat (3)	2,985
skleníkových	472	plyn (447) podmínka (10) emise (3) květina (2) půda (2) rostlina (2)	1,271
stroužky	402	česnek (394)	1,990
střehu	511	v (488) do (17)	1,174
širým	634	nebe (608) svět (12) moře (4)	1,577
takž	341	jakž (336)	
trestného	2185	čin (1933) kop (177) střelení (24) hod (16) bod (5) jednání (4) stíhání (3) území (3)	0,734
trestný	1614	čin (1371) kop (172) bod (22) hod (10) tábor (3) přestupek (2) okruh (2)	1,922
tuzemsku	633	v (631)	0,158
týkající	2707	se (2704)	
týkajících	745	se (743)	
uhličitého	566	oxid (505) kysličník (61)	
úhony	222	bez (199) doznat (10) dojít (8)	2,252
úkor	1346	na (1344)	0,074
upřenýma	213	oko (212)	0,469
určitostí	285	s (280)	1,754
úrokových	454	sazba (407) míra (19) výnos (9) rozdíl (4) náklad (3) platba (2) podmínka (2)	1,762
ustání	346	bez (343)	0,867
uzdě	364	na (353) po (7)	1,099
vězte	205	, (200) : (4)	0,488
víkendech	746	o (726) po (11)	0,540

Slovní tvar	Abs.f.	kolokáty	j. u. (%)
vínku	207	do (199) v (5)	0,488
vlastnického	332	právo (310) bydlení (6) podíl (3) vztah (3) senát (2)	2,410
vrub	310	na (302)	2,581
vybranou	994	na (805)	
vztahující	365	se (359) ruka (3) k (324) na (38)	0,822
zabývající	801	se (800)	
zabývajících	235	se (235)	
záchodě	382	na (371) v (9)	0,524
zalapal	240	po (237)	0,837
zalapala	276	po (276)	
záloze	466	v (440) k (6) o (5) proti (2)	2,575
zásluze	219	po (210) o (7)	0,913
zavrtěl	2325	hlava (2136) se (148) ocas (12) ocásek (3) noha (2)	1,032
zavrtěla	1392	hlava (1256) se (117) ocásek (3)	1,078
zavrtí	260	hlava (234) se (16) ocas (3) zadek (2)	1,923
zavřenýma	427	oko (426)	0,234
zemního	690	plyn (653) těleso (9) val (6) výměník (4) odpor (4) vodojem (2) prostředí (2)	1,306
znalecký	202	posudek (180) odhad (11) ústav (5) úřad (3) pohled (2)	0,495
pohodě	2188	v (2076) k (53) o (12) na (11) po (6)	1,280

### Příloha 3: Monokolokabilní jednotky v ostatních sledovaných jazycích

Pro další jazyky, kde bylo monokolokabilních výsledků méně, jsou lemmata a slovní tvary uvedeny ve společné tabulce. Lemmata jsou barevně vyznačena. Referenční slova vícenásobně monokolokabilních prvků, která se nacházela jinde než na bezprostředně sousedící pozici, jsou označena zeleně. Ani zde však nebyly takové vzdálenější prvky systematicky vyhledávány.

#### 3a. Slovenština

Lemma/Tvar	Abs.f.	kolokáty	j. u. (%)
alia	314	inter (313)	
alio	308	inter (308)	
blížiaci	515	sa (511)	0,584
členské	31991	štát (31766) krajina (153) poplatok (10) podiel (9) príspevok (9) organizácia (9) schôdza (4) vklad (4)	0,038
členského	15100	štát (15055) podiel (5) pestovateľ (2) štatút (2) príspevok (2)	0,046
členskému	1502	štát (1498) producent (2)	
členskom	8013	štát (8002) štatút (2)	0,075
členský	12712	štát (12671) štatút (4) poplatok (2) príspevok (2)	0,039
členským	12223	štát (12172) krajina (30) spoločnosť (3) organizácia (2)	0,074
členskými	12070	štát (11910) krajina (131) organizácia (3)	0,066
decembru	606	k (606)	
dispozícii	6993	k (6972) o (6) na (4) v (2)	0,114
elektrickou	328	energia (321)	2,134
eurách	597	v (592) o (3)	0,168
facto	370	de (336) ipso (13) post (12)	
hazardný	434	hra (402) hráč (11) hranie (3) hračka (2) zmýšľanie cesta (2) činnosť (2)	2,304

