

EFL students' perceptions of corpus-tools as writing references

Shu-Li Lai¹

Abstract. A number of studies have suggested the potentials of corpus tools in vocabulary learning. However, there are still some concerns. Corpus tools might be too complicated to use; example sentences retrieved from corpus tools might be too difficult to understand; processing large number of sample sentences could be challenging and time-consuming; also, not all English as a Foreign Language (EFL) learners know how to induct and deduct information from the concordance lines. So far, not much is known regarding how EFL writers actually perceive such tools as writing aids. To better understand this question, building on the same data set as the one published in 2015 (Lai & Chen, 2015), this study investigated students' perceptions of corpus tools right after they applied such tools to three writing tasks. Four online corpus tools, including monolingual and bilingual concordancers and collocation retrieval systems were provided along with two online dictionaries. After tool-training sections, students performed three timed-writing tasks online in three consecutive months and received individual recall interviews after each writing task. The interviews served as the major source of data. The analysis of the qualitative interview data revealed how the students perceived (1) the corpus tools as writing references, (2) the roles of the authentic example sentences, (3) the bilingual feature of the corpus tools, (4) the presentation of the corpus results, and (5) the features of the collocation retrieval system. Overall, the 14 students greatly valued corpus tools as writing references. The data also revealed how consulting the corpus helped the students to increase their confidence in writing, particularly in terms of wording. By eliciting students' perceptions and comments right after they integrated these tools into their writing tasks, the results obtained revealed not just writer perceptions but also on-site empirical data regarding how corpus tools contribute to EFL writing.

Keywords: reference tools, concordancers, EFL writing.

1. National Taipei University of Business, Taiwan; shulilai@gmail.com

How to cite this article: Lai, S.-L. (2015). EFL students' perceptions of corpus-tools as writing references. In F. Helm, L. Bradley, M. Guarda, & S. Thoušny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 336-341). Dublin: Research-publishing.net. <http://dx.doi.org/10.14705/rpnet.2015.000355>

1. Introduction

Corpus tools have gained increasing attention in the field of English as a Second Language (ESL)/English as a Foreign Language (EFL) learning in recent years. Their unique characteristics, providing abundant example sentences and presenting keywords in context, are believed to be beneficial in L2 learning, especially in regard to vocabulary learning and L2 writing. It has been argued that such tools have an impact on vocabulary learning and ESL/EFL writing. They draw learners' attention to word patterns, collocation information, and contextual environments. They also increase learners' depth of vocabulary knowledge. Teaching or learning vocabulary using corpora assistance is known as the "corpus-based" approach and is often associated with the "data-driven approach" or "Data-Driven Learning" (DDL), coined by Tim Johns (1991). He established the relationship between corpora and language learning, believing that learners should be guided to detect the underlying rules or patterns in language use and to draw conclusion from clues in the data (Johns, 1994, 2002). Then, students will gradually learn how to discover facts about the language from the concordance lines and eventually become independent learners.

Research findings have suggested some benefits of corpus learning. It increases L2 writers' lexical and contextual awareness (Tribble, 2002); it encourages autonomous learning and improves critical thinking skills (Kirk, 2002). It also helps translation and interdisciplinary language studies (Boulton, 2011, 2012). However, such tools also carry some limitations. Observing the large amount of authentic sample sentences can be time-consuming and may frustrate learners (Granger & Tribble, 1998). Also, although the features of keywords in context facilitate the inference and generalization of rules, not all learners know how to induct information from concordance lines (Gabel, 2001). Some of the concordancing tools are complicated to use and present the concordance outputs in formats that learners may find difficult to interpret or generalize (Yoon & Hirvela, 2004). With the potential contributions and limitations of corpus tools, how EFL writers perceive such tools after they actually use them during the writing process is not clear. This study was conducted to further explore this question.

2. Method

2.1. Research design

The study was conducted in an EFL introductory writing class for non-English majors ($N=14$) over a semester. Participants were all college students, mostly with

engineering backgrounds. The students performed three in-class writing tasks in a semester, each followed by a semi-structured interview. The interview notes served as the major data of this study. The writing processes were screen-recorded by a computer program. The video files of the writing process served as stimuli during the interview process.

