Vocabulary Learning through Data-driven Learning in an English as a Second Language Setting[*]

Adem Soruç[1]
Sakarya University

Bilal Tekin[2]

Abstract

The past twenty years have witnessed rapid advances in the field of corpus studies, with some studies, for instance, investigating the effectiveness of data-driven learning (DDL), through which students can discover and learn. However, empirical evidence on this approach to learning is still limited. Therefore, the present study, first, investigates whether DDL is effective in vocabulary learning compared to traditional instruction (TI), and, second, explores students' attitudes toward vocabulary learning when taught with either a DDL or a TI approach. The study was carried out at a secondary school in Kampala, Uganda, and compared students in a DDL group (*N* = 36) to those in a TI (control) group (*N* = 36). The pre-/posttest analyses revealed that, although both groups performed in the posttests better than the pretest, the DDL group achieved significantly higher scores than the TI group on both the immediate and the delayed posttest. Moreover, semi-structured interviews with six students corroborated the quantitative results, revealing that the students favored DDL, based on such expressions as "freedom," "comfortable," "relaxing," "technological," "feeling myself responsible," and so on. The paper gives some implications for the same teaching/learning situation and makes suggestions further research.

Keywords

Corpus • Concordance • Data-driven learning • Vocabulary • Autonomy

With the advent of educational technology, an array of pedagogical approaches or technological uses have filtered into the language classroom, generally bearing different names such as computer-assisted language learning (CALL); multimedia learning (e.g., Mayer, 2001, 2005; Soruç, 2015); e-learning (Peng, Su, Chou, & Tsai, 2009); at the turn of the millennium, m-learning, using mobile devices such as personal digital assistants (PDAs), mobile phones, and laptops (e.g., Hockly, 2013; Şad, 2008; Şad & Göktaş, 2014; Saran, Çağıltay, & Seferoğlu, 2008; Saran, Seferoğlu, & Çağıltay, 2009); and corpus-aided language learning (e.g., Aston, Bernardini, & Stewart, 2004; Çelik & Elkatmış, 2013; Huang, 2011). All these technological pedagogies have played a major role in bringing about greater effectiveness in teaching in language classrooms, but it is corpus-aided language learning, or data-driven learning (DDL), that is the focus of the study discussed in this paper. The paper compares the effectiveness of using concordance lines to search for different types of word meaning with traditional method(s) generally used to teach words in the second/foreign language classroom.

Corpora and Data-driven Learning

Although in the past the first computer corpus—also known as Brown Corpus—was thought to be "a useless and foolhardy enterprise" (Francis, 1992, p. 28), today it is thought that corpora, or concordance lines, have "revolutionized" (O'Keeffe, McCarthy, & Carter, 2007, p. 21) dictionary-making processes in language teaching. Corpus-assisted, or data-driven, learning helps not only learners search for polysemy (different types of meaning); semantic prosody (different types of rhythm, intonation, and stress); phraseology (different styles of words and phrases); and authentic or pattern grammar (Conrad, 2000; Dönük, 2016; Kılıçkaya, 2015; Reppen, 2010; Uysal, Bulut, & Al Hosein, 2013) but also teachers, for instance, design a corpus-based English reading course for academic purposes (e.g., Kırkgöz, 2006). Today, DDL is viewed as the key pedagogical approach for using corpora in language teaching and learning (e.g., Çelik, 2011; Çelik & Elkatmış, 2013; Huang, 2011; Johns, 1991), which was first defined as "the use in the classroom of computer generated concordances to get students to explore the regularities of patterning in the target language" (Johns & King, 1991, p. iii), or, as Johns (1994, p. 297) later put it, "the attempt to cut out the middleman as far as possible and to give direct access to the data so that the learner can take part in building up his or her own profiles of meaning and uses." By doing so, as well as by developing more autonomy (e.g., Çelik, 2011; Huang, 2011), learners themselves can discover the original uses of words or find out how collocations of a particular word work together.

The DDL approach is also theoretically sound. Geluso and Yamaguchi (2014, p. 227) argued that DDL is "firmly grounded in cognitive linguistic theory as learners

analyze masses of input in a quest to become more familiar with structural regularities via inductive means." Using this method, learners are driven by a process of inquiry that provides them to go through an array of "psycholinguistic guessing games thanks to concordance before and while learning vocabularies" (Yılmaz & Soruç, 2015, p. 2628), employing corpora as a "mediational tool" (Vygotsky, 1978). The approach is also an awareness-raising activity that can be viewed within the scope of Schmidt's (1994) "noticing hypothesis," which is discussed in the field of second language development. Boulton (2010) specifically stated that in a DDL class, learners are not taught "overt rules" (p. 535) but instead allowed to discover language-related structures, target words, or collocational groups, and at the end to detect those specified language patterns among multiple samples by getting help either from corpora as a "mediator," as a "portable teacher," or from concordancing lines with words line by line as "an awareness-raising tool."

However, although few, there are some arguments against and for DDL. While, for instance, Kennedy and Miceli (2001) found that learners exposed to DDL-centered tasks become discouraged, Rezaee, Marefat, and Saeedakhtar (2014) found that learners who engaged in DDL gave positive comments on it. DDL was favored as it not only provides rich and authentic input (e.g., Johns, 1991; Sun & Wang, 2003; Yoon, 2011) but also transforms relatively passive learners into active ones, and essentially the subordinate into autonomous (Huang, 2011; Lin & Lee, 2015; Starfield, 2004; Yoon, 2008). Although Boulton (2010) argued that DDL does not fit well into teaching certain aspects of grammar, some studies (e.g., Frankenberg-Garcia, 2012; Gaskell & Cobb, 2004; Gilmore, 2009; Quinn, 2014) found positive effects of DDL in terms of correcting typical L2 mistakes. Additionally, Rezaee et al. (2014) and Smart (2014) favored DDL, claiming that it improves learners' knowledge of collocations and grammar knowledge as they frequently come across multiple examples juxtaposed line by line that are related to the use of an unknown word in different contexts (Thurstun & Candlin, 1998; Wu, Witten, & Franken, 2010).

