Online Corpus Tools in Scholarly Writing: A Case of EFL Postgraduate Student

Sedigheh Shakib Kotamjani1, Ommehoney Fazel Razavi2 & Habsah Hussin1

1 Department of language and Humanities, Faculty of educational studies, Serdang, Selangor, Malaysia
2 Department of Professional Communication, International University of UM-Wales, Kuala Lumpur, Malaysia

Correspondence: Sedigheh Shakib Kotamjani, Department of Language and Humanity, Universiti Putra Malaysia, Malaysia, 43400 Selangor, Malaysia. Tel: 60-166-335-101. E-mail: nshakib58@gmail.com

Received: June 29, 2017   Accepted: August 3, 2017   Online Published: August 5, 2017

Abstract
Some studies have reported the positive outcome of using concordancers and dictionaries in (ESL) context. This study aims to examine how an EFL writer consulted with concordancers and dictionaries along with Google and Google Scholar when engaging in academic writing at university level. The researcher investigated a non-English-major postgraduate student corpus consultation over five months. The researcher provided a toolkit including corpus tools; concordancers, collocation dictionaries, thesaurus, Google, in combination with traditional reference resources such as monolingual and bilingual online dictionaries. The participant received a three-session training to consult with different resources while writing research paper. Real-time data, stimulated recall interview, participants’ writing and query logs served as the main sources of data. Results showed that the participant was aware of the applicability of each corpus tool. He could successfully solve 604 linguistic problems, and promoted his linguistic awareness. It is implied that corpus tools have the potential to assist EFL writers in proofreading and editing the surface levels of their writing.

Keywords: academic writing, concordancing, corpus tools

1. Introduction
Innovative technologies such as personal computers and Internet have revolutionized the process of foreign or second language writing (Stapleton & Radia, 2009; Warschauer, 2007). Particularly, innovations in processing data and exponential increase in data storage capacity have paved the way to provide abundant linguistic information which cannot be available in traditional resources such as dictionaries. Moreover, applied linguistics has creatively provided reference resources with a strong potential to assist foreign language in writing process (Frankenberg-Garcia, 2012; Tono, 2012). One of these recent reference resources is called concordancing or using corpora. Concordancing has gradually been introduced as a language pedagogy tool by Johns (1997), who coined data-driven learning (DDL). Concordancing or corpus consultation is defined as searching in a corpus database and analyzing the concordance lines to elicit the correct usage of words or collocations.

The majority of the studies that examined the use of concordancing as a reference resource had trained learners and assigned them some tasks in a classroom setting. They consulted with corpora to correct errors in their written tasks, revise their writing according to teachers’ feedback, or correct their errors independently (Frankenberg-Garcia, 2005; Sullivan, 2007; Gilmore, 2009; Kennedy & Miceli, 2010). The mentioned studies examined the results of the learners’ use of corpora, while they were doing limited writing tasks in language or translation classroom.

However, several studies (Alharbi, 2012; Park & Kinginger, 2010; Yoon, 2008; Yoon, 2016) investigated the way consulting with corpus had an effect on students’ writing. These studies tracked participants’ independent corpus consultation by employing search logs or screen recording methods over specific time as students were composing their writing tasks. Moreover, the results of previous studies revealed that concordancing served as a useful tool by providing writers with the instances of language use concerning lexicogrammatical patterns and frequency information. However, the success of non-native English language writers in achieving appropriate results varied. The success was determined by several important factors such as language proficiency, learning style, and the nature of the task.
Learner concordancing as an instructional tool has been theoretically related to data-driven learning (DDL) (Johns, 1991), or discovery learning (Bernardini, 2004). The underlying theoretical foundation of DDL and the cognitive tool is derived from social constructivism and distributed cognition. Jonassen (1992), defined cognitive tools as technologies supporting cognitive processes, or those assisting learners’ engagement in higher order thinking. Cognitive tools also referred to applications that help learners generate and test hypotheses in problem-solving context.

In addition to concordancers, Google as a concordance have been considered as one of the highly promising areas to revolutionize language pedagogy and second language writing. Several researchers suggest that in Google-assisted language learning, Internet search engines serve as the concordancing tools (Acar, 201; Conroy, 2010; Panah et al., 2013; Shei, 2008). Moreover, Fletcher (2011), argued that the compiled corpora is supplemented by Web, since the internet provides freshness and spontaneity, scope, linguistic diversity, free access to data. He describes a clear framework for using information on the Web in three simple approaches: ‘hunting’, or directly query for particular information, ‘grazing’, utilizing ready-made data sets and ‘browsing’, or finding valuable information by chance. To relate the contribution of corpus tools to writing, the researcher highlights the role of academic writing in higher education context. The skill of academic writing plays a major role in students’ academic success and career. Nevertheless, academic writing for non-native speakers (NNSs) is cognitively demanding, and students might not produce the target language in a native-like way. EFL students have problems with use of proper collocations, word choice and interference of their mother tongue (L1) (Bloch, 2009; Paquot & Granger, 2012).

Many studies have reported that after having studied English for several years, non-native students still have experienced lots of difficulties in their writings (Hinkel, 2002; Silva & Silva, 2009; Yoon, 2005, 2008). Due to the effect of process-oriented writing pedagogy, the emphasis has been shifted to idea-development and content in academic writing. The majority of non-native speaker students especially those from EFL context struggle with writing in terms of grammar, lexico-grammatical patterns and collocation. Therefore, EFL writers require support concerning language features, namely, appropriate use of vocabulary, collocation and grammar.

To scaffold university students in collocation, grammar, word choice and sentence level errors, some researchers introduce corpus concordancing (Park, 2010; Yoon, 2008; Yoon, 2016). The findings of relevant studies revealed that consulting with corpus tools improved collocation and lexico-grammatical patterns of non-native students in their writing (Todd, 2001; Yoon & Hirvela, 2004; O’Sullivan & Chambers, 2006; Gilmore, 2009). However, little research has tracked the process of interaction with corpus consultation during scholarly writing to investigate the participant’s cognitive processes and whether the corpus consultation has a possible effect on students’ writing. The present study aimed to investigate to what extent recently emerging corpus tools in combination with traditional online resources assist a participant in solving lexical and lexico-grammatical problems in a university setting. Therefore, the following research question is addressed in this study:

To what extent corpus consultation assisted a participant in writing a scholarly article?

2. Methods

2.1 Participant

The focus of this study is to examine to what extent corpus tools assist an EFL postgraduate student to solve his linguistic problems during scholarly writing. Amin is a pseudonym chosen for an Iranian postgraduate student who is doing his Ph.D. in Industrial Engineering. Amin was enthusiastic about improving accuracy and appropriateness of his writing; therefore, he volunteered to become participants of this study.

2.2 Research Design

This study uses a qualitative case study approach to gain a more comprehensive understanding of the participant interaction with corpus tools. Case study method is implemented for being a robust research method, providing detailed and in-depth description of participants’ interaction with corpus tools.

2.3 Instrument

To scaffold the participant in writing, the researcher designed and developed an interface called Onlineconc. The researcher featured five concordancing resources namely Corpus of Contemporary American English, Google, JTW, Flax learning collocation, and Frazheit. Four kinds of dictionaries such as Ozdic, Bilingual and monolingual dictionaries, and Thesaurus featured in each tab on Onlineconc. The participant was required to initially register on the website to start using reference resources.
The first reason for designing this toolkit was to avoid opening different windows on students’ laptops. Another reason was that this website facilitated the process of examining participants’ query logs and tracking search patterns.

2.4 Sources of Data

2.4.1 Document Analysis

Amin was working on his research papers, and after completing his writing using reference resources, he submitted his writing for document analysis. The researcher used computer-generated query logs and screen-recording or real-time data of the writing process to analyze his writing in terms of linguistic aspects such as grammar, collocation and lexical bundles. Amin screen-recorded 200-minute of his writing process over one semester. Consequently, his writing was collected and coded for further analysis concerning linguistic functions. He uploaded his writing on Onlineconc website for document analysis.

2.4.2 Stimulated Recalls and Real-Time Data

Another source of data was stimulated recall. In the field of second language research, stimulated recall method is a useful tool to uncover cognitive processes that might not be evident through simple observation. In this study, the participant’s computer screen was captured during corpus consultation. This type of data is called real-time data. The researcher conducted stimulated recall instantly after screen-recording. The researcher asked the participant some questions to uncover his cognitive processes while consulting with language references and composing his research paper.

2.4.3 Query Log

To trace the participant’s look-ups and the processes in which he interacted with the online concordance website, every query the participant performed in his search box was saved in the server in the form of query log. For each query, the query log displays the following information: (a) reference consulted (b) query (c) date and time. Figure 2 shows an instance of query log on Onlineconc.
2.5 Procedure

In the first session, the researcher asked the participant to take a diagnostic test including ten items to increase his awareness in recognizing collocation, colligation and lexico-grammatical structures. Afterwards, the researcher illustrated and explained the concepts of collocation, colligation, and formulaic phrases using power point presentation and hands-on practice on how to use each corpus tool in detail. In the following session, the participant did hands-on practice to use each resource individually. While using corpus tools, Amin recorded his activities using a screen recording program which was installed on his computer in the second session. Then, the researcher asked participant to use corpus tools during writing his research paper outside university campus and screen-recorded his laptop. Then the researcher conducted stimulated recall sessions with the participant after corpus consultation, and he revealed his intentions for searching each queried word.

This study employed different sources of data such as query logs, screen-recording, stimulated recalls and document analysis of student’s writing. Therefore, different sources of data were triangulated to have a more comprehensive understanding of the participant’s corpus tools consultation processes while writing a scholarly paper.

2.6 Data Analysis

To find an answer to the research question regarding to what extent the use of corpus tools assist the participant in solving his problems, the researcher developed a coding scheme based on analyzing the participants’ writing, stimulated recalls and query logs. Each interaction with corpus tools was coded on the following three dimensions: (a) whether the consultation led to correct text formulation; (b) whether the participant found an incorrect solution to the given problem and (c) whether he abandoned the consultation. Each interaction between participant and corpus was given a set of codes, which consists of ‘P’ (positive) ‘N’ (negative) and ‘NE’ for no effect on writing (See Table 1).

Table 1. Coding scheme for effect of *Onlineconc* on writing

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive effect</td>
<td>For a given problem, the participant finds a correct and appropriate solution and applies it to the writing.</td>
<td>P</td>
</tr>
<tr>
<td>Negative effect</td>
<td>The solution the participant applies to a given problem is semantically, or syntactically incorrect, or stylistically inappropriate in the given context.</td>
<td>N</td>
</tr>
<tr>
<td>No-effect</td>
<td>The participant search the results and gives up performing further queries and did not make any changes to the text</td>
<td>NE</td>
</tr>
</tbody>
</table>

3. Results

In this study, the extent to which corpus consultations led to successful text formulation and accuracy of the participant’s writing is examined. Table 2 illustrates that (79.1%) of the problem-solving instances had positive
effect on participants’ writing and revisions, while nine problem-solving instances affected the quality of writing negatively (5%). These cases can be related to various factors such as formulating the wrong anchor word, returning wrong solutions from concordance lines or applying the wrong solution inappropriately to his writing. For 15.4% of the problem instances, consultation with reference resources did not make any changes in the participant’s writing.

Table 2. Results of Amin’s interaction with corpus tools

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive effects</td>
<td>604</td>
<td>(91%)</td>
</tr>
<tr>
<td>Negative effects</td>
<td>8</td>
<td>0.01%</td>
</tr>
<tr>
<td>Non-effects</td>
<td>46</td>
<td>0.06%</td>
</tr>
<tr>
<td>Sum</td>
<td>658</td>
<td></td>
</tr>
</tbody>
</table>

Amin frequently referred to corpus tools for finding appropriate lexical and lexico-grammatical queries (89.2%). It was followed by confirmation, pattern hunting, and synonym purposes. A total of 604 (91%) instances were found as evidence to show that participant had benefited from using language reference resources. He was more aware of the linguistic aspects of his writing; therefore, he encountered more potential problems and performed more consultations with concordancers. In what follows, some instances of Amin’s interactions with the corpus tools are elaborated to uncover the process of corpus consultation. As can be seen in Table 3, the data revealed that the participant’s interaction with the online reference resources positively influenced his language production in his scholarly writing.

Table 3. Amin’s Instances of interaction with corpus tools

<table>
<thead>
<tr>
<th>Query</th>
<th>Reference resources</th>
<th>Interactions with corpus tools in Stimulated recall session</th>
</tr>
</thead>
<tbody>
<tr>
<td>In regard to 'information'</td>
<td>COCA</td>
<td>I wanted to know ‘in regards to’ is correct or ”with regard to”, so I looked up in COCA. Both phrases are used. To make sure I searched it on Ozdic tool as well. I looked up on Google, both phrases were correct, but it is recommended to use ‘with respect to’ since is more standard than ‘in regard to.’</td>
</tr>
<tr>
<td>Finding verb for the noun ‘development’</td>
<td>Ozdic, Jtw</td>
<td>I looked up in JTW and Ozdic to find proper verb for ‘development’. I found that ‘occur’ or ‘take place’ can be suitable.</td>
</tr>
<tr>
<td>Finding how to generate a sentence with “not only but also’</td>
<td>Google scholar</td>
<td>I tried different ways to generate a proper sentence, finally I found one pattern in Google Scholar, so I decided to use ‘not only, but also.’</td>
</tr>
<tr>
<td>The proper preposition for ‘begin’</td>
<td>Frazeit</td>
<td>I wanted to know whether ‘with’ is a suitable preposition for begin, I found many results related to ‘begin with.’</td>
</tr>
<tr>
<td>It is ‘worth to mention’ or ‘worth mentioning’</td>
<td>Google</td>
<td>I was very confused with the use of ”It's worth&quot; whether a gerund or infinitive follow it. I looked up in Google, I found that It’s worth mentioning is used by native speakers and worth to mention is used by the non-native speaker. Therefore, I understood that the correct usage is worth mentioning.</td>
</tr>
</tbody>
</table>
In what follows, the parts of the text which are underlined present Amin’s instances of correct text formulation in his academic writing. The first query was related to usage of the lexical bundle; items 2, 3, 5 were associated with lexical choices. Other instances were related to grammatical and lexico-grammatical choices. All seven lexical queries led to generating the appropriate linguistic items in his writing. Figure 3 illustrates Amin’s stances of interaction with corpus tools in his research paper.

Although each collocation sounds grammatically correct, Amin aimed to choose the more appropriate collocation. As he evaluated the query results, he refused to use two adjectives of ‘further’ and ‘relevant.’ He chose ‘detailed’, as he thought that in the given context ‘detailed’ matched the formal academic context better. Therefore, his interaction with the corpus tools engaged him in multi-stage cognitive processes including making queries, evaluating the results, making a decision and applying the results to his writing.

Explaining the very nature of cognitive processes in corpus consultation is difficult; nevertheless, the lexical similarity in queries proposed a relationship between his use of corpus consultation and the language production in his paper. He mentioned that he found the pattern of ‘not only but also’ in query results, and he transferred the information obtained from the interaction and generated a novel sentence without consulting corpus tools for the second time. Amin re-used the linguistic items and was able to detect collocation patterns. He did not only restate un-analyzed chunks, but also he embedded the proper collocation in different grammatical contexts.

Concerning the types of linguistic problems addressed with corpus tools, the descriptive statistics revealed that Amin was aware of the applicability of each resource and consulted with them for distinctive purposes. He consulted with bilingual online dictionary for finding an equivalent, COCA, JT W, Flax, and Ozdic for collocation and Google for confirmation of word and phrase usage and monolingual online dictionary for intended meaning. The participant showed a tendency to make lexical queries in the dictionary-type resources while he consulted with the concordancer-type resources more for collocation and stylistic matters.

4. Discussion

A total of 604 (91%) instances were found as evidence to show that the participant had benefited from online corpus consultation resources. The corpus consultation was shown to be the most successful for checking simple grammatical points, collocations and finding proper synonym or antonym. The results are in agreement with several previous studies which reported using corpus helped learners not only to solve their problems in their writing but also enhanced their language awareness (Chambers & O’Sullivan, 2004; Gaskell,2004; Kennedy & Miceli, 2001). In the same venue, the result of this study is in line with Park (2012) who examined the processes in which participants interacted with a corpus, and they developed their language competence through consulting with the corpus.

Henceforth, the results of the current study supported those of Alharbi’s (2012) who investigated Saudi Arabian students to find out how they made use of concordancing together with online dictionaries to improve their second language writing. The results further confirmed that consultation with corpus helped participants to overcome language-related problems through assessing and modifying of the concordance lines of their search outcome. The research findings emphasized that not only learners gained advantages from interacting with corpus, in terms of improving textual performance, but also improved the awareness of grammatical, lexical, and lexico-grammatical choices.

The results of this study are in agreement with previous studies (Gilmore, 2008; Kennedy & Miceli, 2010; O’Sullivan & Chambers, 2006; Park, 2010; Yoon, 2008) which they empirically examined corpus consultation and concordancing as a language reference resource for improving English-as-second- language writing,
especially in university context.

5. Conclusion

In the process of academic writing, even advanced learners require support regarding complexities in language aspects. This lack of support for lexical, grammatical and other surface-level problems can be the primary source of the hindrance for EFL academic writers at universities. Under such circumstances, the inclusion of language reference tools such as concordancing tools and online dictionaries can improve students’ ability to proofread and edit the surface levels of their writing. Therefore, corpus tools enable learners to build up their confidence in writing by checking their hypotheses and going beyond their current linguistic repertoire. The results further indicated that concordancers and other reference resources were useful cognitive tools for solving linguistic problems during scholarly writing. The multiple resources invigorated them to switch between different tools to solve different lexical and lexico-grammatical problems. However, it is worth mentioning that students need training and hands-on practice in concordancing.

Acknowledgment

This paper is based on a segment of a research project funded by the Graduate Putra Grant of Universiti Putra Malaysia. The researchers wish to acknowledge, and are grateful for the support provided by UPM for the research.

References


Yoon, H. (2005). An investigation of students’ experiences with corpus technology in second language academic writing. The Ohio State University, Columbus, OH.


Copyrights
Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).