EMERGING TECHNOLOGIES

A CORPUS APPROACH FOR AUTONOMOUS TEACHERS AND LEARNERS: IMPLEMENTING AN ON-LINE CONCORDANCER ON TEACHERS' LAPTOPS

Jang Ho Lee, Chung-Ang University

Hansol Lee, University of California, Irvine / Korea Military Academy

Cetin Sert, Corsis Research

The present article deals with the issue of how to create and operate a customizable on-line concordancer from viewpoints of language teachers and with their own laptops. It aims to introduce how to use and manage this application without relying on computer engineers for various pedagogical purposes, focusing on the four beneficial dimensions of its interface and technical features: accessibility, simplicity, functionality, and manageability. In addition, the carefully written directions illustrate how to implement this open-source application on laptops using pre-established or customized corpora. For those in the field of language teaching and learning, this application is designed to allow teachers to operate different types or levels of corpora in separate spaces on one server and to enable multiple simultaneous connections in classroom contexts. Ultimately, the authors believe that this application will not only allow students to actively experience data-driven learning anywhere and anytime, but also will help teachers manage their own version of this on-line concordancer by autonomously uploading any kinds of source texts for corpus analysis at their pedagogical discretion.

Keywords: Autonomy, (on-line) Concordancer, Corpus Analysis, Data-Driven Learning

APA Citation: Lee, J. H., Lee, H., & Sert, C. (2015). A corpus approach for autonomous teachers and learners: Implementing an on-line concordancer on teachers' laptops. *Language Learning & Technology*, *19*(2), 1–15. Retrieved from http://llt.msu.edu/issues/june2015/emerging.pdf

Received: May 29, 2014; **Accepted:** July 25, 2014; **Published:** June 1, 2015

INTRODUCTION

Along with the inflow of computer technology into classroom environments, the use of corpora in the teaching and learning of second languages has attracted the interest of the research community, as can be seen in a number of research articles and books since the 1990s (e.g., Bernardini, 2002; Gavioli & Aston, 2001; Granath, 2009; Johns, 2002; Reppen, 2010; Sinclair, 1997). From among the diverse potential uses of corpus linguistics, the use of a concordancer—a computer application that is designed to analyze corpora-based data—has caught the eyes of many to date as a valuable resource tool that contributes to target language (TL) teaching and learning. Concordancers perform two general functions: a word-frequency function, which "provide[s] data on the number of instances of all words in a corpus of text" and a concordancing function, which allows one to "find all instances of a given word in a corpus and to present these instances in their immediate linguistic context" (Flowerdew, 1993, p. 231). Especially, the second function has been highlighted as a potential language-learning resource, given its ability to offer an immense amount of authentic language input to learners. According to Johns (1986), "the multiple contexts offered by a concordance" give learners opportunities for testing a wide

range of "the hypotheses generated by one context to be tested against other contexts" (p. 160). Through this process, learners can develop strategies to correctly guess the meaning of unknown words and analyze the syntactic patterns of contexts where these words are situated. Such an inductive learning or discovery learning process (Nation, 2001) has been acclaimed for enhancing learner autonomy (Godwin-Jones, 2011) as well as learners' motivation (Kettemann, 1995).

Despite their merits, concordancers have been empirically implemented only to a limited degree in classroom environments, the reason for which can be inferred from Römer's explanation (2011): "Teachers and learners have to be provided with access to corpora that are available on the Internet or to off-line corpora and easy-to-use software packages" (p. 216) (see Gavioli & Aston, 2001 for a similar point). As far as the availability of concordancing applications is concerned, the authors point to the limitations of currently available on- and off-line corpus analysis applications. First, corpus-analysis applications such as *MicroConcord* (Scott & Johns, 1993) and *Wordsmith Tools* (Scott, 1996) do not provide an on-line concordancing function. These concordancers, to put it another way, can be used in a limited way with only a single-user mode. Teachers then cannot use these applications for in-class activities with a group of their learners, but rather use them at the pre-class preparation phase (e.g., use them to print out concordance lines associated with a target word for upcoming classes). It should also be mentioned that even if these applications are made available, they are rather difficult to use unless one has been successfully trained and has mastered all their functions and instructions (see Whistle, 1999 for an earlier and similar discussion). On the other hand, on-line concordancers available on Web pages, such as the British National Corpus (BNC) and Brown Corpus, would also be regarded as limited on a practical level in that they only offer results for corpora analyses that are pre-established and pre-determined by their developers. In other words, there is no room for teachers and researchers to select and/or customize source texts for analysis and present them accordingly, despite the fact that these Web applications are free and accessible in classroom environments with Internet access. Although these provide reliable and authentic linguistic data, one cannot guarantee that every user of these Web-based concordancing applications would relish the positive aspects of concordancing for their learning, given the number of users with low TL proficiency or those with a variety of different pedagogical purposes.

To this end, the present article aims to provide an open-source on-line concordancing application along with step-by-step directions for readers to manage this Web application with their own selected corpora, which would allow them to overcome the aforementioned drawbacks inherent in currently available on- and off-line concordancers. This concordancer, which we call an "On-line Concordancer" (OC, henceforth) will help any language teachers with basic computer knowledge manage a customizable concordancer simply by using their laptops as a server for this Web application and uploading any kind of text files as target corpora. Furthermore, this application will enable students to simultaneously log-on to their teacher's laptop to find concordance results relevant to target words. Overall, this paper attempts to address the concerns of others who have reiterated the importance of easy and accessible concordancing applications for teachers and researchers, by suggesting an application which not only provides them with some degree of freedom to operate their own version of a concordancer, but also helps them to rely less on computer science experts or complicated Internet servers. In what follows, the authors will delve into theoretical rationales for using concordancing in the language learning process. Then, the four beneficial dimensions of the suggested concordancing application—accessibility, simplicity, functionality, and manageability—will be discussed in detail, followed by explicit directions for the implementation of this application using a laptop. Lastly, the present paper will suggest several pedagogical and research implications, encouraging teachers and researchers to be more involved in this line of research.

