How Can Search Engines Improve Your Writing?

Adam Acar (adam.acar@gmail.com)
Kobe City University of Foreign Studies, Japan

Joe Geluso (joe.geluso@gmail.com)
Kanda University of International Studies, Japan

Tadaki Shiki (airsheeks@orange.plala.or.jp)
University of Marketing and Distribution Sciences, Japan

Abstract
The basis of the study reported in this article sprang from the assumption that incorporating the use of web search engines such as Google or Yahoo into English as a second language (ESL) writing may be useful for learners in identifying grammatical errors in their texts. The study has measured the effectiveness on reducing the number of grammatical errors by way of a simple method which divides long sentences into smaller word blocks and searches them within quotation marks on the Internet. The results suggest that using the quotation marks search method enables beginning ESL writers to notice and improve grammatical mistakes or unnatural expressions easily by checking the number of results each search query generates and choosing the word block with the highest number.

INTRODUCTION

Based on previous studies, research on writing strategy use seems to follow a certain trend in which importance has been placed upon communicative writing activities with the aim of increasing students' motivation for English composition whilst outright teacher correction of surface errors is considered less significant during the activity (Fregeau, 1999; Massi, 2001; Jarvis, 2002). Fregeau (1999) reported that a method using interactive sessions composed of dialogue journaling, peer reading, clarification questions, idea revisions and instructor/student conferences would be more effective than correcting every mistake in writing for the purpose of motivating students to develop a positive attitude toward it. Writing should also serve as a tool for achieving interactive missions set out for communicative purposes with a focus on the use of real communicative writing activities (Massi, 2001). In a lesson suggested by Jarvis (2002), a writing activity modeled after Process Writing was recommended. In this activity, students are engaged in speaking, reading and writing interactively while the development of their literacy is being promoted through developing a class story, writing diaries and individual conference interviews. Its aim is to build students' confidence in their abilities; therefore, it is more important to let them have fun writing by accepting mistakes, including surface errors. The above frameworks for writing help to promote classroom interaction by giving students opportunities to discover how the target language is actually put to use through various language activities.

LITERATURE REVIEW

Prominence is given to the ways new electronic technology is utilized as an innovative tool to meet interactive and collaborative needs. Computer technologies including the Internet, email, weblogs and so on seem to be well integrated into writing classes, serving as an aid in creating an interactive atmosphere in which students can keep their motivation high. For example, the benefits of using email in an ESL writing class are highlighted by Belisle (1996). In Belisle's study, students worked together with other classmates, peers and teachers via email, and through praxes they began to realize their potential in a collaborative working environment, transforming into more active learners. Mansor (2007) also discusses how email can be implemented in a collaborative learning atmosphere. It is assumed that information technology has an impact on English writing such that students' motivation is so high that usage and proficiency can be increased. Moreover, student-centered language learning is encouraged when topics have links to outside the classroom (McDonald, 2009). Web-based writing was also suggested as an effective strategy by Trokeloshvili and Jost (1997). The method is based upon the idea of displaying student text publicly using the Internet, thus motivating students to put out attractive work because the Internet connects them with an authentic audience.

The studies outlined above offer an account of ESL writing from the viewpoint of promoting collaborative activities in an interactive way to boost students' motivation. One problem that remains, however, is that misgivings have been expressed about effectiveness when students lack linguistic skills in judging writing (Jacobs, 1989). More importantly, another practical aspect of computer-oriented English writing that deserves more attention is the utilization of technology as a tool to train students to broaden their lexical knowledge in order to facilitate their writing. Access to an entire body of information using the Internet may provide unique opportunities to aid in developing writing skills (Kenworthy, 2004). Krajka (2000) explains that materials available on the Internet belong to the outside world; therefore, they are authentic in contrast with traditional textbooks. One of the examples shown in his study includes a writing activity in which students prepared a letter, comparing theirs with sample letters available online. In addition, through email writing, a student replied to the teacher by making full use of received messages as a guideline for writing and combining appropriate phrases picked up from those messages to compose new sentences (Trenchs, 1996). Johnson (2004) also draws our attention to student accessibility of various sources of information in a computer-oriented classroom. He mentions that the advantage of utilizing the Internet would be to look up words they do not know on an online dictionary. A more advanced technique would be to utilize corpora for composition since they can provide an abundance of written sources. Corpora, databases of written or spoken texts collected from authentic materials, can be an effective tool to help students gain lexical knowledge and enhance the quality of vocabulary (Chen, 2004).