Previous Research

Some studies identified conspicuous findings that support the use of DDL-centered tasks. For instance, Thurstun and Candlin (1998) piloted some materials prepared using a concordancing, or microconcord, program—a corpus program containing 1,016,000 words taken from academic texts. Learners were first presented with words used in multiple examples, which were then followed by a series of exercises to ensure that they used the items correctly and appropriately. Although the learners were puzzled by the cut-off sentences of the one-line concordances, the results showed that they found the materials effective as "they provide a helpful, very different, and very innovative approach to vocabulary learning" (p. 277).

Yoon (2008) conducted a study in a graduate-level advanced English as a second language (ESL) academic writing course to find out whether corpus use would influence the language skills of learners in an effective way. The results revealed that DDL-centered tasks "heightened the students' language awareness, which, in turn, affected their approaches to writing and the writing process" (pp. 43–44), enhancing their awareness, especially of "common usage and collocation in writing" (p. 44). Yoon defined a corpus as a tool serving as a "meaningful reference" and/or "catalyst" (p. 44) to mediate learners in their writing.

Çelik and Keser (2010) investigated whether there was any relationship between learners' logs of online corpora and their lexical acquisition of academic words, measuring improvement through a vocabulary knowledge scale and a collocation test. The results showed that, while a positive relationship was established between the length of learners' presence online and vocabulary knowledge, no relationship was found between the time spent online and the collocation test. In another study, Çelik (2011) compared the effects of DDL to those for an online dictionary use group. In this study, 68 students at the tertiary level received two different types of instruction for five sessions, in each of which ten collocations were taught. Pre-/posttest results showed that learners receiving vocabulary instruction through DDL remembered much more than those engaged in online dictionary use. Therefore, he suggested that DDL be incorporated into the syllabi of intensive English language teaching programs to teach words, word groups, or collocations.

Frankenberg-Garcia (2012) gathered data from a group of English as a foreign language (EFL) learners in Portugal, comparing the effectiveness of using dictionary definitions to that of corpus examples. Learners' performance was measured at two levels: whether they learned the meaning of a target word and whether they used the learned word appropriately in a sentence. The results showed that, while dictionary use was more effective at helping learners get to know the meaning of the words correctly, corpus use was more effective at helping learners write the words appropriately on a syntactic level.

A few studies found counterintuitive results related to the effectiveness of corpus-driven learning. For instance, Ünaldı, Bardakçı, Akpınar, and Dolas (2013) recruited 69 Turkish learners of English in the 10th grade to compare the effectiveness of contextualized, decontextualized, and corpus-informed vocabulary instruction. The participants were pre-intermediate and studied at a state school. In the eight-week instructional stage, the learners in the corpus-informed group studied target words by themselves from the concordance lines, while those in the contextualized group learned the words in a meaningful context by reading texts and those in the decontextualized group found word meanings by translating, using dictionaries, and

doing word mapping. The results of a multiple-choice vocabulary test suggested that learners who received words in decontextualized forms did the best, which was contrary to the findings in the literature on the corpus-based approach. The results should be carefully considered, though, especially because only one data collection instrument was used in the study, which was a multiple-choice test that might possibly have favored the students in the decontextualized group.

Geluso and Yamaguchi's (2014) study explored the attitudes of 30 Japanese university students toward a DDL approach, in which the students were asked to investigate words and phrases using the Corpus of Contemporary American English (COCA), a corpus tool. Then, the students were given a questionnaire, interviewed, and asked to write reflection logs. The results showed that, although the students had some "reservations" (p. 240) about whether they could use the words that they searched for appropriately, they still favored learning vocabulary or formulaic sequences using corpora or DDL, which they described as "a good tool" (p. 240).

In a study of the needs of 323 Turkish medical students in the Faculty of Medicine of a Turkish state university, Özdemir (2014) found that a large majority of the students (80%) paid greater attention to the meaning of English words with their collocations in medicine. To this end, Özdemir built a small-scale corpus from research articles in the *Journal of Medical Case Reports*, consisting of 31,731 words in total. The participants had easy access to the specialized corpus, researched independently discipline-specific words with concordance lines, and practiced using the tool both in and out of the classroom for self-initiated projects. Although the study does not make clear how students' perceptions or feelings were collected, it still gives new insights related to English for specific purposes (ESP) pedagogy, because it, for instance, provided sufficient information about corpus use for ESP words that dictionaries in general cannot provide.

In Yılmaz and Soruç's (2015) study of 40 pre-intermediate Turkish EFL learners at the tertiary level, one class was given eight classroom hours of vocabulary instruction, using a COCA corpus program to teach specific vocabulary that the participants did not know prior to the study, and another class was used as a control group. The results showed that the DDL group improved its performance significantly more than the control group. Thus, the study indicated that DDL-centered tasks make the learner both "a linguistic researcher" and "an explorer," and that language-learning activities are "enjoyable" (p. 2629).

In another recent study, Uçar and Yükselir (2015) researched whether prep students at a state university could learn verb-noun collocations through a corpus tool, COCA. The researchers assigned some students to a corpus group ($N = 15$) and others to the control ($N = 15$) group. They chose 15 targeted collocations from the

students' textbook, teaching those word groups in four classroom hours. Although the vocabulary test developed for the study was found to be less reliable, the results showed that those in the corpus group grasped more about verb-noun collocations than those in the control group.

On one hand, Lin and Lee (2015) investigated the effectiveness of DDL as opposed to the grammar translation method, gathering data from six second-year early-career teachers who were studying in a master's program in the English department of a Taiwanese university. After all the teachers attended training sessions related to DDL, they were asked to prepare materials based on both DDL-centered and grammar-translation tasks and to teach three different sets of grammar concepts, such as passive voice, relative pronouns, and a series of lexical phrases in English to three different classes of first-year English majors. Teachers were asked to keep reflective journals, analyses of which showed that the teachers favored DDL-centered tasks, frequently describing DDL with such words as "new," "innovative," "fresh," and "interesting" (p. 269).

