

Independent research project with web-derived corpora for language learning

Yoko Hirata

Hokkai-Gakuen University

Yoshihiro Hirata

Hokkai-Gakuen University

Recent studies have indicated the enormous potential of the web as a source of linguistic data for language learning in the classroom. However, there are comparatively fewer reports on the successful application of corpora compiled from the web as a tool for EFL students to raise their awareness of various lexical phrases. The purpose of this paper is to investigate Japanese students' perceptions of inductive approaches to understanding word combinations and lexical limitations by using their own web-derived corpora and an online concordancer. By explaining the benefits and problems that the students experienced during the corpus-based analyses, this study also determined how the use of corpora can be applied to teaching English effectively.

Vocabulary acquisition is widely considered an important goal for Japanese learners in secondary and tertiary institutions. In recent years, because of their ease of use and usefulness, various electronic dictionaries have become essential reference sources for students (Cobb, 2003). A wide range of pop-up illustrations with sound (Tribble, 2003) and a user-friendly GUI, which is often equipped with gap-filling lexical exercises and games (Sobkowiak, 2002) are the major advantages for students to have these dictionaries (Koren, 1997). However, the biggest challenge that students face using these language resources is the fact that not enough highly specific examples and meaningful contexts are provided for students to improve their lexical competence (Cobb, 2003; Kaur & Hegelheimer, 2005). Students find it difficult to understand how the word is used in real-life situations and which words they usually occur with (Robinson, 1989). Despite the fact that a variety of word combinations, such as fixed expressions and collocations, have played a significant role in effective language teaching (Carter, 1987; Widdowson, 1989; Nattinger & DeCarrico, 1992),

traditional teaching methodologies are unlikely to provide students with ample opportunities to acquire these recurrent features of language use.

Research Background

Corpus-Based Data-Driven Language Learning

As computers have been employed as powerful tools to facilitate foreign language teaching, various applications of corpus and concordance have dramatically expanded the options for understanding lexical, grammatical and structural patterns of language. For example, recent studies on corpus linguistics and language teaching methodologies have shown that having students access various corpora and drawing their attention to concordance evidence provides them with a more objective view of language (Chambers, 2005; Gaskell & Cobb, 2004; Hunston, 2002; Vannestål & Lindquist, 2007). Dodd (1997) emphasizes the importance of using concordancers for language learning, as they allow students to pay closer attention to the target word in rich contexts and to gain an extensive knowledge of words in more effective ways. Stubb (1995) points out that concordance evidence in which the target word is included contains more reliable facts about frequency and typicality of words than lexical information based on the native speaker's intuition. Willis (2000) claims that it is important to create a "pedagogic corpus", which consists of written and spoken texts with transcriptions from a variety of sources, and to provide EFL students with "focus on language form" activities based on the findings of their language analysis. All of these teaching methodologies which draw from corpora and concordancers are based on "data-driven learning" (DDL), as defined by Johns (1991, p.2) who claims that "the language learner is also, essentially, a research worker."

Although these studies have displayed the potentially positive characteristics of corpus based learning, few instructors have actually used corpus data as a valuable resource for successful language learning (Szendeffy, 2005). Specifically, instructors teaching EFL students at lower intermediate proficiency levels have reported difficulty in encouraging students to make explicit generalizations about the target word (Aston, 1995; Wichmann et al., 1997). With large-scale corpora, these students have problems in making reasonable conclusions about the structural rules of the target word (Johns, 1986). This results from the fact that concordance evidence taken from such data-driven learning does not provide situational and communicative contexts for examples (Cheng et al., 2003; Sinclair, 1997). Another problem for these students to conduct corpus analyses might be the lack of user-friendly analytical tools for novice corpus users to provide them with better insight into language use. Many of the concordancers which are currently available have not been especially designed for retrieving lexical and syntactic patterns from the texts and going back to the original texts later. It is indispensable to develop effective approaches in order for even inexperienced EFL students to become more self-reliant and confident in understanding the relationship between the meaning of a word and the context in which it is presented.

