Novitas-ROYAL (Research on Youth and Language), 2011, 5(2), 273-286.

DEVELOPING COLLOCATIONAL COMPETENCE THROUGH WEB BASED CONCORDANCE ACTIVITIES

Serkan ÇELİK*

Abstract: The last decade has witnessed a strong impact of corpora and corpus linguistics on the methodology of language pedagogy. Corpora based language learning claims to provide a discovery based authentic learning environment. However, the research in corpora based language instruction (data-driven learning) is still in progress and needs more effort to draw promising implications for EFL / ESL settings. This quasi-experimental research, which is a comparison between two experimental conditions, was conducted to investigate the effects of data-driven learning (DDL) on EFL learners' achievement and retention of lexical competence comparing to dictionary use. Instruction was delivered to the participants of the study through a learning management system (Moodle). A collocation test was developed to gather data through ANCOVA and descriptive statistics. The results of the study revealed that pre and posttests did not show a significant difference between the two experimental groups. A later 'retention' test did show that the corpora-based learning group had a higher level of retention. Some suggestions for further research were discussed based on the findings.

Keywords: Data-driven learning, Corpus, Concordance, Collocational Competence

Özet: Geride bıraktığımız 10 yıllık süreçte, derlem dilbilimin dil eğitim yöntemleri üzerindeki güçlü etkisinin alan yazında yoğun olarak tartışıldığı gözlemlenmiştir. Derlemlere dayalı dil öğrenme yaklaşımları, keşfe dayalı ve özgün bir dil öğrenim ortamı sunabilme iddiası taşımaktadır. Bununla birlikte, derleme lere dayalı dil öğretim araştırmaları İngilizce'nin yabancı ve/veya ikinci dil olarak öğrenimi noktasında halen yeterli ve etkili çıkarımlar oluşturabilmiş değildir. İki deneysel koşulun karşılaştırıldığı bu çalışmada, İngilizce'yi yabancı dil olarak öğrenen öğrencilerin sözcüksel gelişimi üzerinde derleme dayalı dil öğretim tekniklerinin etkisi çevrimiçi sözlük kullanım yöntemi ile karşılaştırılarak incelenmiştir. Öğrenme içeriğinin bütün katılımcılara Moodle öğrenme yönetim sistemi üzerinden çevrimiçi olarak sunulduğu çalışmada, veriler bir ilişkisel sözcük testi aracılığıyla elde edilmiştir. Verilerin analizinde Kovaryans ve tanımlayıcı istatistik yöntemleri kullanılmıştır. Çalışmanın sonuçlarına göre, deney ve kontrol gruplarının ön ve son test uygulamaları arasında istatistiki olarak anlamlı bir fark bulunmamıştır. Bununla birlikte, uygulanan kalıcılık testi sonuçları, derleme dayalı dil öğretim etkinliklerinin kullanıldığı grubun daha başarılı olduğu belirlenmiştir. Bulgular ve sonuçlar doğrultusunda geleceğe dönük bir takım araştırma önerilerinde bulunulmuştur.

Anahtar Sözcükler: Veri Yönlendirmeli Öğrenme, Derlem, Derlemler, İlişkisel Sözcük Yeterliği

Introduction

Emerging technologies in the field of language pedagogy have provided a wide scope to language researchers, learners and instructors in terms of embedding technology into the language learning curriculum. Computer assisted language learning aims to enhance the learning environment, meet individual learning requirements, enrich learning experiences, and diminish the conventional role of the teacher by overcoming the restrictions of traditional instruction. Hence, it provides language teaching practitioners with innovative points of view for more effective language learning environments (Kasapoglu-Akyol, 2010). This paper is intended to determine how an online language course including concordance activities can enhance the quality of vocabulary learning by focusing on collocational competence.

* Kırıkkale University, Kırıkkale, Turkey, sercelikan@yahoo.com

Language Learning and Technology

Over the past decade or so, all disciplines of the English Language such as phonetics, lexicology, graphology, grammar, and discourse analysis have been deeply influenced by technological developments. Crystal (1995) points out that development in electronic and computer sciences have changed the perspective toward language pedagogy. He underscores the research possibilities gained by improvements in processing and analyzing huge language texts through computer systems.

