

Lenka Šíková Hrejsemnová: *Monokolokabilní slova v češtině a dalších evropských jazycích*

Diplomová práce. Praha: FF UK, 2013.

Prezentovaná práce se zabývá monokolokabilitou v jazyce, tj. specifickými, velmi omezenými možnostmi kombinování některých slov (tzv. monokolokabilních slov či tvarů) s ostatními slovy.

Při zkoumání kolokability slova se sleduje kombinovatelnost ostatních slov (tzv. referenčních slov) se zkoumaným slovem v rámci syntagmatu. Monokolokabilita představuje určitý extrém. Zatímco v případě normální kolokability je množina potenciálních kolokací (referenčních slov) pro zkoumané slovo velmi početná (v různých syntagmatech se mohou vyskytovat různá referenční slova), u čisté monokolokability jde o přítomnost jediného referenčního slova v jediném možném syntagmatu (neboli přítomnost monokolokabilní jednotky v textu implikuje přítomnost specifického referenčního slova v rámci daného syntagmatu). Tento extrémní případ čisté monokolokability je v jazyce vzácný, zajímavé jsou však i další případy omezené kolokability, proto se pojem monokolokability (navzdory jeho etymologii) používá ve všech dosavadních pracích i pro případy, kdy referenční slova tvoří množinu slov až o maximálně sedmi či osmi prvcích (tzv. kolokační paradigma). (Bylo by tedy navzdory české tradici asi vhodnější mluvit o velmi omezené kolokabilitě, což je i termín užitý v anglickém názvu práce.)

Cílem autorčiny práce (která navazuje na její úžeji zaměřené zkoumání z klauzurní práce) je nejprve prozkoumat tento jev z metodologického hlediska, navrhnout způsoby, jak monokolokabilitu vyhledávat v korpusech, následně tento postup aplikovat na data v Českém národním korpusu a na základě toho vytvořit popis monokolokabilních slov a jejich kolokací (včetně klasifikace), který by byl využitelný i na další jazyky. Tuto aplikaci autorka představuje v závěrečné kapitole: monokolokabilitu popisuje na základě dat z paralelních korpusů na třech dalších jazycích.

První kapitola je věnována vymezení monokolokability, v části 1.1. jsou vymezeny související pojmy (zejm. kolokační paradigma, kolokace ve vztahu k zavedeným pojmům syntagma a paradigma). Probrány jsou i termíny související s vyhledáváním v korpusech (např. referenční slovo). V tomto kontextu autorka definuje i *slovo*, za něž považuje "formu oddělenou mezerou z každé strany přímo v korpusu" (s. 13). Část 1.2 rekapituluje dosavadní teorii a výzkum. Monokolokabilita jako taková je podle zjištění autorky poměrně málo studována, přehled literatury proto zahrnuje Čermákovy studie (v nichž monokolokabilní slova popisuje výhradně jako součást frazémů) a dvě tematické diplomové práce (Káninský, Stojčev), vůči nimž se diplomantka vymezuje. Autorka dále komparativně shrnuje i výsledky své vlastní klauzurní práce, v níž kromě frazémů (např. *křížem krážem*) již dříve odhalila i tři další typy monokolokabilních spojení: reflexiva tantum (*smát se*), slova valenčně vázaná (*záležet na*) a víceslovná pojmenování (*nadmořská výška*).¹ Ze zahraniční literatury (v níž není pojem monokolokabilita zaveden) se autorce podařilo objevit dvě studie (Moonová, Richter & Seiler), které se jevu v menší míře věnují.

