The effects of utilizing corpus resources to correct collocation errors in L2 writing – Students' performance, corpus use and perceptions

Yi-ju (Ariel) Wu[1]

Abstract. Data-Driven Learning (DDL), in which learners "confront [themselves] directly with the corpus data" (Johns, 2002, p. 108), has shown to be effective in collocation learning in L2 writing. Nevertheless, there have been only few research studies of this type examining the relationship between English proficiency and corpus consultation. The current study intends to fill the gap by investigating how 140 learners of three different levels of English proficiency from Taiwan utilized Corpus of Contemporary American English (COCA) to correct eight different types of collocation errors adapted from their writing. Data was obtained from three aspects: learners' collocation performance, learners' COCA use, and learners' evaluation toward COCA. A mixed-methods approach that included quantitative statistics and qualitative interviews was used. The results showed that even though learners of higher English proficiency performed better when using corpus, learners across all proficiency levels have improved collocation performance by 30%. Nevertheless, even though lower proficiency learners have received the same amount of assistance from corpus consultation, they did not think corpus as helpful as their higher proficiency fellows did. Teachers should not restrict the use of corpus to higher proficiency learners only because lower proficiency can also benefit from corpus use.

Keywords: data-driven learning, L2 writing, corpus linguistics, collocation teaching, collocation learning.

1. Chinese Culture University, Taipei, Taiwan; yjarielwu@gmail.com

How to cite this article: Wu, Y. J. (2016). The effects of utilizing corpus resources to correct collocation errors in L2 writing – Students' performance, corpus use and perceptions. In S. Papadima-Sophocleous, L. Bradley & S. Thouësny (Eds), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 479-484). Research-publishing.net. https://doi.org/10.14705/rpnet.2016.eurocall2016.610

1. Introduction

DDL, originated from Johns (1990), in which learners approach linguistic data and induce patterns, has shown to be effective in collocation learning. Learners are able to consult a corpus, induce patterns, and correct collocation errors in their L2 writing (e.g. Kennedy & Miceli, 2010; O'Sullivan & Chambers, 2006). While most researchers have argued that student-led corpus consultation can benefit advanced learners the most (Lin, forthcoming; O'Sullivan & Chambers, 2006), there has been little empirical evidence to support that claim (e.g. Boulton, 2009; Tono, Satake, & Miura, 2014).

The current study intends to fill the gap by investigating how 140 learners of three different levels of English proficiency from Taiwan utilized COCA to correct eight collocation errors adapted from their writing. The research questions addressed are as follows:

- How are the collocations enhanced after learners of three different levels of English proficiency consult COCA?
- How do learners of three different English proficiency levels utilize COCA?
- How do the students reflect on their COCA use?

2. Method

2.1. Participants

The participants were non-native English speakers whose first language was Mandarin Chinese. Before taking part, the students had learned English for about eight to ten years, and their English proficiency levels spanned mostly the B1 to B2 levels in the Common European Framework of Reference for language (CEFR). Students were divided into three levels as Group A (46 students, lower-intermediate level, B1 in the CEFR), Group B (47 students, intermediate level, B1+ in the CEFR) and Group C (47 students, upper-intermediate level, B2 in the CEFR).

2.2. Research instrument and materials

First, after the corpus tutorial and collocation instruction, learners completed the paper-based test in which they corrected eight collocation errors and provided

three correct answers for each error. Table 1 shows the eight collocation errors adapted from their writing and were chosen based on three sets of variables: L1 congruency, level of difficulty, adjective+noun collocation and verb+noun collocation (e.g. Nesselhauf, 2003; Sun & Wang, 2003). Learners were not allowed to use any reference tools, which helps researchers to evaluate their original collocation knowledge. Afterwards, learners completed the COCA-based test, in which they corrected the same eight collocation errors by consulting COCA. In the study, learners used the 'LIST' function of COCA which showed the 100 most frequent collocates of the searched word. Learners were allowed to check other reference tools such as dictionaries, and they reported how many collocates they had clicked when checking COCA LIST functions. Learners completed the questionnaire regarding their corpus use after the COCA-based test. Ten students from three proficiency levels out of 140 subjects were chosen to videotape their corpus consultation and be interviewed at the end.