One of the most crucial contributions of the computer sciences to language pedagogy has been observed within applied linguistics in constructing, processing, and analyzing language corpora (Johns, 1990). The term corpora refers to the electronic authentic language database(s) available on the Internet or stored in personal computers (Hasselgård, 2001; McEnery & Wilson, 2001). It is also defined as the body of e-texts considered to be representative or a sample of a language (Ball, 1996) or a collection of materials that has been made for a particular purpose such as texts being analyzed for their linguistic features (Richard, Platt & Platt, 1992). Language corpora can be either written texts using extracts from newspapers, business letters, popular fictions, books, or magazines, published or unpublished school essays, etc.; or spoken texts involving any recorded and transcribed formal or informal conversations, radio and TV shows, weather broadcasts, business meetings, or even birthday parties, etc.

Use of corpora and concordancing in language learning is relatively novel (Stevens, 1995). To access or make use of a corpus, one should use a concordance to look at linguistic patterns. The contemporary version of concordancing includes a software program that makes it possible to analyze all instances of a linguistic form or structure in a corpus with the context in which the words appear. When a word needs to be examined, for example, the program scans the texts in its storage, locates all the occurrences of the word under examination, and lists these words on the screen in a list form within their immediate context (Barlow, 1996). These compiled lists are called concordance lists which enable teachers and learners to examine words in their natural contexts (Biber, Conrad, & Reppen, 1998; Sinclair, 1991; Tribble & Jones, 1990)**,** so that they can see how they collocate with other words, which patterns they follow, which prepositions they go with, and so on (Willis, 1990). Tribble and Johns (1990) point out that a concordance can be utilized to find instances of authentic usage to demonstrate features of vocabulary, collocations, grammar points or even the structure of a text, to generate exercises based on examples drawn from a variety of corpora. Furthermore, using concordance software, language learning may be more learner-centered; learners can be motivated to discover new meanings and to examine lexical and grammatical collocations. Johns refers to learners studying concordance lists as "language detectives" (1997, p. 101) whose task is to discover the rules of the language they are focusing on by finding, identifying and inferring these linguistic implications from context. In this respect, as the proponents of data-driven learning (e.g. Johns, 1994), concordances are superior to conventional grammar books, dictionaries and course books, because they provide easy access to huge amounts of `authentic' language in use, foster the learners' analytical capacities, promote their explicit knowledge of the L2, facilitate critical language awareness, and support the development of learner autonomy (Gabel, 2001). Similarly, Cobb (2003) suggests that language learning is more likely to happen when adequate examples are noticed and processed by learners.

Johns (1990; 1991) defined all of the corpora and concordance based activities as data-driven learning (DDL). DDL implies that a learning activity using concordance outputs can be used by the learners to derive the different features of a word, in terms of semantics and grammar

in various contexts. Hunston (2002, p.170) also points out that, "DDL involves setting up situations in which students can answer questions about language by studying corpus data in the form of concordance lines or sentences".

Kennedy (2003) explains the second or foreign language acquisition process as depending on consciousness raising and in which learners should be exposed to authentic language materials as much as possible. However, the vocabulary instruction methods designed through rules and identical meanings hardly help students to foster their language production. Johns (1991) claims that the traditional dictionary use model is both tiring and not productive as a means of vocabulary learning. Furthermore, while listing the usages of words; traditional dictionaries cannot illuminate how learners can realize the various meanings of a single word. Traditional dictionaries often provide vague, limited, and artificial examples for each definition, which is insufficient to fully understand an unfamiliar word. Respectively, traditional language learners are addicted to dictionaries as the main source for looking up word definitions and examples; however, this task is often too laborious and time-consuming (Cobb, 2003). By using the concordance tool of corpora to search for word contexts, learners are involved in a more speedy and efficient language learning experience. Chen (2004) posits that integration of corpora into vocabulary classrooms not only provides learners with faster searching tools and better quality of contexts than traditional dictionaries which are not likely to achieve but enhance their learning motivation. Sinclair (1997) and Granger (1998) assert that DDL should best be exploited to teach collocations which are sometimes not easily recognized by the native speakers as well. Halliday (1993) underlines the efficient role of corpora in language learning in terms of providing the most reliable data on deciding which grammar rules should be taught first. Biber et al. (1998) indicate that corpus and concordance studies revealed a new viewpoint to language pedagogy and its stakeholders by taking the communicational aspect of language rather than its theoretical structure.