REVIEW OF LITERATURE

Based on the previously discussed pedagogical benefits of using corpora and concordances in TL learning

and teaching contexts, computerized concordancing is well supported by several theoretical frameworks. For example, in light of its capacity to generate a considerable amount of repeated input related to a target word or expression in the Key Word In Context (KWIC) format, in which "each word is centered in a fixed field, and each occurrence of the word is listed on a separate line" (Godwin-Jones, 2001, p. 9), concordancing fits in nicely with the precept of "input enhancement" (Chapelle, 2003). In terms of vocabulary acquisition, this means that learners would be exposed to several different contexts surrounding a target word, which would raise the possibility of their noticing of such an item. Schmidt's *noticing hypothesis* (2001) would then theoretically support a concordancing technique in terms of this characteristic (see Lai & Zhao, 2005; Sprang, 2008 for the value of concordancing methods in this regard).

Concordancing as a way of enhancing vocabulary acquisition is also well grounded in Laufer and Hulstijn's (2001) *involvement load hypothesis*, which is derived from previous studies on incidental TL vocabulary acquisition, and which proposes that "[target] words which are processed with higher involvement load will be retained better than words which are processed with lower involvement load" (p. 15). This hypothesis further suggests that any vocabulary learning task may consist of two cognitive components, *search* and *evaluation*, and one motivational component, *need*, and that a task associated with a higher degree of these components will bring about higher amounts of vocabulary learning. It can be further suggested that any learning activity integrating concordancing, if properly designed, should be conductive to satisfying these factors. For example, the act of concordancing itself is inherently search- and evaluation-oriented, as learners are invited to come to an understanding of the meaning of a target word by referring to its occurrences in different contexts. In terms of *need*, it is the responsibility of teachers or researchers to design a task in such a way that learners are required to search for the meaning or other aspects of target words by using concordancers. This is reminiscent of Wong's pedagogical suggestion (2005) that a target task should be *meaningful* and learners should be *involved* with some activity related to target input, for any input-enhanced activities to be successful.

The value of concordancing in terms of providing a significant amount of input and enhancing learners' cognitive involvement may be qualified, however, by whether this input in its intact form is comprehensible to learners—the main thrust of Krashen's widely-cited *comprehensible input hypothesis* (1985). That is, the question of whether concordance lines extracted from authentic reference corpora such as BNC and Brown are comprehensible and useful for one's students would require one to consider a range of learner factors, presumably with TL proficiency at the top of the list. Allan (2009) suggested that concordance data drawn from reference corpora may not be effective for certain groups of learners. In particular, she noticed that a few randomly selected sample concordance lines contained several low-frequency or highly specialized words that might be unfamiliar to learners around the intermediate level of English (i.e., those in the B1 or B2 range according to the Common European Framework of Reference cited by the Council of Europe, 2001), hindering their comprehension of these extracted sentences. With a similar concern in mind, Cobb (1997) built a corpus himself which consist of source data from learners' own textbooks. This may be a more appropriate approach for TL learners. These graded or customized corpora may not be considered to be on a par with reference-based corpora (i.e., BNC, Brown) in terms of their authenticity; nevertheless, the corpora can be considered to contain a sufficient number of pedagogically meaningful linguistic elements and features (Allan, 2009). The issue of whether learners should be exposed to authentic language data based on corpus analysis as opposed to more concocted language ignited a heated debate among authors in the late 1990s (Carter, 1998; Cook, 1998; Prodromou, 1996), whose discussions dealt with not only the effectiveness of authentic corpora, but also the ownership of English and the legitimacy of producing corpora based on specific varieties of the English language and consequently providing these corpora to learners.

While it is beyond the scope of this paper to discuss the literature on this issue even further, the authors would like to state that providing opportunities to language teachers and researchers to have an experience in building and managing their own concordancers would be of utmost importance to reach

any pedagogically sound decision on the use of concordancers. While some of them have relied on the use of commercially available concordancers for their research and teaching, these applications do not provide enough room for customization. Moreover, others who have used authentic reference corpora have had to be satisfied with their user interfaces and with perhaps less than ideal results, as these corpora also do not allow teachers to freely customize their concordances.

The above-mentioned study by Cobb (1997) was one of the very few which managed to design and build a concordancer along with a corpus based on customized text sources (i.e., the students' textbook in this case). This approach, of course, sounds ideal to all of us. However, the problem lies with the fact that most language teachers and researchers are far from being able to program their own concordancers to their respective tastes. For example, the study conducted by Lee and Swales (2006), which looked at the effects of using specialized corpora for their EFL students' writing course, is worth mentioning due to the fact that even these corpus experts were forced to rely on a widely-used commercial concordancer—Wordsmith Tools—in their research and that this restricted their use of corpora at the materials preparation stage of their experiment before any use in classroom activities. To build and manage a concordancer application, language teachers cannot help but rely on computer experts—"the gatekeepers"—to operate on their own terms (Bloch, 2009, p. 62). This also echoes the point made by Lee and Lee (2013): heavy reliance on computer experts without any relevant background knowledge discourages teachers and researchers from micro-tailoring their class materials in accordance with their goals and their classroom contexts.