2.2. Online reference tools

While students composed their writing online, they had access to two online dictionaries and four web-based corpus tools. Of the four corpus tools, two were concordancers and the others were collocation retrieval systems. For each type, a monolingual and a bilingual tool were provided (see [Table 1](#)).

Table 1. The four corpus tools

Tools	Features	Corpus
TotalRecall ²	Bilingual concordancer	Sinorama (1990-2000)
VLC web concordancer ³	Monolingual concordancer	Brown corpus, LOB, news articles from times on various topics, short stories
Tango V-N collocation ⁴	Bilingual collocational retrieval system	Sinorama (1990-2000)
NTNU collocation retrieval system (NTNU CCRS)	Monolingual collocation retrieval system	British National Corpus

In addition to the four corpus tools, and in order to provide students' with a natural writing setting, similar to their previous writing experience, students had access to two online dictionaries as well (see [Table 2](#)). Most students rely on the bilingual online dictionary (i.e. the online yahoo dictionary) when they have to write in English. Some students even use the monolingual dictionary, though not many. The researcher thus could find out how students used dictionaries and concordancers differently, and how they perceived the two different kinds of tools as writing aids.

Table 2. The two dictionaries

Tools	Features	Outputs
The Yahoo dictionary	Bilingual	Chinese definitions with English examples
Cambridge advanced learners' online dictionary	Monolingual	English definitions with English examples

3. Results and discussion

Overall, students' perceptions of the corpus tools as writing references were positive. They reported that the corpus tool was a complement to the bilingual dictionary, as has been previously discussed (Lai & Chen, 2015). Very often, bilingual dictionaries do not provide enough sample sentences. Several students commented on the crucial roles of the example sentences in their writing. They provided contextual clues and displayed key words in context. They helped students generate the meaning of the word and differentiate apparent synonyms. Additionally, these sentences provided syntactical information and acted as models for usage of the keywords. Students expressed their happiness about having a sentence to imitate. One even related this experience to the L1 learning experience he had where he practiced writing a sentence by imitating how people wrote. According to Martin-Rutledge (1997), exposure to sufficient contexts explains words better than definitions or explicit information in dictionary entries can. In addition to the linguistic help, students reported that they sometimes got inspired by reading the concordance lines. They enriched their content as well.

In this study, students used the bilingual concordancers much more often than the monolingual concordancers. One student commented that providing both Chinese and English examples, in parallel forms, is very important. He even further noted that whether the tools provided bilingual example sentences was his main concern when deciding the type of tool to use. The bilingual tool, on the one hand, allows bidirectional searches; on the other hand, the Chinese parallel concordances provided clues to help the students quickly find the information they needed. In fact, it also gave those who had limited reading proficiency a chance to make sense of the corpora and to make use of such tools. Clues were found in this study as well. This might explain why students with different proficiency levels all managed to use the corpus tools in some way, and all held a positive view, which contradicts the findings of some similar studies (e.g. Yoon & Hirvela, 2004). It is very likely that the bilingual feature of the corpus tools eased the process and made it easier for EFL writers to retrieve information from concordance lines.

Students also liked the way corpus tools presented their results, especially VLC. When VLC presents its results, the keywords are aligned in the middle, highlighted in red; in addition, the words following the keywords are listed in alphabetical order. Students even encountered some incidental learning regarding word usage during the corpus observation process. In fact, students' perceptions of the presentation of the corpus results by VLC carries a number of theoretical and pedagogical implications. The unique format, different from a dictionary,

drew students' attention to the collocation behaviors, grammatical patterns, and related meanings. Thanks to this, students were more likely to notice the patterns. According to Schmidt (2001), SLA is mostly driven by what learners pay attention to, and what learners notice in the target language.