Collocations and Data-Driven Learning

The data driven learning approach tries to provide language teachers and learners with a new perspective to raise lexical competence. This new perspective requires focusing on collocations which are defined as a group of co-occurring words (Bahns 1993; Aston, 1995; Kita & Ogata 1997; Shei & Pain 2000; Altenberg & Granger 2001; Hoey 2000, 2003; McAlpine & Myles 2003; Nesselhauf 2003; Chen, 2004;).

One of the most frequent mistakes in language learning appears through the use of collocations (Dundley-Evans, 1994; McAlpine & Myles, 2003). Gui and Yang (2002) also found out that the mistake in collocations was the most dominant mistake that students face in their study conducted with Chinese EFL students. Altenberg and Granger (2001) and Nesselhauf (2003) show that even students learning English at advanced levels have problems with collocations. During the process of learning, English collocational incompetency was also observed to be one of the main problems for native Turkish speaking students (Koç, 2006). One of the possible explanations for this problem is that students have difficulty with automating the use of words. In conclusion, these studies prove that students need more detailed information on collocations, (McAlpine & Myles, 2003). Corpora and concordance tools can be used to determine the collocational relationships among words. Moreover, corpora based research may present more reliable and quantitative data compared to the individual studies (Hunston, 2002).

Related Studies

Little research in the literature focuses on data-driven learning, though they generally use printed concordance outputs as in-class materials to assess their effects on vocabulary development (Gaskell & Cobb, 2004). On the other hand, any research on corpora and concordancing was identified as utilizing online sources as a learning platform as in the current study. Data-driven learning implies that providing language learners with authentic usage of language through internet based corpora may support a constructivist language learning atmosphere.

In her study, Koç (2006) points out that university-level Turkish EFL students are incompetent in terms of their lexical and collocation knowledge. In conclusion, traditional language instruction methods which heavily depend on grammar and artificial learning materials are not functional in terms of using language on a communicative and productive basis. Anğ (2006) adopted a sight into corpus research by means of a computer-aided approach and further inquired into how influential concordance is acting on first-year learners' learning of collocative knowledge in the department of English Language Learning at Bogazici University. Data was not only confined to the participant views on the use of concordance. The software itself can also deal with editing abstracts, and the use of language clichés in introductory lines of rhetorical steps in a research paper, as well as trying to shed light on the credibility of producing such texts.

In his think-aloud sessions with two groups of tertiary learners at Bilkent University, Kurtul (1999) investigated the differences of two teaching methods in which some words are dragged from long-term memory more often along with contextual references, and some other words lack a proper context between the two groups of participants. Only average Pre and post test results have been used to determine the discrepancy between groups. Kivanc (2003), in a study conducted through face-to-face interactions, displays some similarities with Kurtul (1999) in terms of purpose and data collection tools. Kivanc (1999) has thoroughly studied the differences between contextual and non-contextual vocabulary learning activities and tried to determine the relationship between learners' vocabulary learning habits and strategies, perceptual learning choices, and their native language competence.

Cobb (1997), in his attempt to handle the issue of how it is possible to obtain measurable findings from vocabulary acquisition from concordance output software, discusses to what extent concordance output software relies on vocabulary learning in an offline setting at the University of Kaboos, in Omman. Another research paper on vocabulary teaching based on concordance output software in an online environment was done by Sun and Wang (2003) in Taiwan, with a group of 81 junior high participants. Researchers focused on associated vocabulary acquisition and how it is acted on by three different online concordance sites. Research findings suggest that the experimental group’s score is much higher than that of their counterparts in terms of high-frequency words; however, there seems to be no discrepancy between the two groups when students cope with low-frequency words.