<b>Lemma/Tvar</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
		(2)	
hovädzí	353	dobytok (343) čaj (3) loj (2)	1,136
how	385	know (379)	
how	381	know (379)	
hybný	236	silá (221) páka (8)	2,966
hydínového	694	mäso (687)	0,145
hydínový	1464	mäso (1379) kus (13) vajce (10) tuk (6) výrobok (6) diel (3) trup (3) chov (3)	1,997
inter	325	alia (313) partes (3)	0,940
inter	318	alia (313) partes (3)	0,629
inulínový	297	sirup (283) cukor (13)	0,337
januáru	334	. (331)	0,898
katarální	278	horúčka (278)	
know	383	how (379)	
know	380	how (376)	
krátkosť	355	v (328) na (9) pre (8) kvôli (2)	2,254
krívačka	950	slintačka (931)	2,000
ktorí	28794	, (27847) a (682) alebo (93) - (53) ale (35) ( (20) " (9) ) (7)	0,167
limitné	485	hodnota (481)	0,825
limitný	1040	hodnota (1012) kritérium (5) suma (3) koncentrácia (2) množstvo (2)	1,444
mail	340	e (325)	1,813
mail	223	e (220)	0,901
mailový	213	e (211)	0,939
makrofinančný	269	pomoc (259) stabilita (5)	1,859
mim	321	; (320) vyd (321)	0,312
mimovládny	1500	organizácia (1423) expert (10) sektor (9) agentúra (7) aktér (6) orgán (6) združenie (5) subjekt (4)	1,933
minimis	606	de (604)	
minimis	603	de (601)	
modifikovaných	659	geneticky (648)	1,669
modroplutvého	226	tuniak (226)	
modroplutvý	260	tuniak (260)	
mutandis	418	mutatis (418)	

<b>Lemma/Tvar</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
mutandis	418	mutatis (418)	
mutatis	420	mutandis (418)	
mutatis	418	mutandis (416)	
nadväznosti	485	na (469) medzi (3) na_on (2)	1,863
nachádzajúci	1258	sa (1249)	0,715
nachádzajúcich	401	sa (399)	0,499
nasledovne	657	: (615) . (19) , (4) " (2)	2,588
neoddeliteľnou	328	súčasť (321) časť (6)	0,305
neposlednom	766	rad (763) miesto (2)	
neposledný	777	v (774) na (2) rad (768) miera (6) miesto (2)	
odhliadnuť	200	od (199)	0,500
odchylné	1104	od (1098)	0,543
odstredeného	365	mlieko (365)	
odstredený	730	mlieko (729)	0,137
ohľadu	4925	na (4889)	0,407
olivového	999	olej (983) strom (6) háj (3)	0,402
opt	209	out (185) in (19) outs (2)	
opt	207	out (185) in (19) outs (2)	
organizovanému	390	zločin (384)	1,285
orný	313	pôda (297) plocha (4) plodina (3) poľnohospodárstvo (2) región (2) zem (2)	0,641
osebe	706	sám (661) samý (39)	0,850
palube	1479	na (1472) po (6)	0,068
pitný	320	voda (311) destilát (4) mlieko (2)	0,627
plaviaci	764	sa (763)	0,131
pochádzajúci	4052	z (3975) od (56) s (4)	0,420
poplatník	1153	daňový (1148)	0,434
poplatníkov	719	daňový (717)	0,278
poradový	497	číslo (473) oheň (11) č (4) krok (3) cvik (2)	0,605
pozmeňujúci	7772	návrh (7563) rozpočet (100) smernica (23) nariadenie (22) dohoda (11) opatrenie (4) okolnosť (3)	0,618
poznamenať	1742	, (1681) aj (6) . (5) - (5) : (4) že (2)	2,239
predsedajúca	3338	pani (3330) pán (2) , (3300) . (32) ? (2)	0,150 // 0,120