In a similar study, Özbay and Kayaoğlu (2015) researched six EFL teachers' (all female) perceptions regarding any greater role of corpus tools on their integration into the classroom environment. The teachers were teaching at a prep school of a state university in Turkey and had "little or no contact with the corpus and corpus tools in their previous teaching" (p. 97). They however went through a series of workshop trainings for eight weeks, and data related to their perceptions of corpus use in the classroom came from semi-structured interviews. Although the data were limited only to one instrument, it was still revealed that the teachers "favored the language exploration process" (p. 85) using corpus tools.

Much more recently, Tekin and Soruç's (2016) qualitative study explored the perceptions of 26 students related to corpus-based activities at an international high school in Istanbul, Turkey, collecting data through a personal evaluation checklist, reflection papers, and interviews. Results revealed that the students found corpus-based activities "easy," "fun," "innovative," "autonomous," and "practical" but "complex." In another study, Aşık, Vural, and Akpınar (2016) investigated the attitudes toward and beliefs about DDL of 126 students in an English language teaching department (ELT). To this end, they introduced corpora or a variety of corpus tools as the content of a lexical competence course during the semester. Data came from a questionnaire and focus-group interviews, which indicated the students increased their lexical awareness through DDL-based activities, in particular, improving their knowledge or "depth of vocabulary" (p. 92) by learning words with different collocations.

To conclude, all these studies generally show that learners assigned DDL-centered tasks have at least a chance to analyze collocations, word forms, syntactic patterns, and/or the constraints of the target word (Frankenberg-Garcia, 2014; Leel, 2011).

Notwithstanding the efficiency of vocabulary learning thanks to DDL, which allows learners to discover or learn independently by exploring between and/or within concordancing lines, Mukherjee (2006) argued that a big dilemma still remains concerning what corpus studies *suggest* for language classroom pedagogy and what teachers actually *do* in the classroom. One of the most important reasons for this, as suggested by Conrad (2005) and Flowerdew (2012), is that, due to a paucity of DDL knowledge or awareness related to its benefits, language teachers still do not benefit from the advantages it provides. Thus, more research studies should be conducted in different settings involving, in particular, young learners at different language levels in order to generalize the theoretical construct underlying corpus-based learning or DDL-centered activities. The present study aims to fill this gap by addressing the following research questions:

RQ1: To what extent do DDL-centered tasks or TI-based activities improve ESL learners' vocabulary knowledge?

RQ2: What are the beliefs of ESL learners concerning DDL tasks or TI-based activities?

The Study

Setting

This study was conducted at a private dormitory boys' high school in Kampala, Uganda, at the beginning of the third semester in academic year 2013–14. The school at the time of the study had almost 650 learners and six English language teachers. The school had a computer laboratory, and classes were equipped with one computer, one overhead projector, and a sound system for doing listening and video-based activities. Learners received eight hours of English classroom instruction every week, which were generally taught by two teachers.

Participants

At the beginning of the study, the number of learners was 87, and all were 10th-grade students ranging in age from 12 to 15. Although the participants stated that they knew English since they were studying in ESL context and although they were in fact good at speaking as a productive skill, they were not good at writing. Nor were they proficient enough in terms of vocabulary knowledge in English, which was also established by the vocabulary test that was used for carrying out this study. All learners were conveniently sampled for the study but randomly assigned to the experimental DDL group ($N = 36$) and the traditional instruction group ($N = 36$). Mortality effect was observed in this study. Two students did not want to be involved in the study, although its purpose was explained to the students and they were assured of confidentiality. Moreover, another two missed the pretest, five the

immediate posttest, two the delayed posttest, and another four failed to participate in the instructional stage, thus leaving as the final *N* size 72 students for the pre-/posttest analyses.

Selection of the Target Words

The learners at the time of the study were using an upper-intermediate English textbook, which was not above their language proficiency as English is one of the formal languages in Uganda. In order to determine the vocabulary level of the students as a whole, the Examination for the Certificate of Proficiency in English (ECPE) C2 Michigan Proficiency Vocabulary Test (50 items in total) was used. Words known at and above 70% were excluded, leaving 32 target words to be taught through both DDL- and TI-centered tasks.

Instructional Groups and Process

After the researchers had determined the target words ($N = 32$), four regular classroom hours were used for either of the instructional groups (DDL vs. TI), which were exposed to eight target words in each class. In the instructional period, the groups received different instructions. In the DDL group, learners were first trained on how to use corpus programs (e.g., the British National Corpus) to search for words. In addition, they were also presented with selected multiple examples of the target words used in sample sentences. After that, as well as allowing learners to search for the target words themselves in context, they were additionally asked to read and examine those sample one-line concordances related to the words (Figure 1 below) individually, and then to find and highlight key words around the target word in pairs. The learners discussed different uses and/or meanings of the target words, and were asked to write key sample sentences using the targeted words. Through doing the DDL-centered tasks, learners could

a) read authentic examples related to the target words, deducing and discovering the meaning of them inductively;

b) compare their predictions and negotiate them in pairs, thus allowing them to interact with one another while completing the tasks; and

c) increase their awareness of how the words were used in meaningful and repetitive contexts (e.g., His lips are *parched with* thirst, Don Esteban moistened *his parched lips)*

```
y nailed him up Ah what pain His lips are PARCHED with thirst and they mock him stil
eeds before the ground becomes completely PARCHED To this class belong many of our c
 a most unpleasant manner Her tongue felt PARCHED for a cool drink She came upon him
What do you say Don Esteban moistened his PARCHED lips and with the back of his hand
 up boys Can you not see the gentleman is PARCHED Mayhew I have to get back They all
u like a drink B I d love some coffee I m PARCHED or B No thanks I ll have some coff
on the phone And I m Tegan Walker and I m PARCHED Could I get a water Of course Um f
st around Bombay across immense tracts of PARCHED and sun scorched table land where
faster fell the rain How lovely cried the PARCHED flowers as they raised their droop
```

Figure 1. A sample of concordance lines in BNC for "parched."