Supatronant (2005) elaborated on the purpose of how concordance output software is in direct relation to definitive and productive knowledge of learners at engineering departments, and made use of pre and post tests, monitoring the tests so as to collect data; interview profiles and a student questionnaire for concordance output software also proved beneficial. Supatronant (2005) reported that students achieved greater progress in operating concordance output software. As for their attitudes towards learning, students stated that despite its complicated and extraordinary use, they were content with the end-result. Kennedy and

Mitceli (2001) attempted to figure out how effectively intermediate learners can analyze an Italian corpus with no tutoring at all; however, in light of pre and post evaluations, leaving learners with the corpus alone and expecting them to properly make use of it is not plausible. Kenning (2000) claimed that complicated outputs of the software could be used for spotting problematic patterns in language learning, and through these outputs it would be possible to identify these patterns and incorporate them into a more detailed syllabus. Todd (2001) also aimed at developing the writing skills of graduate students via concordancing and pointed out that concordance output software and corpus use would not be enough to shed light on checking punctuation, making inferences regarding the correct usage of these mistakes and being able to produce correct patterns, but that it would be wise to consider independent variables and structural patterns.

Methodology

This quasi-experimental study reports on the effects of web based concordancing activities (Data-driven learning) on Turkish EFL learners' acquisition of academic words and prepositional phrases in comparison to the online dictionary use method. The dependent variable of the study is learners' achievement on a collocation test developed by the researchers. The independent variables are two experimental groups who study using concordancing activities and an online dictionary. Instruction was delivered to both experiment groups through the Moodle learning management system. The first experimental group delved both into concordance lines selected by the researcher and also searched the Contemporary Corpus of American English (COCA) in order to attain the meaningful deductions. The second experiment group studied the same vocabulary via an online dictionary (The Longman Dictionary of Contemporary English Online) chosen according to the general preferences of ESL students enrolled at Boston University, where the researcher had a Fulbright fellowship.

Research Questions

The research questions of the study are given below:

1. Do web-based concordancing activities have any effect on language learners' collocational competence in comparison to online dictionary use method?

2. Do web based concordancing activities have any effect on the retention of language learners' collocational competence in comparison to the online dictionary use method?

Participants

Participants in the course were 68 students enrolled at the Ankara University School of Medicine. These learners were supposed to have a two-hours-per-week English course, during which this experiment was administered at the faculty lab designated for this purpose.

Data Collection Instrument

A multi-faceted procedure was followed throughout the vocabulary selection and test development processes. Five English language instructors were requested to rate the 570 words belonging to Coxhead's (2000) Academic Word List (AWL) and 130 collocations which were tagged as prepositional phrases through the Contemporary Corpus of American English. These instructors evaluated each of the academic words and prepositional phrases according to Eichholz and Barbe's (1961) scale of a word`s familiarity to the students.

Not Familiar..0…….1…….2……|….3…….4…….5..Familiar
Critical Point

After computing evaluation results through Excel, the means of the five instructors' rates were defined. Those means of these two word types were then listed on an increasing scale, and the first 15 academic words and first 10 prepositional phrases were perceived as the study words. Two measurement and assessment experts and two English language instructors were asked to review the collocational knowledge test in order to meet the requirements in terms of face and construct validity. Finally, a native English speaker with a PhD in language pedagogy reviewed the test.

The collocational knowledge test, which was designed and developed by the researchers, was utilized to assess learners' collocational competence on prepositional phrases. The test required participants to define the most widely used right collocate of the given words or prepositional phrases. All the correct answers and distracters used in the instrument were defined according to searches done through the British National Corpus (BNC). All the correct collocates of the words and phrases were analyzed and the most widely used one was selected as the correct answer. Respectively, all the distracters were checked through all the online corpora and the search engines (Google, Yahoo, etc.), and it was noted that there is no collocational use between the word(s) and the distracter. The piloting of the collocational knowledge test was done with 135 EFL students and a .857 Kuder-Richardson (KR-20) value was obtained. According to the item analysis of the piloting results, only one item was excluded from the test. Pre and post tests were administered one week before and after the experiment, which took five weeks; the exact post test form was also administered as a retention test three weeks after the post test administration.

Procedures

The experiment included two participant groups and two online courses through the Moodle learning management system. The first group (Experiment 1, E1) also called as DDL (Data-driven Learning) course, studied the words and prepositional phrases through concordance and corpora based activities, while the second group (Experiment 2, E2), also labeled as ODU (Online Dictionary Use) course, utilized an online dictionary to acquire the academic words and phrases.