Table 1. Eight collocation errors in the paper-based and COCA-based tests

	Collocation errors and three sample answers	**Verb+noun collocation or adjective+noun collocation**	**Easy or difficult collocation**	**L1 congruency**
Q1.	*tall salary :high, good, nice	Adj.+noun	Easy	L1 congruent
Q2.	*produce money: make, earn, get	Verb +noun	Easy	L1 congruent
Q3.	*appropriate reason :legitimate, justifiable, compelling	Adj.+noun	Difficult	L1 incongruent
Q4.	*destroy attempt :defeat, foil, thwart	Verb +noun	Difficult	L1 incongruent
Q5.	*good desire: genuine, sincere, real	Adj.+noun	Easy	L1 incongruent
Q6.	*talk concern: express, show, convey	Verb +noun	Easy	L1 incongruent
Q7.	*unsatisfying desire : insatiable, unquenchable, overwhelming	Adj.+noun	Difficult	L1 congruent
Q8.	*cancel taxes : reduce, eliminate, abate	Verb +noun	Difficult	L1 congruent

2.3. Data analysis

First, the scores from both paper-based and COCA-based tests were analyzed using Stata's descriptive statistics and a regression analysis to investigate the differences in the learners' scores to answer research question one about learners' performance and improvement in collocation knowledge. To answer research

question two regarding subjects' COCA use, the use record from the 140 subjects, regarding the number of collocates they had clicked on in COCA, was analyzed by using descriptive statistics and a regression analysis in Stata to see how the number of collocates checked on the COCA LIST correlated with learners' performance and improvement. Also, the videotaped videos and interview were analyzed qualitatively to understand the qualitative corpus behavior. In addition, the third question regarding learners' attitudes toward corpus use was answered by analyzing questionnaire results and interview results.

3. Discussion

In total, 140 subjects corrected eight collocation mistakes in the paper-based and COCA-based tests, so in total, 1120 question hits were produced in each test. First, as shown in Table 2, subjects of higher proficiency have better collocation performance in the COCA-based test (Chan & Liou, 2005). Likewise, when learners utilize corpus to learn lexico-grammatical patterns, advanced learners perform the best (e.g. Johns, 1991; Lin, forthcoming).

The current study further shows that the key point which makes learners of higher proficiency level outperform learners of lower proficiency levels is their better analytical and linguistic skills, rather than the better query skills (e.g. O'Sullivan & Chambers, 2006) or more corpus searches, because the behavior log showed that learners of all proficiency levels had no problem conducting corpus query and the frequency of corpus consultation does not differ among learners of different proficiency levels.

Nevertheless, Table 2 also indicates that although learners of higher English proficiency outperformed learners of lower English proficiency in the COCA-based test, learners of all three proficiency levels improved the same amount in collocation by 30%. This shows that to assist learners of lower proficiency in DDL, student-led corpus consultation can be made easier through the following ways, such as giving learners access to a dictionary, providing sufficient corpus training (e.g. Kennedy & Miceli, 2010) and offering adequate teacher support such as underlining the errors (e.g. Mueller & Jacobsen, 2015; Tono et al., 2014) to help them conquer the difficulties in corpus consultation, such as inadequate skills in corpus query (e.g. Charles, 2011) and unfamiliar vocabulary and grammar in concordance lines (e.g. Chang, 2014), as the questionnaire data also showed that learners of lower proficiency did not think corpus use more difficult compared with their high proficiency classmates in the current study.

Nonetheless, even though subjects of lower English proficiency improved the same amount compared to the higher proficiency fellows, they gave corpus use lower evaluation compared with learners of higher proficiency. This also aligns with previous studies that except for advanced learners in Yoon and Hirvela (2004) who showed lower motivation toward corpus use, learners of higher proficiency generally showed more positive feedback to DDL compared to subjects of lower proficiency (O'Sullivan & Chambers, 2006; Tono et al., 2014).