Learners in the TI group, in contrast, engaged in a series of traditional vocabulary-learning activities that were prepared before the study began. They were first given definitions of the target words in each class, were then asked to find the synonyms and/or antonyms of these words using a dictionary, and finally they did mechanical fill-in-the-blank exercises. To prevent teacher variability, the second author himself gave all the instructions throughout the study.

Data Collection Instruments and Procedure

A quasi-experimental pre-/posttest research design was implemented for the study, which provided the advantage of a mixed-method approach (Lynch, 1996; Mackey & Gass, 2005) both to measure, if any, learners' improvement in vocabulary knowledge after receiving the instructions and to explore their attitudes toward the instructional groups. To gather quantitative data, the ECPE C2 Michigan Proficiency Vocabulary Test was used, of which three different versions were produced for pre-/posttests to rule out possible effects of test item familiarity. All the versions of the test were piloted, and coefficient alpha reliability analysis found preferable levels of internal consistency (Cronbach's alpha .87, .85, and .85 for pre-, post-, and delayed posttest, respectively). The study commenced after the learners took the first version of the test (two weeks before the study) and the target words were specified. Each group received an equal number of classroom hours of instruction from the second author. As soon as the instructions had been given, the learners received another version of the same vocabulary test as the immediate posttest to determine whether the instructions were effective and whether they were able to learn the words. Three weeks later, this time to measure whether learners still remembered the words and/or still retained their meaning, they took the third version as the delayed posttest.

To determine further qualitative support, six learners in total (three from DDL, three from TI) were then randomly selected and invited for interviews. The format was semi-structured because it is recognized as being "less rigid" (Mackey & Gass, 2005, p. 173), thus enabling the researcher to have enough freedom to ask further questions and probe for new information. Some of the questions addressed in the interview were as follows:

1. How did you find the DDL- and/or TI-centered tasks?

2. Do you prefer to use DDL-centered tasks or traditional methods for developing your vocabulary? Why?

3. What did you like or dislike most in the instructional class?

4. What kind of difficulty did you have in the instructional period?

Data Analysis

Both quantitative and qualitative data analyses were carried out.

For quantitative analyses, a series of ANOVAs were conducted using SPSS 17 to measure group differences. In addition to the selection of those unknown words for the instructions after the pretest, to ensure that both instructional groups started at the same level of vocabulary knowledge, first, a one-way between-groups analysis of variance was run. It was revealed that the actual difference in mean scores between the groups was not statistically significant, $F(1, 70) = .241$, $p = .625$, thereby confirming that any greater performance gained in the posttests was due to the instructional types at the end (DDL and TI in this study). Next, a one-way repeated measures ANOVA was conducted to identify any significant difference between the instructional groups.

A grounded approach was adopted for the qualitative analyses of the students' comments in the interviews (Dörnyei, 2007). For the initial step of the analyses, after the interviews were all transcribed, students' comments were first listed to examine them for salient themes (i.e., the open coding stage). The lists made by the researchers/authors were then compared. After any types of disparities were resolved through negotiation, Miles and Huberman's (1994) formula was used to determine the inter-rater reliability of the analyses, which reached 96 percent. As the second step of the analyses, students' statements were consolidated around four central concepts (i.e., the axial coding stage). These were factors related to affective domain, sense of autonomy, facilitative effects of technology, and challenges faced. In this case, the overall agreement rate between the two raters was even higher because the responses of the learners were definite and the four themes were clear.

Results

Vocabulary Test Results

A one-way repeated measures ANOVA was conducted to compare scores of the instructional groups as measured by the Michigan vocabulary proficiency Test in the pretest, immediate posttest, and delayed posttest. The means and standard deviations are presented in Table 1, and they indicate a significant effect for time, Wilks's lambda

= .096, F (2, 70) = 325.80, p < .0005, multivariate partial eta squared =.90. The statistics reveal clearly that both the DDL and the TI instructional group increased their scores from the pretest to the posttest, and that the difference is statistically significant and the effect size very large.

Table 1

Descriptive Statistics for Vocabulary Knowledge of Instructional Groups in Tests

	Pretest		Posttest		Delayed Posttest	
	Mean	SD	Mean	SD	Mean	SD
Vocabulary Test						
DDL	24.99	13.07	65.19	14.13	58.77	17.40
TI	23.60	10.82	55.90	17.22	45.75	21.34

Note. DDL = 36, TI = 36.