Many corpora, The Corpus of Contemporary American English (COCA), for example, also have concordancers built into them. A concordance is defined as "a collection of the occurrence of a word-form, each in its own textual environment" (Sinclair, 1991, p. 32). A concordancer, then, allows language learners to view natural occurrences of words in formulaic sequences as they appear in a corpus. When considering the abundance of formulaic language (e.g., collocations, phrasal verbs, idioms, etc.) in everyday speech among native speakers of English (Ellis, 2002; Ellis, Simpson-Vlach, & Maynard, 2008), the usefulness of corpora and concordancers to English language learners becomes readily apparent. In turn, the fields of applied linguistics and second

language acquisition have enjoyed a significant amount of research in recent years examining corpora and concordancers as tools to facilitate second language (L2) learning, particularly in the area of writing (Chen & Baker, 2010; Kennedy & Miceli, 2010; Yoon, 2008; Yoon & Hirvela, 2004).

Most studies on corpora and concordancers focus on either the use of corpora specific to certain institutions, or the use of more publicly accessible corpora, such as the British National Corpus or the Collins COBUILD Corpus (Yoon, 2008). Recently, however, there has been developing interest in using the web as a corpus and concordancer (Wu, Franken, & Witten, 2009). Wu et al. (2009) note the attractiveness of using the web as a corpus due to its sheer size. Additionally, the web is readily available to students and many are already experienced users. To utilize the web as a corpus, one must use a search engine such as AltaVista, Yahoo, or Google. Among search engines, Google appears to offer the most results for any given search, as it accesses over 1 trillion words of text and continues to grow (Sha, 2010).

Some research on the use of Google as a corpus and concordancer has already been done. Shei (2008), for example, developed a method of using Google to identify formulaic sequences in texts on the web. Shei discovered that one can gauge how tightly bound a formulaic sequence is by comparing the number of hits Google generates each time an additional lexical unit is added to a sequence. Shei used the example *have been found to be....* The first search of *have* generated some 3 billion hits; *have been* generated some 1.8 billion; *have been found* some 12 million. With each additional word added to the phrase, the number of hits falls. However, the frequency line, the function of the number of words in a phrase in relation to the number of results generated, remains relatively stable. If you were to plot out the number of hits on a graph as each lexical unit is added to a formulaic sequence, the resulting line would have a fairly linear decline, that is, no extremely sharp drop offs. Shei reasons that as long as the frequency line remains relatively linear, "the overall fragment may be considered formulaic" (p. 71). Shei gives an example comparing two phrases: *have been found to be infected,* and, *have been found to be contaminated.* These phrases generated 15,200 and 22,400 hits, respectively, compared to *have been found to be polluted* which generated only 1,140 hits. Following Shei's reasoning, one could deduce that either of the first two phrases is more tightly bound than *have been found to be polluted.*

Wu, Franken and Witten (2009) note potential pitfalls when using the web as a corpus. For one, most of what is on the web is not subjected to any editorial process, hence ungrammatical language will most likely be encountered at some point. To counter these perceived pitfalls, Wu, Franken and Witten (2009) used an off-line static corpus supplied by Google that had been subjected to an editorial process to clean up any 'dirty results'. Still, Kenworthy (2004) notes that access to an entire body of information using the Internet may provide unique opportunities to aid in developing writing skills. Krajka (2000) explains that materials available on the Internet belong to the outside world; therefore, they are authentic in contrast with traditional textbooks. One of the examples shown in his study includes a writing activity in which students prepare a letter comparing theirs with sample letters available online. In addition, through email writing, a student replied to the teacher by making full use of received messages as a guideline for writing and combining appropriate phrases picked up from those messages to compose new sentences (Trenchs, 1996).

While there are studies that examine the use of web-derived corpora or specific methodologies of gleaning data from web searches to inform writing (i.e., Wu, Franken and Witten (2009); and Shei (2008), respectively), there are no studies that we know of that simply analyze the effect of using Google searches in their raw form on students' writing. This study attempts to fill that gap in the literature.