<b>Lemma/Tvar</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
prejudiciálny	560	začatie (557) konanie (560)	0,536
pričom	12321	, (12048) - (103) ( (64) ; (42) . (14) " (10) ? (2)	0,308
pričom	12300	, (12045) - (103) ( (64) ; (42) " (8) a (3)	0,285
pridanú	810	hodnota (803)	0,864
prihliadnutie	929	s (909) bez (4) na (3) o (3) na (888) k (37)	1,076 // 0,216
psychotropný	231	látka (226)	0,877
ratingový	1000	agentúra (947) systém (9) hodnotenie (7) odvetvie (5) postup (5) proces (5) činnosť (4) spoločnosť (3)	1,500
rovnajúci	601	sa (600)	0,166
rybného	2450	hospodárstvo (2433) priemysel (4) bohatstvo (2)	0,122
rybný	3619	hospodárstvo (3502) zdroj (18) priemysel (12) sektor (11) produkt (9) kvóta (7) výrobok (6) oblasť (4)	0,391
skleníkový	1534	plyn (1428) efekt (76) pestovanie (7) pestovateľ (3) plocha (3) záhradníctvo (3) účinok (3)	0,065
skleníkových	1298	plyn (1294)	0,308
slintačka	934	krívačka (931)	0,214
správcofský	308	spoločnosť (285) subjekt (4) inštitúcia (4) služba (3) schopnosť (3) orgán (2)	2,273
takže	3530	, (3455) - (34) ( (10) " (10) ; (8) . (8) : (2)	0,085
tieňovým	564	spravodajca (544) spravodajkyňa (14)	1,064
týkajúca	1474	sa (1467)	0,475
týkajúceho	1684	sa (1681)	0,119
týkajúcej	1904	sa (1901)	0,158
týkajúcej	18871	sa (18828)	0,228
týkajúci	39955	sa (39865)	0,208
týkajúci	1950	sa (1947)	0,103
týkajúcich	9772	sa (9744)	0,287
týkajúcim	913	sa (912)	0,110
týkajúcimi	929	sa (927)	0,215
týkajúcom	387	sa (387)	
týkajúcou	204	sa (204)	
týkajúcu	1659	sa (1659)	



<b>Lemma/Tvar</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
uhličitý	690	oxid (653) kysličník (26) chlorid (9)	0,290
úkor	962	na (961)	
úkor	962	na (961)	
ústrety	425	v (425)	
usúdi	249	, (242)	2,811
valný	550	zhromaždenie (546)	0,365
vedecko	311	technický (273) výskumný (19) technologický (8) fantastický (4) výchovný (2) technicky (2)	0,965
verejno	268	súkromný (221) právny (33) prospešný (7) propagačný (2)	1,866
verejno	257	súkromný (212) právny (33) prospešný (7)	1,946
všeobecnosť	2790	v (2760) zásada (11)	0,681
vyd	322	mim (321)	0,311
vyd	320	mim (319)	0,313
výrobnému	441	odvetvie (425) závod (2) priemysel (2) proces (2)	2,268
vzdušného	546	priestor (531)	2,747
vzhľadom	12806	na (12379) k (379)	0,313
vzťahujúce	1339	sa (1332)	0,523
vzťahujúci	3108	sa (3097)	0,290
vzťahujúcich	1004	sa (1003)	
zahraničí	754	v (753)	0,133
zahustené	334	maslo (325) mlieko (4)	1,497
zaoberajúci	534	sa (533)	0,187
zárobkovo	1000	samostatne (974) samostatný (2) činný (997)	2,400
zatýkací	276	rozkaz (275)	0,362
zdravotne	605	postihnutý (571) neškodný (8) vyhovujúci (7) závadný (6) poistený (5) spôsobilý (3)	0,826
zemného	611	plyn (608)	0,491
zemný	1128	plyn (1102) práca (5) val (4) násyp (3)	1,154
zemný	233	plyn (231)	0,858
zemným	213	plyn (210) val (2)	
zmluvnou	518	strana (507) ochrana (2)	1,737

<b>Lemma/Tvar</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
zoči	239	voči (239)	
zreteli	222	na (222)	
zúčastňujúci	282	sa (281)	0,355
životne	429	dôležitý (414)      nevyhnutný (4)      potrebný (2)      závislý (2)	1,171