From the descriptive statistics (Table 1 and Figure 2), it is clear that both instructional groups started with equal knowledge of the target words at the beginning of the study, which was statistically confirmed as well, and that both types of instruction helped learners learn the targeted words immediately after the instructions were completed, although the learners receiving the DDL-centered vocabulary tasks (M = 65.19, SD = 14.13) performed much better than those receiving the TI-centered vocabulary exercises (M = 55.90, SD = 17.22) in the immediate posttest. Regarding the delayed posttest mean scores of the groups to determine whether the learning was sustained over time, although retention rates decreased for both the DDL and the TI group (M = 58.77, SD = 17.40; M = 45.75, SD = 21.34, respectively), it was still remarkably higher for both groups when compared to learners' pretest performance. These results showed that the learners, especially those in the DDL group as compared to those in the TI group, were able to maintain their performance at a much higher level over a period of three weeks after completing the instructions.
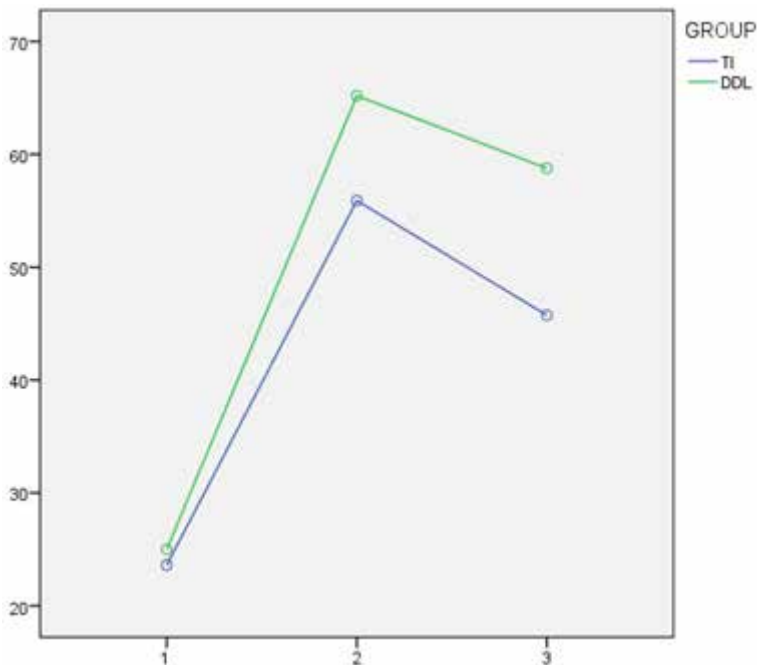
Figure 2. Estimated marginal means of the DDL and TI groups.

Interview Results

On completing the transcriptions and analyses of the qualitative data, the researchers categorized the ideas into four common themes: affective factors, sense of autonomy, facilitative effect of technology, and challenges faced.

Affective factors. Affective factors are emotional factors that may either help or hinder learning, as discussed within the scope of the field of second language development. The present study found the same results. In fact, learners had a general positive attitude toward DDL-centered tasks but a negative attitude toward TI-based activities.

For instance, learners from the DDL group made the following positive statements:

I found DDL very comfortable and relaxing, because when the teacher gave me a task to complete, he was not on my back to teach the new word, so it made it easier and more relaxed.

I enjoy the use of a computer while learning because it makes learning fun and exciting to me and so makes me less stressed.

I felt excited when I looked for the words in the DDL class, because it is a new method of learning and it can involve the use of a computer or smartphone, which helps me stay more focused.

> I liked learning vocabulary through technology as I was not stressed by the teacher.

> I felt happy because after making a guess, I was proud of myself, so it really improved my vocabulary.

However, learners in the TI group had negative feelings toward TI-based vocabulary-learning activities. For instance, one stated that he found the TI-based activities boring, frustrating, and tiring.

> I found it very stressful, because the teacher always imposed on us to learn the word and made us very stressed while learning.

Another felt the approach influenced his motivation in a negative way.

> I think the way the teacher explains and practices the new words with us is not motivating, because often we do not understand his definitions but only memorize them. This is frustrating, and repeating every word again and again makes me tired and bored.

Another student argued that as TI-based activities are mechanical, and the class dry, he forgot the things that he had learned easily.

Sense of autonomy. Autonomy is another important psychological concept that is studied a lot in the field of second language development. It means a learner's ability to take charge of his/her own learning process as he/she feels responsible for learning, which was found exclusively fed by DDL-centered tasks.

For instance, two learners stated the following:

> DDL tasks helped me learn word combinations. They taught me to be independent and to learn new things on my own without the teacher's help.

> I also liked DDL, because the teacher gave me freedom and left me to work on my own.

Another two learners stated the following:

> In the DDL class, I read as many sentences as I need to understand the word. I learn better, because I like learning by myself.

> I felt responsible because DDL helped me learn things by myself.

However, it was found that TI-based activities were not much favored as they were generally teacher centered, without giving much responsibility to the learners to do things independently. Illustrating this result, learners from the TI group expressed the following:

> Because the teacher demanded from us to learn those words in sample sentences without context, although we did our best to memorize them, I did not like it.

> The class was teacher dominated, without giving us much time to do things. He was teaching, but I want to learn myself.

Facilitative effect. Technology and its tools have facilitative effects in second language development as they provide learners multiple lenses or ways into the learning process, which was clearly apparent from learners' statements, particularly in the DDL group.

For instance, two learners stated the following:

Concordance line provided me so many examples so that I could learn how the words are used in different contexts.

DDL-centered tasks gave me straightforward, fast, and lucid answers.

Another two supported technology use and stated the following:

It was an enjoyable learning experience for me. I found it fun, because it involves computers and I really like to use computers.

I use a computer for games, to read books etc. I like learning from a computer.

Challenges. This theme emerged for both instructional groups. While learners in the DDL group found the use of the corpus program difficult at the beginning (although they liked it later), those in the TI group considered the activities boring and tiring.

For instance, learners from the DDL group stated the following:

It was quite hard for me to get around the website or the program at the beginning, but later I learned and found it rather interesting.

At first I found it difficult, and I resisted using it, but then my friends praised it, so I tried it and realized that it was actually a great learning experience.

Learners in the TI group stated, for instance, the following:

It is a very time-consuming method as some people did not understand the words, and the teacher made a lot of repetitions to explain the meaning of words himself.

The examples were without context and isolated. I felt bored.

Discussion

The present study set out to investigate whether DDL-centered tasks or TI-based activities are more effective in helping students develop vocabulary in a second language (first research question), and to explore students' attitudes in both instructional groups (second research question) when learning vocabulary in English. The results for the first research question indicated that, although both groups improved their performance from the pretest to the posttests, the performance of learners in the DDL group was superior to that of learners in the TI group, not only in the immediate posttest after the instructions but also in the delayed posttest three weeks after the instructions (see Figure

2). The results of the second research question were similar, lending much support to the greater effectiveness of the DDL group compared to that of TI. Put simply, the results in general revealed clearly that the DDL group outperformed the TI group in the tests and that the learners experienced more fun, more freedom, more relaxation, less boredom, less rigidity, and less teacher dominance when they started to learn vocabulary through the DDL approach. According to the results of this study, although they cannot be overgeneralized, as an overall theme it can be suggested that, although learners in the DDL group found cut-off sentences in concordancing lines ambiguous and thus felt puzzled at the beginning, they nevertheless easily accommodated later as the corpus program or DDL tasks provided an "innovative approach" (Thurstun & Candlin, 1998, p. 277) to vocabulary development in the second language and "a rich experience" (p. 277) with multiple examples related to words, word groups, or collocations in authentic examples.