The Web as a Corpus

As the World Wide Web becomes an enormous language database at students' fingertips, the potential benefits and advantages of web-extracted data for linguistic research and lan-

guage learning have recently been discussed. For example, a growing number of web-based corpora, such as WebCorp (Kehoe & Renouf, 2002) and KWICFinder (Fletcher, 2007), have been readily available free of charge for linguistic research. Todd (2001) claims that it is effective for students to create their own small concordances from the Internet and to understand various lexical items. Ghani et al. (2001) suggest the successful application of a commercial web search engine for locating pages relevant to the target word and retrieving collections of written texts as a corpus file. Kilgariff and Grefenstette (2003) go further in claiming that web-derived data serve as an important substitute for existing authentic corpora.

For Japanese EFL students who do not have much opportunity to be exposed to vast amounts of authentic English information in everyday contexts, the web data could be valuable information that is easily available. Despite the rapid growth of web-based corpora, there is still a general lack of research concerning how the web data can best be presented to EFL students at all different levels of proficiency. In addition, little research has been done so far concerning how students, who lack previous corpus learning experience, can be trained to discover what works and what does not work for their own language learning using the web as a source of language and lexical information.

The purpose of this study is to investigate how Japanese EFL students at intermediate levels of English proficiency perceive the effectiveness of setting up and exploring web-derived corpora, and whether they gain a good understanding of the collocational and recurrent lexical patterns of target words. Particularly, the benefits and problems students experience during the web-based corpus consultation are examined. The focus of this project was on encouraging individual students to compile their own corpus from the web and to observe how words behave in tangible examples of actual usage with limited assistance from the instructor.

The Case Study

Student Profiles

The subjects of this study ($N=69$) were lower intermediate learners of English enrolled in a university reading program. They had been familiar with the traditional textbook-based, teacher-directed language learning approach taken in a large lecture-type classroom where the teacher tended to adopt a predominantly by rote mode of instruction (Cooker & Torpey, 2004). The students had neither experience in online independent language learning nor a corpus based analysis in or beyond the conventional classroom, and 7.2% of the students had no previous experience in their daily lives of using commercial web search engines even in their native language. Almost all of the students put vocabulary acquisition in a central role in improving reading skills and, therefore, expressed great interest in improving their vocabulary skills. All of the students had studied English for six years primarily in the classroom.

The Setting and Procedures

The project, which was based on the inductive DDL activities (Mishan, 2004), was implemented in a half-term English reading course that the students were required to take as

one of their compulsory classes. The reading course was designed to mainly focus on the understanding of a variety of texts, sentence structures, and paragraph organization. In order to improve students' reading competence, the course also aimed to develop students' lexical skills, including the teaching of collocations and phrasal verbs, as well as the explanation of word structures, such as prefixes and suffixes. The classes were all scheduled for one and a half hours each week in a computer assisted language learning (CALL) classroom and students were given four weeks to complete the project. Before the project started, there was a period of guidance which was offered by the instructor for the purpose of making students understand the basic concepts and techniques for corpus studies. Also, sample activities were given to the students about how to understand various language forms from concordance lines. The project was divided into the following two stages: a. Compilation of a Corpus from the Web and b. Analysis of Concordance Evidence.

Compilation of a Corpus from the Web

The initial stage of this project aimed at building students' own database of web documents with the use of commercial web search engines for their own individual linguistic investigations. First of all, before the compilation of the database, students worked in pairs and decided what words to search for. They identified some target words that they didn't fully understand the meaning and usages from the previous reading lessons. Each pair of students was required to select as many as six target words.

Among the many commercial search engines, Google (<http://www.google.co.jp/>) was chosen for this project, since it did not require expertise or experience. Students entered a target word as a search term into Google and identified the search results on the Google return page. By reading a variety of electronic texts on different subjects using skimming and scanning techniques, the students tried to understand the various meanings and lexical patterns of the key word.

At the beginning, without any control over what kinds of texts should be in the corpus, students browsed web pages of various genres such as diaries, blogs and news. The students were then asked to open up a prepared text document and to assemble the linguistic data by dragging the web page into the document. After they became accustomed to searching using Google and understood which subject directory they should use, they were encouraged to use other web search engines, such as Yahoo! (<http://yahoo.com>), which took them to other particular genres or categories. By the end of this stage, a set of web data containing the key words had been retrieved and their own corpus was constructed.

