

STUDENTS' PERCEPTIONS OF USING CORPORA AS REFERENCE TOOLS IN ESSAY ERROR CORRECTION AND EXPLORING WORDS WITH AFFIXES: A PILOT STUDY

Emin Idrizi

International Balkan University, Skopje, Macedonia, eminidrizi@gmail.com

Igballe Miftari

International Balkan University, Skopje, Macedonia, igballe_bale@yahoo.com

Abstract: Computer corpora are seen as very useful reference tools that can be used as alternatives and complementary resources to modern dictionaries. This is because advanced corpora can now be used to carry out enquiries and find linguistic evidence that are not easily found in dictionaries (Flowerdew, 2009). Most importantly, they are accessible to anyone, including language teachers and learners, who would like to investigate various language aspects, register, punctuation, to name just a few. This paper reports on a pilot study conducted with a group of university students engaged in essay error correction and exploring words with affixes using corpora. The study primarily explores participants' experiences and attitudes towards engaging in more autonomous language investigations with little preliminary corpus training. The results generally indicate that participants have positive attitudes toward making use of corpora for error correction and exploring new vocabulary. The paper also discusses some possible obstacles and considerations practitioners should take into account before incorporating corpus activities in the language education.

Keywords: corpora, corpus consultation; concordancing lines; error correction; words with affixes

INTRODUCTION

Modern day corpora are no longer tools accessible only to researchers, but they have also become available to teachers and students in language education settings. This was clearly pointed out by the pioneer of Data-driven learning (DDL) technique, Tim Jones, in his well-known saying: "Every student a Sherlock Holmes" (1997, p. 101), referring to a language learner who is seen as someone who can directly engage in language investigation by the help of corpus consultation. This direct involvement of language learners in language enquiries is typically referred to as *direct* use of a corpus as opposed to *indirect* (Leech, 1997) in which learners get to enjoy the benefits of corpus but not by engaging in direct use of the tool, such as in the case of using corpus-based dictionaries. This paper is mostly concerned with direct use of concordancing lines.

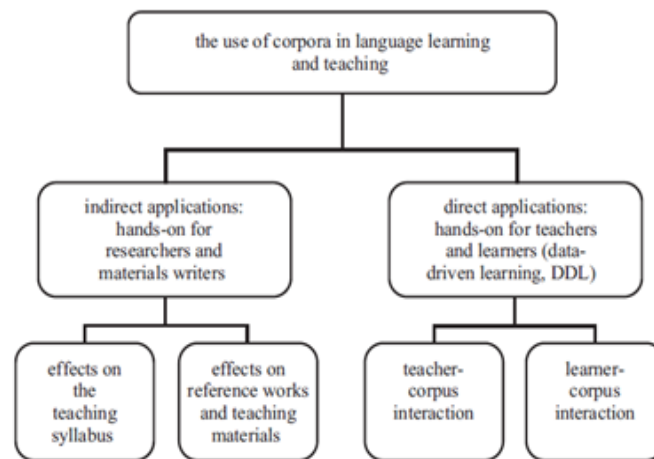


Figure 1 Direct and indirect use of corpora (Römer, 2010, p. 19)

Corpora are seen as very useful and practical tools for learners of language. For Gilquin and Granger (2010), having learners use concordancing lines makes their learning more motivating and exciting as it includes "an element of discovery" (p. 359). The discovery nature of corpus activities, as a result, may lead to more autonomous learning (Warren, 2016). Corpus consultation is typically seen as a learning technique in which students are considered to have a leading role in investigating, assessing, and inferring linguistic rules based on the data they collect or see in the corpus (Ebeling, 2009). Corpora are also seen as tools in which numerous linguistic investigations can be carried out. Leech (1997), for instance, argues that "the study of corpora can illuminate virtually all areas of linguistic study" (p.9). That is to say, a learner can utilize a corpus to

research words, patterns, various sentence structures, collocation, punctuation, and other necessary linguistic issues related to the language he is studying.