V druhé kapitole je probírána metodologie vyhledávání. Za zdroj dat byl zvolen korpus SYN2010. Následně podkapitoly jsou věnovány způsobu, na jehož základě kolokabilitu odhalit. Autorka se rozhodla (oproti dřívějším pracím) zaměřit na statistické vlastnosti referenčního slova vzhledem k monokolokabilnímu. Tyto vlastnosti mohou být získány z korpusu jednak pomocí výpočtu entropie, jednak pomocí Herfindahl-Hirschmannova indexu (dále HHI, což je metoda užívaná v ekonomii; pro výpočet kolokací v lingvistice ji používá V. Cvrček). Obě metody autorka následně srovnává, představuje základní seznamy výsledků a jejich vlastností a dochází k tomu, že rozdíly mezi nimi nejsou pro výzkum zásadní. Dále proto počítá již jen s jednou z nich (HHI). Autorka zároveň na základě pilotní analýzy dat stanoví i detailní kritéria pro monokolokabilitu, která počítají i s přípustným 3% počtem referenčních slov, která jsou nad rámec monokolokabilního kolokačního paradigmatu. V návaznosti na to je také řešen problém, jak vysoké hodnoty HHI při ručním dohledávání mohou ještě poskytovat monokolokabilní výsledky. Ukazuje se, že žádná univerzální hranice neexistuje a prohledávání je třeba podřídít vlastnostem zkoumaných dat.

Třetí kapitola se věnuje podrobné analýze českých dat. Výskyty, které jsou výsledkem ručního prohledání monokolokabilních slov, autorka roztrídila podle několika hledisek do různých kategorií: 1) Kolokační paradigma referenčních slov může být složeno a) z jediného referenčního slova; b) z několika synonymních slov; c) referenční slova mohou tvořit množiny, které odpovídají polysémii zkoumaného slova; d) ke zkoumanému slovu se může vztahovat více kolokačních paradigmat. 2) Referenční slova mohou nabývat různé slovnědruhové podoby, která je buď a) plnovýznamová,

¹ Monokolokabilní slova zde podtrhávám.

b) gramatické povahy. 3) Monokolokabilní slova mohou být a) zcela neohebná, b) částečně ohebná, c) plně ohebná; monokolabilní jednotky je také možné rozlišovat podle toho, zda je monokolokabilní celý lexém, nebo jen některý z jeho tvarů, autorka rozlišuje i jejich slovní druh (na základě jejich systémové podobnosti s normálně kolokabilními slovy). Monokolokabilní jednotky autorka dále popisuje z hlediska jejich formální a sémantické charakteristiky a blíže představuje a rozebírá zajímavé problémy, které s sebou nese jejich polysémie/homonymie. Část 3.3 je věnována teoreticky podnětným otázkám. Jednak řeší, do jaké míry je koncept monokolokability vhodný pro popis vlastních jmen, terminologie a zkratk. Také se zabývá vlivem gramatiky, otázkami, zda je vhodné uvažovat o monokolokabilních slovech jako samostatných jednotkách, a problematikou interpunkce (interpunkce je při vyhledávání v korpusu považována za pozici a může být brána za referenční slovo). Autorka upozorňuje, že na základě přijaté definice jsou monokolokabilní všechna slova, která vyžadují doplnění přes předložku a také automaticky všechny tvary substantiv v lokálu (protože se musí pojit s předložkou). Poslední jmenovanou skupinu tvarů mezi monokolokabilní jednotky nezařazuje. Autorka tvrdí, že v tomto případě jde o "čistě gramatická pravidla", zatímco předchozí případy jsou lexikálního typu nebo na hranici lexikálního a gramatického typu. V podkapitole 3.3.6 si autorka všimá různé sémantické a gramatické povahy slov v rámci monokolokabilního spojení: zatímco u frazémů je obtížné připsat formální a sémantické vlastnosti monokolokabilnímu slovu, u reflexiv tantum má takovou charakteristiku naopak referenční slovo. Další podkapitola 3.4 je věnována praktickým aplikacím, které by výzkum kolokability mohl/měl přinést do lexikografické praxe (návrh, jak by měla vypadat hesla ve slovnících oproti současnosti).

Čtvrtá kapitola klasifikuje případy monokolokability ve třech dalších jazycích, které autorka ovládá (vzhledem k povaze výzkumu byl výběr jazyků limitován nutností porozumění): slovenštině, francouzštině a angličtině. Metodika a popis odpovídají vyhledávání v češtině, povaha dat je však odlišná. Jde totiž o korpusy InterCorp, které obsahují v první řadě právní texty a zápisy z jednání Evropského parlamentu. Autorka data použila alespoň k základnímu postižení rozdílů mezi jazyky. Ukázalo se jí, že analýza je aplikovatelná i na jiné jazyky a rozdíly odpovídají odlišnostem daným především rozdílnou morfologickou stavbou (což se projevuje hlavně u reflexiv tantum).