Analysis of Concordance Evidence

Once a corpus was created from the web, the students independently consulted various lexical patterns by using a concordance program called **Lex**, the algorithm of which was developed by the co-author. Its implementation as a web application was carried out by a volunteer who had wide experience in constructing server-side JAVA applications. **Lex** was installed in an Apache web server with Tomcat on a Linux computer. The system structure of **Lex** is outlined as follows.

Lex

Lex is a user-friendly concordancer for retrieving and displaying lexical combinations from any kind of text data. This program performs the simple function of searching and extracting all the occurrences of a certain key word or phrase in a corpus, in order to find collocational and lexical patterns which are associated with the key word. The search results can be displayed in Key Word In Context (KWIC) mode in a plain text document. Key words are displayed with approximately six words on either side. This program, with a simplified easy-to-use interface, has been designed specifically for students without any corpus investigation experience. Its access is restricted to registered students. Figures 1 and 2 demonstrate two different interfaces of **Lex**. The letters and the numbers on the left-hand column provide the origin of each concordance line.

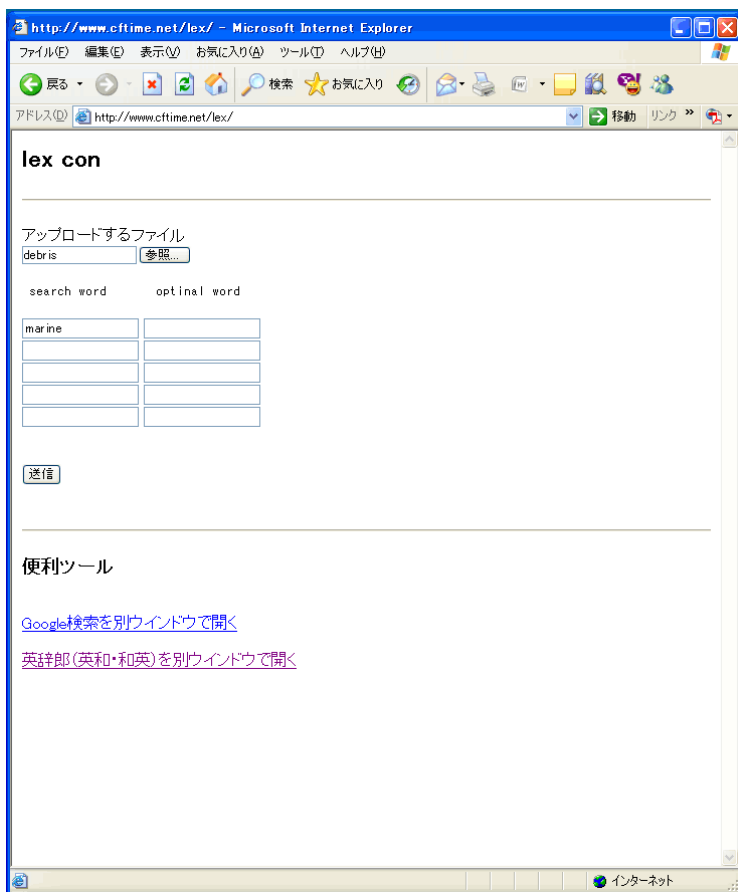


Figure 1. Interface design of Lex

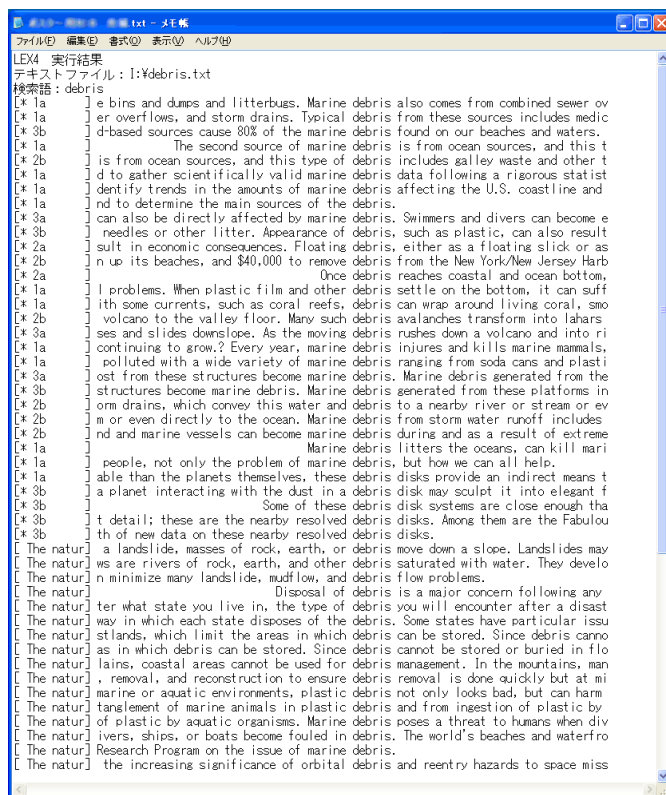


Figure 2. Search results by Lex

Based on the research questions listed below, the students were encouraged to seek out the data for themselves and classify their search results to look for specific word combinations or usages.

1. Did you identify what type of context is associated with the key words?
2. Did you notice what kinds of words are used with the key words?
3. Did you see what phrases and typical collocations are included with the key words?
4. Did you identify other language features that are associated with certain kinds of texts?

The students were instructed to organize their findings by focusing on various language features such as the use of prepositions, verbs and pronouns, and to examine collocations and other lexical combinations.

Homemade

- homemade ice cream 手作りアイスクリーム
- homemade soap 手作り石鹸
- homemade gift wrapping 自分でする贈り物の包装
- homemade yacht 手製のヨット
- homemade air conditioner 手製のエアコン
- homemade water heater 自家製湯沸かし器
- homemade tattoo 自分でした入れ墨
- homemade puzzle 自分が作ったなぞなぞ

Collocations

- verbs: make, try, enjoy
- adjectives: natural, delicious, simple
- nouns: ice cream, food, family, recipe, soap, gift wrapping, tool, yacht,
- air conditioner, water heater, tattoo, puzzle

Debris

- noun: 破片、瓦礫、残骸 がらくた、ごみ (scattered pieces of rubbish or remains)