Since corpora can be very flexible in terms of the range of linguistic features it could examine, it is reasonable to conclude that there are also numerous hands-on activities that language learners can engage in with their help. One of the most frequent corpus activities that is mostly noted in the literature is using concordancing lines or corpus data in writing (Gaskell and Cobb, 2004; Gilmore, 2009). For instance, learners can be tasked to use a corpus as a supporting tool during the writing of their second or third draft of their essay. Or, they can be given a task in which they would be provided with an essay which contains deliberate errors. Learners, thus, can be tasked to do error correction of the essay by drawing insights directly from the corpus. Another hands-on activity that could be carried out is investigating words in the language. For instance, learners can be given a task to draw various word-formation insights by consulting a corpus (Moon, 2010). Other activities may include: having learners explore register in the corpus; using corpus in speaking activities in terms of written / spoken language differences (Reppen, 2010), and so on.

Although corpus-based activities seem to be quite beneficial and promising for language education, a valid question to ask is: what do learners think of corpora as reference tools? According to Chambers (2005), literature and studies are not sufficient to prove that language learners are willing to use corpora as reference tools in addition to the standard dictionaries or textbooks. The answer to this question is important for two reasons: (1) one is that there is a growing interest on how the learner learns and perceives language teaching and learning, rather than focusing only on how the instructors should teach, as well as what techniques and materials should be used in the language classroom (Yule, 2014); (2) only by confirming learners' positive attitudes towards corpora we can ensure that corpus-based activities will benefit our language classrooms.

The main aim of this pilot study, therefore, is to explore a group of university students' experiences and attitudes towards corpora as reference tools after engaging them in two corpus activities: using concordancing lines for error correction in L2 writing and exploring words with certain affixes in the corpus.

METHODOLOGY

Participants and materials

The pilot study included eleven university students, nine of which were 2nd year students and two 4th year students. The students were all studying in the Department of English Language Teaching at International Balkan University in Skopje, Macedonia. The selection of students was random and no criteria, such as: language proficiency, gender, age etc., was used during this process. In addition, the participants in this pilot study had no previous knowledge on corpora and ways to use them.

Two data collection instruments were used during this study. One was a questionnaire which consisted of three rating questions (ranging on a scale from *very easy* to *very hard*) and seven Likert scale-type statements (ranging on a scale from *strongly agree* to *strongly disagree*). It was designed to explore the perceptions and attitudes of the participants towards corpora as reference tools and it was intended to contribute to the quantitative part of the study. The second data collection was a self-reflective essay which was planned to complement the questionnaire as well as to add to the qualitative aspect of the study. The essays included some guiding questions which were intended to orient the participant to expressing their views on topics valid to the study.

Procedures

This pilot study was carried out in two sessions on two separate days. The sessions lasted about an hour each and included all the participants in one single group. Both sessions were held within the IBU premises, in the university Computer Lab. The participants had their individual computer and internet connection available to carry out the tasks planned for the sessions.

In the first session, the participants, who had no previous knowledge on corpora, were firstly introduced to corpora, their role, and underwent brief corpus training in order for them to get familiar with some of the basic corpus features and processes. In this particular stage, they were taught how to find words, sort the concordancing lines, find collocation, notice grammatical patterns etc. The online corpus interface named *Skylight*, administered by Dickinson and Francis (2009), and the *British National Corpus* were used during this session. After the short drill, the participants were given a task which involved a written essay with deliberate errors. Thus, their task was to improve the essay by correcting the mistakes, but by finding linguistic evidence with the help of *Skylight*. The essay mainly included: mistakes in terms of syntax, punctuation, and collocation. In the end of the session, the participants shared their findings and corrections they had done with the whole class and the instructor.

The second session had the participants involved in another corpus based enquiry. This time, however, another online corpus interface named *IntelliText* (Sharoff, 2014), run by University of Leeds, was used in their

new queries. Initially, the participants went through some preparation for the upcoming task which included exploring words with affixes in the corpus. Namely, some preliminary exercises with *IntelliText* and ways to use the software features to find words with affixes were done with the class. Afterwards, a hands-on activity followed, in which participants were tasked to explore and find words with certain affixes in the corpus. In addition, they were also asked to look at concordancing lines for some of the words found in order to extract some examples of their choice. Finally, participants were given the chance to share their corpus findings with the instructor and the class, which included words with the affixes included in the study and some examples for those words extracted from the corpus.

Upon finishing the first session, all the participants were asked to complete a questionnaire in which they would provide their perceptions and attitudes towards the corpora and the activities they did in the two sessions. After the questionnaire was completed, the participants were asked to write a self-reflective essay, in which they would go into a more detailed description of their experiences with corpora. The data collected from the questionnaire was statistically processed with SPSS whereas the self-reflective essays were analyzed qualitatively by drawing important conclusions on the students' attitudes and perceptions of corpus use. These are discussed in more detail in the following section.

RESULTS

The questionnaire

The questionnaire revealed mixed results in terms of how participants viewed *Skylight* for essay error correction. When asked to rate the difficulty of using this particular corpus interface (in error correction task), the majority of them (72%) found it easy while the rest of the participants found it either very easy (18%) or medium difficult (9%).

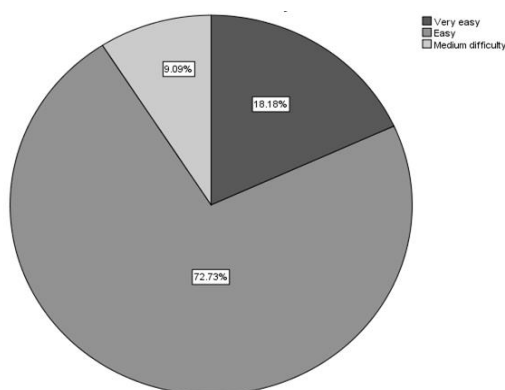


Figure 2 Participants' rating of the difficulty of using corpus for error correction of essays

The participants' responses, however, differed when it comes to using *IntelliText* interface to explore words with affixes. When asked to rate the difficulty of using this particular interface to find new vocabulary, the vast majority of respondents (81%) rated it as medium difficult, while for the rest of them it was found easy. It is evident from these figures that participants found working with *IntelliText* more difficult than working with *Skylight*.

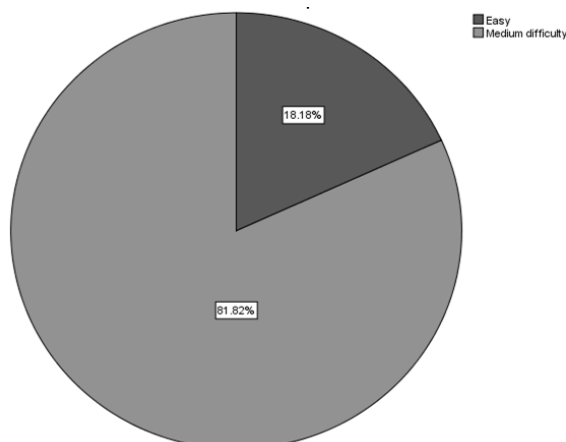


Figure 3 Participants' rating of the difficulty of using corpus to explore words with affixes in the corpus

Although students received a brief training in both sessions in both corpora, participants still felt a need for more training in corpus use. As the chart in *Figure 4* below illustrates, almost half of the participants felt that more corpus training was necessary for them in order to handle the corpora better. Some 45% of the participants, on the other hand, felt uncertain on the issue whereas the rest 9% saw no need for more training.