Pátá kapitola obsahuje shrnutí a nejdůležitější závěry.

Kritické zhodnocení

Diplomová práce představuje hodnotnou empirickou studii, podloženou daty z Českého národního korpusu (a částečně také z paralelních korpusů InterCorp). Autorka vyšla z definice monokolokability tak, jak ji dříve vymezil F. Čermák, a se zkušenostmi ze své dřívější klauzurní práce podrobila jeho koncept podrobnému empirickému zkoumání; autorka zároveň klasifikovala jednotlivé případy, které pod definici monokolokability spadají. V práci byla aplikována metoda pomocí HHI, která poskytuje dobrá vodítka k určení monokolokability. Autorka získaná data pečlivě zpracovala a ve výsledku podala komplexní pohled na to, co si lze v češtině představit pod pojmem "monokolokabilita". Lze souhlasit s jejími slovy ze závěrečné kapitoly, že jde o zatím asi nejrozsáhlejší českou korpusovou studii na téma monokolokability.

Práce však zároveň obsahuje i některé myšlenky, postupy a nejasnosti, které si žádají kritické výtky. Těm se nyní budu věnovat.

1. Autorka hovoří o tom, že monokolokabilita je v literatuře jen okrajově zkoumaný jev. Ve svém přehledu literatury kromě několika českých prací ukazuje i dvě zahraniční práce na angličtině, které se monokolokability dotýkají. Postrádám však zakotvení autorčina zkoumání monokolokability v širším (nejen specificky českém) metodologickém rámci: alespoň základní přehled metod, které jsou zkoumání monokolokability příbuzné. Zvláště se nabízí srovnání s postupy tzv. *collostruction analysis*, kterou vyvíjejí Gries a Stefanowitsch. Tato analýza se zabývá kolokacemi uvnitř gramatických konstrukcí. Podnětná by mohla být například úvaha nad tím, jaký přínos by studiu monokolokability mohla přinést *varying collexeme analysis* a naopak. Výzkum monokolokability tak i po této autorčině diplomové práci zbytečně zůstává ve zdánlivě výlučném vakuu.

2. Zároveň je mi zatěžko smířit se s tím, že autorka hlouběji nerozebrala určité relevantní koncepty z obecnělingvistického hlediska. Teoretické zhodnocení pojmu slova by si v práci zasloužilo více než jen odkaz na Haspelmathovu kritiku o intuitivnosti slova, zvláště proto, že tento pojem je naprosto klíčový pro vymezení monokolokability a vlastnosti jednotlivých čtyř typů monokolokability (klasifikovaných již v klauzurní práci) s pojetím slova velmi úzce souvisejí. Práce se této problematiky částečně dotýká až na straně 71, ačkoliv čtenář si podobné otázky musí klást nejspíše od stránky 16, kde jsou představeny výsledky klauzurní práce.

Také bych uvítal, kdyby se do rozboru vztahů mezi studovanými jazyky v 4. kapitole promítly i úvahy odborněji typologické.

3. Role interpunkce v rámci studia monokolokability mi zůstává nejasná. Autorka sice interpunkci nepovažuje za slovo (s. 13). Na s. 31 ale uvádí, že s interpunkcí bude pracovat jako se specifickým druhem kolokátu (referenčního slova). I v další analýze je interpunkce průběžně uváděna společně s dalšími kolokáty. Ačkoliv je vždy důsledně připomínán zvláštní status interpunkce, zůstává i po vysvětlení na s. 69 nejasné, proč nebyla interpunkce popsána odděleně od zbytku kolokací. Také není z práce zřejmé, zda byly kolokace prohledány i s "přeskočením" interpunkce (to by mohlo být relevantní např. pro (zcela hypoteticky monokolokabilní) spojení *době, kdy*).