Since both courses were designed to be taught for five weeks, there were five sessions in each course. At each of these sessions, 10 identical words were taught. The piloting of the study, conducted with four learners having the same course as the original study participants, suggested that all sub sections require nearly 80 minutes to complete the specific activities. Each of these weekly programs includes three main activity types. The initial part of the sessions was a learner guideline describing what the learners were supposed to do during the session and giving them the list of the words they should study. Moodle allows its users to link to any site from the Internet. Thus, the second part of the sessions was unique in terms of experimental groups. While the second section of the DDL course included a concordance web page designed by Mark Davies, a corpus linguist working at Brigham Young University, the second section of each ODU course sessions included a query page of Longman English Dictionary Online. At this point, the DDL course had numerous online corpora; however the learners were asked to work with only the Corpus of Contemporary American English, which

is probably the biggest corpus in the world with nearly 400 million words. This web page is also a significant example of an online data-driven learning environment, because it provides users and language researchers with many linguistic search opportunities and capabilities such as restricting your queries with text and genre types and production time.

The third part of the sessions in the courses was designed to assess the students in terms of definitional and productive vocabulary knowledge types. In addition, all reliability and validity issues of these tests were checked and covered before putting them into use. In this respect, after .916 and .790 Kuder-Richardson (KR-20) values were obtained for the definitional and productive vocabulary knowledge tests, items that were not over .20 were excluded from the final analysis of the data. For definitional knowledge, a matching exercise was developed and implemented into Moodle via a quiz module. The second exercise was a cloze test aimed at evaluating learners' vocabulary knowledge transfer skills for new contexts. In the vocabulary knowledge test, learners are expected to match the given words with the correct definition. The second quiz type was a productive vocabulary knowledge test; learners were asked to find the correct word to complete the sentences. All of the test scores of the learners in each group were stored via the grades section of Moodle and converted into Microsoft Excel files to pursue further analysis.

Results and Discussion

The data gathered throughout the study were analyzed using the Statistical Package for Social Sciences (SPSS version 17.0). Since the main research goal of the current study was to determine the effect of treatment on independent variable, the most convenient statistical method was decided as Covariance Analysis (Ancova). The pre-test of the experimental groups were taken as the covariate of the study. In addition, the statistical assumptions of the Ancova were checked before initiating the final analysis process and assured to pursue the analysis.

The first table depicts the results of descriptive statistics of the pre, post and retention test results of the collocation test. The participant groups are labeled as E1 (Data-driven Learning Group) and E2 (Online Dictionary Use Group). According to the means of the pre-test results, both groups performed at a very similar rate (E1: $\overline{X}$ = 5.56 / E2: $\overline{X}$ = 4.97). These results were meaningful, as they revealed that there was no significant discrepancy between the groups pre-experimentally in terms of pre-learning backgrounds. However, the means of the post-test results showed a positive variance in favor of group E1 which studied through corpora and concordance based activities (E1: $\overline{X}$ = 15.53 / E2: $\overline{X}$ = 14.38). On the other hand, the retention test results pointed out that while there was a decrease in both groups in comparison to those of the post-test, the decrease in the group studying through data-driven learning was less than the one studying through the online dictionary (E1: $\overline{X}$ = 14.31/ E2: $\overline{X}$ = 11.38). The obtained descriptive statistical findings are illustrated in Table 1.

Table 1. Descriptive Statistics of Pre-test, Post-test and Retention Test Results of Collocation Test

Group		Pre-test	Post-test	Retention Test
E1	$\overline{X}$	5.56	15.53	14.31
	Sd	3.31	3.64	3.44
E2	$\overline{X}$	4.97	14.38	11.38
	Sd	3.62	3.77	3.70

Descriptive statistics of the pre, post and retention test results in terms of lexical type are presented in Table 2. When the pre, post and retention test results of the academic words and prepositional phrases were examined, it was found that the means of both groups were very comparable (E1: $\overline{X}$academic words(aw): 3.68 / $\overline{X}$prepositional phrases (pp): 1.87), (E2: $\overline{X}$aw: 3.11 / $\overline{X}$pp: 1.85). The means of post-test results indicated a positive variance in favor of group E1 (E1: $\overline{X}$aw: 9.96/ $\overline{X}$pp: 5.56), (E2: $\overline{X}$aw: 9.11/ $\overline{X}$pp: 5.26). On the other hand, despite the decline in the means of retention test results of both groups as compared to those of post-test results, the study revealed that the decline in the group studying through data-driven learning activities approach (E1) was less than the one studying through online dictionary use (E2) (E1: $\overline{X}$aw: 9.59 / $\overline{X}$pp: 4.71), (E2: $\overline{X}$aw: 7.52/ $\overline{X}$pp: 3.85). In other words, it can be concluded that the maintenance of learning through a data-driven approach has a more long-term effect on learning. The findings obtained are shown in Table 2.

