

# Monokolokabilita ve dvou typologicky odlišných jazycích: srovnání češtiny a italštiny\*



Zora Obstová (Praha)

## MONOCOLLOCABILITY IN TWO TYPOLOGICALLY DIFFERENT LANGUAGES: A COMPARISON OF CZECH AND ITALIAN

The present study deals with the phenomenon of extremely restricted collocability (monocollocability) in Czech and Italian. On the basis of a list of words with the highest degree of monocollocability extracted from corpora with the aid of the so called Herfindahl-Hirschman Index (HHI), we try to analyse this little-explored phenomenon in both languages. Monocollocable words (MW, often referred to as *cranberry words*) and the fixed combinations in which they occur are investigated in terms of syntactic and collocation structures, frequency and diachronic development. In the light of corpus evidence, the phenomenon of extremely restricted collocability appears more complex than originally believed. The paper shows that MW in both languages form a very heterogeneous category and that many differences between Czech and Italian monocollocable structures can be explained by typological (e.g. the degree of nominal inflection) or historical factors.

### KEYWORDS

monocollocable words, cranberry words, language typology, Czech, Italian, corpora

### KLÍČOVÁ SLOVA

monokolokabilní slova, jazyková typologie, čeština, italštiny, korpusy

## 1. KOLOKABILITA A MONOKOLOKABILITA

Kolokabilitu bychom mohli obecně definovat jako schopnost slova (či dalších jednotek) spojovat se v textu s jednotkami jinými (srov. Čermák — Holub, 2005, s. 12). To, s kterými jednotkami se slovo pojí a jakým způsobem v textu funguje, o něm zároveň mnohé vypovídá a umožňuje mezi jiným i vymezit jeho lexikální význam. O důležitosti kontextu se zmiňuje už Saussure („téměř všechny jednotky jazyka závisejí buď na tom, co je v mluvním řetězu obklopuje, nebo na po sobě jdoucích částech, jimiž jsou samy skládány“), který „souhru“ po sobě jdoucích jednotek přirovnává k „fungování stroje, jehož součástky na sebe působí navzájem, třebaže jsou rozmístěny jen v jediné dimenzi“ (Saussure, 1996, s. 156–157). Vztah kontextu, v němž se určitá jednotka vyskytuje, a lingvistickými vlastnostmi této jednotky je tudíž vzájemný. Kom-

\* Tato studie vznikla za podpory projektu Univerzity Karlovy Progres č. 4, *Jazyk v proměnách času, místa, kultury*. Chtěla bych poděkovat oběma anonymním recenzentům za cenné podněty a připomínky k textu článku.



OPEN ACCESS

binovatelnost je základní vlastností a přirozenou funkcí všech lexémů: jak podotýká Čermák (2010, s. 279), snad s výjimkou některých interjekcí nezůstávají mimo tuto kombinatoriku žádné systémové jednotky jazyka.

Kolokabilita má v jazyce povahu kombinatorické škály. Některá slova mají kolokabilitu velmi vysokou, na první pohled téměř neomezenou (např. české adjektivum *velký*); na opačné straně této pomyslné kombinatorické stupnice potom stojí slova s extrémně restringovaným, uzavřeným kolokačním paradigmatickým, která označujeme za monokolokabilní (např. české *[na] úkor [někoho, něčeho]*). Často se v této souvislosti hovoří o kontrastu tzv. principu volného výběru (*open-choice principle*) a principu předurčeného výběru (*idiom principle*)<sup>1</sup>, je však zřejmé, že zcela volná, neomezená kombinovatelnost lexikálních jednotek ve skutečnosti neexistuje (i kolokabilita zmíněného adjektiva „velký“ je podmíněna sémantickou kompatibilitou a některá spojení — jako například *\*velký spánek* nebo *\*velký vzduch* — běžně použít nemůžeme).

V této stati se budeme zabývat slovy s anomálně omezeným kombinatorickým potenciálem, a tedy slovy monokolokabilními. Taková slova pravděpodobně existují ve všech přirozených jazycích. Zdá se však, že tyto velmi zajímavé a specifické útvary mají v různých jazycích jen málo společného, neboť — jak podotýká Čermák — „their combinatorial uniqueness is due to a specific development of each word in these languages“.<sup>2</sup> Náš výzkum se pokusí tuto hypotézu ověřit a zaměří se na monokolokabilní slova ve dvou typologicky odlišných jazycích: v češtině a italštině. Po stručném souhrnu dosavadních přístupů k problematice představíme metody, jimiž lze tato slova identifikovat; zmíníme především výhody, ale i problémy, které s sebou nese systematické vyhledávání těchto slov v jazykových korpusech. V následujících kapitolách budeme na vzorku českých a italských monokolokabilních slov, identifikovaných v korpusu, analyzovat syntaktické a sémantické vlastnosti těchto jednotek i celých spojení, v nichž se vyskytují, jejich kolokační struktury a frekvenci v obou jazycích. Krátce se budeme věnovat i jejich diachronnímu vývoji. Článek si klade dva hlavní cíle: 1. Představit monokolokabilní slova v češtině a italštině z nejrůznějších úhlů a upozornit na nové pohledy, které korpusový výzkum vnaší do tradičního pojmání monokolokability; 2. zjistit, zda lze v oblasti monokolokability v češtině a itaštině vysledovat nějaké společné tendence a nakolik se do této problematiky promítají typologické odlišnosti mezi oběma jazyky či specifika historického vývoje.

## 2. RŮZNÉ PŘÍSTUPY K PROBLEMATICE

Termín monokolokabilní slova (MS), zavedený do české lingvistiky F. Čermákem (1982), víceméně odpovídá pojmu, které se pro slova s výrazně omezenou kolokabilitou užívají v jiných jazykových kontextech: v angličtině jsou označována jako *cranberry words* (např. Aronoff, 1976)<sup>3</sup>, *bound words* nebo *unique words*, v němčině

1 Srov. Sinclair, 1991, s. 109–115.

2 Čermák et al., 2016, s. 7.

3 Aronoff chápe *cranberry words* analogicky ke *cranberry morphs*, tj. morfémům, které se podobně jako morfém *cran-* vyskytují pouze v jediné kombinaci: „They are words which, like

jako *unikale Komponenten* (Fleischer, 1989) či *Unikalia* (Dobrovoľskij — Piirainen, 1994), v italštině jako *componenti/vocaboli a collocazione unica* (Veland, 2004, 2006).

MS byla dosud zkoumána především v rámci frazeologie: jako na nevydělitelnou součást frazému na ně pohlížejí např. Filipec — Čermák (1985, s. 173); Dobrovoľskij — Piirainen (1994, s. 449) je považují za *phraseologisch gebundene Formative*. V nejnovějších pracích se ovšem poukazuje na to, že MS se vyskytují nejen ve frazémeh, ale také v běžných kolokacích (srov. např. Konecny, 2010, s. 305) a především ve víceslovných pojmenováních terminologické povahy (srov. Čermák, 2014, s. 13; Čermák, F. — Čermák, J. — Obstová — Vachková, 2016, s. 7 a 13). Čermák (2014, 2016 aj.) definuje MS jako slova, která vykazují kombinatoricky anomální chování, a stojí tudíž na kombinatorické periférii lexikonu. Poukazuje na to, že „jejich jazykový, resp. textový život je striktně omezen na jedinou, popř. na několik málo kombinací s jinými slovy, a bez nich toto slovo neexistuje“ (Čermák 2016, s. 11). Ve většině případů se MS nekombinují pouze s jediným lexémem: taková slova jsou velmi vzácná. Obvykle mívají 1 až +/- 7 kolokátů (srov. např. Čermák 2016, s. 12), které téměř vždy tvoří uzavřené kolokační paradigma. Čermák na rozdíl od jiných badatelů vyzdvihuje fakt, že častěji než lexémy lze považovat za monokolokabilní konkrétní slovní tvary. To pochopitelně souvisí se specifickou povahou MS, z nichž mnohá jsou natolik defektní, že se — přinejmenším ze synchronního pohledu — vyskytují pouze v jediném tvaru ([*nechat na*] *holičkách*, [*pozdě*] *bycha* [*honiti*], [*jít k*] *duhu* apod.). Taková slova zcela postrádají sémantickou i syntaktickou autonomii, mnohdy u nich nelze jednoznačně určit ani slovní druh, jejich lemma je pouze hypotetické (např. analogií předpokládané substantivum *duh* v současném jazyce de facto neexistuje), a Filipec s Čermákem je označují za *verboidy* (1985, s. 174). Je zřejmé, že určitá defektnost je pro mnohá MS skutečně příznačná a činí z nich — spolu s absencí vlastního významu a nevydělitelností z příslušné kombinace — velmi zvláštní a nesporně zajímavé útvary, které lze jen velmi obtížně klasifikovat, a nesnáze působí i jejich lexikografické zpracování. Čermák (2014, s. 10–12) kritizuje běžnou lexikografickou praxi, kdy české slovníky (např. SSČ, SSJČ) místy poskytují zavádějící informace o významu i užití MS, a navíc uměle (re)konstruují jejich lemmata (*dutat*, *duh*, *bych* apod.). Není bez zajímavosti, že připravovaný *Akademický slovník současné češtiny* (ASSČ) na tyto výtky do jisté míry reaguje a předpokládané reprezentativní tvary těchto slov nerekonstruuje, neuvádí jejich slovnědruhovou charakteristiku ani nevykládá jejich význam (srov. Kochová — Opavská, 2016, s. 171, 174).

