

Posudek diplomové práce

Univerzita Palackého v Olomouci, Filozofická fakulta

Autor práce: *Anna Boková*

Název práce: **Building and Exploring a Corpus of Academic Writing by Czech Students of English**

Vedoucí práce: *Mgr. Michaela Martinková, PhD.*

Autor posudku: *Mgr. Michaela Martinková, PhD.*

Bodování: 0 = nedostačující, 1 = slabé, 2 = přijatelné, 3 = průměrné, 4 = dobré, 5 = výborné

Body

1. Originalita a přínos práce, aktuálnost a důležitost tématu.

Podnět vytvořit srovnatelný korpus akademické angličtiny vzešel od vedoucí, množství času a práce, kterou autorka věnovala jeho vytvoření, je obrovské. Ve vytvořeném korpusu spočívá hlavní přínos práce – korpus může být sdílen skrze Sketch engine a lingvisté s ním mohou dále pracovat.

2. Kvalita přehledu poznatků o tématu: použití literatury, kvalita zpracování, kritické zhodnocení literatury.

*Autorka vychází ze dvou základních typů zdrojů: jednak literatury týkající se žákovských korpusů (zejména stati Grangerové), pro výzkum samotný pak z Biberova 1999 přístupu k žánrům, tedy skrze lexical bundles. Lit. review je zpracováno solidně a množství literatury je dostatečné. Za problematické považuji dva aspekty. Prvním je to, co autorka píše hned v úvodu práce: „The corpus based research into academic writing (e.g. Biber et al. 1999, Cortes 2004, Hyland 2008) has shown that this register is characterized by the usage of recurrent conventionalized multi-word expressions (“lexical bundles”, “clusters” or “n-grams”) that signal “competent use of language in a particular context” (Dontcheva-Navratilova 2012, 38)“. Tato pasáž naznačuje, že bundles se v jiných typech registrů nevyskytují, a to autorka určitě nechce říct. Naopak zkoumání lexical bundles je jedním ze způsobů jak k rozdílu mezi typy textu přistoupit, to autorka naznačuje citací z Hylanda (na str. 6 píše o žánrech, na str. 22 pak o registrech – může se autorka k rozdílu vajídřit?) Biber sám ve zmíněné gramatice porovnává conversation a academic language – může se autorka zamyslet nad tím proč právě tyto dva? Druhým problémem je, když autorka ztotožňuje lexical bundles s „prefabricated chunks” that are stored in our lexicon“ (str. 18), konkrétně říká: „she [Wray] uses the term **formulaic sequence** which is: “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray 2002, 3-9). Lexical bundles studied by Biber et al. (1999) fit this definition“ (tamtéž). S tím nesouhlasím. Může autorka osvětlit, jak chápe mental lexicon?*

3-4

3. Formulace záměrů práce, výzkumných otázek, hypotéz.

Záměrem bylo jednak korpus vytvořit, což autorka splnila, a jednak provést drobnou sondu. Otázky i hypotézy jsou stanoveny na str. 6: „what structural types of lexical bundles are used by Czech learners of English, secondly, whether they use them more or less often than professionals and thirdly, whether they use the same lexical bundles as professional linguists or not. Building on Hyland’s research, it is assumed that professional linguists may use different repertoire of lexical bundles than L2 students of English (L1 Czech), since there exist genre variation between students” writing and professional writing.“ Otázky i hypotéza samozřejmě mohly být konkrétnější, výzkumu na datech českých mluvčích je však poskrovnu, proto toto jako pilotní studie postačí.

4

4. Metody práce. Jejich popis a adekvátnost pro zodpovězení výzkumných otázek.

Kompilace korpusu je v pořádku. Namítnout by se dalo, že korpusy nejsou v jednom ohledu srovnatelné. Mohla by se autorka zamyslet nad tím v jakém?

Lexical bundles autorka vyhledávala jako n-gramy, atributem je slovní tvar. To muselo nutně znamenat zařazení interpunkce, jak se autorka s tímto problémem vypořádala?

4-5

5. Zpracování výsledků a jejich interpretace.

Výsledky jsou přehledně zpracovány do tabulek. Více viz bod 6.

