

Posudek na diplomovou práci kol. **Lenky Šíkové Hrejsemnové** **Monokolokabilní slova v češtině a dalších evropských jazycích**

Diplomová práce kol. Lenky Šíkové Hrejsemnové s titulem *Monokolokabilní slova v češtině a dalších evropských jazycích* si vzala za cíl prozkoumat dosud nedostatečně zmapovanou oblast anomální lexikální kolokability a obecně syntagmatiky jazyka. Autorka v práci navazuje na své dvě předchozí nepublikované klauzurní práce, které jí sice posloužily jako pilotní studie, ale zároveň musela ve světle nových zjištění, dat a postupů jejich výsledky od základů revidovat. Materiálovou základnou jí byl zejména reprezentativní korpus současné psané češtiny SYN2010, dále pak paralelní korpus InterCorp.

Monokolokabilní slova jsou většinou chápána jako takové jednotky, které ve svém kontextu vyžadují přítomnost konkrétního referenčního slova. V drtivé většině případů se ovšem tento fenomén neomezuje pouze na jevy s jedním obligatorním kolokátem, ale s nějakou velmi malou a většinou omezenou skupinou (kolokační paradigma). Termín – monokolokabilní – tak může být zavádějící a za úvahu tak stojí neologismus „minikolokabilní“ slovo, který práce sice nevyužívá, který ale vystihuje podstatu zkoumaného lépe.

Vzhledem k tomu, o jak málo probádané téma v lingvistice jde, není ani rozbor předcházejících studií na dané téma příliš rozsáhlý. Autorka však obsáhla téměř vše, co bylo k tématu za posledních 30 let napsáno (v našem kulturním okruhu). Jednotlivé přístupy kriticky hodnotí a upozorňuje na jejich nedostatky.

Excerpci monokolokabilních slov založila na teorii pravděpodobnosti a teorii informace. K identifikaci monokolokabilních jednotek použila jak entropii, tak zejména Herfindahl-Hirschmanův index (HHI) diversity, který není v lingvistických pracích příliš obvyklý. Právě v odklonu od mechanického pojetí monokolokability, jako vlastnosti jevu, který se pojí pouze s jedním nebo s maximálně 8 různými referenčními slovy spatřuji největší přínos celé práce. Autorka na příkladech ukazuje, že nezáleží na absolutním počtu kolokátů (na který kladly důraz dřívější přístupy), ale na jejich frekvenční distribuci, která musí být vztažena k celkové frekvenci hledaného slova (tj. na jejich diversitě). Hodnota normalizovaného indexu (HHI*) posloužila k vytvoření seznamu, který byl poté podroben ruční analýze. Ve výsledku tak autorka dospěla k seznamu monokolokabilních slov, který slouží jako výchozí materiál pro formulování hypotéz a klasifikaci.

Analýza byla vedena paralelně jednak po linii lexémů (lemmat) a jednak slovních tvarů (které v práci poněkud nestandardně nazývá lexikální formy). Porovnání výsledků vede k postulování jevu, který nazývá dílčí monokolokabilita, tj. monokolokabilita, která je vztažena pouze k jednomu členu (nebo několika málo členům) paradigmatu (např. *zásluha* vs. monokolokabilní *zásluze*). Toto zjištění je tak ve shodě s metodologickým směřováním v současné lexikografii, která čím dál častěji akcentuje názor, že význam je vlastností především slovního tvaru (nikoli celého lexému).

Práce obsahuje celou řadu různých klasifikací monokolokabilních slov, za všechny jmenujme třídění na základě počtu referenčních slov a kolokačních paradigmat. Toto kritérium rozděluje monokolokabilitu na absolutní, monokolokabilitu s několika (málo) referenčními slovy stejného druhu, polysémií monokolokabilitu a vícenásobnou monokolokabilitu. V klasifikaci se u každého druhu monokolokabilních jednotek uvádí jeho typová frekvence. Stejně zajímavý by byl i údaj o frekvenci tokenové, který by ovšem byl nutně zkrácen faktem, že průzkum probíhal pouze na jádrové slovní zásobě s frekvencí vyšší než 200 výskytů.

Analýza přispívá několika pozorováními i do obecnějšího kvantitativního studia kontextu. Výsledky k nimž kol. Šíková Hrejsemnová dospěla potvrzují, že je třeba rozlišovat různé strany kontextu a každou analyzovat zvlášť, a to i v jazyce s relativně volným slovosledem jako je čeština (analýzy

prováděné na obou kontextových stranách dohromady jsou nepřesvědčivé, viz s. 40n).

Autorka věnuje pozornost i využití svého výzkumu v lexikografické praxi, která se s problémem monokolokability musí vypořádat. V práci je obsažena jak kritika současného stavu v českých slovnících, tak i nástin možného řešení.

V poslední kapitole se práce věnuje srovnání se slovenštinou, angličtinou a francouzštinou. Výsledky ukazují, že klasifikace vypracovaná pro češtinu je vcelku použitelná i na jiné jazyky, ačkoli se ve shrnutí konstatuje, že data paralelního korpusu nejsou pro tento typ analýzy optimální.