Similarly, Schmidt (1990; 2001) concluded that "subliminal language learning is impossible, and that noticing is the necessary and sufficient condition for converting input to intake" (1990, p. 129), which might be done only "when the demands of a task focus attention on what is to be learned" (p. 129). The results reported in the present study seem to be consistent with Schmidt's (1994) noticing hypothesis, as the DDL tasks helped learners focus on the word itself in the concordancing lines (see Figure 1). That is, the DDL approach enhanced learners' awareness of the multiple uses of the words in context, which subsequently led to "a habit" (Yoon, 2008, p. 44) of looking up different meanings using the concordancing program.

Moreover, as seen in the results, learners in the DDL group developed greater autonomy than those in the TI group, as the DDL tasks encouraged learners to, as Dörnyei and Skehan (2003, p. 611) put it, self-regulate their own decision-making processes by/for themselves. They felt more independent and more responsible for their vocabulary development. Crabbe (1993, p. 443) claimed that "learning is more meaningful, more permanent, more focused on the processes and schemata of the individual when the individual is in charge." In other words, as learners get used to discover different meanings of words or structural forms in an inductive way from authentic data (i.e., corpus), they start to gain more autonomy, thus enabling them to use self-regulating strategies or gradually to control their own learning process. According to a recent study, Bozpolat (2016) found a positive relationship between self-regulated learning strategies and having high self-efficacy and high academic performance. In the interviews conducted in the present study, learners in the DDL group made similar arguments, lending greater support to DDL, such as "I felt responsible myself" and "DDL left me to work and to learn on my own." However, those in the TI group complained about teacher dominance over them when learning vocabulary. All these findings suggest that, regarding the teaching/learning

situation, when learners are taught how to regulate their own learning process rather than being introduced to rules or definitions in a deductive manner, they feel more autonomous, more responsible, and more efficacious in terms of what they can achieve independently.

Finally, the current study also observed the importance of technology use in the language classroom as today's generation employs technology from a young age and thus they are called "digital natives." Learners in the DDL group liked the corpus program as they were ready to study online using technological equipment. Therefore, it can be argued that DDL can be one of the ways of scaffolding or mediating the growth of learners' motivation in the class and/or toward tasks (Vygotsky, 1978). One of the learners, for instance, stated in the interview, "As I am good at computer use, I did not have any difficulty understanding the concordancing program." This fact can be explained by learners' online learning readiness level, which was based on Horzum, Kaymak, and Güngören's (2015) study that found a positive relationship between learners' online learning readiness and their motivation. Thus, as the learners who received DDL instruction in the present study were willing and competent enough to work using a computer tool, they were found to be motivated as well. Another learner in the interviews stated, "In an anxiety-free environment, I got help from the computer to learn and to discover for myself." In the words of Johns (2002, p. 108), learners in the DDL group showed a similarity to the fictional private detective Sherlock Holmes: DDL drove learners to act more like a linguistic researcher, an explorer, an observer, a hunter, and a detective. In other words, while identifying common patterns, chunks, or phrases regarding key words, the learners were, in fact, unconsciously engaged in "pattern hunting" (Geluso & Yamaguchi, 2014, p. 239).

To conclude, the results of the present study should be carefully considered when making generalizations, bearing in mind that the study set out to investigate whether there was any effect of DDL-centered or TI-based activities on both learning and retaining vocabulary in a limited context. The main results revealed that, although both instructional groups performed better in the posttests than the pretest, only the DDL group maintained substantially more over time what they had learned, as shown in the delayed posttest. In fact, this study showed that corpus-assisted vocabulary learning can be one of the alternative and "promising" pedagogical approaches (Çelik & Elkatmış, 2013, p. 1092).

Implications for the teaching/learning situation

Although research results generally may well be regarded for the setting(s) where the research is carried out, the results of the present study might still be extended to an EFL and/or ESL context with a similar teaching/learning situation to the one described in the present study. First, this study investigated any possible effectiveness

of corpus-based learning as "a tool" when teaching vocabulary in English as well as students' possible improvement in the DDL group. Because only words unknown to the learners were selected and taught through corpus-based learning activities, and the learners both learned and remembered significantly more in comparison to TI learners, it can be surmised that teachers teaching in similar settings can gain advantages from using corpus-assisted learning activities as they mediate learning (Yoon, 2008). Moreover, the study adopted a mixed-method approach, thus providing both test results and verbal data gathered through the interviews. Albeit there is a lack of an overall generalizability of the test results to other settings (this is difficult when true randomization is not achieved), the results of the present study based on the verbal interview data nevertheless seem consistent with the results of some earlier research studies carried out in the Turkish setting (e.g., Çelik, 2011; Özdemir, 2014; Tekin & Soruç, 2016; Uçar & Yükselir, 2015; Yılmaz & Soruç, 2015). Consequently, teachers teaching in an ESL setting and/or in an EFL context (such as Turkey) can likewise benefit from the following suggestions:

- English language teachers should find more different ways to increase learner autonomy so that learners can take charge of or self-regulate their learning process, which can be achieved to a greater extent through, among many other methods, the DDL approach.

- Vocabulary development can occur through both DDL- and TI-based activities. But bearing in mind the greater effectiveness of the former over the latter, especially in the delayed posttest as shown in this study, teachers can use concordancing programs more often to facilitate learners' vocabulary development.