Table 2. Distribution of number and percentage of answers from subjects of various levels in the paper-based and COCA-based test

		Performance
Group A	Paper-based	124 (33.96%)
	COCA-based	240 (65.28%)
	Improvement	116 (31.32%)
Group B	Paper-based	168 (44.7%)
	COCA-based	285 (76.1%)
	Improvement	117 (31.4%)
Group C	Paper-based	173 (46.88%)
	COCA-based	293 (77.97%)
	Improvement	120 (31.09%)

4. Conclusions

This paper showed that corpus consultation is beneficial for learners' collocation enhancement, for both higher and lower proficiency learners, because learners of low proficiency have improved the similar amount as their fellows of higher English proficiency; but they were not aware of the efficacy corpus use has brought to them, as their evaluation toward the utility of COCA was statistically lower. This can be implied rather than restricting the corpus consultation use to learners with high English proficiency, as many researchers have suggested, teachers should, instead, be more open-minded about allowing all learners with different levels of English proficiency to obtain assistance from this resource.

5. Acknowledgements

I would like to thank the Ministry of Education in Taiwan for funding my PhD dissertation writing in which this paper has been adapted from.

References

Boulton, A. (2009). Testing the limits of data-driven learning: language proficiency and training. *ReCALL, 21*(1), 37-51. https://doi.org/10.1017/S0958344009000068

Chan, T. P., & Liou, H. C. (2005). Effects of Web-based concordancing instruction on EFL students' learning of verb–noun collocations. *Computer Assisted Language Learning, 18*(3), 231-251. https://doi.org/10.1080/09588220500185769

Chang, J. Y. (2014). The use of general and specialized corpora as reference sources for academic English writing: a case study. *ReCALL, 26*(2), 243-259. https://doi.org/10.1017/S0958344014000056

Charles, M. (2011).Using hands-on concordancing to teach rhetorical functions: evaluation and implications for EAP. In A. Frankenburg-Garcia, L. Flowerdew, & G. Aston (Eds), *New trends in corpora and language learning* (pp. 26-43). London: Continuum.

Johns, T. (1990). From printout to handout: grammar and vocabulary teaching in the context of data driven learning. *CALL Austria, 10*, 14-34.

Johns, T. (1991). Should you be persuaded? Two samples of data-driven learning materials. *English Language Research Journal, 4,* 1-16.

Johns, T. (2002). Data-driven learning: the perpetual challenge. In B. Kettemann & G. Marko (Eds), *Teaching and learning by doing corpus analysis* (pp. 107-117). Amsterdam: Rodopi. https://doi.org/10.1163/9789004334236_010

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology, 14*(1), 28-44

Lin, M. H. (forthcoming). Effects of corpus-aided language learning in the EFL grammar classroom: a case study of students' learning attitudes and teachers' perceptions in Taiwan, *TESOL Quarterly* (Early View). https://doi.org/10.1002/tesq.250

Mueller, M., & Jacobsen, N. D. (2015). A comparison of the effectiveness of EFL students' use of dictionaries and an online corpus for the enhancement of revision skills. *ReCALL, 28*(1), 1-19.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223-242. https://doi.org/10.1093/applin/24.2.223

O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: corpus consultation and learner evaluation. *Journal of Second Language Writing, 15*(1), 49-68. https://doi.org/10.1016/j.jslw.2006.01.002

Sun, Y. C., & Wang, L. Y. (2003). Concordancers in the EFL classroom: cognitive approaches and collocation difficulty. *Computer Assisted Language Learning, 16*(1), 83-94. https://doi.org/10.1076/call.16.1.83.15528

Tono, Y., Satake, Y., & Miura, A. (2014). The effects of using corpora on revision tasks in L2 writing with coded error feedback. *ReCALL, 26*(2), 147-162. https://doi.org/10.1017/S095834401400007X

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257-83. https://doi.org/10.1016/j.jslw.2004.06.002

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

CALL communities and culture – short papers from EUROCALL 2016
Edited by Salomi Papadima-Sophocleous, Linda Bradley, and Sylvie Thouësny

Typeset by Research-publishing.net
Cover design by © Easy Conferences, info@easyconferences.eu, www.easyconferences.eu
Cover layout by © Raphaël Savina (raphael@savina.net)
Photo "bridge" on cover by © Andriy Markov/Shutterstock
Photo "frog" on cover by © Fany Savina (fany.savina@gmail.com)
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-43-8 (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: 978-1-908416-44-5 (Ebook, PDF, colour)
ISBN13: 978-1-908416-45-2 (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 2016.