- space debris 地球の周りにあるロケットや衛星の破片
- marine debris 海で浮遊しているごみ
- debris flow 土石流

Collocations

- verbs: create, pollute, affect
- adjectives: orbital, plastic, man-made
- nouns: space, marine, flow, research, issue, environment, disk, population, mitigation

Figure 3. Examples of student's work

During this stage, the students compared their findings from concordance output (see Figure 3) with the correspondent entries and their examples in dictionaries. The students were also encouraged to identify the differences between closely related words, for instance, "debris", "refuse" and "garbage" which were all included in the previous reading lessons, and to give examples to show how they are used differently. After students completed their project, they submitted the list of their findings as an assignment.

Data Collection and Analysis

At the end of the project, all the students were asked to reflect on the projects by completing a questionnaire (see APPENDIX) concerning their attitudes and perceptions of the project. The questions mainly focused on the benefits and problems of implementing the project. Data collected from the questionnaire including open-ended questions which were transcribed and analyzed. The rating scale used in the questionnaire was a 10-point Likert scale with 1 representing *strongly disagree* and 10 representing *strongly agree*. In order for students to fully understand the questions, the questionnaires were written in Japanese. For the purpose of attaining a mean response for each question, the responses were totaled and averaged. Standard deviation was then obtained for the purpose of examining statistically significant differences between students' responses. The questionnaire was also analyzed using Spearman's correlation to determine correlations between responses and significant factors underlying their responses. Correlation is significant at the .01 level (2-tailed). The students' comments were also utilized to investigate their attitudes toward the implementation of the project.

Findings

The assignments the students submitted showed that they had successfully discovered common lexical patterns that were associated with the key words. While 59.4 % of the students were using commercial web search engines, such as Google and Yahoo, everyday or almost everyday, 7.2% of the students had never used any of them before. As for electronic dictionaries, although many students are satisfied with their own dictionaries ($M = 6.55$ $SD = 1.73$), almost all the students stated that dictionaries do not provide enough highly concrete examples in meaningful contexts. They claimed that this problem could lead to an inappropriate use of words in many different situations.