6. Diskuze práce (o tom, jak výsledky zodpovídají výzkumné otázky).

*Autorka poukázala na rozdíly mezi 4-gramy v korpusu publikovaných lingvistických statí a v korpusu žákovském, škoda je, že rozdíly v užití jednotlivých 4-gramů nejsou seřazeny podle signifikance. Zjištění o nadužití spojovacího výrazu *On the other hand* na začátku věty potvrzuje výsledky jiných výzkumů, které poukazují na nadužívání konektorů v úvodu věty. Některá autorčina vyjádření jsou sporná, např. když vyloučila více 4-gramů, které se vyskytují v méně než v pěti různých dokumentech, ze studentského než profesionálního korpusu: „One of the plausible explanations would be simply the fact that students repeat themselves more while writing about a certain topic“ na str. 39. To je vysvětlení kruhem, které neosvětlí ani odkaz k další literatuře: „This phenomenon is supported by Hyland who suggests that students seem to rely more on prefabricated lexical bundles when they propose their arguments“. Některá tvrzení jsou příliš silná: „This finding may suggest that students tend to use these bundles [it is necessary] more than professional linguists.“ Nejde právě o bundle, který Biber zmiňuje v souvislosti s akademickou angličtinou? Autorčino vysvětlení na str. 51 navíc neobstojí: „Students may also want to stress the information that follows their proposition.“ Za zajímavé zjištění považují naopak fakt, že konstrukce s pasivem se zdá být over-represented ve studentském korpusu. Dokáže autorka toto nějak vysvětlit?*

3-4

7. Formální zpracování. Jazyková a stylistická úroveň, rozsah práce, grafická úprava, dodržení publikační normy.

Citační norma je dodržována, práce je psána slušnou angličtinou, která je lepší v teoretické než v praktické části práce, kde se autorka místy opakuje (např. popis studentského korpusu na str. 30). Překlepy jsou ale řídké: Killgarriff str. 29, The fact that the; data na str. 16. Občas se objeví prohřešky proti gramatice či neidiomatická vyjádření: there exist genre variation str. 6, systematicky the chapter 4 na str. 7 i jinde, the choice of control native corpus na str. 15, as Granger states thatna str. 16, there is no strict definition of formulaic language is str. 18, Lexical bundles in conversation are not complete structural units, too (str. 20), with greater frequency (str. 23), The aim is to find out what bundles are there str. 24, since misto while na str. 25. % of lexical v tabulce 8 – nechybí bundles?

4

Pro vedoucího práce:

Hodnocení spolupráce diplomanta s vedoucím práce:

Autorka pracovala poměrně samostatně, konzultace probíhaly standardně.

Prohlášení vedoucího kvalifikační práce o kontrole plagiátorství:

Na základě kontroly vysokoškolských kvalifikačních prací systémem Theses.cz ani jiným způsobem nebyla u výše uvedené práce zjištěna taková shoda s jinými pracemi, dokumenty nebo texty, která by zakládala odůvodněné podezření z porušování autorských práv.

Celkové hodnocení práce.

Autorka prokázala schopnost samostatně pracovat se Sketch Engine, vytvořit subkorpusy, sebrat z nich data a vyvodit z nich závěry. Přes uvedené výtky práce splňuje požadavky kladené na diplomovou práci, a proto ji doporučuji k obhajobě.

Poznámka: Celkové hodnocení ani návrh klasifikace **NELZE vypočítat jako prostý průměr** bodů za jednotlivé hodnotící kategorie.

4

Doplňující otázky k obhajobě:

Otázky viz v bodech výše. Dále:

1. Autorka píše že „Another contribution of learner corpora is a methodology called data-driven learning. Students are encouraged to use corpora of native speech and via its examples they may learn more about how the language works. In other words, they are encouraged to explore and investigate a given corpus on their own which may help them with the process of learning the language (2002, 26-27).“ Je toto skutečně contribution of learner corpora? Co přesně se na straně 26-27 píše?

2. Objevila autorka u akademické angličtiny studentů nějaké lexical bundles, které jsou příznačný pro jiný registr?

*3. Na str. 34 autorka píše, že „the **Students' Theses Corpus** is considered as a reference corpus“, což nepovažuji za správné. Co je reference corpus?*

Závěrečné hodnocení

Práci k obhajobě doporučuji ~~nedoporučuji~~

Návrh klasifikace: **B-C**

Datum: 30.8.2015

Podpis: