

The Use of Online Corpora as Reference Resources for Revision Tasks in Writing

Qinqin Luo^(✉), Weizhen Huang, and Ying Liao

School of Foreign Languages, Southwest Petroleum University, Chengdu, China
qinqinluo@126.com, 363086681@qq.com

Abstract. The current study reports on an experiment with 60 Chinese college students, aiming to explore the effects of using online corpora BFSU CQP web for revision tasks in EFL writing. Quantitative data about the 60 lower level EFL learners' scores in the post writing test show that there is a significant difference between the experimental group and the control group in terms of the lexical accuracy and the use of cohesion devices, which prove that corpora as reference resources are more useful than the online dictionary in helping learners improve their linguistic competence in writing. The follow-up interviews on the selected participants in the experimental group reveal different learners' perceptions of using online corpora in the process of revising essays and the reasons behind their opinions, which can provide guidance for teachers to apply corpora effectively in teaching writing.

Keywords: Online corpora · Reference resources · Revision tasks · Writing

1 Introduction

The process writing approach that has been prevailing throughout the West in the past thirty years emphasizes the significance of revision. The process of revising is not only helpful to improve the overall writing quality but also crucial to promoting learners' writing competence. But it is found that a considerable number of English as a foreign language (EFL) learners ignore this process and they seldom correct all the underlined errors, leading to the occurrence of similar errors again and again. The reasons for the phenomenon may partly lie in the fact that it is difficult to correct all the errors by learners themselves especially those with lower level of proficiency. However, providing the correct form directly to them may result in learners' over-reliance on teachers. Just as Corder [1, p. 11] states, "making a learner try to discover the right form could often be more instructive to both learners and teachers". That is to say, it is more effective for learners to self-correct errors with the help of some reference resources.

Since Data-driven learning was proposed by Johns [2], corpora have been gradually used as reference resources to assist learners in solving their language problems in writing in western world. Many empirical studies [3–7] have proved the positive effects of using corpora for error correction in writing, but some factors such as error types, learners' training and language proficiency may, to some extent, influence the effects of

corpus use. And it can be also found from studies [8–10] that learners' search behavior and types of corpora also had impacts on the outcome of corpus consultation. Besides correcting errors, Corpora as reference resources could be also used to learn linking adverbials [11, 12], to differentiate synonym adjectives [13], and to assist learners in writing creatively [14]. Thus corpora can be used for various purposes in writing.

However, there are still some limitations of previous studies concerning corpus use in writing. Firstly, the studies concerning lower-level EFL learners in China are relatively few, which is partly due to the fact that corpus tools are not sufficiently user-friendly. Secondly, most studies using corpora for error correction just examine whether the accuracy rate of corrections can be improved but seldom explore the impact of corpus consultation on learners' long-term writing competence. Thirdly, there is still a lack of studies comparing the roles of corpora and traditional resources such as dictionaries in improving learners' linguistic competence in writing. Therefore, the present study focuses on Chinese lower-level EFL learners and selects the user-friendly online corpora as the platform aiming to examine the long-term effects of using corpora for revision tasks, that is, to check whether corpora can play more effective roles in promoting EFL learners' writing than online dictionaries in the following aspects: to improve the linguistic accuracy and lexical complexities, and to make learners more skilled in using cohesion devices such as linking adverbials thus to improve the cohesion of the essay.

2 Method

2.1 Research Questions

In this study, the following research questions are investigated:

- (1) Which one is more useful in helping learners improve linguistic competence in free writing, consulting corpora or consulting the online dictionary in the process of revising essays?
- (2) What are the differences between learners who make apparent progress and those who make no progress or even behave worse in post-writing test in terms of their attitudes toward corpus use in the process of revising essays?

Learners' linguistic competence is mainly examined in the writing test in terms of the linguistic accuracy, complexity, and the use of cohesion devices. And EFL learners' attitudes are investigated in the form of interview.

2.2 Participants

60 Chinese sophomores with 7 years' experience of learning English participated in the whole experiment, and they finished all the designated writing and revision assignment under the teacher's instruction. Among them, 27 (3 females, 24 males) from one college English class are in the control group and 33 (4 females, 29 males) from another class are in the experimental group. All of them were lower-level learners since their scores in College English Test Band 4 (CET 4 for short) in July 2014 ranged from 350–399 (The total score is 710). At the end of the term they would take CET 4 again thus they

had strong motivation for improving their writing competence, which may guarantee their enthusiasm in continually writing and revising essays. A writing test conducted before the experiment showed no statistically significant difference between the two groups in English writing competence.

2.3 The Online Corpora and Online Dictionary

To supervise them better and confirm they really exploited the reference resources, the two groups were supposed to use the designated online corpora or online dictionary.

The online corpora used by the experimental group was **BFSU CQP web** (<http://124.193.83.252/cqp/>), which was set up by Mr. Wu and maintained by Dr. Xu and Mr. Wu of the National Research Centre for Foreign Language Education. Compared with other online corpus, BFSU CQP web has the following advantages: Firstly, it is user-friendly and easy to handle enabling technically less competent learners to exploit corpora just like browsing web pages [15]. Secondly, EFL learners can get multiple corpus examples from different corpus. When learners log on to BFSU CQP web, they can have quick access to several native speakers' corpora such as Brown corpus, CLOB corpus. So when one corpus fails to provide sufficient examples about a key word they can conveniently consult another corpus to get more corpus evidence. Thirdly, the 4 parallel corpora can help learners a lot when learners are confused about some new words in the corpus examples or when they are unable to convey what they want to describe in target language.

The control group in this study preferred <http://dict.cn/> as the online dictionary because a majority of them have been already accustomed to using it. This dictionary can be stored in mobile phones and it not only provides functions of the monolingual dictionaries but also has the advantages of bilingual dictionaries.

2.4 Procedure

The experiment lasted for a whole term. The procedure was as follows.

Step 1: Training About Corpus Use. Since all the participants had no prior experience in consulting corpus, instruction was given to the experimental groups about how to use corpora on revision tasks before the writing assignment. As Smart [16] states, interacting with corpora directly may be a challenge for learners who are not used to inductive activities and for those who have never used corpora in language learning. Thus the training started from the paper-based materials without requiring learners to access corpora directly. The teacher prepared reference sheets containing concordances selected from BFSU CQP web. For example:

We should always remind ourselves the importance of learning in the university.

Figure 1 provided information for students to correct their mistakes. When they were able to correct errors in this way, they were required to get direct access to corpora to finish some revision tasks. 3 types of revision tasks were included here.

Type 1: Correct the lexico-grammatical errors in the sentences.

More and more people dependent on computers.

- Type 2: Find an alternative to the underlined words.
Besides the good effects, it also has bad effects.
- Type 3: Use appropriate cohesion devices (linking adverbials).
It is going to rain now. I think we should go back home.

he told himself. She no longer wanted anything about him to	<u>remind</u>	her of the circumstances of their meeting that first night in Parioli
that some of the children , in spite of their color ,	<u>remind</u>	him of children he has known and loved , perhaps even of
had mounted it on velvet and hung it over my desk to	<u>remind</u>	me always to use the power of the paper in a Christian
I hung up I was reasonably certain that Francesca had wanted to	<u>remind</u>	me of our town meeting the next evening , and how important
Washington Square seems not part of a free land . It may	<u>remind</u>	one of Russia , China or East Berlin ; but it ca
of Russia , China or East Berlin ; but it ca n't	<u>remind</u>	one of the freedom that Washington and the Continental soldiers fought for
their proportionate numerical strength would have warranted . We do well to	<u>remind</u>	ourselves that from men and women of New England ancestry also issued
little credit for the invention " . It is hardly necessary to	<u>remind</u>	students of covered bridges that Timothy Palmer was born in 1751 in

Fig. 1. Eight concordance lines of REMIND retrieved from BFSU CQP web

The teacher showed how to do simple queries as well as complex queries in BFSU CQP web in the classroom. If they just conducted the simple query, they may hardly find out the appropriate alternatives. For example, to revise the second type of the errors, learners had to input [pos = "J.*"] "effect", and then they can get the concordance evidence as follows (Fig. 2).

meteorologists the drought has had a	<u>limited effect</u>	on crop production . " Rainfall
, said the extra spending will have a	<u>dramatic effect</u>	. " With these efforts , China w
improve the practice and beef up the	<u>supervisory effect</u>	of the committees , Zhao said .
e registration ban is good but of little	<u>significant effect</u>	this year . He said figures from
i Lama 's retirement will only have a	<u>limited effect</u>	on the autonomous region whil
ent document , which has a universal	<u>binding effect</u>	, is not subject to court hearing;
ddling of international affairs and the	<u>positive effect</u>	the country has had on the prom
China and India will not only have a	<u>profound effect</u>	on our two countries , but also
prosperous Central Asia would have a	<u>positive effect</u>	on China 's underdeveloped Xi
ability and social service will have a	<u>negative effect</u>	on building efficient governme

Fig. 2. Ten concordance lines of EFFECT extracted from BFSU CQP web

After the training, some revision tasks were given to the experimental group to make the participants more skilled in consulting corpora independently. One thing to note is that there was no special training on the control group for they knew how to use the online dictionary to revise their essays.

Step 2: Writing and Revision Assignment. After the training, the participants were required to write 6 essays in total and revise the marked essays based on teachers' requirements. The procedure about every essay was as follows. Firstly, the teacher gave students the writing assignment and provided some cohesion devices (linking adverbials) relevant to the essay for them without telling them the specific usage before

writing. Secondly, the participants finished the writing assignment after class and handed in them on time. Thirdly, the teacher underlined all the places that should be revised by consulting corpora or the online dictionaries. Sometimes the teacher also offered some synonyms or words that were easily confused to learners, in order that they could differentiate and use these words or phrase better by consulting reference resources by themselves. Fourthly, the experimental group revised the essays and kept a record of the concordance evidence while the control group finished the revision tasks by referring to the online dictionary. After that, they exchanged their revised essays with each other to examine whether the revision is correct or appropriate and then handed in the essay again to the teacher. Finally, the teacher examined the revised essays and gave a comment in the class based on their correction work.

2.5 Data Collection and Analysis

In order to examine whether the use of corpora have positive effects on learners' linguistic competence in writing, the study employed pre- and post-writing test design. And the topic of the writing tests were both about the internet that the participants were familiar with. Both tests included measures of learners' linguistic accuracy, complexity, and the use of cohesion devices without considering the contents of the essay. The linguistic accuracy was marked by the number and seriousness of the lexico-grammatical errors. The fewer errors they committed, the higher scores they got. The mechanical errors such as spelling errors, capitalization errors were regarded as minor errors which has less negative influence on the score than the serious grammatical errors such as verb forms, tense. The linguistic complexity was assessed in terms of lexical and syntactic complexity. The lexical complexity was measured by the sophistication and array of learners' productive vocabulary [17]. The syntactic complexity was judged by how varied and sophisticated the construction elements or linguistic structures are [18]. The use of cohesion devices was mainly marked globally based on the frequency, variety and appropriateness of the linking adverbials in the essay. Three college English teachers were invited to mark the essays respectively in terms of the above three aspects to ensure the objectivity of the test. The average score of the 3 teachers about each aspect would be calculated, and then we put the data in SPSS17.0. An independent T-test was conducted to examine whether there was significant difference between the two groups in the three aspects of linguistic competence in writing. To know the participants' attitudes toward corpus use in the process of revising essays, a semi-structured interview was conducted on 6 participants in the experimental group.

3 Results Analysis and Discussion

3.1 Results for the Pre-writing Test

It can be revealed from Table 1 that there was no significant difference between the control group and the experimental group in terms of their linguistic accuracy, complexity, and the use of cohesion devices. So the two groups were equivalent in

linguistic competence in writing at the beginning of the experiment ($p > 0.05$). The following problems existed equally in both of the two groups' essays: frequently appeared lexico-grammatical errors, lack of cohesion devices such as linking adverbials, an obvious lexical poverty reflected by their use of too many superordinates or general words such as good, bad, important, etc.

Table 1. Comparison between the two groups about the pre-writing test

Category	Group	N	M	Std.	Mean Diff.	t	Sig. (2-Tailed)
Accuracy (5)	Control	27	2.8770	0.5149	0.0228	0.198	0.844
	Ex.	33	2.8542	0.3769			
Complexity(5)	Control	27	2.6974	0.5052	-0.0244	-0.214	0.831
	Ex.	33	2.7218	0.3789			
Cohesion devices (5)	Control	27	2.9267	0.4194	-0.1688	-1.681	0.098
	Ex.	33	3.0955	0.3584			
Total (15)	Control	27	8.4996	0.8814	-0.1719	-0.827	0.412
	Ex.	33	8.6715	0.7296			

3.2 Results for the Post-writing Test

The results of the post-test in Table 2 showed that there were significant differences between the two groups in terms of linguistic accuracy, the use of cohesion devices and the overall writing quality ($p < 0.05$) but no significant difference existed between them in linguistic complexity ($p = 0.121 > 0.05$).

In terms of linguistic accuracy, the experimental group had superior results than the control group, showing that the experimental group made fewer lexico-grammatical errors in essays. In the process of correcting errors, they had to analyze concordance lines and induce the lexical or grammatical rules and eventually applied the rules into the correction work. It's obvious that they were actively involved in the learning process, undertaking the researchers' role, to solve linguistic problems. This time-consuming process may make them deeply aware of the wrong linguistic forms and the repeated linguistic pattern in concordance line may help them master the right usage better. Thus they would try to avoid the previous errors and naturally recalled the right pattern which frequently appeared in corpus examples. This supports Huang's [19] opinion that corpus-based inductive learning helps learners notice and acquire collocational patterns better, which in turn enables learners to generate more accurate patterns. However, the limited examples in the online dictionary were not sufficient to make learners in the control group fully notice the errors. Due to the native language interference, they always committed the similar errors, which could be found in their post-test. This result shows

support for a number of papers [3, 4] which study the use of corpora as a reference resource for error-correction.

Table 2. Comparison between the two groups about the post-writing test

Category	Group	N	M	Std.	Mean Diff.	t	Sig. (2-tailed)
Accuracy (5)	Control	27	2.9944	0.4444	-0.6668	-5.521	0.000**
	Ex.	33	3.6612	0.4818			
Complexity (5)	Control	27	2.9937	0.5136	-0.2084	-1.573	0.121
	Ex.	33	3.2021	0.5081			
Cohesion devices (5)	Control	27	3.3707	0.4921	-0.3059	-2.366	0.021*
	Ex.	33	3.6767	0.5033			
Total (15)	Control	27	9.3581	1.3596	-1.1970	-3.318	0.002**
	Ex.	33	10.5552	1.4149			

**p < 0.001, *p < 0.05

Regarding linguistic complexities, the experimental group also had a higher mean score than the control group, but the difference was not significant. The reasons may be as follows. Firstly, it may take a long time for learners to improve their linguistic complexities. One term's training about writing and revision may not be enough for them to make significant progress. Secondly, some participants didn't know how to skillfully consult corpora to improve the complexities. In most of the cases they had to conduct complex queries or consult the parallel corpus for finding better alternatives to replace the underlined simple word or phrase, but some of them seldom did the complex query and the parallel corpus was not exploited fully.

Concerning the use of cohesion devices, the experimental group also behaved significantly better. They used the linking adverbials more frequently and more skillfully in the post-test. A lot of participants in the experimental group tried to put some linking adverbials such as however in the middle of a sentence as a parenthesis since however was often put in the middle of a sentence in corpus examples. But most participants just habitually put them at the beginning or the end of a sentence. This result is consistent with Boulton's [11] opinion that corpus data as reference resource are more effective in learning linking adverbials than the traditional resources.

3.3 Results for the Interview

The follow-up interview included 3 questions: (1) What are the positive and negative sides of using corpora in revision tasks? (2) Do you think the use of corpora can improve your overall writing competence? (3) Will you keep the habit of using corpora in the future writing activities? There were altogether 6 participants in the experimental group

being interviewed. Among them, 3 were randomly selected from those who made apparent progress in the post-test and they were assigned to the first group being interviewed in the first round. The other 3 were the only 3 learners who got lower scores in the post-test and they were assigned to the second group being interviewed in the second round. The two different groups showed divergent opinions about the use of corpora.

The first group generally had a favorable attitude towards corpus use for revision tasks in writing. Firstly, they believed it could help them improve their linguistic accuracy by saying that “it’s really helpful for detecting and correcting errors, especially errors about collocation and preposition.” This supports O’Sullivan’s [20] opinion that corpus consultation can present an opportunity for learners to notice their errors and provide evidence that learners can use to correct their linguistic output. Besides the positive effects on error-correction, they also emphasized some other benefits of using corpora for revision tasks such as “raising the awareness of the importance of collocations”, “mastering the usages of words and phrases better”, “acquiring new words incidentally”, “learning synonyms in a better way” and so on. Although some negative sides such as “too time-consuming” and “failure to understand the incomplete sentences” were also mentioned, they generally agreed that the positive sides outweighed the negative sides and they believed that the use of corpora was beneficial to improving their linguistic competence in writing as long as they could persist in using corpora. Thus when being asked about whether they would keep using corpora, two of them gave a definite answer “Yes”, saying that they would firstly consult corpora when they encounter some expressions that they are not sure of. And the other one was a little hesitated by saying “it depends” for he believed that dictionaries and corpora could complement each other. He revealed that he would firstly refer to the dictionary and when the dictionary fail to provide what he wants then he would consult the corpora.

The second group revealed less positive attitudes toward corpus use. Although they all admitted there were some positive effects of using corpora in writing, they expressed their reluctance to use corpora sometimes for they were usually more willing to be told how to correct the errors directly rather than take much time to explore the right form by themselves. They have already been used to being told what to do by the teacher since the primary school. This may be one of the reasons why a small number of students in the experimental group made no improvement. In their eyes, the negative sides outweighed the positive sides, which could be reflected from the following comments: “too many examples in concordance lines”, “too much time spent in searching, analyzing and inducing”, “too many new words”, “unable to induce the right grammatical rules”, “failure to get what I want”, etc. Due to these factors, while being asked about whether the use of corpora improved their linguistic competence in writing they were not certain. One of them responded that sometimes he consulted corpora just for finishing the teachers’ assignment but not for finding out the best alternatives or the most appropriate word. And the other two also expressed the similar meaning that if they could spend more time in learning from corpora and persist in using them, their linguistic competence in writing would be improved. Finally, only one of them implied that he would still keep on using corpora for reference purposes in writing and the other two responded that they will firstly refer to the online dictionary which can save much time and then go to corpora for more detailed information if time permits.

The results of the interviews revealed that the more benefits learners got, the more positive attitudes they had. Thus, to make lower-level learners actively involved in learning from corpus consultation, teachers can start with some simple tasks since the accomplishment of these tasks can help them feel the sense of achievement, which may stimulate them to keep on using corpora. In the meanwhile, appropriate guidance should be offered to them to reduce their anxiety in using corpora especially when they get confused in analyzing the concordances.

4 Conclusion

The present study explored the effects of using online corpora BFSU CQP web in the process of revising essays, finding that the use of corpora is more useful than the online dictionary in improving EFL learners' linguistic competence in writing, especially the linguistic accuracy and the use of cohesion devices. However, the interviews revealed that learners' attitudes toward corpus use were closely related to the benefits learners got from corpora. If they got no progress from using corpora they would naturally show negative attitudes and reluctant to use them in the future. Thus teachers should help them overcome the difficulties technically and psychologically to make them benefit more from using corpora. If learners can skillfully use corpora and get achievements, they will voluntarily exploit corpora for their own purposes, which will definitely promote learner autonomy and improve their writing competence. Nevertheless, there are still some limitations of the present study. Firstly, the learners' essays were merely marked by three teachers, the automated writing assessment system can be used in the further study to ensure more objectivity and transparency in the assessment. Secondly, this study just focused on three aspects of learners' linguistic competence regardless of the contents of the essay, the following study can concentrate on the effects of corpus use on the improvement of content since the content also has great impact on the overall writing quality. Thirdly, the experiment lasted for only one term, a longitudinal study should be conducted to know whether the corpus use can make learners form the habit of using corpora for writing purposes.

Acknowledgements. This work was supported by Southwest Petroleum University Science and Technology Foundation (Grant No. 2012XJR013) and Web Culture Project Sponsored by the Humanities and Social Science Research Base of the Sichuan Provincial Education Department (Grant No. WLWH15-28).

References

1. Corder, P.: *Error Analysis and Interlanguage*. Oxford University Press, Oxford (1981)
2. Johns, T.: Should you be persuaded: two examples of data-driven learning. In: Johns, T., King, P. (eds.) *Classroom Concordancing*. ELR, Birmingham (1991)
3. Gaskell, D., Cobb, T.: Can learners use concordance feedback for writing errors? *System* **32**(3), 301–319 (2004)
4. Gilmore, A.: Using online corpora to develop students' writing skills. *ELT J.* **11**, 1–10 (2009)

5. Chambers, A., O'Sullivan, Í.: Corpus consultation and advanced learners' writing skills in French. *ReCALL* **16**(1), 158–172 (2004)
6. O'Sullivan, Í., Chambers, A.: Learners' writing skills in French: corpus consultation and learner evaluation. *J. Second Lang. Writ.* **15**(1), 49–68 (2006)
7. Tono, Y., Satake, Y., Miura, A.: The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL* **26**, 147–162 (2014)
8. Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J.M.: Learners' search patterns during corpus-based focus-on-form activities. *Int. J. Corpus Linguist.* **17**(4), 483–516 (2012)
9. Chang, J.Y.: The use of general and specialized corpora as reference sources for academic English writing: a case study. *ReCALL* **26**(2), 243–259 (2014)
10. Lee, D., Swales, J.: A corpus-based EAP course for NNS doctoral students: moving from available specialized corpora to self-compiled corpora. *Engl. Specif. Purp.* **25**(1), 56–75 (2006)
11. Boulton, A.: Testing the limits of data-driven learning: language proficiency and training. *ReCALL* **21**(1), 37–51 (2009)
12. Cotos, Elena: Enhancing writing pedagogy with learner corpus data. *ReCALL* **26**(2), 202–224 (2014)
13. Yeh, Y., Liou, H.C., Li, Y.H.: Online synonym materials and concordancing for EFL college writing. *Comput. Assist. Lang. Learn.* **20**(2), 131–152 (2007)
14. Kennedy, C., Miceli, T.: Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. *Lang. Learn. Technol.* **14**(1), 28–44 (2010)
15. Xu, J.J., Wu, L.P.: Web-based fourth generation corpus analysis tools and the BFSU CQP web case. *Comput.-Assist. Foreign Lang. Educ.* **5**, 10–15 (2014)
16. Smart, J.: The role of guided induction in paper-based data-driven learning. *ReCALL* **26**, 184–201 (2014)
17. Wolfe-Quintero, K., Inagaki, S., Kim, H.Y.: Second language development in writing: measures of fluency, accuracy, and complexity (Report No. 17). Second Language Teaching and Curriculum Center, University of Hawai'i, Honolulu (1998)
18. Ortega, L.: Syntactic complexity measures and their relationship to L2 proficiency: a research synthesis of college-level L2 writing. *Appl. Linguist.* **24**(4), 492–518 (2003)
19. Huang, Z.P.: The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL* **26**, 163–183 (2014)
20. O'Sullivan, I.: Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. *ReCALL* **19**(3), 269–286 (2007)