The results of the questionnaire revealed the students' various perceptions of the present project. Overall, the students reacted very positively towards the outcome of the project and the majority felt that what they had accomplished was very productive and satisfying. The mean ($\pm SD$) of this response was 6.06 (± 1.90). The results also indicated that the students thought the project helped them understand even small nuances of the key words and their various usages that English dictionaries never offered. The mean ($\pm SD$) of this response was 7.35 (± 1.89). In conjunction with using dictionaries and examining the key words with their own web-derived corpus, the students found they were able to understand the meanings and usages of the key words more clearly than ever before. The average ($\pm SD$) of this response was 7.81 (± 1.92). The students were able to examine various example sentences containing the key word in a very effective way. The average of this response was 6.33 (± 2.31).

As shown in Table 1, there was a moderate correlation between the effectiveness of understanding even the small differences in the meanings of the key words and the validity of using the web-derived corpus in combination with dictionaries ($r = .49$ $p < .01$). Also, the correlation of .64 ($p < .01$) is significant because it illustrates relationship between the effectiveness and the students' high appreciation of the project. The students found the project an interesting way of looking at actual language use. Following the table are some comments from students' responses to the questionnaire.

Table 1. Correlation Between Factors for Effectiveness of Project

	Understanding of the meaning of the key words	Effectiveness of the web-derived corpus with dictionaries	High appreciation of the project
Understanding of the meanings of the key words	—		
Effectiveness of the web-derived corpus with dictionaries	.49**	—	
High appreciation of the project	.49**	.64**	—

Notes: Correlation Matrix (N = 69), **p < .01.

Lex is a useful tool for improving vocabulary skills. It made it easier for me to identify and understand some groups of words that are related to the key word within a context. I've found several examples of language features which are not described in traditional dictionaries.

It took several hours until I finally decided on the key words for my project. Once I started using Lex, I easily discovered common lexical patterns which are associated with the key words. I realized that the key words are used in many different contexts.

Although most of the web pages were familiar to me as I scanned them, I was able to understand the meaning of the key word in context. I've found that the meaning of the key word can be clearly classified by paying attention to the other surrounding words.

Although the dictionaries provided only limited example in which the key word was included, I found this project useful in that I could understand how a key word and its associated words were used in many different contexts.

By the end of the project, I think I acquired some skill in identifying the difference between the most and least useful information I needed for my analysis in a wide range of different text types.

In addition, many students stated that, as the term progressed, they acquired the skills to infer various meanings of the key words from the web-derived texts, even though the texts contained unfamiliar subjects and topics for students. The mean ($\pm$ SD) of this response was 6.48 ($\pm$ 2.2). There was a moderate correlation between high appraisal of the project

and the effectiveness of inferring the meanings from the contexts ($r = .49, p < .01$). One of the students, for example, reported as follows:

I didn't need to understand every word in order to extract the necessary information. I found that the online reading differed from the traditional paper-based reading. It was not easy for me to get used to this difference.

Also, some students needed the improvement of the system structure of **Lex**. The average ($\pm SD$) of this response was 6.2 (± 2.05). It is important to note that those, who had rated the project highly, expected an improvement in the system structure of **Lex**. There was a weak correlation between high appraisal of the project and requests for the development of **Lex** ($r = .34, p < .01$). One of the students made the following comment.

"The key words that I selected for this project had relatively low frequency of occurrence and therefore downloading web pages to build the corpus took me more than a week. I think the instructor should improve the system to sort out this problem."

It was speculated that there was a close relationship between students' previous online experiences and the benefits and problems they would experience during the project. However, there was no correlation between students' previous experiences of using search engines and their high appraisal of using **Lex**.

The students focused on some of the shortcomings of this project. Many students stated that it had taken too much time to construct the web-derived corpus and to consult the key words from it. The mean ($\pm SD$) of this response was 6.75 (± 2.06). These students expressed their need to get familiar with the process of language consultation in some degree. The average ($\pm SD$) of this response was 7.32 (± 1.92). As shown in Table 2, those who pointed out the laborious aspect of constructing the web-derived corpora and consulting the key word from it had a weak correlation with those who acknowledged the need for accustoming themselves to this analytical approach of language learning.