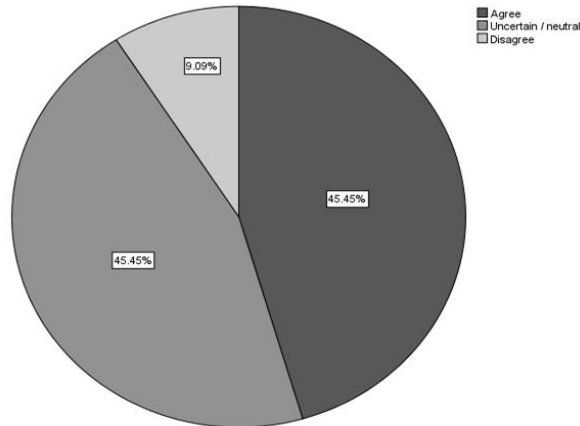


Figure 4 Participants' views on the necessity for more corpus training

The questionnaire responses indicate that students had positive perceptions of *Skylight* and its use in essay error correction. When asked to share their views on the usefulness of the corpus, over half of the respondents agreed with the fact that the tool was valuable, while the other ones strongly agreed with this idea.

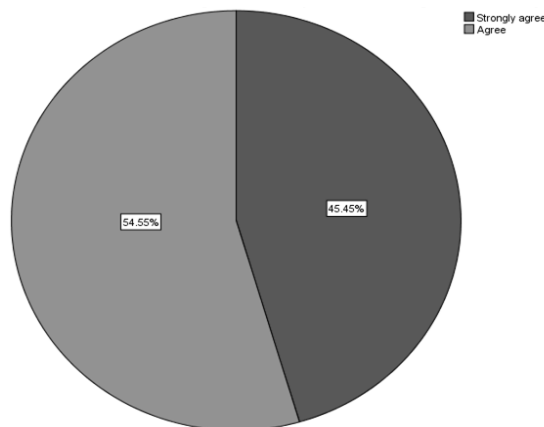


Figure 5 Participants' perceptions of the usefulness of corpus in essay error correction

Similarly, responses also showed that students had positive perceptions of *IntelliText* and its use in exploring words with affixes. When asked to share their views on the usefulness of the corpus, about 80% of the respondents strongly agreed with the idea that the tool was valuable, while the rest agreed with this idea.

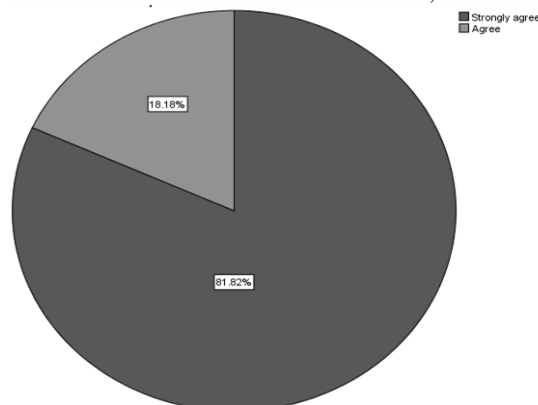


Figure 6 Participants' perceptions of the usefulness of corpus in exploring new words with affixes

The self-reflective essays

Remarks on these self-reflective essays indicate that participants in the study unanimously expressed their positive perceptions on the corpora used during the sessions and the actual activities in the sessions. Some participants clearly showed that they were astonished to find out that a tool with such enormous capabilities exists. For instance, one student wrote ‘this tool is a revolutionary invention’; while another expressed her surprise to learn ‘that such a useful tool for English language exists’. The participants also shared their opinions on whether they saw corpora as tools that need to be incorporated in the language teaching and learning. In this sense, their remarks indicate that they were strongly supportive of the idea that corpus should be part of the language classroom.

DISCUSSION AND CONCLUSION

The collected data from the questionnaire and the self-reflective essays used in this small-scale study reveal useful insights on how university students perceive corpora as reference tools for language investigation. Below we discuss the findings in more detail.

The questionnaire figures discussed in the previous section clearly show that participants found working with *Skylight* relatively easy in contrast to *IntelliText* which, based on the responses, they found more difficult. In fact, this was also observed during the actual sessions in which participants were noticed to cope better with the former compared to the later. A rational explanation behind this distinct perception of these two corpora may be on the fact that *Skylight* was designed with simpler layout and simplified corpus queries for the purpose of serving teachers and students of English (Dickinson and Francis, 2009). The data here thus suggests that this actually may be appropriate. That is, user-friendly corpora may work better for language students.