METHODOLOGY

Procedure
During the first step of the data collection, participants were gathered in the computer lab and asked to perform 4 different writing activities in 40 minutes. The activities included introducing family members, explaining what is happening in a picture, describing the most memorable moment in their life, and translating a recipe from Japanese to English. These four writing activities were specifically chosen since they were commonly used in writing exams. After completing the writing tasks, subjects submitted their writings via an online survey web site.

Upon submission of the writings, subjects were taken to a classroom and given a 20-minute long training session on how to use Google's "quotation mark" search feature to detect writing mistakes. Subjects were told that they should check all of their sentences by entering in Google's search box as word blocks (four words) in quotation marks and keep in mind that fewer than 100 Google hits for any word block means a potential mistake within the block. Additionally, subjects were instructed to pay special attention to prepositions and articles and search multiple times for the same word block by replacing prepositions and articles with their alternatives (See Appendix II). Subjects were then given one week to correct their mistakes on their own by using the Google quotation mark search method. All of the subjects submitted revised writings within a week.

Subjects, data collection
A total of eight subjects, who were students at the engineering department of a major public university in Southeast Japan at the time, participated in this study. All of the participants reported having a formal English language education of at least six years. The data was collected during the 2nd week of November, 2009 and there was no incentive (extra-credit or other forms of compensation) to participate. Students completed this assignment as a part of their homework.

Coding
There were only two types of improvements made use of in the coding, namely "writing clarity improvement" (phrase or sentence sounds more natural compared with the original) and "grammatical improvement" (phrase or sentence has fewer or less significant grammatical errors), since there was not a commonly accepted form of coding writing accuracy in the literature. After reviewing more than 30 different error classification methods, Polio (1997) concludes that there are many discrepancies between linguists in terms of coding writing errors and error free units. Thus, it was decided that using this kind of simple coding can be easier to understand and replicate for other scholars in the discipline. The coding was performed solely by the main author of this study. (See Table 2 for a sample coding).

FINDINGS

After the comparison of pre-training and post-training sentences, it was found that students improved their writing dramatically. The average of improved sentences to total sentences ratio was about 24% for the 8 subjects. In other words, by using this simple method, students showed some sort of improvement in a quarter of all sentences they wrote. Table 1 illustrates number of sentences, improved sentences and improvement rate for each subject. Additionally, readers should note that the number of sentences remained the same in both pre-post texts: students were advised to not add or delete any sentences, but just make changes within their original sentences.

Table 1
Writing Improvement after Using the Google Quotation Mark Search Feature

Subject #	Total Sentences	Improved Sentences	Improvement (Total Sentences/Improved Sentences)
Subject 1	17	4	24%
Subject 2	19	4	21%
Subject 3	30	9	30%
Subject 4	35	11	31%
Subject 5	27	5	19%
Subject 6	41	10	24%
Subject 7	19	3	16%
Subject 8	27	7	26%
Average	26.88	6.63	24%

Readers should note that this study provides a quantitative assessment of writing improvement as a result of the Google quotation mark search method, but does not report any statistical test (e.g., Chi-square, z-scores, etc.). That is because we only compared the sentences in pre and post conditions and counted the number of cases when there was a "writing clarity" or "grammatical" improvement. In other words, we did not code errors in pre and post conditions separately which would have allowed us to run a Chi-Square test. As can be seen in the following Table 2, almost all of the post training sentences still contain serious errors but phrases in italics sound better or closer to natural English. If pre and post errors were coded separately there would be zero change in the results (reduction in the number of sentences with errors) which is not true.

Table 2
Coding Form, Subject #3

Before	After	Improvement Type
My most unforgettable moment is my last chorus concert.	*The most unforgettable moment* is my last chorus concert.	Writing Clarity

They are Â father,mother,elder brother, grandmother and me	*They are father, mother, older* brother, grandmother and me.	Grammatical Improvement
Secong, putting a pat of thin slice ginger and beef,and cooing them at high heat in 30 second	Second, *put a pat* of thin slice ginger and beef. Then, boil them at high heat in 30 seconds.	Writing Clarity
.Fourth,boiling down theã€broth in the pot.	Fourth, *boil down* the broth in the pot.	Writing Clarity
At the end,putting them on the bow,andã€adding something green.Â	At the end, *put them* on the bowl, and add something green. Itâ€™s ready.	Writing Clarity
My father has already retired the job and now he is fermerÂ	My father has already *retired from work* and now he is farmer.	Writing Clarity
My mother engaged part time job	My mother *takes on a part-time job*.	Writing Clarity
Althou she is 90, she is very buzzing	Although she is 90, *she is very fine*.	Writing Clarity
I have present for you	I *have a present* for you.	Grammatical Improvement