Table 2. Correlation Between Factors for Task Work and Need for Training

	Requiring time and effort	Need for training and experience
Requiring time and efforts	—	
Need for training and experience	.26**	—

Notes: Correlation Matrix (N = 69), ** $p < .01$.

It is also important to note that many students realized that the corpus they built did not represent all the language in use. The identification of specific collocational patterns was not necessarily easy. Some students specifically mentioned as follows:

I did my best to keep downloading the web pages until I found satisfactory results to build a much larger and more comprehensive corpus. Despite my efforts, the types of texts that I created were not exactly what I had expected.

I realized that when I downloaded the web pages which were not particularly useful for my study, such as texts full of advertisements and promotional sentences, the corpus I had compiled by myself did not show the context I needed for my analysis.

Also, there were some students who were not fully confident if the retrieved data were valid and reliable. The average ($\pm SD$) of this response was 6.48 (± 2.05). One student reflected as follows:

It was difficult for me to find out whether the retrieved information was reliable or not for my own learning resources.

There was a weak correlation between those who felt the retrieved data were unreliable and those who thought that the language data were unequally retrieved from the web ($r = .24$, $p < .01$).

Pedagogical Implications

Benefits

Although this study was conducted within a relatively short period of time, there were some educational benefits of this discovery learning approach being used in the classroom. The results of the project indicate that two major pedagogical advantages of using web-derived corpus in language learning have been identified: raising students' awareness of actual examples of language and promoting their independent language learning.

The first advantage of this project is that many students benefited from searching for various meanings and usages of the target words. The comments made by the students suggest that many students were able to identify that each of these texts addressed particular topics and contained a certain range of information that was directly relevant to the target word. Before the project was implemented, it had been anticipated that it would be challenging for students, who have a lower intermediate proficiency in English, or no prior online experience, to make their own discovery about actual language use with a web-derived corpus. However, the process of browsing a large amount of data and scanning to locate information by themselves allowed even untrained students to develop their awareness of the features of different text types and identify various lexical patterns. Unlike "decontextualized language," which was claimed by Widdowson (2000, p.7) as a limitation of corpus studies, students' constructed web-derived texts were not decontextualized for students, since they were already familiar with the context the target word was used in. Without solely relying on the limited number of examples provided by dictionaries, this hands-on activity of constructing a web-derived corpus helped students experience a wider range of English than they would ever encounter in the classroom and, therefore, enlarged their lexical competence.

Another important benefit of this project is the fact that the project has enhanced students' capacity to take charge of their own learning. Students' comments have shown that, at a later stage of the project, many students began to make critical evaluations of their

entire corpus experience. It was also evident that a majority of students had developed an awareness of the web information which was not particularly useful for their analysis. In addition, *Lex* lessened the students' frustration to some extent in searching the key word from a vast amount of information and, therefore, encouraged them to be more independent when learning English. Although students were novices in the use of the web as a source of lexical information, the project assisted students in considering not only what and how they should search for identifying the various patterns which are associated with the key word, but also how they will be able to conduct lexical searches by themselves in the years ahead. As a number of studies focus on this area (Braun, 2007; Kaltenbock & Mehlmauer-Larcher, 2005; O'Sullivan, 2007), the project generally enhanced their self-confidence and made students autonomous participants of the language learning process. This learning process would be valuable resource for their future language learning.

Subjects of Future Study

Despite the benefits highlighted in the previous section, several challenges are involved in implementing the present project. The major issue to be considered concerning future implementation of this project can be largely divided into the following two groups:

First of all, some students did not have positive views about the technical aspect of the analytical process in this project. The findings suggest that, despite the fact that the present project did not require students to have any specific knowledge or skills of corpus consultation, more than half of the students expected to improve the system structure of *Lex* so that they could avoid tedious corpus constructing tasks and easily retrieve the information they needed for their own learning resources. These students stated in their comments that retrieving collections of lexical information from various types of texts on the web was laborious and time consuming. At the beginning of the project, some students, specifically those who had selected key words with a low frequency of occurrence, had a problem locating the desired websites manually. As a result, these students had difficulties in analyzing recurrent features in the corpus and drawing their own conclusions about language use. For these students, it can be assumed that utilizing excerpts from online texts for their own corpus study was a real challenge.