### 3b. Francouzština

<b>Lemma/Tvar</b>	<b>Abs.f.</b>	<b>kolokáty</b>	<b>j. u. (%)</b>
adjudicateur	376	pouvoir (374)	
adjudicateurs	610	pouvoir (605)	0,329
adjudicatrices	327	entité (322)      autorité (5)	
a fin	35189	de (30122)      que (4244)	
alcoométrique	343	titre (334)      table (6)      degré (2)	0,292
alcoométrique	322	titre (319)      degré (2)	0,311
angulaire	342	pierre (334)      déplacement (3)	1,462
angulaire	242	pierre (235)      déplacement (3)	1,653
aphteuse	702	fièvre (701)	
avérer	601	se (597)	0,334
axé	283	sur (281)      autour (2)	
axée	491	sur (486)      vers (2)      autour (2)	0,204
axées	347	sur (346)	0,288
axer	1446	sur (1417)      vers (6)	0,698
axés	210	sur (208)      vers (2)	
basant	233	sur (233)	
basées	706	sur (653)      dans (19)      en (11)      à (5)      au (5)	0,992
basés	497	sur (434)      dans (28)      en (15)      à (10)      au (6)	1,006
biais	2604	par (2555)      de (17)      en (9)	0,845

Lemma/Tvar	Abs.f.	kolokáty	j. u. (%)
biais	2603	par (2553) de (17) en (9)	0,807
bivalve	492	mollusque (487)	1,016
c'	32764	être (32725)	0,021
catarrhal	386	fièvre (386)	
catarrhale	386	fièvre (386)	
conformement	255	se (253)	
contenues	1633	dans (1566) à (29) au (11)	1,653
décortiqué	245	riz (241)	
demie	316	et (311)	1,266
dénuer	237	de (227) du (5)	
dépît	1805	en (1756) de (8) avec (7) par (5)	0,838
déroule	435	se (430)	0,922
déroulent	350	se (348)	0,571
désireux	532	de (511)	0,584
destiner	22811	à (19315) au (13155) auquel (160)	0,658
détriment	1037	au (1023) à (8)	0,386
détriment	1037	au (1021) à (8)	0,387
digitales	276	empreinte (268)	1,456
dioxyde	780	de (770) carbone (577) soufre (70) azote (68) titane (48) silicum (6)	1,155 // 0,130
écrier	588	se (567) me (17) nous (3)	0,170
efforce	495	se (472) me (22)	
efforcent	400	se (400)	
efforcer	2458	de (2401) à (35) au (2)	0,286
égide	273	sous (273)	
égide	273	sous (273)	
échéant	7928	cas (7925)	0,013
échéant	7928	, (7410) . (202) et (117) ; (56) ) (52) : (38) " (8) _ (4)	0,517
échéant	8764	cas (8748)	
emblée	233	de (233)	
encéphalopathie	384	spongiforme (382)	0,260
enclin	280	à (261) au (18)	

Lemma/Tvar	Abs.f.	kolokáty				j. u. (%)
encontre	3572	à (3569)				0,084
encontre	3572	à (3569)				0,084
ensuivre	280	se (275)				
exclama	371	se (371)				
exclamer	447	se (430)	me (7)	nous (6)		0,225
exempt	430	de (364)	du (10)	en (3)	dont (2)	0,524
exergue	288	en (288)	mettre (281)	mise (6)		0,347
exergue	288	en (288)	mettre (281)	mise (6)		0,347
facto	362	de (332)	ipso (20)			
férié	249	jour (242)				2,811
fur	406	au (402)				
fur	406	mesure (404)				
hoc	711	ad (705)	post (4)			0,141
hoc	711	ad (705)	post (4)			0,141
hocher	254	tête (250)				1,575
change	2463	de (2223)	du (132)			1,175
influenza	614	aviaire (571)	A (15)	aviaire (8)		1,980
inoxydable	224	acier (209)				0,948
instar	655	à (654)				
instar	655	à (654)				
instauration	1245	de (1109)	du (107)			2,329
intégrant	1340	partie (1334)				0,299
intégrante	1325	partie (1324)				0,075
inuline	275	de (264)	en (4)			2,545
inuline	275	de (264)	en (4)			2,545
laps	244	de (243)	temps (240)			0,410 // 1,639
laps	241	de (240)	temps (237)			0,415 // 1,660
mesdames	406	, (383)	- (19)			0,985
mesdames	389	, (383)	- (2)	et (362)	, (24)	1,028 // 0,771
minimis	663	de (663)				
minimis	663	de (633)				