In view of the shortcomings associated with currently available concordancers, including off-line single-licensed applications and pre-designed on-line reference concordancers, our on-line concordancing application, the OC, is primarily designed to overcome these limitations. With this application, teachers with Internet access and laptops with Microsoft Windows 8 (or any later version), which provides fundamental libraries as well as a Web server function for the OC, are able to build customizable on-line concordancers (details will be discussed in *Manageability* section). To be more specific, teachers have the ability to upload any selection of text files as a target corpus, which can then be analyzed by this application in accordance with their pedagogical intentions and purposes. It also allows multiple simultaneous connections via the Internet, whereas previous PC-based concordancers are limited to a single user at one local computer. Another merit of this program lies in the fact that students do not have to download corpus data or application (which makes use of corpus data) into their computers, and they do not need to learn complicated procedures to use a concordance program. Instead, they can simply log on to the Web site which is run and operated by their teacher, and type a word or expression into this application with a simple user interface to see relevant concordance results. In addition, this application is free from any licensing issues because the authors started this research as an open-source project. In what follows, four beneficial dimensions of the OC will be discussed, along with instructions for how to install and use it in various computer settings.

FOUR BENEFICIAL DIMENSIONS OF THE ON-LINE CONCORDANCER

In describing the beneficial dimensions of the OC, the authors have borrowed Bloch's (2009) three pillars for the design of a user interface: accessibility, simplicity, and functionality. In addition to these three dimensions, they add one more—manageability—in light of the fact that the OC is designed to be manageable for teachers and researchers with no advanced computer knowledge. Although these dimensions do not constitute a theory of building courseware design *per se*, as Bloch acknowledged in his paper, the authors argue that they are nevertheless useful guidelines in evaluating any newly-developed application. An overview of these dimensions and how the OC fares in terms of each of them is summarized in Figure 1. Detailed explanations of each dimension will be given individually below.

Accessibility

Accessibility is an important feature of the OC, in particular from students' points of view. What students need to do before accessing this application is to type the IP address of teachers' own computers into the address bar of their respective Web browsers (e.g., http://175.192.110.161). As long as students' laptops and smartphones (or other electronic devices) have access to the Internet, the technical specifications of their devices do not pose any significant problem in using this application, and neither do their operating systems or type of Web browser. Furthermore, the OC follows the "anytime and anywhere principle" (Kukulska-Hulme & Shield, 2008) in terms of student accessibility, as students can log onto the Web page of this application on their mobile devices such as iPads, Galaxy Notes, or even on smartphones as long as their teachers' laptops continue to act as servers for the application. In addition, since the OC is an on-line application, it is capable of forming simultaneous connections, which would allow multiple students to access the application at the same time using their own individual connections. While the expected time for this application to produce concordance lines related to a target word or expression is largely subject to the number of users simultaneously accessing the server computer as well as its systemic capacity, our piloting work has found that it works smoothly with up to approximately 45 students without any performance limitations (see Appendix for detailed information on the performance testing).

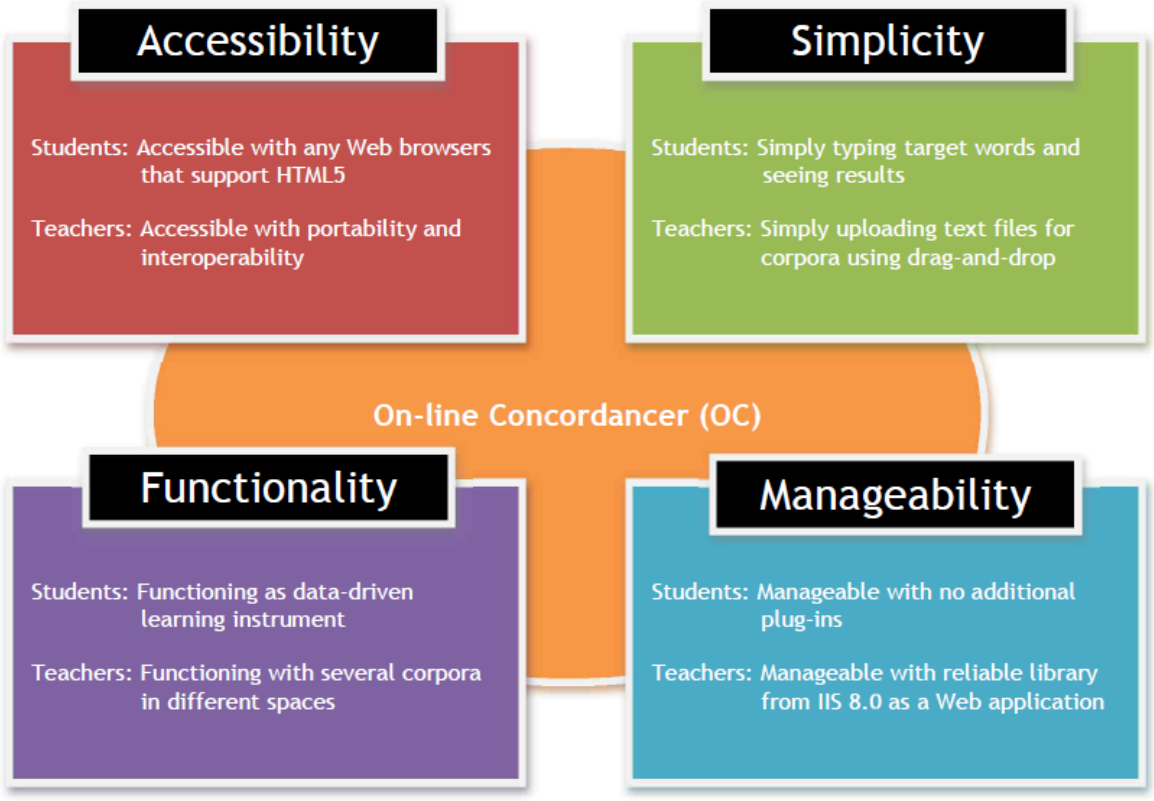

Figure 1. Four beneficial dimensions of the OC.