It can't be denied that the amount of information students had gained from the web depended on their investment of time and effort. However, as stated in the previous section, an important part of the present research project is to engage students in tackling and analyzing ample contextualized examples by themselves and to gain the ability to understand the use and pattern of certain words for their own studies. For the purpose of enhancing efficiency, time-saving search methods and techniques should be developed for the future implementation of the project.

The second issue to be considered is related to students' concerns about the reliability of the web as a language resource. As stated in the comments made by students, during the process of searching the web, a majority of students seemed to understand different text types and categories of the web sites. However, even those who had successfully compiled their own corpus were not confident whether the information retrieved was likely to be accurate and/or reliable. It has been acknowledged that many web pages have been created without thinking seriously about correctness and corpora derived from this web data do not

represent error-free language use (Kilgariff & Grefenstette, 2003). One possible solution to this problem could be encouraging students to construct a thorough and well-balanced corpus. As suggested in Vannestål & Lindquist (2007), another way of solving this problem is to provide students with the instructor's full support and firm guidance. Although these techniques could be included in the future studies, analyzing linguistic data based on the process of trial and error is another particularly important factor if data-driven methodologies are intended to play a major role in their language learning. In order for students to make web data more suitable and efficient for language learning, this analytical approach in combination with supplementary activities, which will help students acquire skills in identifying unreliable sources of information, is another option available to the instructor.

Conclusion

The overall purpose of this study was to investigate how less-advanced EFL students perceived the effectiveness of a corpus consultation project based on the web-derived data. The findings suggest that, in combination with the use of students' dictionaries, hands-on experience of corpus compilation and consultation can enhance students' awareness of how to take advantage of web data to facilitate their language learning. It has also become clear that there will be a need for further research regarding the way in which web-derived corpora have been used in the classroom in order to promote students' self-motivation and self-direction in their language learning in general. Although there were a number of issues which required further exploration, there is no doubt the web offers infinite possibilities as another language learning resource, stimulating students' intellectual curiosity and deepening their knowledge of language.

Acknowledgement

The authors would like to express their heart-felt gratitude to Mr. Hideki Ooyane for his dedication and tremendous help in developing *Lex*.

References

- Aston, G. (1995). Corpora in language pedagogy: matching theory and practice. In G. Cook & B. Seidlhofer (Eds.), *Principle & practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 257-270). Oxford: Oxford University Press.
- Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19(3), 307-328.
- Carter, R. (1987). *Vocabulary: Applied linguistic perspectives*. London: Allen & Unwin.
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning and Technology*, 9(2), 111-125.
- Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as a language researcher: Putting corpus linguistics on the timetable. *System*, 31, 173-186.
- Cobb, T. (2003). Do corpus-based electronic dictionaries replace concordancers? In B. Morrison, G. Green, & G. Motteram (Eds.), *Directions in CALL: Experience, experiments, evaluation* (pp. 179-206). Hong Kong: Polytechnic University.

- Cooker, L., & Torpey, M. (2004). From the classroom to the self-access center: A chronicle of learner-centered curriculum development. *The Language Teacher*, 28(6), 11-16.
- Dodd, B. (1997). Exploiting a corpus of written German for advanced language learning. In A. Wichmann, S. Fligelstone, G. Knowles, & T. McEnery (Eds.), *Teaching and language corpora* (pp.131-45). London: Longman.
- Fletcher, W. H. (2007). Concordancing the web: Promise and problems, tools and techniques. In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web* (pp. 25-45). Amsterdam: Rodopi.
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301-319.
- Ghani, R., Jones, R., & Mladenic, D. (2001). Mining the web to create minority language corpora. *Proceedings of the 10th International Conference on Information and Knowledge Management (CIKM)*.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, 14 (2), 151-162.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom concordancing* (pp.1-13). Birmingham: Birmingham University English Language Research Journal 4.
- Kaltenbock, G., & Mehlmauer-Larcher, B. (2005). Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching. *ReCALL*, 17(1), 65-84.
- Kaur, J., & Hegelheimer, V. (2005). ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer Assisted Language Learning*, 18(4), 287-310.
- Kehoe, A., & Renouf, A. (2002). WebCorp: Applying the web to linguistics and linguistics to the web. Retrieved September 29, 2007, from <http://www2002.org/CDROM/poster/67/>.
- Kilgariff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as a corpus. [Special issue] *Computational Linguistics*, 29(3), 333-347.
- Koren, S. (1997). Quality versus convenience: Comparison of modern dictionaries from the researcher's, teacher's and learner's points of view. *TESL-EJ*, 2(3). Retrieved March 9, 2007, from <http://tesl-ej.org/ej07/a2.html>
- Mishan, F. (2004). Authenticating corpora for language learning: A problem and its resolution. *ELT Journal*, 58(3), 219-227.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- O'Sullivan, I. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19(3), 269-286.
- Robinson, P. J. (1989). A rich view of lexical competence. *ELT Journal*, 43(4), 274-282.
- Sinclair, J. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, G. Knowles, & T. McEnery (Eds.), *Teaching and language corpora* (pp.27-39). London: Longman.
- Sobkowiak, W. (2002). What can be, but is not (and why), in learners' MRDs. *Teaching English with Technology*. 2(3). Retrieved March 9, 2007, from http://www.iatefl.org.pl/call/j_article9.htm