DISCUSSION AND CONCLUSION

The use of the Internet in ESL is gaining popularity in part because it can promote collaborative and interactive activities which motivate students. However, there is presently little research available in the area of using the Internet searches as a tool to help second language writers make decisions about their writing. Therefore, this applied study was undertaken in order to ascertain the effectiveness of the Google quotation mark search method with a simple experiment and to provide a practical way to effectively take advantage of unlimited sources of information through the use of free search engines.

This study has a number of limitations; therefore, the results should be interpreted very carefully and replicated before adopting this method. First of all, there's no inter-coder reliability score as there was only one coder--the main author of this article. Secondly, this study was solely conducted for practical purposes and does not test or build any theory. The operationalization and the measurement of "writing error" were totally arbitrary. The literature review served as a supplementary source and did not lead to any research question or hypothesis. Finally, we recognize that not all sources on the Internet are written by native English speakers or follow perfect grammar. It is possible that some linguistically accurate word blocks might still end up with fewer than 100 hits while nonstandard word blocks might generate millions of hits (e.g., Hip Hop song lyrics).

Nevertheless, this study has shown that using a search engine as a corpus can dramatically decrease the number of writing mistakes. It can also teach learners about correct usage of prepositions, articles and more. By analyzing the number of errors in student essays before and after using Google quotation mark search method, we found that many of the mistakes were deleted or corrected by the subjects on their own. Additionally, we observed that students not only fixed a considerable amount of grammatical writing errors but also had a chance to check if their phrases were natural (e.g. "making homework" might not be classified as a grammatical error but it surely is not a natural phrase in English). Indeed, revising what they wrote in comparison with written information available online may help learners become more effective writers as they go through the process of searching and scrutinizing language usage through authentic materials. Thus, finding and correcting mistakes through the use of web search engines can be considered both an effective solution for common writing problems, and a training tool for ESL learners to identify errors in their own writing.

Despite the fact that Google can be extremely helpful for ESL learners, as we found in our study, only a few academic studies have referred to it as a writing source. We believe this is mostly because, until recently, Google was never defined as a corpus (Sha, 2010). On the other hand, Google has a record of millions of books, magazines, newspapers, blogs and academic archives updated daily which in a sense makes it the world's largest corpus. Most importantly, Google has a very simple interface and available to everyone who has Internet access unlike other subscription based corpuses. We are confident that the Google quotation mark search tool can help second language writers detect unnatural and incorrect phrases easily. However, we would not go so far as to say using Google or other search engines improve learners' writing skills. Future studies should evaluate the effect of frequent online corpus usage on ESL learners' writing skills.

REFRERENCES

Belisle, R. (1996). E-mail activities in the ESL writing class. *The Internet TESL Journal.* Retrieved 29 December, 2010 from http://iteslj.org/Articles/Belisle-Email.html

Chen, Y. H. (2004). The use of corpora in the vocabulary classroom. *The Internet TESL Journal.* Retrieved 29 December, 2010 from http://iteslj.org/Techniques/Chen-Corpora.html

Chen, Y. H., & Baker, P. (2010) Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology, 14*(2), 30-49.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24,* 143-188.

Ellis, N.C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375-396.

Fregeau, L.A. (1997). Preparing ESL students for college writing: Two case studies. *The Internet TESL Journal.* Retrieved from http://iteslj.org/Articles/Fregeau-CollegeWriting.html

Jacobs, G. (1989). Miscorrection in peer feedback in writing class. *RELC Journal, 20*(4), 68-76.