Ukazuje se však, že situace je mnohem komplexnější, neboť celá řada MS si navzdory extrémně omezené kolokabilitě zachovává vlastní význam a nevykazuje žádné anomálie ani z hlediska slovnědruhové kategorizace (srov. *stroužek* [*česneku*]; *přimhouřit* [*oči*]). Jindy narážíme na rozdíly v kolokabilitě tvarů téhož lemmatu (např. kolokačně výrazněji neomezený lexém *soulad* versus monokolokabilní tvar *souladu*, objevující se téměř výhradně v rámci víceslovné předložky *v souladu s ně-*

---

cranberry morphs, concatenate only with specific words and not with syntactic classes. For example, the noun *headway* occurs only as the direct object of the verb *make*, just as *cran* occurs only in *cranberry*“ (Aronoff, 1976, s. 15). O morfému *cran-* se ovšem zmiňuje už Bloomfield (1933, s. 160), který takové komponenty nazývá *unique elements*.



čím). Je ovšem nasnadě, že rozdíly v kolokabilitě lemmatu a jednotlivých slovních tvarů jsou patrné především v jazycích s rozvinutou nominální flexí,<sup>4</sup> jako je právě čeština. U jazyků, které nominální flexi nemají vůbec nebo jen v redukované míře, se tyto rozdíly pochopitelně neprojevují nebo se omezují např. na odlišnou kolokabilitu singuláru a plurálu, popř. neexistenci či neužívanost singulárového či plurárového tvaru ([*under/through/outside the*] *auspices* [*of sth*]; [*campo di*] *concentramento*; [*agli*] *antipodi*, srov. níže kap. 5).

### 3. IDENTIFIKACE MS

Aby bylo možné monokolokabilní slova zkoumat či lexikograficky zpracovávat, je nutné je nejprve v jazyce identifikovat. Dosud se k tomuto účelu využívalo především textových analýz, dotazníkových šetření mezi rodilými mluvčími, introspekce a nezdědka se lingvisté jako ke zdroji uchýlovali právě ke slovníkům, kde se na omezenou kombinovatelnost slova obvykle poukazuje poznámkami „pouze ve spojení“, „nejčastěji ve spojení“ atp. Vzhledem k tomu, že se kolokabilita slov výrazně mění v čase a téměř všechna MS prošla delším či kratším monokolokabilizačním procesem, v jehož průběhu lexém postupně ztrácel schopnost volně se kombinovat s dalšími slovy, je zřejmé, že se v jazyce neustále objevují „nová“ MS, která tradiční slovníky nedokážou zachytit. Jedinečnou možností, jak taková MS v současném jazyce identifikovat, nabízejí jazykové korpusey, v nichž lze slova s extrémně omezenou kolokabilitou vyhledávat systematicky. První prací, která se snaží o systematické zmapování MS v jazykových korpusech, a to hned ve čtyřech typologicky odlišných jazycích, je publikace *Language Periphery, Monocollocable words in English, Italian, German and Czech* (Čermák et al., 2016). Autoři se zde pokusili identifikovat MS ve čtyřech srovnatelných korpusech (BNC, CORIS, SYN 2010 a ad hoc sestavený německý korpus textů z Wikipedie) o velikosti cca 100 milionů slov pomocí analýzy na základě Herfindahl-Hirschmanova indexu (HHI). Tento index, který se používá v ekonomii k měření koncentrace (zejména při zkoumání monopolního postavení podniků na trhu), může sloužit i ke kvantifikaci diverzity kontextu (podrobněji k indexu a jeho použití v lingvistice viz Cvrček, 2013, s. 119), a tedy k identifikaci jednotek s anomálně omezenou kolokabilitou. Výsledky, které byly tímto způsobem získány, musely být podrobeny důkladné manuální selekci, aby se např. odstranila slova, u nichž je omezená kolokabilita způsobena spíše faktory lexikálně gramatickými (zvrtná slovesa, složené slovesné časy apod.).<sup>5</sup> Autoři se také museli vyrovnávat s obtížně řešitelným (nejen na úrovni automatické analýzy) problémem homonymie a poly-

4 Většinu MS v češtině a patrně i v dalších jazycích lze přiřadit ke kategorii substantiv či adjektiv. Monokolokabilní verba se zdají být obecně velmi vzácná (viz níže, kap. 6).

5 Monokolokabilitu lze ovšem pojmát i širěji: Cvrček (2013, s. 119–128) a také Šíková-Hrejsemnová, která z jeho názorů vychází ve své diplomové práci (2013, s. 16), považují za monokolokabilní i reflexiva tantum či slova charakterizovaná konkrétní valencí, spjatou s užitím určité předložky (*pomýšlet na*).

sémie.<sup>6</sup> Navzdory všem těmto nesnázím i určitým omezením, která s sebou analýza pomocí statistické míry nese (vzhledem k tomu, že HHI musí být aplikován na určité kvantum dat, aby byly výsledky průkazné, byla do výzkumu zahrnuta pouze slova s absolutní frekvencí převyšující 200 výskytů v korpusu, a na méně frekventovaná MS nebyl brán zřetel), vznikl čtyřjazyčný seznam nejfrekventovanějších současných MS, který je užitečným východiskem pro další bádání v této oblasti.



## 4. ANALÝZA VZORKU MS S NEJVYŠŠÍM HHI V ČEŠTINĚ A ITALŠTINĚ

### 4.1 IDENTIFIKACE PODLE LEMMAT A SLOVNÍCH TVARŮ

Korpusovou analýzou dle HHI byla získána i česká a italská MS, jimž se věnujeme v této stati. Kolokabilita slov má, jak už bylo zmíněno, povahu škály, a nejenak je tomu i u jednotek, které lze (v souladu s výše uvedenou Čermákovou definicí) označit za monokolokabilní. Hodnoty Herfindahl-Hirschmanova indexu, na jehož základě byla analýza korpusů provedena, by nám měly poskytnout informaci o velikosti kolokačního paradigmatu výchozích slov. HHI může nabývat hodnot od 0 do 1: naprostá většina lemmat i slovních tvarů v češtině i italštině dosahuje hodnot velmi nízkých (průměrná hodnota HHI u všech slovních tvarů v italském korpusu je 0,16), hodnoty vyšší než 0,3 jsou výjimečné (z 23 395 italských slovních tvarů, které byly podrobeny analýze, tuto hodnotu překračuje pouze 2650 slov a v češtině je situace obdobná). Při identifikaci monokolokabilních jednotek nás samozřejmě zajímají hodnoty nejvyšší, je však velmi obtížné určit hranici, od níž lze slova již považovat za monokolokabilní. MS s větším počtem podobně frekventovaných kolokátů mohou mít i hodnoty relativně nižší (kolem 0,5), na následujících stránkách se však — především z praktických důvodů<sup>7</sup> — zaměříme pouze na slova s nejvyššími hodnotami HHI, tedy na ta, u nichž se (nejvyšší) hodnota indexu pohybuje v rozmezí 1 až 0,7. Tuto podmínku v italštině splňuje 95 MS, v češtině 108 MS.<sup>8</sup>

Čeština je z hlediska Skaličkovy strukturní typologie flektivnější než italština, především v oblasti jmenné flexe; italština má navzdory bohaté morfologii slovesných časů a způsobů (z nichž mnohé jsou ovšem tvořeny opisy) i četné rysy izolační, především využívání předložek k vyjadřování pádových vztahů, relativně pevný slovosled apod.<sup>9</sup> Rozdíly v rozvinutosti jmenné flexe lze zohlednit už při korpusové

6 K problémům manuální selekce monokolokabilních kandidátů v jednotlivých jazycích srov. příslušné kapitoly in Čermák et al., 2016.

7 Spodní hranici jsme zvolili především z praktických důvodů, abychom soubor zkoumaných slov omezili, ale také proto, že pod touto hranicí výrazně přibývá MS s větším počtem kolokátů, případně slov s jedním dominantním kolokátem, jejichž zařazení mezi MS může být diskutabilní.