Table 2. Descriptive Statistics of Pre, Post and Retention Test Results in terms of Lexical Type

Group		Pre-test (aw)	Pre-test (pp)	Post-test (aw)	Post-test (pp)	Retention (aw)	Retention (pp)
E1	$\overline{X}$	3.68	1.87	9.96	5.56	9.59	4.71
	Sd	2.27	1.43	2.41	1.60	2.32	1.54
E2	$\overline{X}$	3.11	1.85	9.11	5.26	7.52	3.85
	Sd	2.57	1.39	2.62	1.62	2.63	1.59

The means of the post-test results the groups obtained from the overall collocation test regardless of lexical type are presented in Table 3. During this study, it was observed that the means of the post-test results ($\overline{X}$= 15.39) of the learners studied through data-driven learning approach (E1) were higher than those of online dictionary use group ($\overline{X}$= 14.50).

Table 3. Descriptive Statistics of Pre and Post Test Results

Group	N	Means of Pre-test	Std. Deviation	Means of Post-test	Std. Deviation	Means of Corrected Post-test
E1	32	5.56	3.31	15.53	3.64	15.39
E2	34	4.97	3.62	14.38	3.77	14.50

The fourth table aims to explain the covariance analysis results of the post-test results in comparison with the pre-test results. A covariance analysis was conducted to determine whether the observed variance in the means of the post-test results was statistically significant. The results revealed that the variance observed in favor of the data-driven learning group in the corrected means of post-test results was not significant. {F (1,63) = 1.114, p>.05}. Respectively, the variance observed in favor of the data-driven learning group was very comparable with the other group (.892). However, when the corrected means of post-test results were assessed, a distinctive increase (nearly ten points) was found in post-test results as compared to pre-test results in both groups. Thus, it can be commented that regardless of lexical type, the groups' learning performance during the practice period in the context of collocation was positive.

Table 4. Covariance Analysis Results of Post-test Results in comparison with Pre-test Results

Source	Sum of Squares	Sd	Means of Sq	F	p
Corrected Model	166.813	2	83.407	7.130	.002
Pre-test	145.054	1	145.054	12.400	.001
Group	**13.033**	**1**	**13.033**	**1.114**	**.295**
Error	736.944	63	11.698		
	903.758	65			

The means of the post-test results of prepositional phrases of collocation test indicated a positive variance in favor of the data-driven learning group ($\overline{X}$= 5.55) as compared to the online dictionary use group ($\overline{X}$= 5.27) (Table 5).

Table 5. Descriptive Statistics of Pre and Post Test Results in terms of Prepositional Phrases

Group	N	Means of Pre-test	Std. Deviation	Means of Post-Test	Std. Deviation	Means of Corrected Post-Test
E1	32	1.87	1.43	5.56	1.60	5.55
E2	34	1.85	1.39	5.26	1.62	5.27

The covariance analysis results (Table 6) obtained from the post-test results of prepositional phrases within the collocation test by taking pre-test results as the covariate (to determine whether the variance in the means of post-test on prepositional phrases of collocation test was statistically significant) proved that the difference in the corrected post-test results observed in favor of data-driven learning group was not statistically significant {F (1,63) = .642, p>.05}. Although a slight mean variance (.286) was observed in favor of the data-driven learning group, a clear increase was not found in the corrected means of the post-test in both groups as compared to those of the pre-test. The increase rate (of means) in both groups was nearly four points. Thus, it is obvious that the groups' learning performance in terms of collocational competence was successful during the practice period.