- Teachers can use traditional ways of teaching vocabulary in the classroom as well, but the facilitative effect of technology and/or its tools should not be forgotten, and therefore teachers can at least give homework or projects to learners to complete at home independently and thereby be engaged in DDL (for instance, www.lextutor.ca).

- When introducing concordancing programs, teachers should keep in mind learners' language proficiency, (which was not a problem for the present study because Uganda is an ESL country), and learners' technological knowledge.

- As teacher-centered activities may generally lead to "dry" classes, language teachers should find alternatives to TI-based activities to make learning more engaging and less intimidating, for instance, captioned movie clips (e.g., Yüksel & Tanrıverdi, 2009) as well as DDL-centered tasks.

- Corpus-aided language learning activities should not be limited only to improving learners' word and/or grammar knowledge in the second language,

but rather they can be used effectively in the first language environment, for instance, to teach punctuation in Turkish (Çelik & Elkatmış, 2013).

- The online learning readiness level of learners is important and should be carefully raised by teachers, because the more time learners are engaged in corpora or related programs, the more motivated they feel (Horzum, Kaymak, & Güngören, 2015) and the more vocabulary they learn (Çelik & Keser, 2010).

Overall, this study showed that the corpus-based approach to vocabulary development seems to be effective for language learners. Corpora, or concordancing lines, as O'Keeffe et al. (2007) put it, have revolutionized traditional ways of learning second language vocabulary, thereby assisting learners in searching for different types of meaning or different styles of words and phrases (Reppen, 2010). And it is a well-known fact that the more words a learner knows, the more L2 reading comprehension occurs. Such a relationship has been established by another recent study conducted by Güngör and Yaylı (2016) in Turkey, which found that vocabulary coverage is a better precursor to L2 reading comprehension, and thus the authors suggest that as well as using traditional methods, students should be taught "how to benefit from corpus-approaches in the classroom" (p. 1185) to enhance their vocabulary coverage and their reading comprehension alike. Therefore, corpus-assisted vocabulary learning activities (or DDL) can be viewed as one of the alternative ways that L2 teachers can effectively use, at least for the sake of some learners' choices, when teaching either in an ESL context (the setting in which the present study was carried out) or in a similar teaching/learning situation.

Conclusion

This study explored the effectiveness of two instructional groups on second language vocabulary development. It found that, although both groups improved their performance relative to their level at the start of the study, those in the DDL group remembered more than those in the TI group at the end of the study. However, the circumstances in which the results were obtained should be carefully taken into account when considering the study's implications. First, the study was carried out in an ESL context with Ugandan high school learners in only 10th grade. Second, the instructional period was short, being limited to four classroom hours for each instructional group. Finally, as the study gathered data only from boys (as the school itself was a boys' high school), the results may not be generalizable to female learners. In the future, cross-sectional studies should be conducted involving different grade levels and considering gender as well. In addition, the incorporation of DDL-centered tasks into the teaching of language skills (e.g., writing) could be researched so that the use of corpora can be extended beyond lexis or grammar. Future studies could also be carried out to investigate the effectiveness of corpus-based learning on teaching

both grammar and vocabulary to the same group of participants. In line with the DDL approach, as well as developing materials and training teachers, learners should also be trained to understand how they can get the most out of corpora because it is like a portable teacher from which learners can, by themselves, learn words or word groups, discover authentic patterns, and receive corrective feedback.

References

Aşık, A., Vural, A. S., & Akpınar, K. D. (2016). Lexical awareness and development through data driven learning: Attitudes and beliefs of EFL learners. *Journal of Education and Training Studies, 4*(3), 87–96.

Aston, G., Bernardini, S., & Stewart, D. (Eds.). (2004). *Corpora and language learners* (Vol. 17). Philadelphia, PA: John Benjamins Publishing.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572.

Bozpolat, E. (2016). Investigation of the self-regulated learning strategies of students from the faculty of education using ordinal logistic regression analysis. *Educational Sciences: Theory & Practice*, *16*, 301–318. http://dx.doi.org/10.12738/estp.2016.1.0281

Çelik, S. (2011). Developing collocational competence through web based concordance activities. *Novitas-ROYAL (Research on Youth and Language)*, *5*(2), 273–286.

Çelik, S., & Elkatmış, M. (2013). The effect of corpus assisted language teaching on the learners' proper use of punctuation marks. *Educational Sciences: Theory and Practice*, *13*(2), 1090–1094.

Çelik, S., & Keser, H. (2010). The correlation between learners' logs of navigations through online corpora and lexical competence. *Ankara Üniversitesi Egitim Bilimleri Fakultesi Dergisi, 43*(2), 149–170.

Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *Tesol Quarterly, 34*(3), 548–560.

Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393–409). Mahwah, NJ: Lawrence Erlbaum.

Crabbe, D. (1993). Fostering autonomy from within the classroom: The teacher's responsibility. *System, 21*(4), 443–452.

Dönük, D. (2016). Revisiting language learning through the self: Discovery learning in the 21st century. *Turkish Online Journal of English Language Teaching, 1*(2), 66–74.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford, UK: Oxford University Press.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 589–630). Oxford, UK: Blackwell.

Flowerdew, L. (2012). *Corpora and language education*. New York, NY: Palgrave Macmillan.

Francis, W. N. (1992). Language corpora B. C. *Directions in corpus linguistics*: *Proceedings of Nobel Symposium*, *82,* 17–32.

Frankenberg-Garcia, A. (2012). Learners' use of corpus examples. *International Journal of Lexicography*, *25*(3), 273–296.

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL*, *26*(2), 128–146.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, *32*(3), 301–319.

Geluso, J., & Yamaguchi, A. (2014). Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL*, *26*(02), 225–242.

Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, *63*(4), 363–372.

Güngör, F., & Yaylı, D. (2016). The interplay between text-based vocabulary size and reading comprehension of Turkish EFL learners. *Educational Sciences: Theory & Practice, 16*, 1171–1188.

Hockly, N. (2013). Interactive whiteboards. *ELT journal*, *67*(3), 354–358.