8 Všechna analyzovaná MS jsou uvedena v Příloze 1 na s. 241–243.

9 Zmiňujeme zde jen ty typologické odlišnosti, které se nám zdají být podstatné pro korpusové vyhledávání MS. Některé další rozdíly, na něž při studiu MS narážíme, budou zmíněny v dalších kapitolách (např. odlišná pozice adjektiva v české a italské nominální skupině, srov. kap. 8.3)



OPEN ACCESS

analýze: HHI je totiž možno aplikovat jak na lemmata, tak na slovní tvary (druhého postupu bylo využito in Čermák et al., 2016). Je zřejmé, že oba postupy přinesou poněkud odlišné výsledky.<sup>10</sup> V případě italštiny jsou rozdíly minimální, vyhledávání podle slovních tvarů nicméně umožňuje odhalit i slova, která jeví monokolokabilní tendenci pouze v některém z tvarů, např. [*occhi*] *sbarrati*, popř. apokopované tvary jako [*in*] *particular* [*modo*]). Může se ovšem stát, že — vzhledem k tomu, že zde pracujeme jen se slovy s vyšší frekvencí — se do výběru nedostanou tvary některých méně četných sloves, která by jako lemmata stanovenou hranici překonala (totéž by teoreticky mohlo platit o některých, zejména dvojvýhodných adjektivech). Protože se však slovesa mezi MS (v obou jazycích) vyskytují minimálně, považujeme toto riziko za zanedbatelné.

U češtiny je tomu naopak: přehlednější výsledky zde přináší analýza podle lemmat. Při vyhledávání podle slovních tvarů totiž narazíme na týž problém, který v italštině pozorujeme u sloves, také u substantiv a adjektiv: protože se zohledňují pouze jednotlivé alolexy, a nikoli celé deklinační paradigma, mají kandidátské tvary (zejména u těch MS, kde lze za monokolokabilní považovat celý lexém) mnohem nižší frekvenci a často nepřekojí stanovenou hranici 200 výskytů. Při analýze podle slovních tvarů se tak do výběru například vůbec nedostanou adjektiva s velmi vysokým HHI jako *bobkový* nebo *hákový*, která v soupisu lemmat zaujímají přední místa. Vzhledem k tomu, že právě substantiva a adjektiva tvoří mezi MS zdaleka nejpočetnější skupinu, jeví se tedy pro češtinu jako výhodnější analýza podle lemmat; tímto způsobem samozřejmě odhalíme i většinu frekventovaných defektních MS (verboidů), jen je v seznamu najdeme pod jejich (často uměle rekonstruovaným) reprezentativním tvarem (*holička*, *vínek*, *vybraná*). Chceme-li ovšem identifikovat i oněch několik MS, která jsou kolokačně omezena jen v některém ze svých tvarů, popř. ta, kde hraje výraznou roli homonymie nebo polysémie (např. [*se*] *zatajeným* [*dechem*] versus *zatajený* [*plán*, *informace* atd.]), je vhodné přihlídnout i k analýze podle slovních tvarů.<sup>11</sup>

Při identifikaci MS s nejvyšším HHI jsme tedy pro italštinu použili seznam sestavený na základě analýzy slovních tvarů, kdežto u češtiny jsme pracovali jak se seznamem lemmat (primárně), tak se seznamem jednotlivých *word-forms*.<sup>12</sup> To je hlavní důvod, proč se mezi níže analyzovanými MS, zejména těmi českými, objevuje řada slov, která soupisy monokolokabilních slov in Čermák et al. (2016) neobsahují, a některá z nich nenajdeme ani v Čermákové *Periferii jazyka*.<sup>13</sup>

10 Aby byly výsledky ještě přesnější, bylo by výhodné zohledňovat rozdíl mezi lemmatem a slovním tvarem i u jednotlivých kolokátů MS.

11 Výsledky získané pro češtinu analýzou slovních tvarů s sebou navíc přinášejí i další praktické obtíže při manuální selekci, neboť kromě výše zmíněných případů (srov. kap. 3 a pozn. č. 3) je nutno odfiltrovat i všechny tvary, jejichž omezená kolokabilita je způsobena častým souvýskytem s některou z předložek, které se v češtině obvykle pojí s určitým pádem ([*v*] *předprodeji*).

12 Za vytvoření korpusových výstupů pro účely této studie děkuji P. Vondříčkovi z ÚČNK FFUK.

13 Dalším důvodem případných drobných odlišností mohou být i rozdílné přístupy při manuální selekci dat, zvláště v případě MS s vyšším počtem kolokátů a u již zmíněné homonymie a zejména polysémie.

## 4.2 MONOKOLOKABILITA LEXÉMU/SLOVNÍHO TVARU

Jak už bylo řečeno výše, v některých případech lze za monokolokabilní požadovat celý lexém, jindy však omezenou kolokabilitu pozorujeme jen u některého/některých z alolexů. Rozdíly mezi monokolokabilitou lexému a slovního tvaru jsou samozřejmě patrné jen u ohebných slovních druhů, především u substantiv a verb;<sup>14</sup> u adjektiv nemá takovéto rozlišování valného smyslu, protože se shodují v rodě, čísle i pádě se substantivem, s nímž tvoří monokolokabilní kombinaci. Některá adjektiva se tudíž v obou jazycích vyskytují převážně či výhradně v jednom z rodů (*polohrubá [mouka], odpadkový [koš], bobkový [list]; [filo] spinato, [capro] espiatorio, [anidride] carbonica*), a v češtině ve výjimečných případech dokonce jen v jediném pádě (*[v] neposlední [řadě]*).

V italštině jsou vzhledem k výrazně chudší jmenné flexi rozdíly mezi monokolokabilními lexémy a slovními tvary, zejména v kategorii substantiv, méně patrné, neboť zde můžeme rozlišovat nanejvýš mezi singulárem a plurálem.<sup>15</sup>

Česká i italská flektivní MS, tedy zejména substantiva a verba, můžeme z tohoto pohledu rozdělit do tří skupin:

- a) **Monokolokabilní je lexém**, tj. celé konjugační nebo deklinační paradigma (v italštině omezené na tvary singuláru a plurálu substantiv), např. *přimhouřit [oči], zalapat [po dechu], [kruhový] objezd, [státní] zastupitelství, stroužek [česneku]; rendersi [conto di qc.], datore [di lavoro]*.
- b) **Monokolokabilní je pouze jeden tvar** lexému (výjimečně tvarů několik), ostatní alolexy výraznější omezení kolokability nevykazují, např. *[v] rámci [něčeho], [v] souladu [s něčím]* apod. V češtině jsou taková slova často komponenty víceslovných prepozic. V italštině bývá kolokačně omezen buď pouze singulár, nebo plurál substantiv: *[andare su tutte le] furie, [fare la] spola*.
- c) **MS existuje jen v jediném tvaru** (výjimečně ve dvou) a většinou nemá lemma, resp. musí být uměle (re)konstruováno, např. *[na] rozloučenou, [na] úkor [někoho, něčeho], [na] oplátku, [brát/vzít něco v] potaz, [od] nepaměti*. V italštině lze do této kategorie zařadit defektní MS, která nemají singulár nebo plurál (častěji), popř. se v jednom z čísel v současném jazyce (již) vůbec neuzívají, např. *[cantare/gridare a] squarciagola, [battere/correre all'] impazzata, [mettere qc. a] repentaglio*.

Hranice těchto kategorií ovšem nejsou ostré a v mnoha případech, zejména v italštině, je obtížné rozhodnout, zda MS již považovat za defektní (c), nebo zda je ještě

14 Přiřazení MS slovním druhům je pouze orientační a řídí se především formálními kritérii; k obtížnosti či dokonce nemožnosti určit synchronně slovní druh zejména u tzv. verboidů viz níže, kap. 6.

15 Nabízí se tedy otázka, zda má v italštině takovéto rozlišování vůbec smysl. Zmiňujeme je nejen kvůli možnosti symetrického porovnání s češtinou, ale i protože se domníváme, že mezi jednotlivými skupinami lze vysledovat určité rozdíly, které souvisejí mj. i s mírou lexicizace příslušných monokolokabilních spojení.



přičíst ke skupině (b): např. plurál substantiva *vigore* (*vigori*) se v korpusu vyskytuje pouze dvakrát (z celkem 8278 výskytů), plurál substantiva *concentramento* (*concentramenti*) šestkrát (z 525 výskytů slova). Situaci mnohdy ještě komplikuje polysémie MS, a navíc se zdá, že zejména kategorie (b) a (c) mohou z diachronní perspektivy představovat — přinejmenším u některých MS — dvě stadia vývoje, při němž dochází k postupnému omezování kolokačního paradigmatu u některého tvaru slova (b) a jehož výsledkem může být situace, kdy se slovo dochová pouze v jediném tvaru, právě v rámci příslušného monokolokabilního spojení (c).<sup>16</sup>

Monokolokabilitu celého lexému (a) lze v obou jazycích pozorovat zejména u sloves a dále v případech, kdy MS se svým kolokátem (kolokáty) tvoří víceslovné pojmenování terminologické povahy. Defektní MS (c), kterých je v našem vzorku — v obou jazycích — relativně vysoký počet (přes 25%), se naopak objevují výhradně ve frazémeh.<sup>17</sup>

#### 4.3 ČESKÁ A ITALSKÁ MS Z HLEDISKA SLOVNÍCH DRUHŮ

Ačkoli výše zmiňujeme, že u některých MS (verboidů) je přiřazení ke slovnímu druhu obtížné, ba dokonce nemožné (*bycha*, *pré*, *bacha* atd.), v našem vzorku se takových slov vyskytuje jen velmi málo a určení slovního druhu — přinejmenším formálně — možné je. Výjimku představuje např. české MS *krážem* či italské výrazy *frattempo*, *squarciagola* a *malincuore*, u nichž většina slovníků slovní druh neuvádí a odkazuje pouze na spojení, v němž se vyskytují (*nel frattempo*, *a squarciagola*, *a malincuore*); jiné slovníky, např. GRADIT, považují první dvě MS za substantiva a třetí za adverbium. V následující tabulce uvádíme slovní druh pouze u českých a italských slov. Přejímky vyčleňujeme zvlášť, neboť tato MS byla většinou přejata v rámci víceslovných jednotek a určení slovních druhů u jednotlivých komponentů by bylo místy značně problematické (srov. české *de facto*).