In addition to being accessible to students, the OC is also designed to be accessible in terms of its specifications for the teachers and researchers who would run this application on their computers, which are meant to function as servers. That is, the OC only occupies 6 MB or so of disk space, whereas Wordsmith Tools 5.0 takes up around 69 MB. This feature of the OC reflects its high level of "portability and interoperability" (Godwin-Jones, 2004, p. 9), which in turn relieves the CPU burden of its server computers. The remarkably small size of the installed version is due to the fact that the OC does not

include its set background processing as well as user interface libraries, but uses a .NET Framework, which is a modern software framework that provides the essential features required to build and manage the OC. This aspect of accessibility is enhanced by the fact that the OC was written using Visual F#, which was developed by Microsoft Research and uses the .NET Framework. This programming language was originally intended to "empower programmers and domain experts to write simple, robust code to solve complex problems" (Syme et al., 2012, p. 1), and this characteristic was well reflected in the developmental procedure of the OC.

Simplicity

The interface of the OC aims to be simple enough for users to instantly try the application out themselves. As Figure 2 shows, the main page presents nothing but a blank box on which users type a target word or an expression.

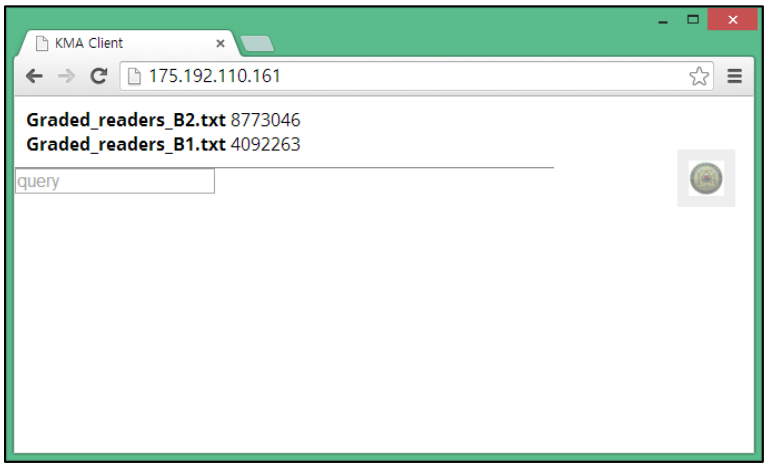

Figure 2. The main page of the OC.

When the analysis procedure of this application is finished, concordance data associated with a target word or expression is consequently displayed on the user's Web browser, as Figure 3 illustrates. Following a typical format for concordance analysis, KWIC design was adopted in the OC as a way to present the concordance results to users. In addition, the keyword in each concordance line is highlighted in red to distinguish itself from the surrounding text. The text box remains on the page for subsequent searches.

Figure 3. The result page of the OC.

Apart from the one to which students are exposed, the OC provides another simple Web page in order for teachers to upload source text files for concordancing analyses (For more information in terms of collecting source data, please see reference lists from Godwin-Jones, 2014, p. 9). For the sake of accuracy, this application encourages teachers to upload texts files in ANSI or UTF-8 format. Typing "/?level=teacher" at the end of the Uniform Resource Locator (URL) (i.e., Web address for the OC, e.g., http://175.192.110.161/?level=teacher) allows teachers to access a Web page such as the one in Figure 4, after passing a verification procedure implemented through the Mozilla Persona login system.[1] On this page, teachers can upload source texts from the local drive to the OC by using the drag-and-drop function—one of the versatile features of HTML5 (Godwin-Jones, 2014) (see the *Manageability* section for detailed information).[2] That is, teachers can drag target text files from the desktop and drop them to the dashed box on the screen that says, "Drop files here." In this way, these selected and dropped text files serve as the basis for the concordancing analyses of the OC. Another simple way of uploading source text files is to copy and paste them into the folders where the OC is installed in their local disk drive.

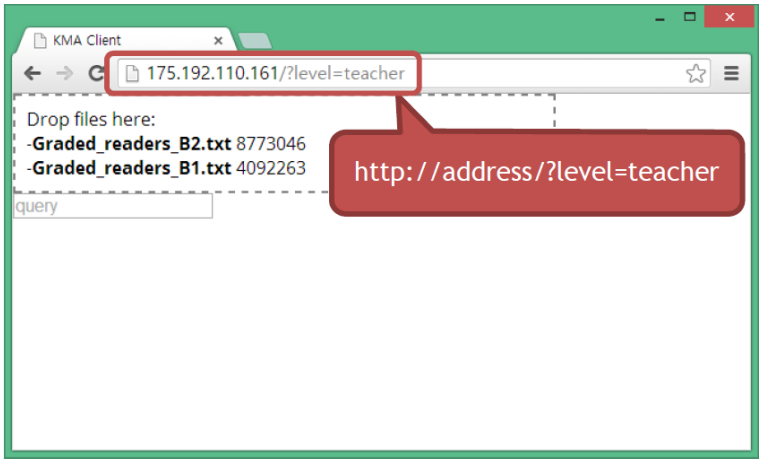

Figure 4. The teachers' page of the OC.

In order to make sure that target source files are properly uploaded and are ready at hand for forthcoming analyses, teachers may simply log onto the main page of the OC with a teachers' authority, which shows a list of the target text files at their disposal (see Figure 4). In addition to uploading or adding files to this list, teachers can also opt to exclude one or more uploaded text files from the list of concordancing analyses by clicking the negative symbol "–" at the head of each file name, which is then not be shown in the students' Web page (see Figure 2). Readers should be advised that the list of uploaded files (in particular, those uploaded via dragging and dropping) is refreshed automatically when the server computer restarts due to users' mistakes or any other reason (e.g., Windows automatic updates).

Functionality

Like any of its precedents (e.g., Bloch, 2009; Cobb, 1997), the OC aims to serve as a pedagogical device through which "data-driven learning" is expected to occur (Johns, 1994). In this way, the OC was designed only to perform a concordancing function as its name indicates. In other words, the OC does not have two other functions that Wordsmith Tools is equipped with (i.e., keywords and wordlists) for the sake of simplicity. While these two functions of this program are inarguably impressive and useful for specific purposes, the authors did not opt to include them in the OC, as students may be discouraged or overwhelmed in the face of more functions than are necessary for TL learning purposes.