Students' perceptions of the collocation tool, NTNU CCRS, is also positive. It was reported to be helpful and user-friendly. Through several mouse clicks, the student writers could find the collocation behaviors of the target words and the frequency of each collocation. As one student commented, this gave him some clues and helped him to make a better decision about which word to use in his writing. As found in the other study, students need more reference tools in addition to the dictionary, especially tools that allow collocation research (Lai & Chen, 2015). One student mentioned that he just relied on his intuition before he learned about such tools. Although the frequency counts carried information that was important and useful for the EFL writers, one student reported that people should be cautious about these numbers. He commented that the most frequent collocates may not be the right words for the writing context. On the one hand, students saw the strength of such tools, and were able to make good use of the frequency information; on the other hand, they were very cautious when selecting a word to use. They took the frequency into consideration and looked into each instance to evaluate the contexts.

4. Conclusions

In general, students believed that corpus tools provided information that was not likely to be found in a bilingual dictionary. Corpus tools were found to provide more example sentences and give information on collocations. Overall, the students greatly valued them and stated that having access to corpus tools helped them to improve their wording and increase their confidence in their own work.

5. Acknowledgements

I would like to thank the developers of the corpus tools mentioned in the study for designing them and for generously giving all users open access to them.

References

- Boulton, A. (2011). Bringing corpora to the masses: free and easy tools for language learning. In N. Kübler (Ed.), *Corpora, language, teaching, and resources: from theory to practice* (pp. 69-96). Bern: Peter Lang.

- Boulton, A. (2012). Beyond concordancing: multiple affordances of corpora in university language degrees. *Procedia-Social and behavioural science*, 34, 33-38. doi:10.1016/j.sbspro.2012.02.008
- Gabel, S. (2001). Over-indulgence and under-representation in interlanguage: reflections on the utilization of concordancers in self-directed foreign language learning. *Computer Assisted Language Learning*, 14, 269-288. doi:10.1076/call.14.3.269.5792
- Granger, S., & Tribble, C. (1998). Learner corpus data in the foreign language classroom: form-focused instruction and data-driven instruction and data-driven learning. In S. Granger (Ed.), *Learner English on computer* (pp. 199-209). New York: Longman.
- Johns, T. (1991). Should you be persuaded: two examples of data driven learning. *ELR Journal* (New Series), 4, 1-16.
- Johns, T. (1994). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 293-313). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524605.014
- Johns, T. (2002). Data-driven learning: the perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis: proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz, 19-24, July, 2000* (pp. 107-117). Amsterdam: Rodopi.
- Kirk, J. (2002). Teaching critical skills in corpus linguistics using BNC. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp.154-164), Amsterdam: Rodopi.
- Lai, S. L., & Chen, H. J. (2015). Dictionaries vs concordancers: actual practice of the two different tools in EFL writing. *Computer Assisted language Learning*, 28(4), 341-363. doi:10.1080/09588221.2013.839567
- Martin-Rutledge, V. (1997). *Use of examples in the bilingual dictionary: an empirical study*. Unpublished master's thesis, University of Ottawa, Canada.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524780.003
- Tribble, C. (2002). Corpora and corpus analysis: new windows on academic writing. In J Flowerdew (Ed.), *Academic discourse* (pp. 131-149), Harlow: Longman.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257-283. doi:10.1016/j.jslw.2004.06.002

Published by Research-publishing.net, not-for-profit association
Dublin, Ireland; info@research-publishing.net

© 2015 by Research-publishing.net (collective work)
© 2015 by Author (individual work)

Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy
Edited by Francesca Helm, Linda Bradley, Marta Guarda, and Sylvie Thouéšny

Rights: All articles in this collection are published under the Attribution-NonCommercial -NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Under this licence, the contents are freely available online (as PDF files) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.



Disclaimer: Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it is not under consideration for publication elsewhere. While the information in this book are believed to be true and accurate on the date of its going to press, neither the editorial team, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

Trademark notice: product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Copyrighted material: every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-28-5 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

ISBN13: 978-1-908416-29-2 (Ebook, PDF, colour)
ISBN13: 978-1-908416-30-8 (Ebook, EPUB, colour)

Legal deposit, Ireland: The National Library of Ireland, The Library of Trinity College, The Library of the University of Limerick, The Library of Dublin City University, The Library of NUI Cork, The Library of NUI Maynooth, The Library of University College Dublin, The Library of NUI Galway.

Legal deposit, United Kingdom: The British Library.
British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2015.