	čeština	italština
substantiva	42,6 %	49,5 %
adjektiva	41,7 %	16,8 %
adverbia	11,1 %	5,3 %
verba	3,7 %	4,2 %
přejatá (a v č. nesklonná) slova	0,9 %	24,2 %

TABULKA 1. Slovní druhy MS

V obou jazycích převažují substantiva, následovaná — zejména v češtině — překvapivě četnými adjektivy. Nejméně zastoupena jsou verba.<sup>18</sup>

<sup>16</sup> K diachronnímu vývoji MS viz níže, kap. 10.

<sup>17</sup> K frazémům a termínům srov. níže kap. 8.

<sup>18</sup> Toto zjištění odpovídá tomu, co Fleischer (1989, s. 121n) předpokládá pro němčinu. Stejná tendence byla ostatně potvrzena také na větším vzorku pro italská a německá MS (srov. Obstová — Vachková, v tisku).



Zajímavý rozdíl mezi češtinou a italštinou je v počtu přejatých MS: zatímco v češtině se do našeho vzorku dostalo pouze jediné takové slovo (*[de] facto*), v italštině tvoří přejímky významnou skupinu.

#### 4.4 MONOKOLOKABILNÍ KOMBINACE (MK)

Ustálená slovní spojení obsahující MS můžeme nazývat monokolokabilními kombinacemi (MK).<sup>19</sup> Tyto kombinace lze v zásadě rozdělit do dvou skupin podle toho, zda mají charakter frazému, nebo víceslovného pojmenování terminologické povahy.

##### 4.4.1 FRAZÉMY VERSUS TERMÍNY

Čermák (2014, s. 13; Čermák et al. 2016, s. 7) konstatuje, že naprostá většina MS tvoří v kombinaci se svými kolokáty široce pojímaný frazém nebo — rovněž široce pojímaný — termín.<sup>20</sup> Tuto tendenci potvrzují i monokolokabilní kombinace v našem vzorku.<sup>21</sup> V obou jazycích lehce převažují frazémy, které v češtině tvoří více než 50 % a v italštině více než 60 % všech zkoumaných MK (např. *chodit křížem krážem*; *prendere il sopravvento*). Za víceslovné termíny lze v češtině považovat kolem 40 % a v italštině pak kolem 30 % všech kombinací (např. *tržná rána*; *cordone ombelicale*). Přesná procenta zde neuvádíme, neboť jsme si vědomi neostré hranice mezi široce pojímanými frazémy či termíny a běžnými kolokacemi a zároveň i terminologické neustálenosti pojmu kolokace. Do skupiny frazémů tak řadíme např. i gramatické (kvazi)frazémy, jako jsou frazémy prepoziční (v souladu s něčím či *in prossimità di qc.*), do skupiny termínů potom spojení jako *památkově chráněný*, která by jiní badatelé mohli označit za běžné kolokace (srov. např. Konecny, 2010, s. 310). Výše uvedená — byť jen přibližná — čísla jsou zcela v souladu s Čermákovým tvrzením, že výskyt MS se neomezuje na členství ve frazémeh a že extrémně omezená kolokabilita je příznačná i pro komponenty mnoha víceslovných termínů. To platí i pro přejaté výrazy: v češtině lze jediné cizí spojení (*de facto*) ve vzorku přiřadit k frazémům, v italštině je přejatých frazémů sedm (v naprosté většině jde o latinismy, např. *a priori*), termínů potom šestnáct (především anglicismy, např. *mailing list*).<sup>22</sup>

<sup>19</sup> Srov. např. termín *Cranberry Expressions* (Trawiński et al., 2008).

<sup>20</sup> Frazém zde vymezujeme ve shodě s Čermákem jako jedinečnou a ustálenou kombinaci minimálně dvou prvků, z nichž některý/žádný nefunguje stejným způsobem v jiných kombinacích, popř. se vyskytuje pouze v kombinaci jediné (či několika málo). Termín potom jako jednoznačný, ustálený, deskriptivní, systematický název užívaný pro denotáty v určitém vědním oboru, ale i řemeslu či speciálním povolání (srov. např. Čermák, 2010, s. 225 a s. 132).

<sup>21</sup> Počty MS a MK si přesně neodpovídají, neboť jedno MS může být komponentem několika MK (např. *in precedenza vs. avere/dare la precedenza*). Výjimečně může být MS součástí frazému i termínu zároveň: takový případ se sice mezi zkoumanými slovy, kde převládají MS s jediným kolokátem, nevyskytl, mezi MS s nižším HHI jich však několik najdeme (např. termín *midollo osseo/spinale/allungato* a frazém *fino al midollo*, srov. Obstová, 2016b, s. 38).

<sup>22</sup> V češtině (resp. v korpusu SYN 2010) se termíny přejaté z angličtiny vyskytují také, je jich však méně a mají mnohem nižší frekvenci než v korpusu italském, takže se do našeho vzorku nejfrekventovanějších MS nedostaly.



OPEN ACCESS

Převážná část identifikovaných frazémů má verbální, popř. adverbialní charakter, skoro všechny gramatické frazémy plní funkci prepozic. Termíny mají téměř výlučně povahu nominální (srov. následující podkapitulu).

#### 4.4.2 FORMÁLNÍ STRUKTURY MK

Monokolokabilní kombinace, do nichž spolu s MS vstupují i další komponenty, jsou funkčně ekvivalentní téměř všem slovním druhům. České i italské MK v našem vzorku jsou převážně nominálního, verbálního, adverbialního a prepozičního charakteru, jak je zřejmé z následujících tabulek:<sup>23</sup>

typ MK	zastoupení ve vzorku	převládající form. struktura	příklad
nominální	41,2 %	<u>A</u> + S	<i>klíčn<u>í</u> kost</i>
verbální	22 %	V + prep + <u>S</u>	<i>mít (něco) za lubem</i>
adverbialní	22 %	prep + <u>S</u>	<i>bez ustání</i>
prepoziční	7 %	prep + <u>S</u> prep + <u>S</u> + prep	<i>na úkor (někoho/něčeho) v souladu s (něčím)</i>
adjektivní/atributivní	5,3 %	<u>ADV</u> + A	<i>zbrusu nový</i>

**TABULKA 2.** Nejčastější formální struktury MK v češtině

typ MK	zastoupení ve vzorku	převládající form. struktura	příklad
nominální	31 %	S + <u>A</u> S + prep + (det) + <u>S</u>	<i>filo spinato</i> <i>fiore all'occhio</i>
adverbialní	29 %	prep + (det) + <u>S</u>	<i>a casaccio</i>
verbální	23 %	V + prep + (det) + <u>S</u>	<i>andare a braccetto</i>
adjektivní/atributivní	8,7 %	prep + <u>S</u>	<i>a tracolla</i>
prepoziční	7,8 %	prep + (det) + <u>S</u> + prep	<i>sotto l'egida di (q.)</i>

**TABULKA 3.** Nejčastější formální struktury MK v italštině

K číslům uvedeným v tabulkách je třeba podotknout, že ani zde si počet MS a MK zcela neodpovídá (srov. poznámku 15), protože jedno MS může figurovat v několika kombinacích. V některých případech je navíc obtížné rozhodnout, zda danou strukturu přiřadit k verbálnímu, nebo adverbialnímu typu, tj. zda považovat sloveso za nedílnou součást monokolokabilního spojení (např. [*dodat/řící*] *na vysvětlenou*; [*venire/essere/restare*] *a galla*).<sup>24</sup> Jindy se zase určitá kombinace používá také samostatně, bez slovesa, např. ve

<sup>23</sup> Uvádíme zde jen nejčastější typy kombinací; podrobný rozbor formálních struktur italských MS viz Obstová, 2016b, s. 38–42.