Lemma/Tvar	Abs.f.	kolokáty			j. u. (%)
mutandis	895	mutatis (889)			0,224
mutandis	895	mutatis (889)			0,224
mutatis	890	mutandis (887)			
mutatis	890	mutandis (887)			
navals	508	chantier (495)	constructeur (5)		1,575
navigable	551	voie (532)	réseau (9)	canal (2)	1,452
navigable	263	voie (251)	réseau (9)		1,141
navigables	288	voie (281)	canal (2)		1,736
occurrence	876	en (856)			2,283
opté	220	pour (218)			0,909
opter	705	pour (677)	faveur (4)		1,447
opter	255	pour (248)	entre (3)	en (2)	0,784
ores	433	de (433)			
ores	433	de (433)	et (433)	déjà (432)	0,231
ouvrable	2055	jour (2045)	semaine (7)		0,146
ouvrable	712	jour (702)	semaine (7)	lundi (2)	0,140
ouvrables	712	jour (702)	semaine (7)		0,421
paddy	230	riz (230)			
parce	17857	que (16880)			
passible	569	de (536)	dont (17)	du (7)	0,709
plupart	5710	le (5709)			
polyesters	528	de (520)	fibre (501)	filament (10) fil (3)	1,515 // 2,652
posteriori	644	a (642)			
posteriori	644	a (637)	à (5)		
prenantes	753	partie (750)	multi-partie (3)		
qua	201	sine (197)			1,005
régi	1225	par (1212)			1,061
rechange	387	de (385)			0,517
rechange	387	de (385)			0,517
réjouis	2500	me (2476)	se (8)	nous (8) te (3)	0,080
reposant	470	sur (457)			2,766

Lemma/Tvar	Abs.f.	kolokáty	j. u. (%)
sine	220	qua (197) die (10) lege (2)	0,948
sine	220	qua (197) die (10) lege (2)	0,948
situe	858	se (841)	1,752
situent	310	se (302)	2,265
spongiforme	408	encéphalopathie (400) encéphalite (8)	
statu	333	quo (333)	
statu	333	quo (333)	
sterling	278	livre (276)	0,361
susceptible	6919	de (6575)	0,770
suspens	452	en (448) ce (3)	
suspens	452	en (448)	0,665
tandis	4619	que (4457)	
tandis	4145	que (4006)	0,125
ultrapériphériques	918	région (904) zone (4) territoire (2)	0,871
ultrapériphérique	986	région (967) zone (4) situation (4) statut (2) territoire (2)	0,710
valorem	328	ad (328)	
valorem	328	ad (328)	

### 3c. Angličtina

Lemma/Tvar	Abs.f.	kolokáty	jiná užití (%)
accordance	56231	in (56205) into (3) with (56183) to (27) therewith (6)	0,027 // 0,023
accordance	56228	in (56203) with (56180) to (27) therewith (6)	0,034 // 0,021
according	10126	to (10079)	0,454
adhere	646	to (640)	0,929
alia	2445	inter (2440)	0,204
alia	2434	inter (2430)	0,164
amounted	469	to (466)	0,640

Lemma/Tvar	Abs.f.	kolokáty	jiná užití (%)
amounting	556	to (556)	
annum	332	per (332)	
annum	331	per (331)	
attributable	390	to (384)	1,538
auspice	306	under (302) through (3)	0,327
auspices	304	under (300) through (3)	0,329
behalf	12112	on (12092) in (14)	
behalf	12111	on (12091) in (14)	0,050
bluefin	273	tuna (271)	0,368
bluefin	273	tuna (271)	0,368
capita	656	per (651)	0,762
capita	670	per (664)	0,896
complied	1604	with (1587)	1,060
complies	962	with (960)	0,208
complying	1134	with (1123)	0,970
conducive	331	to (322)	2,719
conducive	331	to (322) environment (6)	0,906
dealt	1716	with (1574) in (52) blow (47)	2,506
depending	1241	on (1209) upon (30)	0,161
durum	268	wheat (255)	
eh	222	, (211) ! (3) " (2) - (2)	1,802
eh	222	, (211) ! (3) " (2) - (2)	1,802
expectancy	270	life (263)	2,593
expectancy	264	life (257)	2,652
facto	551	de (504) ipso (23) post (15)	
facto	551	de (504) ipso (23) post (15)	
forefront	274	at (213) to (35) in (22)	1,460
forefront	274	at (214) to (35) in (22)	1,095
fourthly	361	, (332) and (9) : (5) - (5)	2,770
hoc	785	ad (779) post (3)	
hoc	785	ad (779) post (3)	