Data-driven learning in the pedagogical context of corpora and concordancing refers to the process by which learners are exposed to and encouraged to analyze an extensive list of contexts surrounding a target word or expression, and consequently are expected to derive the meaning of this target item or find some

regularities and patterns related to it (Johns, 1986). As discussed in the previous section, this would be made possible due to the OC's characteristics of input enhancement (e.g., salient and repeated input) and raising learners' awareness. With a target keyword highlighted in red, learners are encouraged to notice this item, infer its meaning based on the numerous contexts provided, and hopefully even figure out some of the grammatical patterns related to this item. While learning TL vocabulary and grammatical patterns may be of primary concern to a high proportion of learners, other aspects of the TL such as collocation patterns or semantic prosody are also subject to acquisition for more advanced and skillful learners (see Lee et al., 2012 for some examples). It may also be suggested that these different levels of TL knowledge may be acquired better with a different grade of corpora, a point that Allan (2009) has already alluded to.

With regard to the above point, it is another strength of the OC that teachers and researchers can run more than one corpus dataset simultaneously by using different *spaces*. The concept of spaces can be construed as uploading and making modifiable objects (i.e., a set of text files for the corpus analysis in this context) available independently at different on-line addresses. In other words, using spaces fulfils the function of separating learners into groups with similar TL knowledge or shared study interests (e.g., business, medical, military, and engineering). That is, a teacher can run several different Web pages uploaded with different corpora simultaneously, with his or her single computer acting as a server. Students can simply put "/?space=SPACENAME" in the query of a URL in order to access each space of the OC (e.g., http://175.192.110.161/?space=SPACENAME), which can be distinguished from other spaces by names defined by the user (see Figure 5).

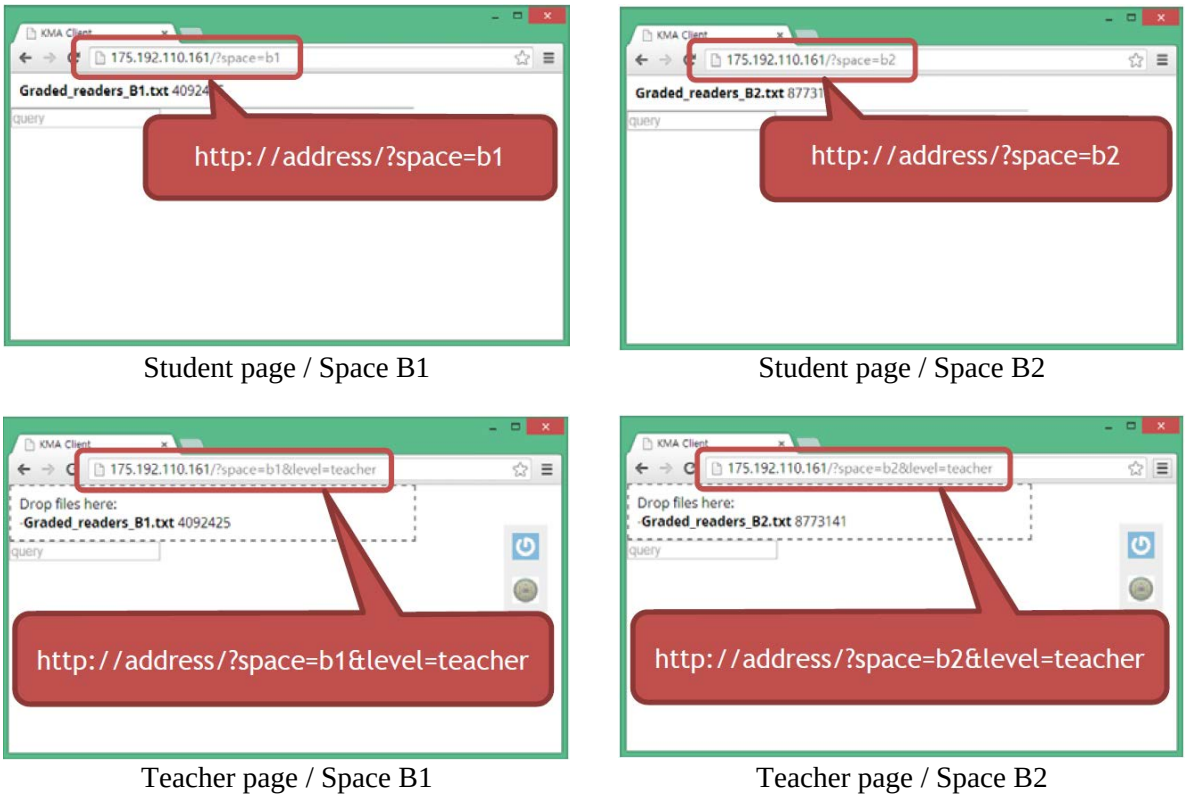

Figure 5. The pages for the different spaces of the OC.

The procedure for uploading additional text files to different spaces remains the same. As shown in Figure 5 above, teachers can log on to the Web site by adding "&level=teacher" at the end of the new IP

address for each space after logging in with Mozilla Persona account verification (e.g., http://175.192.110.161/?space=SPACENAME&level=teacher). The source data, selectively uploaded by teachers on this particular Web page, would therefore not affect other spaces. In this way, teachers can build several different spaces in accordance with their pedagogical purposes or goals. However, it should be noted that those text files in the *local folders* of the OC will be included in the corpus analyses of all the activated spaces.

Manageability

The concept of *manageability* is a widely used concept in the field of computer science and has been recently highlighted in that its key role is to "maintain smoother service delivery and increase business continuity and availability", along with reliability, availability, and serviceability of an application (Radle et al., 2013, p. 1). Given the purpose of this paper and its intended audience (i.e., teachers, students, and researchers with no advanced computer knowledge), the manageability of a newly developed program would be defined by its easy and continuous use. Toward this end, the OC is designed to foster an environment where students would use this application smoothly. This on-line application requires clients to use any kind of a wide range of Web browsers that support HTML5 (e.g., the latest versions of Chrome, Firefox, Opera, and Safari, and Internet Explorer 10 or above). In other words, the architecture of this application adopts HTML5 features "the native environment of the Web browser" (Godwin-Jones, 2014, p. 8) so that students do not have to install additional plug-ins. If a Web browser that does not support HTML5 is used, however, the connection will not be ideal for accessing the OC. Since this application applies Web Sockets libraries from HTML5, which abstracts server-to-client data push techniques, it will make users with these browsers (e.g., the Internet Explorer 9 or below) fall back to AJAX or data polling at certain intervals. Web browsers' compatibility with HTML5 features in terms of Web Sockets can easily be checked by visiting the "Can I use Web Sockets?" site.

Furthermore, this on-line application allows even teachers with no advanced computer knowledge to manage their own Web-based concordancers based on their individual pedagogical purposes or goals. Basically, the OC only requires a laptop with Windows 8 or a later version. The main reason for this condition is that previous versions of Windows do not provide the functions on which this application heavily relies (e.g., Web Sockets library). In addition, Windows became stable in terms of converting a laptop into a Web server when the Internet Information Service (IIS) supported this function, and Windows 8 started to include the IIS version 8.0 that provides a number of functions that support the OC in terms of its high degree of manageability (e.g., Web Socket protocol support). The authors have tested most free libraries that promise the same functionality, but it seems that none of these come close to the one from Microsoft in terms of manageability. Although one may find alternatives to these libraries, it would be a burden and challenge to have ordinary language teachers deal with the cryptic errors of random software libraries instead of spending their time on pedagogical issues that really matter to their teaching and research under a manageable and stable system. In addition, Windows 8 is equipped with Internet Explorer 10 as a default, lightening the possible burden of finding another Web browser that supports HTML5.

DIRECTIONS: IMPLEMENTING THE OC WITH YOUR LAPTOP

A fundamental operation underlying the following directions is to set up your laptop as a Hypertext Transfer Protocol (HTTP) server, meaning that the local Internet Protocol (IP) address thereof would be the URL for the OC. As a first step, you should enable the IIS function and Microsoft .NET Framework in the control panel. Your Windows usually does not operate the IIS as a default condition unless the user activates it, while the .NET Framework is normally activated. However, readers are advised to make sure that the .NET Framework is clicked as an option, which provides the critical library source for the OC as mentioned above. After the configuration process, the IIS function will then be enabled in Windows and the IIS manager application can be selected on the memo interface of Windows 8.

After setting up the IIS manager, your default Web site can be accessed by typing "http://127.0.0.1" or "http://LocalHost" in the address bar of your Web browsing applications. If your computer is on-line and has its own IP address, the default Web site can also be accessed through its IP address as well (e.g., http://175.192.110.161). If you keep failing at this step, you should check the condition of your connection thoroughly (e.g., checking your Windows virtual firewall setting). However, you will need to change extra settings to implement the OC when you are equipped with a router to share your Internet service or when your connection is physically firewalled. In particular, you should set up the router in order to forward the incoming connections to your computer, leading those connections to your IP address to the MAC address or internal IP address of your computer directly (e.g., the Port forwarding function and DMZ/Twin IP function). Though the possibility to implement the OC is low when you are behind a physical firewall of service providers, you are still able to continue if you can get permission from its operators to adjust the settings for the firewall to open a port for your connections. In most cases, Wi-Fi connections or 3G/4G tethering services by smartphone or relevant electronic devices are not able to support the Web server function for your computer.

As a third step, you should download the source package of the OC to your laptop, which will turn your PC into the main server for this application. Since this program is an open-source project, anyone can visit the Web site (https://bitbucket.org/corsis/kmac) and download the package at no cost. After building the solution using Visual Studio's Visual F# language, you run the execute file as an administrator, making your PC as a server that provides concordancing functions. Then you should link this implemented server and the clients' connections from outside your laptop: "site binding" —a combination of an IP address, a port for the service, and a domain (host header or the type of scheme) in order to guide clients' connections through the designated route to the target Web application in the local hard drive. Your computer is now the server for this on-line application and is ready to be accessed by your students.

CONCLUSION AND OUTLOOK

As discussed in the previous section, the OC can be equipped with any corpus dataset in view of target students' proficiency levels and other relevant factors, and present concordance results accordingly. This would mean that, in a relatively low-level classroom, teachers can still make data-driven learning available to their learners with corpora more fine-tailored to the level of participants. One suggestion for teachers in such a classroom is that they obtain access to the text file of the textbook (or one above the current level of their students) that they are using, and expose them to their learners for discovery learning opportunities. Arguably, corpus data sets based on the students' current textbook will serve as much more manageable analytic tools for vocabulary acquisition and discovering grammatical patterns for these learners than those based on authentic reference corpora.

In the same sense the OC may also be useful in classrooms where a TL is being taught for specific purposes such as business, military, engineering and many others. This is because the language used in a specific field is obviously distinguishable from that used for general communication purposes in terms of its rhetoric, vocabulary, and collocation patterns (e.g., Lee et al., 2012; Walker, 2011). Consequently, corpus data compiled for general purposes would not be as effective as specialized corpus data for teaching and learning a TL for specific purposes. While gaining access to these corpora or collecting relevant corpus data may be another practical issue for teachers, they will serve as valuable pedagogical materials, if obtained, in this type of classroom.