<sup>24</sup> S tímto problémem se setkáváme zejména u monokolokability, kterou bychom mohli označit za progresivní (či „na druhou“), a tedy v případech, kdy je určité MS kompen-



funkci neshodného atributu (*dům je na spadnutí/je to na spadnutí × chatrč na spadnutí; essere in voga × uno slogan in voga*): takové případy lze považovat v širším smyslu za transformace. Je-li některé MS v korpusu významně doloženo ve více strukturách, posuzujeme každou takovou strukturu zvlášť a řadíme ji do příslušné kategorie.

Jak je z údajů v tabulkách patrné, v obou jazycích převládají struktury nominální: v češtině bývají tvořeny spojením monokolokabilního adjektiva a substantiva, v italštině substantivem a monokolokabilním adjektivem nebo (méně často) substantivem postmodifikovaným předložkovou vazbou s monokolokabilním substantivem. Zřejmá preference češtiny vyjadřovat přívlastek adjektivem se odráží i ve výrazně vyšším počtu českých monokolokabilních adjektiv v tabulce slovních druhů (tab. 1). Následují struktury verbální a adverbialní (v italštině jsme zaznamenali vyšší počet adverbialních frází). Mezi nejméně zastoupené patří v obou jazycích struktury prepoziční, které jsou ovšem velmi frekventované (nejfrekventovanějšími monokolokabilními kombinacemi jsou právě víceslovné prepozice, srov. níže kap. 9).

Je třeba zmínit, že analýza celého souboru italských MS (tedy i MS s nižším HHI, která v této studii zkoumány nejsou, srov. Obstová, 2016b, s. 38–42) vedla k mírně odlišným výsledkům: především zaznamenala o něco vyšší počet italských nominálních kombinací (nominální struktury 38,8 %, verbální 18 %, adverbialní 13,6 %, prepoziční 4 %). Tato čísla se velmi blíží hodnotám, které Vogherová uvádí pro všechny víceslovné jednotky (*polirematiche*) v korpusu LIP (Voghera, 2004, s. 61), a naznačují, že se italské monokolokabilní kombinace z formálního hlediska příliš neliší od běžných víceslovných pojmenování.

#### 4.4.3 KOLOKAČNÍ STRUKTURY

Ze stručného nástinu syntaktické typologie kombinací s monokolokabilním komponentem v předchozí kapitole vyplývají i určité rozdíly v kolokačních strukturách českých a italských MK. V obou zkoumaných jazycích stojí kolokáty obvykle v bezprostředním sousedství MS, rozdíl je však v zastoupení pravé a levé strany kontextu, jak je patrné z následujících tabulek:

kolokát vlevo od MS	42 %	<i>vzít v potaz</i>
kolokát vpravo od MS	51 %	<i>bobkový list</i>
kolokát vlevo i vpravo	7 %	<i>s přihlédnutím k něčemu</i>

**TABULKA 4.** Kolokační struktury v češtině

kolokát vlevo od MS	73 %	<i>anidride carbonica</i>
kolokát vpravo od MS	15 %	<i>alter ego; datore di lavoro</i>
kolokát vlevo i vpravo	12 %	<i>alla stregua di qc.</i>

**TABULKA 5.** Kolokační struktury v italštině

tem adverbialního spojení (*na odiv*), které se dále kombinuje s velmi omezeným počtem sloves (*stavět/vystavovat/dávat na odiv*).



V češtině jsou víceméně rovnoměrně zastoupeny obě strany kontextu. Vpravo od MS bývá kolokát především u nominálních a adjektivních spojení, vlevo u frází verbálních a adverbálních. V italštině stojí kolokát (kolokáty) téměř vždy vlevo od základového slova: hodnota uvedená v tabulce u pravostranných kolokátů (15 %) je dána relativně vysokým počtem přejatých spojení, která jsme zde rovněž brali v potaz (srov. *alter ego*, *mailing list* aj.). Původní italská spojení s kolokátem vpravo jsou velmi vzácná, zaznamenali jsme pouze šest případů (*datore di lavoro*, *lasso di tempo*, *rendersi conto*, *stragrande maggioranza* a dvě spojení s apokopovanými formami: *pian piano* a *quartier generale*). Kolokát na obou stranách v obou jazycích vyžadují MS, která jsou součástí komplexních předložek. Nejčastější kolokační struktura (*collocational span*) analyzovaných MK v češtině tedy je MS +1 a MS -1, v italštině převládá struktura MS -1.

Zkoumaná MS mají v obou jazycích v naprosté většině případů (přes 86 %) <sup>25</sup> povahu modifikátoru. Čeština i italština upřednostňují právě větvení: rozdíl v kolokačních strukturách je tedy dán především vysokým podílem monokolokabilních adjektiv, jejichž pozice se v obou jazycích liší (v češtině stojí obvykle před substantivem, v italštině častěji za ním). Případy, kdy MS tvoří hlavu syntagmatu, jsou v obou jazycích spíše výjimečné (max. 5 %): v našem vzorku lze nalézt pouze pět takových českých spojení (*stroužek česneku*, *ublížení na zdraví*, *kruhový objezd*, *pokřčit rameny*, *zalatat po dechu*) a – nebereme-li v úvahu přejaté kombinace — čtyři italská (*datore di lavoro*, *lasso di tempo*, *quartier generale*, *rendersi conto*).

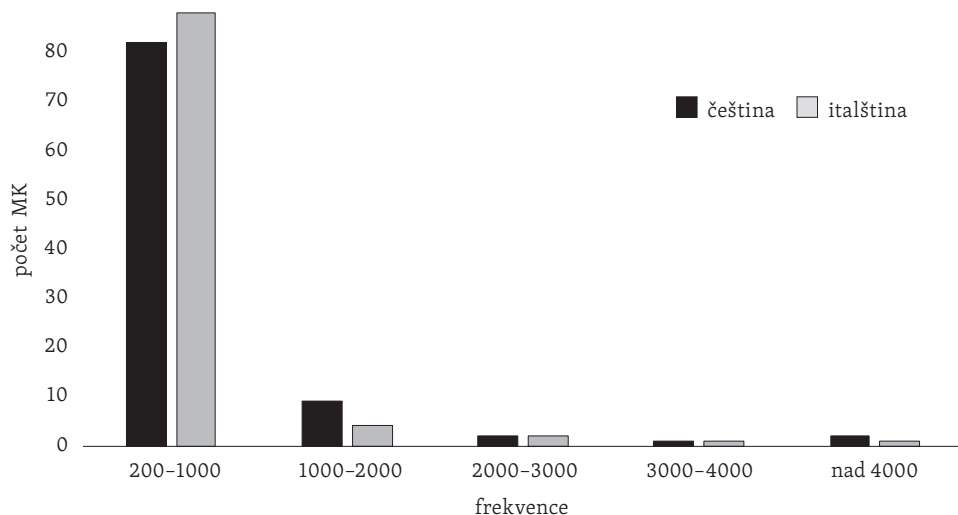
#### 4.5 FREKVENCE MS

Vzhledem k extrémně omezené kolokabilitě představují MS anomální útvary náležející kombinatorické periférii jazyka. Tradičně bývají vímána jako okrajová i z hlediska frekvence: značná část slov, jejichž omezená kolokabilita je v českých i italských slovnících signalizována sémantizační poznámkou („jen ve spoj.“, „solo nella loc.“ apod.), má v korpusech, které jsme použili, frekvenci nižší než 200 (a nebyla tedy, jak zmiňujeme výše, do našeho výzkumu vůbec zahrnuta; srov. např. *pozdě bycha honiti*, *fare lo gnorri*). <sup>26</sup> Identifikace MS v korpusech (a tedy i „odhalení“ takových MS, na jejichž monokolokabilitu slovníky obvykle neupozorňují) však umožnila tento pohled přehodnotit: některá z těchto slov se totiž v korpusu vyskytují velmi často. Nejfrekventovanější český slovní tvar *rámci* má ve stomiliónovém korpusu absolutní frekvenci 16 708, nejfrekventovanější italské MS *vigore* 8274. Frekvenci českých a italských MS v našem vzorku znázorňují následující grafy: <sup>27</sup>

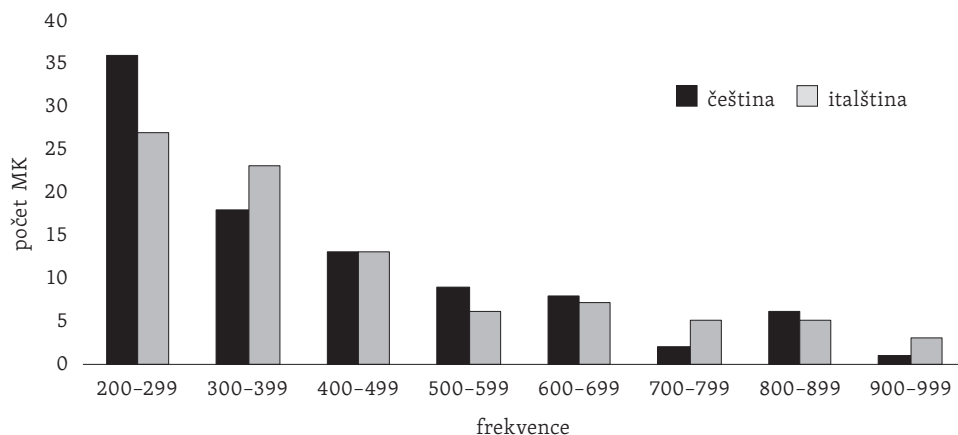
<sup>25</sup> V rámci komplexních předložek ani ve frazémeh typu *chtě nechtě*, *jakž takž* hlavu a modifikátor neurčujeme.