Table 6. Covariance Analysis Results of Post-test Results in terms of Prepositional Phrases in comparison with Pre-test Results

Source	Sum of Squares	Sd	Means of Sq	F	p
Corrected Model	35.117	2	17.558	8.327	.001
Pre-test(Reg.)	33.655	1	33.655	15.961	.001
Group	**1.353**	**1**	**1.353**	**.642**	**.426**
Error	132.838	63	2.109		
	167.955	65			

Table 7 depicts the descriptive statistics of pre, post, and retention test results of the collocation test. The means of the retention test results of the collocation test indicated a variance in favor of data-driven learning group ($\overline{X}$= 13.99) compared to online dictionary use group ($\overline{X}$= 11.67).

Table 7. Descriptive Statistics of Pre, Post, and Retention Test Results of Collocation Test

Group	N	Means of Post-test	Std. Deviation	Means of Retention Test	Std. Deviation	Corrected Means of Retention Test
E1	32	15.53	3.64	14.31	3.44	13.99
E2	34	14.38	3.77	11.38	3.70	11.67

The covariance analysis (Table 8) conducted on retention test results by taking pre and post test scores (to determine whether the variance in the means of post-test on prepositional phrases of collocation test was statistically significant) proved that the difference in the corrected retention test results observed in favor of data-driven learning group was statistically significant {F (1,62) = 10.031, p<.05}. Moreover, a mean difference of more than two points was observed in favor of data-driven learning group (.231). In other words, the experiment showed that the students using the data-driven learning approach had a higher level of retention in terms of collocation in comparison with those using an online dictionary.

Table 8. Covariance Analysis Results of Retention Test in comparison with Pre and Post Test Results

Source	Sum of Squares	Sd	Means of Sq	F	p
Corrected Model	428.322	3	142.774	16.573	.001
Pre-test (Reg.)	79.369	1	79.369	9.213	.004
Post-test (Reg.)	91.234	1	91.234	10.590	.002
Group	**86.419**	**1**	**86.419**	**10.031**	**.002**
Error	534.118	62	8.615		
	962.439	65			

The means of groups' retention test results attained from the prepositional phrases within the collocation test pointed out a variance between groups in favor of data-driven learning group compared to online dictionary use group (Table 9).

Table 9. Descriptive Statistics of Post and Retention Test Results

Group	N	Means of Post-test	Std. Deviation	Means of Retention Test	Std. Deviation	Corrected Means of Retention Test
E1	32	5.56	1.60	4.71	1.54	4.67
E2	34	5.26	1.62	3.85	1.59	3.89

The covariance analysis (Table 10) conducted on the retention test results of prepositional phrases within the collocation test by taking pre-test and post-test results (to determine whether the variance in the means of post-test on prepositional phrases of collocation test was statistically significant) proved that the difference in the means of corrected retention test results observed in favor of data-driven learning group was statistically significant {F (1,62) = 10.078, p<.05}. In other words, the study showed that the students using the data-driven learning approach had a higher level of retention in learning in comparison with those using the online dictionary in the context of prepositional phrases oriented collocation.

Table 10. Covariance Analysis Results of Retention Test Results in comparison with Pre and Post Test Results

Source	Sum of Squares	Sd	Means of Squares	F	p
Corrected Model	45.012(a)	3	15.004	7.378	.001
Pre-test	10.022	1	10.022	4.928	.030
Post-test	7.987	1	7.987	3.928	.052
Group	**10.078**	**1**	**10.078**	**4.956**	**.030**
Error	126.078	62	2.034		
	171.091	65			

The main implication derived from the data analyzed revealed that learners who studied through online concordance activities performed better at defining collocational relations among words than those studied through online dictionaries. This outcome can be understood by the facts that concordance based activities lead learners to do research, give them their own learning responsibilities (Johns, 1988; 1990; 1991) and expose them to authentic language (Mindth, 1996). The results of this study also support the theoretical background of the data-driven learning approach positing that corpora and concordance based vocabulary learning activities have a positive role in enhancing lexical competence (Widdowson, 1990; Coady, 1997a; 1997b).