The authors would like to note that the OC can be made available for both in-class activities as well as students' personal use outside the class, as long as the main server computer (i.e., a teacher's laptop) is switched on and has an Internet connection. Those keen and motivated learners, then, will be able to experience data-driven learning even after school, and use it for their vocabulary and grammar learning, TL composition, and for other language learning-related explorations. This is what previous literature has described as "autonomous learning", which has several merits and deserves more attention from the field

(Godwin-Jones, 2011).

In terms of research, the OC offers a versatile opportunity for researchers to design an empirical study which aims to compare the effects of different corpus datasets on a wide range of TL knowledge for a particular population, using the space function of this application. One example of this would be, in light of Allan's (2009) descriptive comparison of graded and authentic corpora, a format of experimental study where corpora consisting of language data more finely-tuned to the level of participants and authentic reference corpora are compared in terms of their effects on the vocabulary acquisition of a target participant group. Another example of a study using this function of the OC would be to examine learners' attitudes and reactions towards different types of corpora, an area which has until now remained seldom studied.

To conclude, the authors hope that the OC will rejuvenate research on concordancing and TL teaching practices based on corpora. It is also hoped that the OC will result in an increase in the number of action research on the part of TL teachers. Their experiments with this application will invaluably contribute to our understanding of how corpora and concordancing should be used in and out of classrooms.

APPENDIX. Performance Testing of the OC.

A pilot study was conducted in order to evaluate the performance of the OC in terms of its speed and stability from the point of view students. It was the purpose of this study to examine whether the current application can function efficiently and effectively in real classroom environments. This piloting was conducted under the following circumstances: The teacher's laptop was an LG 15U530-KH50K Ultrabook with an Intel® Core™ i5-4200U Processor at 1.60 GHz (3MB Cache, up to 2.60 GHz); 4GB memory (DDR3L 1600 MHz); and a 129GB SSD Hard drive (SATA3 6Gbps). Microsoft Windows 8.1 was the operating system of this laptop. Client connections were established in two computer laboratories at a university in Seoul, Republic of Korea. The technical specifications of these computers (desktops) in these laboratories were as follows: HP Compaq 6000 Pro SFF PC with Intel® Core™2 Duo Processor E7500 (3M Cache, 2.93 GHz); 2GB memory (DDR3 1033 MHz); 250GB Hard drive (Pocket Media Drive); and Microsoft Windows 7 as the OS. The Web browser used in this pilot study was Google Chrome version 35.0.1916.114m, which was confirmed to be compatible with Web Sockets. The detailed information on technical specifications above is given for readers to roughly estimate the performance of the OC in their pedagogical contexts. If the technical specifications of their devices are similar to those in the present study, they can expect similar results.

The authors aimed to create a large-size corpus in order to determine whether the OC worked efficiently and without any technical drawbacks even with a bulky dataset. In building a military English corpus for this study, the authors collected 211 US Army field manuals in PDF format from official US Army Web sites. The size of this corpus was 124,699,768 bytes (118 MB), while the number of tokens (running words) was 15,341,757 (15.3 million) and the number of types was 104,218. Considering the features of the downloadable Open American National Corpus (OANC)—79,267,448 bytes (75.5 MB), 11,694,214 tokens (11.7 million), and 154,083 types—this corpus should be sufficiently large enough in comparison with any future in-house corpora built by language teachers and researchers for various pedagogical purposes.

Since the server performance of the OC was assumed to be subject to the number of simultaneous connections on the part of the computers within the system, the capability of this server-based concordancer needed to be verified through a quasi-experimental pilot testing by creating a range of conditions with different numbers of simultaneous connections. To this end, the authors recruited 155 university-level EFL students who were assigned to groups with 15, 23, 30, 42, and 45 clients accessing the OC at the same time respectively. These numbers were set up in light of typical sizes of classroom in

a wide range of pedagogical contexts. After a brief introduction to the pilot study, the participants in each condition were asked to access the OC Web site via the IP address of the author's laptop, and were given a printed handout which contained 10 different military-related jargon words. A mock vocabulary exercise was distributed to the participants, which asked them to type these target words into the OC and infer the meaning of these words by reading through the concordance output on each target word. In doing so, the authors made sure that all the participants made simultaneous connections and interacted with the OC application run by the server computer. Although a majority of participants were not able to finish the exercise due to time constraints, the time period of the experiment, the authors believe, was lengthy enough to measure server performance in terms of the server's stability and search speed.

The results of this pilot study revealed that the server computer maintained its stability in terms of CPU usage as well as network traffic overload, indicating that a laptop similar in profile to the one used in the present study can perform efficiently as a server computer operating the OC and allow approximately 45 computers within the system to simultaneously access to the OC simultaneously without any lagging or other technical issues. Furthermore, it was revealed through a post-exercise questionnaire that the participants generally experienced a rapid and stable search condition in light of their previous heuristics with Google, one of the world's finest search engines. A questionnaire item which used a five-point Likert Scale asked the participants to rate the speed and stability of the OC in comparison with Google (1 being "*OC is much slower than available search engines—Google*" and 5 being "*OC is much faster than available search engines—Google*"), the mean score of their responses was 2.72, indicating that the participants experienced a similar level of speed with the OC in comparison with Google. More interestingly, the result of a one-way ANOVA among the mean scores of the five groups mentioned above showed that the participants' perception of the speed of the OC was not affected by the number of simultaneous connections (F (4, 150) = 1.19, p = .319).

NOTES

1. The OC adopts Mozilla Persona as a login system to distinguish teachers and students in order to establish the authority in managing the list of corpus source data and for acknowledging bidirectional channels with identifiable clients. With this system, users can easily login with their Yahoo or Google mail account. The OC, of course, admits anonymous logins as well.
2. The use of HTML5 Web Sockets establishes efficient, always-on, bidirectional channels which would then allow a teacher to push texts, search queries and results to all connected students in each space. Currently, the OC supports a form of text push in the form of file uploads to spaces.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Dr. Robert Godwin-Jones, the editor of the Emerging Technologies Column in *Language Learning & Technology*, for his invaluable comments and feedback. Also they would like to deliver special thanks to Dr. Tom Cobb for his insightful suggestions and ideas in regards to this article.