<sup>26</sup> Pro italštinu tato MS analyzuje — i z frekvenčního hlediska — Veland, 2006.

<sup>27</sup> Údaje o frekvenci MS jsou pouze orientační: situace je, zejména v češtině, komplikována tím, že jsme nuceni porovnávat frekvence lemmat s frekvencemi slovních tvarů.



**GRAF 1.** Frekvenční pásmo 200 až 4000



**GRAF 2.** Frekvenční pásmo 200 až 1000

Z grafů je patrné, že MS v našem vzorku se skutečně rekrutují z různých frekvenčních pásem. Slova s velmi vysokou frekvencí jsme ovšem identifikovali jen výjimečně: převážná většina analyzovaných MS se nachází pod hranicí 1000 výskytů v korpusu a s klesající frekvencí jejich počet narůstá. Vysokých frekvencí — v obou jazycích — dosahují zejména některé víceslovné prepozice, byť z hlediska formální struktury patří k nejméně zastoupeným (srov. výše, kap. 8.2): to je zcela v souladu s druhým Zipfovým zákonem (obecně: čím vyšší frekvenční hladina, tím méně typů, a naopak). Náš výzkum se ovšem zaměřil pouze na slova s frekvencí nad 200; k méně frekven-



tovaným MS zatím nemáme srovnatelné údaje.<sup>28</sup> Stejně tak by bylo třeba prozkoumat frekvenci MS v závislosti na složení korpusů, resp. funkčních stylech: v korpusu odborných textů se pravděpodobně setkáme s vyšším počtem termínů s monokolokabilním komponentem než v korpusech reprezentativních a vyvážených, v korpusech mluveného jazyka budou mít zřejmě větší zastoupení monokolokabilní frazémy (srov: *Dej si bacha!*).

#### 4.6 NĚKOLIK POZNÁMEK K DIACHRONII ČESKÝCH A ITALSKÝCH MS

Slova identifikovaná pomocí HHI v korpusech SYN 2010 a CORIS můžeme považovat za monokolokabilní z hlediska současné češtiny a italštiny. Je však zřejmé, že v diachronní perspektivě nabývá otázka monokolokability mnohem komplexnějších obrysů. Mnohá MS byla v dřívějších stádiích jazyka součástí běžné slovní zásoby a jejich kombinatorický potenciál nebyl nijak výrazněji omezen, časem však došlo — v rámci procesu, který bychom mohli označit za monokolokabilizační — k výrazné redukci jejich kolokačního paradigmatu (např. proto, že skutečnost, kterou tyto výrazy označovaly, zcela zanikla, nebo je vytlačila jiná, často novější pojmenování). Na taková slova, resp. slovní tvary lze nahlížet jako na „tradované jazykové relikty“ (srov. Dobrovolskij, 1988, s. 87). To ovšem neznamená, že by kombinace (MK), jež tato MS obsahují, musely být vnímány jako zastaralé nebo byly v lexikonu okrajové: z výše uvedených údajů o frekvenci je patrné, že řada takových spojení se v dnešním jazyce vyskytuje velmi často. U většiny těchto MS není ani zastřena jejich motivovanost. V této stati pochopitelně není dost prostoru pro detailní etymologickou analýzu jednotlivých MS, resp. celých MK (diachronní výzkum tohoto typu by navíc nutně narážel na četná úskalí, od nedostatku dokladů přinejmenším pro určitá období vývoje po nemožnost usuzovat z dochovaných textů na to, jakým způsobem a s jakou frekvencí se jednotlivá MS užívala např. v mluveném jazyce), přesto považujeme za zajímavé alespoň letmo pohlédnout na MS v našem vzorku i z hlediska jejich stáří, resp. jejich prvního doloženého výskytu. Údaje o roce, popř. století prvního doloženého výskytu čerpáme pro italštinu z *Dizionario etimologico della lingua italiana* (DELI), tyto informace jsou ovšem obsaženy i v některých italských výkladových slovnících (např. GRADIT nebo Il Sabatini-Coletti). V češtině tak přesnou dataci k dispozici nemáme, mnohá z analyzovaných MS se nám však podařilo dohledat v elektronických slovnících a databázích přístupných z webových stránek ÚJČ (Vokabulář webový, Lexikální databáze humanistické a barokní češtiny, Kartotéka lexikálního archivu), o něž se opírají např. i údaje v Rejzkově *Českém etymologickém slovníku*.

Navzdory tomu, že se nám v češtině nepodařilo ověřit stáří všech slov a že i u těch MS, která jsme dohledali, lze období jejich prvního výskytu stanovit jen přibližně, můžeme konstatovat, že velká část zkoumaných českých MS (více než třetina) náležela už k nejstarší slovní zásobě. Ve staročeských slovnících jsou doložena slova jako *bobkový* (nejčastěji používáno ve spojeních *bobkový olej*, *bobkové dřevo*), *očitý* (v různých významech, např. *mající oči*, *očima zprostředkovaný*, *očima viditelný*, *očividný*,

<sup>28</sup> Frekvenci některých méně častých MS lze porovnat např. in Čermák, 2014, a Veland, 2006).



ale i v dnešním významu opřený o vidění na vlastní oči), *lub* (obednění kolem mlýnského kamene: *mít něco za lubem* znamenalo mít něco schováno pro sebe — mlynářovi zbylo za lubem vždy trochu mouky, třebaže mlečovi odevzdal zdánlivě všecku),<sup>29</sup> *vrub* (zářez, dluh: při neplacení se dělaly vruby na holi), *úkor* (příkoří, pohana), *ohled*, *třetice* (již ve spojení *do třetice*) a mnoho dalších. V našem vzorku se vyskytuje i celá řada verbálních substantiv, která jsou rovněž doložena v nejstarších textech, např. *domnění* (*domnēnie*), *činění* (*čiňenie*), *spadnutí* (*spadnutie*, ve střední češtině pak již ve spojení *na spadnutí*).

Několik MS pochází pravděpodobně z období středněčeského (*potaz*, *zbrusu*, resp. *z brusu*). Zajímavý je rovněž častý výskyt dějových jmen se sufixem *-ná* ve spojení s předložkou *na* s akuzativem (*na shledanou*, *na rozloučenou*, *na rozmyšlenou*): tento typ se objevuje ve střední době v souvislosti s nivelizací pádových koncovek singuláru verbálních substantiv (srov. Lamprecht — Šlosar — Bauer, s. 275), často již (byť ne výhradně) ve zmíněné akuzativní formě, a dále jej rozvinulo zejména národní obrození.

Mnohá MS jsou naopak doložena teprve po r. 1900, resp. po r. 1950: většinou jde o komponenty ekonomických popř. sportovních termínů, nezřídká přejatých z jiných jazyků (*přesilová*, *antimonopolní*, *dispozice*). Vliv historických událostí na proces monokolokabilizace některých spojení ilustruje adjektivum *hákový*: substantivum *hák* (ze střhn. *hāke(n)*) najdeme už v nejstarších textech, v 16. století je doloženo adjektivum *hákový* ve spojení *háková cihla*, v 19. století a počátkem století dvacátého se toto adjektivum objevuje v řadě technických termínů (*hákové střechy*, *hákový sochor/klíč/žebřík*), a teprve od 30. let 20. století začíná v souvislosti se zneužitím svastiky jako nacistického symbolu převládat spojení *hákový kříž*, které se dnes jeví jako monokolokabilní (pouze u šesti z celkových 248 výskytů má toto adjektivum v korpusu jiný kolokát (*hákový úchyt* — 3×, *hákový nos*, *hákový nosič*, *háková nástavba* — 1×).