The discussed statistics are also consistent with the participants’ responses on their need for more corpus training. As figures in *Figure 4* clearly show, only a small number of participants felt that there was no need for further training. Therefore, it is valid to point out that sufficient corpus training is a must for one who decides to incorporate corpus-based activities in the language classroom (Gaskell and Cobb, 2004). Corpus training, on the other hand, entails not only the ability of students to cope with corpus features, but also to interpret the data found (Gavioli, 1997). Another valid point to make in addition to the ones above is the fact that not all the students are capable of dealing with corpus tasks autonomously. During the sessions, for instance, it was noticed that a few participants had more difficulties coping with the corpus tasks than the others, although this was not very apparent in the figures provided earlier. This leads us to the idea that purely inductive and completely autonomous corpus enquiries may not be suitable for all learners (Leńko-Szymańska & Boulton, 2015).

Despite the relative corpus difficulties experienced during the sessions, the participants unanimously showed positive attitudes towards using corpora for the purpose of investigating language. This can be clearly seen from the statistics in *Figure 5* and *Figure 6* given earlier. The same was also noticed in participants’ remarks given in their essays in which a great stress was put on the usefulness of using corpora. In agreement with their positive perceptions of corpora as reference tools, the participants also favored the idea of incorporating corpus use in the language classroom, and this was particularly stressed in these self-reflective essays.

In conclusion, it is evident from the data presented in this paper that corpora are perceived positively by language learners. Corpora seem to attract as well as motivate learners while doing language investigations in a more autonomous way. However, two considerations should be taken into account when one decides to incorporate corpus in the class: (1) students should undergo adequate training in corpus use; and (2) a more balanced inductive/deductive approach should be used to ensure all students cope with the corpus-based activities in the language classroom.

REFERENCES

- [1] Chambers, A. (2005). Integrating corpus consultation in language studies. In *Language learning and technology*, 9(2): 111-125.
- [2] Dickinson, A., & Francis, G. (2009). A classroom corpus resource for language learners. In *Proceedings of the Corpus Linguistics Conference 2009 (CL2009)*, (p. 414).
- [3] Ebeling, S. O. (2009). Oslo Interactive English Corpus-driven exercises on the Web. *Corpora and language teaching*, 33, 67.
- [4] Flowerdew, L. (2009). Applying corpus linguistics to pedagogy. *International Journal of Corpus Linguistics*, 14(3), 393-417.
- [5] Flowerdew, L. (2010). Using corpora for writing instruction. In A. O’Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 444-457). London: Routledge.

- [6] Gaskell, D., Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301-319.
- [7] Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In *Teaching and language corpora* (pp. 83-99). London: Longman
- [8] Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, 63(4), 363–372
- [9] Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching. *The Routledge handbook of corpus linguistics*, 359-370.
- [10] Johns, T. (1997). Contexts: the background, development and trialling of a concordance-based CALL program. In *Teaching and language corpora* (pp. 100-115).
- [11] Leech, G. (1997). Teaching and language corpora: A convergence. In *Teaching and language corpora* (Vol. 123).
- [12] Leńko-Szymańska, A., & Boulton, A. (Eds.). (2015). *Multiple affordances of language corpora for data-driven learning* (Vol. 69). John Benjamins Publishing Company.
- [13] Moon, R. (2010). What can a corpus tell us about lexis. *The Routledge handbook of corpus linguistics*, 197-211.
- [14] Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- [15] Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. *Corpus-based approaches to English language teaching*, 18-35.
- [16] Sharoff, S. (2014). *IntelliText Corpus Queries*. Retrieved from
- [17] <http://corpus.leeds.ac.uk/itweb/htdocs/Query.html#>
- [18] Warren, M. (2015). *Introduction to data-driven learning* (pp. 337-48). The Routledge Handbook of Language Learning and Technology.
- [19] Yule, G. (2014). *The study of language*. Cambridge University Press.