Práce je po formální stránce na velmi dobré úrovni, všechny výsledky jsou přehledně uvedeny v přílohách. Z hlediska obsahu tato práce dalece přesahuje požadavky na diplomové práce kladené a to především v tom, že zpracovává téma obtížné, dosud nezpracované a to inovativně a metodologicky čistě. V dalším se proto omezím na několik drobných výtek a pár obecnějších poznámek, které přesahují rámec práce a můžou posloužit jako téma k diskuzi.

- A) Inovativnost použití HHI je poněkud zastíněna několika drobnými nedostatky: 1) minimální hodnota HHI je uvedena na s. 28 špatně (nulové hodnoty může dosahovat pouze normalizovaný HHI*). Celkově by pochopení tohoto nového statistického nástroje prospělo, kdyby v práci byl uveden vzorec pro obě varianty HHI (standardní i normalizovanou). 2) Na s. 35 autorka uvádí průměr HHI*, ale neposkytuje informaci o disperzi dat: není zde uvedena ani směrodatná odchylka ani třeba mezikvartilové rozpětí. Tyto veličiny by přitom jasně a přehledně pomohly určit míru příznakově vysokých a nízkých hodnot HHI*. Kladně je naopak potřeba hodnotit srovnání HHI a entropie formou korelace, která ukazuje na srovnatelnost obou způsobů měření diversity.
- B) V úvodu autorka shrnuje problém monokolokability jako porušení jedné ze základních vlastností jazyka, totiž jeho schopnosti kombinace jednotek (s. 9). V této souvislosti se samozřejmě nabízí i jiné, komplementární vysvětlení, které nechápe kolokabilitu jako problém jazyka, ale jako problém jeho deskripce, konkrétně pak vymezení základní jednotky a fascinaci mezerou, jako určujícím faktorem pro segmentaci psaného jazyka na slova. Pokud by lingvisté vnímali arbitrárnost dělení proudu řeči na grafická slova, možná by jednotkám, účastnícím se monokolokabilních spojení, vůbec status slov nepřisoudili (např. při segmentaci textu na jednotky významu by monokolokabilní spojení bylo chápáno jako jeden celek, což také v některých případech děje naznačuje i jejich forma – *mermo mocí* vs. *mermomocí*). Takovýto postoj můžeme najít v čermákovsky pojaté frazeologii již od jejich počátků a jejím výrazem je fakt, že frazémy (k nimž mají monokolokabilní jednotky svojí povahou blízko) jsou jen obtížně segmentovatelné a jejich části pak nejsou samostatně analyzovatelné.
- C) Práce chápe monokolokabilitu jako vlastnost jevů, která překračuje hranice frazeologie a idiomatiky (tím se liší od pojetí Čermákova a jeho žáků). Připomeňme, že Čermákovu pojetí IF právě na omezení kolokability staví. Závěry, ke kterým na základě průzkumu dat v ČNK autorka dospěla, je tak možné chápat i v tom smyslu, že dosavadní vymezení IF bylo příliš úzké a je třeba do této oblasti zahrnout i celou řadu dalších jevů, které právě v důsledku potvrzené monokolokability mezi frazémy a idiomy patří (ačkoli se třeba v jiných attributech s typickými frazémy neshodují).
- D) Zajímavá je rovněž úvaha o monokolokabilních spojeních, v nichž referenční „slovo“ je zastupováno interpunkcí. V této souvislosti není bez zajímavosti, že český pravopis přisuzuje některým pozicím obligatorní interpunkci (např. před některými spojkami). V takových případech samozřejmě čárka žádný význam nenese a dostáváme se tak k paradoxní situaci, že tam, kde čárka být musí, nemá valný smysl. Práce by tak mohla poskytnout zajímavé východisko pro zkoumání účelnosti české psací soustavy z poněkud jiné perspektivy.

E) Další otázkou, která překračuje možnosti předložené práce, je, jak by vypadala spojení vyššího řádu (více než dvě jednotky). Některá spojení slovesa s valenční předložkou naznačují, že by monokolokabilní spojení mohlo být delší (např. *zalapat + po + dech/vzduch*). Omezení na bigramy je sice v práci tohoto druhu oprávněné a pochopitelné, výsledky ovšem naznačují, že celý problém monokolokability není omezen na dvojice jednotek a zasahuje tak do obecnější analýzy diversity kontextu.

Lenka Šíková Hrejsemnová napsala inovativní, přemýšlivou a velmi potřebnou práci. Dokázala se vypořádat jak s nedostatkem výzkumu v dané oblasti, na kterém by mohla stavět, tak s předchozími nedostačujícími koncepcemi, k nimž si postavila se zdravým a kritickým odstupem. Navrhla důkladnou a precizní klasifikaci monokolokabilních slov na metodologicky čistém základě, který doplnila o velmi cenné srovnání s jinými jazyky. V přepracované a zkrácené podobě by si práce zasloužila publikaci ve formě článku. Doporučuji tedy práci k obhajobě s hodnocením výborně.

Václav Cvrček, Ph.D.

V Praze, 2. 9. 2013