Jarvis, D. J. (2002). The process writing method. *The Internet TESL Journal.* Retrieved from http://iteslj.org/Techniques/Jarvis-Writing.html

Johnson, A. (2004). Creating a writing course utilizing class and student blogs. *The Internet*

TESL Journal. Retrieved 29 December, 2010 from http://iteslj.org/Techniques/Johnson-Blogs/

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

Kenworthy, R. C. (2004). Developing writing skills in a foreign language via the internet. *The Internet TESL Journal.* Retrieved from http://iteslj.org/Techniques/Kenworth-WritingSkills.html

Krajka, J. (2000). Using the internet in ESL writing instruction. *The Internet TESL Journal.* Retrieved from http://iteslj.org/Techniques/Krajka-WritingUsingNet.html

Mansor, N. (2007). Collaborative learning via email discussion: Strategies for ESL writing classroom". *The Internet TESL Journal.* Retrieved from http://iteslj.org/Techniques/Mansor-EmailDiscussion/

Massi, M. P. (2001). Interactive writing in the EFL class: A repertoire of tasks. *The Internet TESL Journal.* Retrieved 29 December, 2010 from http://iteslj.org/Techniques/Massi-WritingTasks.html

McDonald, W. C. (2009). Using the e-mail précis for advanced learners: Teacher-student interactions. *The Internet TESL Journal*. Retrieved 29 December, 2010 from http://iteslj.org/Lessons/McDonald-EmailPrecis.html

Sha, G. (2010). Using Google as a super corpus to drive written language learning: a comparison with the British National Corpus. *Computer Assisted Language Learning, 23*(5), 377-393.

Shei, C. C. (2008). Discovering the hidden treasure on the Internet: using Google to uncover the veil of phraseology. *Computer Assisted Language Learning, 21*(1), 67-85.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford university Press.

Trenchs, M. (1996). Writing strategies in a second language: Three case studies of learners using electronic mail. *The Canadian Modern Language Review, 52*(3), 464-497.

Trokeloshvili, D. A., & Jost, N. H. (1997). The internet and foreign language instruction: Practice and discussion. *The Internet TESL Journal.* Retrieved 29 December, 2010 from http://iteslj.org/Articles/Trokeloshvili-Internet.html

Wu, S., Franken, M., & Witten, H. (2009). Refining the use of the web (and web search) as a language teaching and learning resource. *Computer Assisted Language Learning, 22*(3), 249-268.

Yoon, H. (2008). More than a linguistic reference: The influence of corpus technology on L2 academic Writing. *Language Learning and Technology, 12*(2), 31-48.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing, 13*(4), 257-283.

Appendix I
Google "Quotation Mark" Search Technique

Accuracy rule

Step 1: Group your sentences into 4-word blocks.

Eg. **I went to Bali / with my husband**

Block 1 "I went to Bali"

Block 2 "Bali with my husband"

Step 2: Search for each block on Google with quotation marks.

Results **1 - 10** of about **348,000** for **"I went to Bali"**

Results **1 - 10** of about **10,600** for **"Bali with my husband"**

Accuracy rule depending on the number of hits found

More than 1000 results: OK

Between 100-1000 results: Maybe wrong, check the examples on the first page

Less than 100 results: Probably wrong. Detect and change the mistake.

How to Correct mistakes?

Pay Attention to Articles!!!
-a. –the.

E.g., I like catching the fish

● 1- Search for the sentence with –the: **"I like catching the fish".** 10 results.

● 2- Search for the sentence with –a: **"I like catching a fish"**. 1 result

● 3- Search for the sentence without –a and –the: **" I like catching fish"**. **152,000 results**

Pay Attention to Articles!!!
-a. –the.

Eg. I watched from window of my house.

- 1- Search for the sentence with –the: **"from the window of my house". <u>12,000,000 results</u>**.

- 2- Search for the sentence with –a: **"from a window of my house"**. 272,000 results

- 3- Search for the sentence without –a and –the: **" from window of my house"**. 4 results

---

Pay attention to Prepositions
-in,-on, -at, -under, -…

Eg. There was a wound at my skin.

- Search with similar propositions. "was a wound **on** my skin". **<u>72,000 results</u>**.

- Search with similar propositions. "was a wound **at** my skin". 0 result.

- Search with similar propositions. "was a wound **in** my skin". 3 results.

Appendix II
Writing Task

1- Write about the most memorable moment in your life (maximum 10 minutes)
2- Briefly introduce your family members (maximum 10 minutes)
3- Describe what is happening in this cartoon (maximum 10 minutes)
4- Translate this recipe (in Japanese) to English (maximum 10 minutes)