U italských MS je situace obdobná: i zde je více než třetina slov doložena už v duecentu či trecentu. Tato MS většinou pocházejí z latiny a jejich kolokační paradigma nebývalo ve starších textech výrazněji omezeno (srov. *dosso*, *erta*, *bilico*, *spinato*, *pallio*, *dismissura*). Například slovo *datore* se objevuje už v r. 1292 u Giambonih a po staletí bylo užíváno ve významu „dárce“<sup>30</sup>. Setkáme se s ním u Danta: „Questa letizia non può dare altro che utilitate, che rimane nel *datore* per lo dare, e che viene nel ricevitore per lo ricevere,“ i u D'Annunzia, zde dokonce v přechýlené podobě: „Veniva verso di noi quella che m'era stata rappresentata come la *datrice* di forza.“ V dnešní italštině se toto slovo vyskytuje téměř výhradně ve spojení *datore di lavoro* (zaměstnavatel),<sup>31</sup> které je kalkem z německého *Arbeitsgeber* a jehož první výskyt byl zaznamenán v r. 1923 u Mussoliniho. Podobně i slovo *vigore*, doložené už u Danta či Boccaccia („un monaco, il *vigore* del quale né la freschezza [...] potevano macerare“), bylo užíváno bez výraznějších kolokačních omezení, zejména ve významu „životní síla, energie“. Toto slovo má i další významy (Battagliův slovník jich uvádí celkem devět), v kor-

29 Srov. Machek, 2010.

30 Srov. S. Battaglia, *Grande dizionario della lingua italiana*: „chi dà, chi concede, chi elargisce (doni, ricchezze, elemosine, beni, ecc.); chi procura (pace, gioie, felicità, grazie, ecc.)“.

31 Jiné kolokáty tohoto MS (které má v CORISu frekvenci 3489) jsou v korpusu ojedinělé, např. *datore di vita* (8 výskytů), *datore d'ipoteca* (2 výskytů).



pusu CORIS se ovšem objevuje téměř výhradně ve významu „účinnost, platnost“, doloženém v právních dokumentech od 15. století, a to ve spojení *entrare/essere in vigore* (vstoupit v platnost, být v platnosti). Další MS, která byla součástí nejstarší slovní zásoby, jsou arabského, langobardského či jiného původu (*auge, spola, briga*).

Druhá významná skupina MS pochází z 16. století a počátku století sedmnáctého (např. *impazzata, ridosso, unisono, ombelicale, tracolla, stregua*), některá MS byla zaznamenána na konci století osmnáctého (*malapena, carbonica, lasso, vertebrale*).

V našem vzorku jsou hojně zastoupena i slova nová, doložená v jazyce od druhé poloviny 20. století. Velkou část těchto výrazů tvoří přejímky z angličtiny (*tilt, task, line, seller, mailing*); přestože byla tato slova většinou přejata již v rámci příslušného víceslovného pojmenování (*task force, mailing list*), domníváme se, že je z hlediska přijímajícího jazyka lze posuzovat samostatně a na základě korpusových dat je rovněž považujeme za monokolokabilní.

Výše uvedené příklady naznačují, že v procesu monokolokabilizace hraje často významnou úlohu polysémie: slovo se dochová jen v jednom ze svých významů, v rámci určitého ustáleného spojení (termínu nebo frazému), zatímco ve významech ostatních postupně zastarává a z běžného jazyka se vytrácí.<sup>32</sup>

## 5. ZÁVĚR

Slova s extrémně omezenou kolokabilitou byla dosud zkoumána především v rámci frazeologie jako nevydělitelná součást frazémů. Vzhledem k určité archaičnosti, která je pro některá MS typická, byla nezdědká vnímána jako slova v současném lexikonu nepřilíš frekventovaná, hovořilo se o nich jako o „jazykových reliktech“, poukázvalo se na to, že se obvykle vyskytují pouze v určitém tvaru a jejich lemma je nutno uměle rekonstruovat. Korpusový výzkum však ukázal, že problematika monokolokability je mnohem komplexnější a že MS tvoří (přínejmenším v jazycích, jako je čeština a italština) poměrně heterogenní skupinu. Kromě frazémů bývají velmi často také komponenty víceslovných termínů i běžných kolokací, některá MS jsou v současném jazyce značně frekventovaná, jiná se vyskytují vzácně, některá prošla dlouhým historickým vývojem postupné monokolokabilizace, jiná jsou doložena teprve v nedávné době a mnohdy byla přejata, často v rámci celé monokolokabilní kombinace, z jiného jazyka. Některá MS se dochovala jen v určitém tvaru, u jiných lze za monokolokabilní považovat všechny tvary lexému.

Z porovnání českých a italských MS s nejvyšším HHI vyplynuly některé **podobnosti**, ale také celá řada rozdílů. MS v obou jazycích bývají srovnatelně často komponenty frazémů i termínů a vykazují podobné vlastnosti z hlediska slovnědruhového (nejčastěji je lze přiřadit ke kategorii substantiv či adjektiv, velmi vzácná jsou

32 Tuto tendenci můžeme pozorovat nejen u slov, které prošla dlouhým historickým vývojem, ale i u některých výrazů novějších (srov. it. technický termín *ammortizzatore* [tlumič], vyskytující se v korpusu s překvapivou četností — v přeneseném významu — ve spojení *ammortizzatori sociali* [doslova „sociální tlumiče“, tj. zejména podpora v nezaměstnanosti], srov. Obstová, 2016a, s. 20).





naopak monokolokabilní verba). Až na výjimky mají v monokolokabilním spojení povahu modifikátoru a kolokát, s nímž tuto kombinaci tvoří, se téměř vždy nachází v bezprostředním sousedství MS (na rozdíl třeba od němčiny, kde může kolokát stát i na pozicích vzdálenějších). Obdobná je i frekvence analyzovaných MS v současné češtině a italštině (resp. v příslušných korpusech). Letmý dicharonní pohled odhalil, že značná část (přinejmenším 1/3) zkoumaných MS byla součástí nejstarší slovní zásoby obou jazyků a kolokační paradigma těchto lexémů nebylo původně nijak zvlášť omezeno; teprve později došlo k jeho postupné redukci.

**Rozdíly**, které jsme mezi českými a italskými MS zaznamenali, jsou dány především **typologicky**. Výrazně se do problematiky monokolokability promítá zejména odlišná míra flektivnosti: zatímco v češtině, která má silně rozvinutou jmennou flexi, lze MS rozlišovat podle toho, zda monokolokabilní tendenci vykazuje celý lexém, pouze určitý slovní tvar, anebo zda MS dokonce existuje pouze v jediném tvaru a de facto nemá lemma, v italštině, kde se jmenná flexe omezuje na odlišné tvary singuláru a plurálu, je takové rozlišování obtížnější (rozdíl mezi lemmatem a slovním tvarem je zde méně patrný, a je tudíž problematictější označit určitá MS za *verboidy*). Typologicky jsou dány i některé rozdíly ve formálních strukturách (např. častější vyjadřování přívlastku adjektivem v češtině) a strukturách kolokačních (vzhledem k obvyklému postavení adjektiva před substantivem jsou v češtině víceméně rovnoměrně zastoupeny obě strany kontextu, zatímco v italštině stojí kolokáty téměř vždy vlevo od MS). Další odlišnosti mezi oběma jazyky lze vysvětlit **historicky**, např. přítomnost řady verbálních substantiv či dějových jmen se sufixem *-ná* v češtině nebo vysoký počet přejatých MS v současné italštině.<sup>33</sup>

V této studii jsme zkoumali jen slova, která dosahují nejvyšších hodnot indexu a navíc jsou v korpusu poměrně frekventovaná (nad 200 výskytů). Aby byl synchronní obraz MS v obou jazycích přesnější a úplnější, bylo by třeba zohlednit i slova s nižšími hodnotami HHI (jejich kolokační paradigma bývá širší, přesto je však lze stále považovat za monokolokabilní)<sup>34</sup> a především slova s nižší frekvencí a podrobit analýze také korpusy mluveného jazyka, kde předpokládáme významný výskyt frazémů s monokolokabilním komponentem.

#### PŘÍLOHA 1. Abecední seznam českých monokolokabilních slov s HHI (L/P) 1–0,7

antimonopolní	mezipohlavní	pokrčit	takž
berná	načase	polohrubá	točivý
bobkový	nadmořská	potaz	tržná
čerpací	naskrz	pozměňovací	třetice
činění	nedávnem	požadání	ublížení
činžovní	nedílná	provždy	uhličitý

33 Historickým procesem monokolokabilizace ovšem prošlo každé MS a u každého byl tento proces (i jeho příčiny) pochopitelně jiný, srov. výše kap. 1 a kap. 10.

34 Stanovit přesnou hodnotu HHI, od níž už lze slovo považovat za monokolokabilní, vzhledem ke škálovité povaze fenoménu nelze: to je i jedním z důvodů, proč jsme analyzovali pouze slova s nejvyššími hodnotami.



