Dr. Eva Maria Luef
E-mail: evamaria.luef@ff.cuni.cz

Charles University
Univerzita Karlova

Faculty of Arts
Department of English Language and ELT Methodology
nám. Jana Palacha 2
116 38 Prague 1
Czech Republic

June 7, 2020

MA Thesis Report
Klára Divišová (Bc)

This thesis compares the use of phrasal verbs (referred to as "multi-word verbs" or "MWVs") in native speakers of English and Czech ESL learners. Three investigative leads were pursued: the question of MWV avoidance on the part of Czech learners of English, whether or not specific types of MWVs are preferred by the speakers, and the existence of "phrasal teddy bears" or overused and overgeneralized phrases. Results were varied but seemed to confirm the second hypothesis of the favored used of prepositional verbs in the learners (as opposed to English native speakers). The topic of MWV use in ESL learners whose native languages do not use such verbal constructions is highly relevant and sufficiently original for an MA thesis.

The introduction/ literature review is detailed, complete and the student shows a good understanding of the topic. Sufficient background information is provided so that the research problem can be contextualized. It is discussed how other researchers have addressed the problem of MWVs in ESL learners, and it becomes clear to readers how hypotheses are derived from previous literature.

Even though the abstract and introduction state that the study is of quantitative as well as qualitative nature, the quantitative part is clearly under-developed. At times, statistical probabilities are stated (e.g. p. 38: "*When assuming the significance value, for p<0.05, the final number of type and tokens of MWV in the LOCNECB corpus is significantly higher than the number of MWV tokens in LINDSEI_CZB. Therefore,*

the difference is statistically significant between the corpora", or p. 40 *"PhV are represented by 191 tokens in the corpus, whereas there are 398 tokens of PrV, indicating the dominance of PrV, which is statistically significant at p<0.0001"*, also see p. 54, 54, 55, 56…), however, no information is available on how these p values were calculated. Generally, the statistical test needs to be stated clearly (ANOVA, linear regression, …) and the variables entered into the analysis need to be explained in detail. I would suggest that the student reconsiders the analytical nature of the study, and possibly revises the thesis to be a purely qualitative study.

The main weakness of the thesis concerns the methods and results sections. I will briefly list some of my concerns here.

First, normalized data were calculated but many analytical interpretations are based on raw data (e.g., p. 38, 54, etc.). If these references to the raw data were removed (which may include some tables as well), the analyses would be more profound and reliable. I realize that the two corpora are of similar size and the normalized and raw data do not differ by much in many cases, however sometimes differences between raw and normalized data emerge that impact the interpretation of the data (e.g. p. 54). Normalized data should be the basis of any analysis and interpretation.

Next, data skew should be avoided by carefully choosing what data to include in each sample of the native and non-native speakers. As a general rule of thumb, each multi-word verb that is included should have a somewhat acceptable distribution among the speakers of the compared groups, meaning that outliers should be removed from the data set before analysis. Said differently, only MWVs that were used by a sufficient number of subjects (which could be once per subject, for instance) should be entered into the analysis. Otherwise one subject's repeated use of a multi-word verb could skew the data in a certain direction (which seems to have been the case in some instances). Alternatively, if one wanted to compare the (uneven) distributions of MWVs within the groups, a data-distribution coefficient could be calculated (e.g. Gini coefficient or Shannon's entropy). Results would then show whether a small number of subjects used a large number of the MWVs or whether the MWVs are distributed equally among all speakers. Such an analysis would be informative with regards to how idiosyncratic the use of MWVs is among native and non-native speakers.

Investigating so-called "phrasal teddy bears" is a very interesting possibility and certainly relevant to the current study. Literature is not quite clear on the definition of these phrases (as pointed out in the introduction) but I think the concept could be quanitified given the data that was collected for this thesis. If one were to define the concept through (for instance) frequency of use and overgeneralization (correct and incorrect uses), and calculate a ratio between them, a quantitative measure could be achieved that would make for an interesting aspect of the study.

In addition, I was wondering whether lexical frequency of some parts of the MWVs plays a role in their use in non-native speakers. For instance, the verbs "go" and "come" are certainly more frequent than "draw" or "ring", leading me to the assumption that MWVs involving "go" and "come" may be more prevalent. Maybe a difference in lexical frequency rate (calculated from a different data set) is relatable to the frequency of MWVs involving frequent or infrequent verb parts. There could be differences between the two speaker groups.

It is claimed that Czech ESL learners tend to avoid MWVs but that conclusion rests purely on frequency counts (p. 63). Avoidance of forms could be demonstrated by showing that the learners prefer to use other, non-MWVs, when formulating sentences. For example, is the use of "ponder/ consider" instead of "think about" more prevalent in the learners?

The conclusion is rather brief and would benefit from slightly more reflective discussion of the results and potential for further work. Aside from a reiteration of the results, some context should be provided for them. Results of the present study should be compared to previous studies (for instance, the ones cited in the introduction) and some generalization of MWV use in language learners should be formulated, embedding the present study in the existing body of literature on the topic.

Minor comments:

Reconsider the use of the expression "the natives". I suggest using "native speakers" or "first language speakers/ users". Similarly, instead of "mother tongue" I suggest using "first language".

When two parentheses meet (which is quite often the case in the introduction), you could merge them and separate the citation at the end by a semi-colon. For instance, in the introduction: *(i.e. in documents of the European Union in English; Trebits, 2009: 470).*

Graphs displaying some of the results could be included. Especially when data is complex or particular parts of it are highlighted (such as in the case of Table 20).

References:
- Sometimes the journal name is followed by "vol 1, no 5" (e.g. first citation), sometimes "vol" and "issue" (Nassaji), at other times only volume number is given without "vol." (e.g. second citation or Ellis 2012). Gilquin 2015 and Houshyar (and others) show a third type of format: 11(1). Citation style should be consistent throughout the references list.

- Periods may be missing at the end of citations (Biber et al. 1999, Quirk). There should be none following URLs
- Titles and subtitles: sometimes separated by semi-colon, sometimes hyphen
- Capitalization of words in titles should be consistent (either all words or only the first – right now it varies, see e.g. Trebits)
- "Available at" and "last accessed" are not necessary for journal articles that appear in online form when you have volume and issue number (only if you cite specific online documents or newspapers etc.). Journal articles may include the DOI link.

Questions for the defence:

1. What can you say about lexical diversity and the use of MWVs in (Czech) ESL learners? Acquisition-wise, do you think MWVs precede the (often more formal) simple verbs?
2. What insights did you gain regarding semantic diversity (i.e. semantic categories) of phrasal verbs in your sample of ESL learners?
3. Please elaborate on your decision to use frequency counts as an indicator of MWV avoidance. What are the advantages/ shortcomings of this measure?

In summary, this is a well-planned and well-written thesis which investigates interesting hypotheses. As I have indicated above, I believe that there are some areas in which the thesis could be strengthened. I suggest the thesis to be accepted with a grade of *výborně* or *velmi dobře*, depending on the student's performance at the defence.

Eva Maria Luef, PhD