This research was supported by 2013 research funds of the Hwarang-dae Research Institute at Korea Military Academy (Research No. 20130509).

ABOUT THE AUTHORS

Jang Ho Lee received his DPhil in education from the University of Oxford. He is presently an assistant professor in the Department of English Education at Chung-Ang University. His areas of interest are teachers' code-switching in English classrooms, learners' attitudes towards teachers' language uses, and using mobile technology for language learners.

E-mail: jangholee@cau.ac.kr

Hansol Lee is a Ph.D. student in the School of Education at University of California, Irvine, specializing in Language, Literacy and Technology, and an assistant professor in the Department of English at Korea Military Academy. His research interests include computer-assisted language learning, corpus linguistics, and language assessment. All correspondence regarding this publication should be addressed to him.

E-mail: hansol6461@gmail.com

Cetin Sert is a freelance developer in Heidelberg, Germany. His interests include functional programming languages, large-scale document intelligence and computer networks.

E-mail: cetin.sert@gmail.com

REFERENCES

Allan, R. (2009). Can a graded reader corpus provide 'authentic' input? *ELT Journal*, *63*(1), 23–32.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165–182). Amsterdam, The Netherlands: Rodopi.

Bloch, J. (2009). The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology*, *13*(1), 59–78. Retrieved from http://www.llt.msu.edu/vol13num1/vol13num1.pdf#page=66

Carter, R. (1998). Orders of reality: CANCODE, communication and culture. *ELT Journal*, *52*(1), 43–56.

Chapelle, C. A. (2003). *English language learning and technology*. Amsterdam, The Netherlands: John Benjamins.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, *25*(3), 301–315.

Cook, G. (1998). The uses of reality: A reply to Ronald Carter. *ELT Journal*, *52*(1), 57–63.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, *21*(2), 231–244.

Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal*, *55*(3), 238–246.

Godwin-Jones, R. (2001). Tools and trends in corpora use for teaching and learning. *Language Learning & Technology, 5*(3), 7–12. Retrieved from http://llt.msu.edu/vol5num3/pdf/emerging.pdf

Godwin-Jones, R. (2004). Learning objects: Scorn or SCORM? *Language Learning & Technology*, *8*(2), 7–12. Retrieved from http://llt.msu.edu/vol8num2/pdf/emerging.pdf

Godwin-Jones, R. (2011). Autonomous language learning. *Language Learning & Technology, 15*(3), 4–11. Retrieved from http://llt.msu.edu/issues/october2011/emerging.pdf

Godwin-Jones, R. (2014). Towards transparent computing: Content authoring using open standards. *Language Learning & Technology, 18*(1), 1–10. Retrieved from http://llt.msu.edu/issues/february2014/emerging.pdf

Granath, S. (2009). Who benefits from learning how to use corpora? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 47–65). Amsterdam, The Netherlands: John Benjamins.

Johns, T. (1986). Micro-concord: A language learner's research tool. *System, 14*(2), 151–162.

Johns, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 27–45). Cambridge, UK: Cambridge University Press.

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, The Netherlands: Rodopi.

Kettemann, B. (1995). On the use of concordancing in ELT. *Arbeiten aus Anglistik und Amerikanistik, 20*(1), 29–41.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications.* New York, NY: Longman.

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL, 20*(3), 271–289.

Lai, C., & Zhao, Y. (2005). Introduction: The importance of input and the potential of technology for enhancing input. In Y. Zhao (Ed.). *Research in technology and second language learning: Developments and directions* (pp. 95–101). Charlotte, NC: Information Age.

Laufer, B., & Hulstjin, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26.

Lee, H., & Lee, J. H. (2013). Implementing glossing in mobile-assisted language learning environments: Directions and outlook. *Language Learning & Technology, 17*(3), 6–22. Retrieved from http://llt.msu.edu/issues/october2013/emerging.pdf

Lee, H., Shin, K., & Lee, J. H. (2012). How a corpus-based study can help in ROK-US combined operations: A corpus-based study on the military vocabulary. *Korean Journal of Military Arts and Science, 68*(3), 49–72.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes, 25*(1), 56–75.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, UK: Cambridge University Press.

Prodromou, L. (1996). Correspondence. *ELT Journal, 50*(1), 88–89.

Radle, B., Bradicich, T., Anderson, M., & Prewitt, J. (2013). *What is manageability?* New York, NY: National Instruments. Retrieved from http://www.ni.com/white-paper/14415/en/pdf

Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge, UK: Cambridge University Press.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205–225.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, UK: Cambridge University Press.

Scott, M. (1996). *Wordsmith tools*. Oxford, UK: Oxford University Press.

Scott, M., & Johns, T. (1993). *MicroConcord.* Oxford, UK: Oxford University Press.

Sinclair, J. M. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (pp. 27–39). London, UK: Longman.

Sprang, K. A. (2008). Advanced learners' development of systematic vocabulary knowledge: Learning German vocabulary with inseparable prefixes. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 139–162). New York, NY: Routledge.

Syme, D., Granicz, A., & Cisternino, A. (2012). *Expert F# 3.0*. New York, NY: Apress Media.

Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English for Specific Purposes, 30*(2), 101–112.

Whistle, J. (1999). Concordancing and learner autonomy: An experiment with first and second year undergraduates. In K. Cameron (Ed.), *CALL and the Learning Community* (pp. 443–454).

Wong, W. (2005). *Input enhancement: From theory and research to the classroom.* Boston, MA: McGraw-Hill.