



FACULTY OF ARTS  
Charles University

Department of English and ELT Methodology

## A Review of a Final Thesis

submitted to the Department of English and ELT Methodology,  
Faculty of Arts, Charles University

**Name and titles of the reviewer:** Dr. Eva Maria Luef

**Reviewed as:**  a supervisor  an opponent

**Author of the thesis:** Martin Sotona

**Title of the thesis:** Examining lexical complexity in the written production of L2 proficient learners of English

**Year of submission:** 2022

**Submitted as:**  a bachelor's thesis  a master's thesis

**Level of expertise:**

excellent  very good  average  below average  inadequate

**Factual errors:**

almost none  appropriate to the scope of the thesis  frequent less serious  serious

**Chosen methodology:**

original and appropriate  appropriate  barely adequate  inadequate

**Results:**

original  original and derivative  non-trivial compilation  cited from sources  copied

**Scope of the thesis:**

too large  appropriate to the topic  adequate  inadequate

**Bibliography (number and selection of titles):**

above average (scope or rigor)  average  below average  inadequate

**Typographical and formal level:**

excellent  very good  average  below average  inadequate

**Language:**

excellent  very good  average  below average  inadequate

**Typos:**

almost none  appropriate to the scope of the thesis  numerous



**Brief description of the thesis (by the supervisor, ca. 100-200 words):**

Mr Sotona's thesis investigates lexical complexity and sophistication in newspaper commentaries written by L2 students. Comparisons with similar texts written by L1 users of English (newspaper journalists) yielded surprising results. L2 writers showed similar lexical complexity as the L1 writers. However, lexical diversity was different between the two experimental groups, with L1 writers displaying a wider range of lexical knowledge in their texts. The structure and organization of the thesis is good. Problems with the comparability of the study groups and some methodological issues with the analyses prevent the thesis from reaching more conclusive findings

**Review, comments and notes (ca. 100-200 words)**

***Strong points of the thesis:***

The topic itself is very interesting and highly relevant for second language research. Lexical complexity is difficult to research and many of the problems that arose for Mr Sotona have plagued this type of research for a long time. I applaud his efforts to use automated methods for measuring lexical diversity and sophistication, rather than rely on personal judgement of L2 learner texts. This makes his work transparent and reproducible for other researchers.

The comparison of lexical diversity in terms of percentages of A1, A2 etc. words in the texts is a very interesting idea. I find it quite informative and a good measure of lexical sophistication. This makes for an intriguing comparison of L1 and L2 in Figure 15.

***Weak points of the thesis:***

The lack of comparability of professional journalistic articles with L2 learner comments based on those articles is certainly a weak point. It would have been advantageous to keep the compared texts similar in length and style, gather some information on the writers of the L1 texts, and introduce a measure of lexical imitation on the part of the student writers. More information should have been provided on the participants (both L1 and L2), their language learning history, and the experimental set-up.

***Questions to answer during the Defence and suggested points of discussion:***

1. When inspecting Figure 15, what could be an explanation for the differences in vocabulary decline from A1 to C2 in the two study groups? The L1 group shows steady decrease but the L2 group shows upticks in B and B2.
2. How could you measure lexical accommodation (or imitation) to the original L1 texts in your L2 group?

***Other comments:***

In the empirical section, a better description of the actual experimental task would have been appreciated. I was a bit confused about what exactly the students had to do. Did you tell the students to work hard on their commentaries, revise them many times, and turn in a publication-ready piece? That would have been a good emulation of the process that the L1 writers went through. Did you grade the



students on their commentaries and thus make them put more effort into it? This would surely have influenced their lexical sophistication.

A problem with the data collection concerns possible imitation effects of lexical complexity, as you acknowledge yourself in the discussion. One would need to make sure that the students are not lifting too many phrases/ words/ sentences from the original work. This could have been controlled for by including a measure of “lexical accommodation” toward the original article.

I am not sure your written samples/ groups are really comparable. On page 27 you write:

*“Both groups of written samples (L1 and L2 texts) are homogenous by their extent (the mean of L1 texts – 654 tokens, the mean of L2 texts – 781 tokens), genre (economics), and language (English), but heterogeneous by their authorship (L2 English students with L1 Czech vs. L1 English adults) and by their aim (internal assessment written for examiners vs. news-media article written for readers).”*

But a few points are neglected here: 1. There is too much variation/ st. deviation in the L1 texts, so I would argue they are not “homogenous by their extent”. You acknowledge this point yourself a few times when describing the results. Would it have been possible to find articles of similar length in the newspapers? One could have pre-selected articles and given to the students to comment on. 2. About the authorship: the Economist is known for publishing anonymous texts which may be written by a small number of people. So, it could theoretically be possible that your Economist texts were written by only a couple of people. It’s impossible to control for authorship (and, moreover, first language of authors) without bylines. Cooperations of different authors, which are common in newspapers, can also not be excluded.

The detailed descriptions of the Text Inspector functions over-burden the methods section a little. It is not necessary to list/ describe functions that were not used in the thesis (such as BNC, COCA or Scorecard...).

On page 33 you write:

*“By using two independent indices of lexical diversity in the thesis, instead of only one, the reliability of results will be increased”*

It would be interesting to see if the two are correlated (within and across participants).

Figure 5: one could have normalized the number of sentences (using the number of words per text). This would have made for a more informative comparison (and possibly different statistical results).

In the case of number of tokens (p. 37), I suspect that a parametric comparison may have yielded more reliable results, given the data has the right distribution (this applies to all other variables as well). The difference between L1 and L2 seems rather significant, judging from Figure 7.

The TTR analysis is quite interesting, especially the comparison of the repetitions. You could have made a composite variable of numeral count and TTR to account for the effect.



Page 41: a brief reiteration of the three different analyses would have been beneficial for readers at this point.

Table 8 highlights are misplaced: the lowest value of MTLN in the L2 group is not the 10C participant, as indicated, but the 11C. There seems to be less of a correlation between the two variables in the L2 group. That is an interesting fact. It would have been good to see the correlation coefficients here.

On page 51 you write:

*“3/ The overall trend of word types, using EVP frequency lists, is the decrease from A1 to C2 type (with a small exception of A2 type, according to Zipf’s law.”*

I would say that the steady decrease is only clear in the L1 group. The L2 learners show higher B1 and B2 percentages. Also, the distribution does not look Zipfian at first glance, but one would have to test the distribution to be able to define it.

I appreciate the discussion of the limitations. As shown by previous research, measuring lexical sophistication/ complexity is difficult and many factors have to be taken into account. The differences in length and style between your L1 and L2 texts probably had an effect on the results. Additionally, the Text Inspector needs some afterwork – your example of the word “mark” and its classification on page 55 perfectly illustrates that.

Minor:

- Capitalization inconsistencies in Table of content and abbreviation list
- Section 2.4.2.1.: It would have been more informative to display the transformed equation where D is calculated, as the focus of the following passages is on the D value
- The section that is entitled “3.2. Methodology” is followed by “3.2.1. Method”
- Why bold print of statistical results on page 42 and after?
- Figure 13: the % symbol should be near the y-axis, not with the barplot labels

**Proposed grade:**

excellent  very good  good  fail

Place, date and signature of the reviewer:

Prague, 19.01.2022