Horzum, M. B., Kaymak, Z. D., & Güngören, Ö. C. (2015). Structural equation modeling towards online learning readiness, academic motivations, and perceived learning. *Educational Sciences: Theory & Practice, 15*(3), 759–770.

Huang, L. S. (2011). Corpus-aided language learning. *ELT journal*, *65*(4), 1–4. http://dx.doi.org/10.1093/elt/ccr031

Johns, T. F. (1991). Should you be persuaded: Two samples of data-driven learning materials. *ELR Journal, 4,* 1–16.

Johns, T. F. (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. 293–310). Cambridge, MA: Cambridge University Press.

Johns, T. F. (2002). Data-driven learning: The perpetual challenge. *Language and Computers*, *42*(1), 107–117.

Johns, T. F., & King, P. (1991). Classroom concordancing. *English Language Research Journal*, *4*, 27–45.

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, *5*(3), 77–90.

Kılıçkaya, F. (2015). Computer-based grammar instruction in an EFL context: Improving the effectiveness of teaching adverbial clauses. *Computer Assisted Language Learning, 28*(4), 325–340.

Kırkgöz, Y. (2006). Designing a corpus based English reading course for academic purposes. *Reading, 6*(3), 281–298.

Leel, H. (2011). In defense of concordancing: An application of data-driven learning in Taiwan. *Procedia-Social and Behavioral Sciences*, *12*, 399–408.

Lin, M. H., & Lee, J. (2015). Data-driven learning: Changing the teaching of grammar in EFL classes. *ELT Journal*, *69*(3), 264–274.

Lynch, B. K. (1996). *Language program evaluation*. Cambridge, MA: Cambridge University Press.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. New Jersey, NJ: Lawrence Erlbaum Associates.

Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University press.

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge, MA: Cambridge University Press.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). California, CA: Sage.

Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art - and beyond. In S. Braun, J. Mukherjee, & K. Kohn (Eds.), *Corpus technology and language pedagogy*: *New resources, new tools, new methods* (pp. 5–24). Frankfurt: Peter Lang Publishing Group.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom*. Cambridge, MA: Cambridge University Press.

Özbay, A. S., & Kayaoğlu, M. N. (2015). EFL teacher's reflections towards the use of computerized corpora as a teaching tool in their classrooms. *Çukurova University Journal of Education Faculty, 44*(1)*,* 85–104.

Özdemir, N. O. (2014). Using corpus data to teach collocations in medical English. *Journal of Second Language Teaching & Research, 3*(1), 37–52.

Peng, H., Su, Y. J., Chou, C., & Tsai, C. C. (2009). Ubiquitous knowledge construction: Mobile learning re-defined and a conceptual framework. *Innovations in Education and Teaching International*, *46*(2), 171–183.

Quinn, C. (2014). Training L2 writers to reference corpora as a self-correction tool. *ELT Journal, 69*(2), 165–177.

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

Rezaee, A. A., Marefat, H., & Saeedakhtar, A. (2014). Symmetrical and asymmetrical scaffolding of L2 collocations in the context of concordancing. *Computer Assisted Language Learning*, *28*(6), 532–549.

Şad, S. N. (2008). Using mobile phone technology in EFL classes. *English Teaching Forum*, *46*(4), 34–39.

Şad, S. N., & Göktaş, Ö. (2014). Preservice teachers' perceptions about using mobile phones and laptops in education as mobile learning tools. *British Journal of Educational Technology*, *45*(4), 606–618.

Saran, M., Çağıltay, K., & Seferoğlu, G. (2008, March). *Use of mobile phones in language learning: Developing effective instructional materials*. Paper presented at the 5th International Conference on Wireless, Mobile and Ubiquitous Technologies in Education (pp. 39–43). Beijing, China: Institute of Electrical and Electronics Engineers (IEEE).

Saran, M., Seferoğlu, G., & Çağıltay, K. (2009). Mobile assisted language learning: English pronunciation at learners' fingertips. *Eurasian Journal of Educational Research*, *34*, 97–114.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–58.

Schmidt, R. W. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 165–210). London, UK: Academic Press.

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

Smart, J. (2014). The role of guided induction in paper-based data-driven learning. *ReCALL*, *26*(02), 184–201.

Soruç, A. (2015). What makes redundant presentation of multimedia learning difficult? *Procedia-Social and Behavioral Sciences, 191*, 2844–2848.

Starfield, S. (2004). Why does this feel empowering? Thesis writing, concordancing and the corporatizing university. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 138–157). Cambridge, MA: Cambridge University Press.

Sun, Y., & L. Wang. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning, 16*(1), 83–94.

Tekin, B., & Soruç, A. (2016). Using Corpus-assisted learning activities to assist vocabulary development in English. *TOJET: The Turkish Online Journal of Educational Technology*, *Special Issue*, 1270–1283.

Thurstun, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, *17*, 267–280.

Uçar, S., & Yükselir, C. (2015). The effect of corpus-based activities on verb-noun collocations in EFL classes. *TOJET: The Turkish Online Journal of Educational Technology, 14*(2), 195–205.

Ünaldı, I., Bardakçı, M., Akpınar, K. D., & Dolas, F. (2013). A comparison of contextualized, decontextualized and corpus-informed vocabulary instruction: A quasi-experimental study. *Journal of Language and Literature Education, 2*(8), 78–95.

Uysal, H., Bulut, T., & Al Hosein, Y. (2013). Using concordances as supplementary materials in teaching grammar. *Studies of Foreign Languages, 22,* 113–118.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wu, S., Witten, I., & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL*, *22*(1), 83–102.

Yılmaz, E., & Soruç, A. (2015). The use of concordance for teaching vocabulary: A data-driven learning approach. *Procedia-Social and Behavioral Sciences, 191*, 2626–2630.

Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, *10*(3), 130–139.

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.

Yüksel, D., & Tanrıverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *TOJET: The Turkish Online Journal of Educational Technology*, *8*(2), 48–54.