- Stubbs, M. (1995). Corpus evidence for norms of lexical collocation. In G. Cook & B. Seidlhofer (Eds.), *Principle & practice in applied linguistics: Studies in honour of H. G. Widdowson*. Oxford: Oxford University Press.
- Szendeffy, J. (2005). *A practical guide to using computers in language teaching*. Ann Arbor: The University of Michigan Press.
- Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System*, 29, 91-102.
- Tribble, C. (2003). Five electronic learners' dictionaries. *ELT Journal*, 57(2), 182-197.
- Tribble, C., & Jones, G. (1990). *Concordances in the classroom: A resource book for teachers*. Harlow: Longman.
- Vannestål, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL*, 19(3), 329-350.
- Wichmann, A., Fligelstone, S., Knowles, G., & McEnery, T. (Eds.), *Teaching and language corpora*. London: Longman.
- Widdowson, H. G. (1989). Knowledge of language and ability for use. *Applied Linguistics*, 10, 128-137.
- Widdowson, J. G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21(1), 3-25.
- Willis, J. (2000). A holistic approach to task-based course design. *The Language Teacher*, 24(2), 7-11.

Appendix

Post-questionnaire

The aim of the following questionnaire is to collect some data on students' attitudes and beliefs concerning the benefits and problems of the present project.

1. Before you started this project, how often did you use commercial web search engines such as Google and Yahoo in your everyday life?
Never 1 2 3 4 5 6 7 8 9 10 Quite often
2. How much were you satisfied with your own dictionaries in your classroom when you consulted the meaning and usage of a particular word?
Not satisfied 1 2 3 4 5 6 7 8 9 10 Very satisfied
Write your opinions concerning what you were not satisfied with your own dictionaries.
3. Do you think this project was very productive and satisfying? How would you rate the success and efficiency of this project as part of my reading course?
Not successful 1 2 3 4 5 6 7 8 9 10 Very successful
4. This project helped me catch even small nuances of the key words and their various usages that English dictionaries never offered.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree
5. I was able to examine various example sentences containing the key word in a very effective way.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree

6. As I keep retrieving the data from the web, I was able to infer the meaning of the key word even from the contexts I was not familiar with.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree
7. By using together with dictionaries, this project helps me clearly understand various meanings and usages of the key word.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree
8. The language data and the meaning of the key word were unequally retrieved from the web.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree
9. It takes too much time to construct the web-derived corpus and consult the key word from it.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree
10. In order to consult the key word effectively, you need to get used to the process of the language consultation to a certain extent.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree
11. My concern remains about whether the usages of the key word retrieved from the web are correct or not.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree
12. It is indispensable to use both the commercial web search engines and Lex to examine the usages of the key word using the web data.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree
13. I would like to keep using Lex as a tool to examine lexical information, but I need the improvement of the system structure of Lex.
Strongly disagree 1 2 3 4 5 6 7 8 9 10 Strongly agree

Overall Evaluation

Describe what you think about this project. Write problems and benefits of the project in short sentences.

Problems

Benefits