Lemma/Tvar	Abs.f.	kolokáty	jiná užití (%)
insofar	885	as (874)	1,131
insofar	827	as (817) that (2)	0,967
inter	2459	alia (2421) partes (5) pares (3)	0,287
inter	2441	alia (2421) partes (5) pares (3)	0,246
irrespective	1167	of (1162)	0,428
irrespective	1062	of (1059) wheter (2)	0,094
majeure	711	force (710)	
meantime	940	in (928)	1,277
meantime	932	in (927)	0,536
midst	605	in (570) into (23) through (6) from (3)	0,496
midst	605	in (570) into (23) through (6) from (3) to (2)	0,165
minimis	663	de (662)	
minimis	659	de (658)	
mutandis	894	mutatis (894)	
mutandis	894	mutatis (894)	
mutatis	892	mutandis (892)	
namely	4007	, (3586) - (178) : (73) ( (54) ; (38) . (22) -- (15) -- (6)	0,873
namely	3984	, (3586) - (178) : (73) ( (54) ; (38) -- (15) -- (6) " (2)	0,803
ought	2347	to (2300)	2,003
outermost	1111	region (1044) location (29) area (5) town (2) part (2) component (2)	2,430
outset	628	from (418) at (201) since (8)	0,159
outset	628	from (418) at (201) since (8)	0,159
owing	1122	to (1112)	0,891
pave	266	way (260) path (2)	0,376
pertaining	439	to (400) thereto (27)	2,733
psr	333	wine (327)	1,802
pursuance	244	in (242) of (235) thereof (6)	0,820 // 1,230
pursuant	17962	to (17902) thereto (43)	0,078
quo	419	status (372) pro (43)	0,480
quo	414	status (372) pro (39)	0,243
rata	271	pro (248)	



Lemma/Tvar	Abs.f.	kolokáty	jiná užití (%)
revert	223	to (220)	1,345
seekers	626	asylum (560) job (37) fortune (10) Grail (3)	2,396
spite	1461	in (1400) of (1397)	
spite	1458	of (1399)	
tends	360	to (341) towards (16)	0,557
tens	450	of (441)	2,000
thirds	297	two (292)	1,351
unable	3256	to (3249)	
unable	3203	to (3196)	0,219
valorem	323	ad (323)	
valorem	323	ad (323)	
verge	241	on (216) to (13) upon (3)	1,695
versa	295	vice (293)	
versa	294	vice (292)	
vitally	329	important (307) necessary (9) dependent (2) urgent (2)	2,736
vitally	328	important (306) necessary (9) dependent (2) urgent (2)	2,744

## Příloha 4: Přehled klasifikace

V této příloze je pro orientaci o jejich přibližném zastoupení uveden počet monokolokabilních slov v češtině zařazených do jednotlivých typů podle klasifikace načrtnuté v této práci.

### 4a. Klasifikace podle velikosti a struktury kolokačního paradigmatu

	absolutní mon.	víceprvkové par.	strukturované par.	vícenásobná mon.
lemmata	58	55	34	3
slovní tvary	46	51	32	5

### 4b. Klasifikace podle druhu referenčního slova

	S	Adj	Pron	Num	Verb	Adv	Prep	Conj + int.	nejasné
lemmata	58	4	25	0	6	3	47	7	7
slovní tvary	47	2	18	1	6	1	59	10	5

### 4c. Klasifikace podle druhu monokolokabilního slova

	S	Adj	Verb	Adv	nejasné
lemmata	30	72	23	10	15
slovní tvary	53	48	20	3	10