4. Jestliže je v práci monokolokabilita zpracována na základě zvoleného přístupu víceméně formálním způsobem (spadá do ní několik různých skupin spojení a svoji roli může zastávat i interpunkce), je obtížné porozumět faktu, že spojení *záviset na* je považováno za monokolokabilní, zatímco předložkový lokál *autě* nikoliv. Tento přístup k monokolokabilitě nepovažuji za konzistentní. Nebylo by vhodnější mluvit o rozlišení mezi lexikalizovanou a gramatikalizovanou monokolokabilitou?

5. Menší nejasnosti, nápady a otázky:

S. 30: Zde se píše, že vyhledána byla slova s frekvencí přesahující 200 výskytů. Chápu-li to správně: proč byla zvolena stejná frekvenční hranice pro tvary i pro lemmata?

S. 41: Zde by patrně bylo vhodné se vyjádřit i k dosaženým výším korelačního koeficientu (tj. z hlediska jeho velikosti, nikoliv jen signifikance).

S. 57: Citace: "Je-li nějaké doplnění obligatorním argumentem slova, tvoří vztah k němu vlastně součást významu tohoto slova." Nápad: mohli bychom tedy o předložce *na* v kolokaci *záviset na* uvažovat jako o specifickém případě head-markingu (značení vztahu na hlavě) v češtině?

S. 75: Citace: "[M]onokolokabilita není jevem vymezeným jen v rámci teorie, ale existujícím v realitě jazykových projevů." Neměla by být monokolokabilita (i vzhledem k výše uvedené diverzitě) považována spíše za evidenci pro soubor různých jevů v jazykových projevech, než za jeden určitý jev?

S. 79: K rozdílu mezi češtinou a slovenštinou: je otázka, jak by dopadlo porovnání češtiny se slovenštinou, kdyby byla data pro češtinu odebrána ze stejného korpusu InterCorp.

6. K formální stránce práce (k některým obtížím):

a) V práci není jasný rozdíl mezi osobami, za něž a k nimž autorka promlouvá. Autorka kombinuje autorský singulár (typ *udělala jsem*), autorský plurál (typ *udělali jsme*) a čtenářský plurál (typ *podívejme se*). Navíc v některých pasážích autorka to, co pravděpodobně udělala, popisuje jako *je třeba udělat* (viz s. 67, 2. odst.). Čtenář tím nabývá nejistoty o tom, co skutečně bylo s daty provedeno a co je pouze navrhováno.

b) Autorka používá jen výjimečně odrážkové seznamy; jejich užití by se hodilo zvláště při rozřizování monokolokabilních spojení, kde není užít ani jakýkoliv jiný způsob výčtu.

c) Bylo by přívětivé, kdyby ve všech příkladech bylo nějak odlišeno monokolokabilní slovo (autorčin slib umisťovat referenční slova do závorky není důsledně dodržován). Podobně: v příloze č. 2 by bylo usnadnilo orientaci, kdyby kolokáty byly označeny podle toho, zda jsou nalevo, nebo napravo od slova. Práce rovněž předpokládá vedle znalosti angličtiny a slovenštiny také znalost francouzštiny (překlad spojení v příkladech by byl čtenářsky vstřícný).

d) Autorka v práci používá smysluplné obraty, srozumitelnosti by ovšem na mnoha místech napomohlo konciznější vyjadřování a používání jednodušších větných konstrukcí (např. bod B na straně 25) společně s častějším ilustrováním pomocí příkladů.

e) Kapitoly 2.3.3-5 jsou nadepsány poněkud nelogicky (s ohledem na jejich obsah a vztahy mezi nimi).

f) Čtenář by přivítal seznam zkratk a také tabulková data poskytnutá v elektronické podobě, v nichž by měl možnost si data z diplomové práce seřadit nejen abecedně, ale i podle HHI.

g) Ve výsledku je práce po stránce formální zpracována pěkně, jen s malým množstvím překlepů.

I přes uvedené kritické poznámky práci doporučuji k obhajobě a navrhuji hodnocení: *velmi dobře*.