OPEN ACCESS

dálný	nedohlednu	předvečer	úkor
diplovová	nepaměti	přesilová	ustání
dispozici	neposlední	přestání	uzdě
dobírku	nepoznání	převážná	valná
dohledné	nevíře	přihlédnutím	vínku
domnění	nynějška	přimhouřit	visací
dostání	objezd	příspěvková	vrub
facto	objízdná	radno	všudy
fakultní	oblíbou	rámci	výkladní
hákový	obývací	rozloučenou	výkupní
holičkách	očitý	rozmyšlenou	vyměřovací
hydrometeorologický	odiv	sdělovací	vysvětlenou
chtě	odkvětu	shledanou	vyvolávací
jakž	odněti	sklonku	zalatpat
klíčn	odpadkový	skrytu	zastupitelství
koncentrační	ohledem	souladu	zatajeným
krážem	onak	spadnutí	zavazadlový
laskav	oplátku	stroužek	zbrusu
lícni	ostnatý	střehu	zdaňovací
lubem	památkově	širým	zoologická
mání	pokraji	štítná	životně

**PŘÍLOHA 2.** Abecední seznam italských monokolokabilních slov s HHI (L/P) 1–0,7

alter	frantumi	particolar	spinato
auge	frattempo	personam	spola
autonomasia	furie	pian	squarciagola
bilico	galla	pick	stento
bracchetto	immunitario	precedenza	stragrande
briga	impazzata	precindere	stregua
capolino	incirca	pressi	subordine
carbonica	insù	primis	suon
casaccio	integrante	priori	talk
cipollina	intenderci	prossimità	tantino
concentratione	lasso	quartier	tantum
condicio	line	quinte	task
court	lizza	rado	tempore
datore	made	rendersi	tentoni
decorrere	mailing	repentaglio	tilt
dismisura	malapena	rettorale	tracolla
disuso	malincuore	ridosso	unisono
dosso	mountain	ritroso	vedenti

editrice	novero	sbarrati	vertebrale
egida	occhiello	seller	vigore
erta	ombelicale	sitter	vitro
espiatorio	operator	sommato	voga
extravergine	paliò	sopravvento	week
extremis	part	spasso	



## LITERATURA:

- ARONOFF, M. (1976): *Word Formation in Generative Grammar*. Massachusetts: The MIT Press.
- CVRČEK, V. (2013): *Kvantitativní analýza kontextu*. Praha: NLN.
- ČERMÁK, F. — ČERMÁK, J. — OBSTOVÁ, Z. — VACHKOVÁ, M. (2016): *Language Periphery. Monocollocable Words in English, German, Italian and Czech*. Amsterdam: John Benjamins.
- ČERMÁK, F. — HOLUB, J. (2005): *Syntagmatika a paradigmatika českého slova. 1. Valence a kolokabilita*. Praha: Karolinum.
- ČERMÁK, F. (2010): *Lexikon a sémantika*. Praha: NLN.
- ČERMÁK, F. (2014): *Periferia jazyka. Slovník monokolokabilních slov*. Praha: NLN.
- DOBROVOĽSKIJ, D. (1988): *Phraseologie als Objekt der Universalienlinguistik*. VEB Verlag Enzyklopädie Leipzig. Linguistische Studien.
- DOBROVOĽSKIJ, D. — PIIRAINEN, E. (1994): Sprachliche Unikalia im Deutschen. Zum Phänomen phraseologisch gebundener Formative. *Folia Linguistica*, 28, 3–4, s. 449–473.
- FILÍPEČ, J. — ČERMÁK, F. (1985): *Česká lexikologie*. Praha: Academia.
- FLEISCHER, W. (1989): Deutsche Phraseologismen mit unikalener Komponente. Struktur und Funktion. In: G. GRÉCIANO (ed.), *Europhras 88, Phraséologie Contrastive. Actes du Colloque International Kligenenthal — Strasbourg, 12–16 mai 1988*. Strasbourg: Université des Sciences Humaines, Département d'Etudes Allemandes, Collection Recherches Germaniques 1, s. 117–126.
- FLEISCHER, W. (1997): *Phraseologie der deutschen Gegenwart*. Tübingen: Niemeyer.
- KOCHOVÁ P. — OPAVSKÁ Z. (eds.) (2016): *Kapitoly z koncepcie Akademického slovníku současné češtiny*. Praha: Ústav pro jazyk český AV ČR.
- KONECNY, Ch. (2010): *Kollokationen. Versuch einer semantisch-begrifflichen Annäherung und Klassifizierung anhand italienischer Beispiele*. München: Meidenbauer [Forum Sprachwissenschaften 8].
- LAMPRECHT, D. — ŠLOSAR, D. — BAUER, J. (1986): *Historická mluvnice češtiny*. Praha: SNP.
- OBSTOVÁ, Z. (2016a): A frequency dictionary of Italian monocollocable words“ a „An alphabetical dictionary of Italian monocollocable words. In: F. ČERMÁK et al., *Language Periphery. Monocollocable Words in English, German, Italian and Czech*. Amsterdam: John Benjamins, s. 55–68 a 89–97.
- OBSTOVÁ, Z. (2016b): Fenomeno della collocabilità ristretta nell'italiano di oggi. *Linguistica Pragensia*, 26, 2, s. 33–46.
- OBSTOVÁ, Z. — VACHKOVÁ, M. (v tisku): Aktuelle Korpusforschung zu monokollokabelen Wörtern im Italienischen und Deutschen: Bericht und Ausblick. In: Ch. KONECNY — E. AUTELLI — A. ANDREA — L. ZANASI (eds.), *Lexemkombinationen und typisierte Rede im mehrsprachigen Kontext*. 2 vols. Tübingen: Stauffenburg [Stauffenburg Linguistik].
- SAUSSURE, F. de (1996): *Kurz obecné lingvistiky*. Pfl. F. Čermák. Praha: Academia.
- SINCLAIR, J. (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SKALIČKA, V. (2004, 2006): *Souborné dílo, I.–III. díl*. Ed. F. ČERMÁK — J. ČERMÁK — P. ČERMÁK — C. POETA. Praha: Karolinum.
- TRAWIŃSKI, B. — SAILER, M. — SOEHN, J.-P. — LEMNITZER, L. — RICHTER, F. (2008):



Cranberry Expressions in English and in German. In: *Proceedings of the LREC 2008 Workshop: Towards a Shared task for Multilingual Expressions (MWE 2008)*. Marrakech, Morocco, 2008, s. 35–38.

VELAND, R. (2006): Il concetto di collocazione unica e il valore di predizione della dicitura “solo nella loc.” in uso nella pratica

lessicografica. *Zeitschrift für Romanische Philologie*, 122, s. 260–280.

VOGHERA, M. (2004): Polirematiche. In GROSSMANN, M. — RAINER, F. (eds.), *La formazione delle parole in italiano*. Tübingen: Max Niemeyer Verlag.

## KORPUSY A SLOVNÍKY:

KŘEN, M. — BARTOŇ, T. — CVRČEK, V. — HNÁTKOVÁ, M. — JELÍNEK, T. — KOCEK, J. — NOVOTNÁ, R. — PETKEVIČ, V. — PROCHÁZKA, P. — SCHMIEDTOVÁ, V. — SKOUMALOVÁ, H.: *SYN2010: žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z <<http://www.korpus.cz>>.

CORIS/CODIS, Università di Bologna, dostupný z <[http://corpora.dsl.unibo.it/coris\\_ita.html](http://corpora.dsl.unibo.it/coris_ita.html)>.

BATTAGLIA, S. (1961–2002): *Grande dizionario della lingua italiana*. 21 sv. Torino: UTET.

CORTELAZZO, M. — ZOLLI P. (1999): *Dizionario etimologico della lingua italiana (DELI)*. Bologna: Zanichelli.

DE MAURO, T. (2000): *GRADIT, Grande dizionario italiano dell'uso*. Torino: UTET, 2000.

IL SABATINI-COLETTI (2005): *Dizionario della Lingua Italiana — versione CD-ROM*, Milano: Rizzoli Larousse.

*Kartotéka lexikálního archivu* [on-line]. Oddělení současné lexikologie a lexikografie Ústavu pro jazyk český AV ČR, v. v. i., 2007–2008 [citováno ze dne 1. 12. 2016]. Dostupný na <[http://www.ujc.cas.cz/electronice-slovníky-a-zdroje/kartoteka\\_lexikalniho\\_archivu.html](http://www.ujc.cas.cz/electronice-slovníky-a-zdroje/kartoteka_lexikalniho_archivu.html)>.

MACHEK, V. (2010): *Etymologický slovník jazyka českého*. Praha: NLN.

NEJEDLÝ, P. et al.: *Lexikální databáze humanistické a barokní češtiny* [on-line]. Praha: Ústav pro jazyk český AV ČR, v. v. i., 2010 [citováno stav z 1. 12. 2016]. Dostupné z WWW: <<https://madla.ujc.cas.cz>>.

REJZEK, J. (2015): *Český etymologický slovník*. Praha: Leda.

*Vokabulář webový* [on-line]. Verze 1.1.1.. Oddělení vývoje jazyka Ústavu pro jazyk český AV ČR, v. v. i. [citováno ze dne 1. 12. 2016]. Dostupný na <<http://vokabular.ujc.cas.cz>>.

**Zora Obstová** | Ústav románských studií, Filozofická fakulta Univerzity Karlovy | nám. Jana Palacha 2, 116 38 Praha 1  
zora.obstova@ff.cuni.cz