The findings of the study are in a broad compatibility with the similar research in the field of study (Stevens, 1995; Kurtul, 1999; Supatranont, 2005; Ang, 2006). However, the main difference between the current study and Ang's (2006) work is that Ang exploited the concordance outputs for the sake of improving language learners' writing skills rather than vocabulary. In brief, both of the studies underline the valuable contribution of concordance use on collocational development. The current study also reached the same conclusions as Cobb's (1997) study which was conducted in f2f learning environment. The results of the current study also promote the relevant literature on the positive effect of incorporating authentic language materials into language curriculum. (Johns, 1991; Cobb, 2003).

Conclusion

In conclusion, both of the experiment groups performed better in terms of collocational competence according to the statistically positive difference between pre and post test results. However, the post test results showed no significant difference between the groups. The retention test results indicated that data-driven learning group performed better than online dictionary use group. The potential implications of data-driven learning to language pedagogy are something that would have been unthinkable even a decade ago. The software, online corpora and freely available programs have only recently appeared on the Internet, and a whole new world of opportunity is available for all teachers and students to explore.

This paper has focused on the importance of a data-driven approach to vocabulary instruction. The discussions raised point out the role and place of explicit vocabulary instruction in vocabulary development, as well as issues of possible pedagogical implications and specific methods and techniques for fostering vocabulary development. The paper has also presented an important issue; the role of technology in bringing a more systematic and data-based approach and innovative methods and techniques to vocabulary instruction and learning. As Sökmen (1997) mentions, in the following years we will get more help from technology in developing and implementing more ways of explicit vocabulary practice. We will then need to train both teachers and learners so as to enable them to take a more principled and systematic approach to vocabulary instruction and learning.

The Data-driven learning approach can be implemented into intensive English language programs focusing on lexical and collocational competence. New corpora can be compiled for the sake of language learning purposes. Pre-service teachers should be provided with required methodology to exploit the data-driven learning approach in order to assess learners and materials design and development. Vocabulary and grammar teaching materials designed and developed according to data driven learning approach should be incorporated into foreign language classrooms. Corpora and concordance based language learning materials and activities can also be delivered to learners through leaning management systems.

References

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of make in native and non-native student writing. *Applied linguistics*, 22(2), 173-194.

Anğ, F. (2006). *Effectiveness of corpus consultation through concordancing on the formulaic academic language use of freshman ELT students.* Unpublished Master's Thesis. Bogazici University. Istanbul.

Aston, G. (1995). Corpora in language pedagogy: matching theory and practice. In G. Cook and B. Seidlhofer (eds.) *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*. (pp. 257-270). Oxford: Oxford University Press.

Bahns, J. (1993). Lexical collocations: a contrastive view'. *ELT Journal,* 47(1), 56-63.

Ball, C. N. (1996). *Tutorial Notes: Concordances and Corpora.* Retrieved from http://www.georgetown.edu/cball/corpora/tutorial.html.

Barlow, M. (1996) Corpora for Theory and Practice. *International Journal of Corpus Linguistics,* 1(1)1-37. John Benjamins Publishing Co.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Chen, Y. H. (2004). The Use of Corpora in the Vocabulary Classroom. *The Internet TESL Journal,* 10(9). Retrieved from http://iteslj.org/Techniques/Chen-Corpora.html. (accessed 06.05.2007).

Cobb, T. (1997). Is There Any Measurable Learning From Hands-On Concordancing? *System,* 25(3), 301-315.

Cobb, T. (2003). Do Corpus-Based Electronic Dictionaries Replace Concordancers? In Morrison, B., Green, G.. & Motteram, G. (Eds.) *Directions in CALL: Experience, experiments, evaluation.* Polytechnic University:Hong Kong.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly,* 34(2), 213-238.

Crystal, D. (1995) *The Cambridge Encyclopedia of Language.* Cambridge: Cambridge University Press.

Dundley-Evans, T. (1994). Genre Analysis: An approach to text analysis for ESP. In M. Coulthard (eds.), *Advances in written text analysis*, (pp. 219-228). London: Routledge.

Eicholtz, G., & Barbe, R. (1961). Vocabulary development. *Elementary School Journal* 61, 414.

Gabel, S. (2001). Over-indulgence and under-representation in interlanguage: reflections on the utilization of concordancers in self-directed foreign language learning. *Computer Assisted Language Learning,* 14 (3-4), 269-288.

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

Granger S. (1998). *Learner English on computer.* Addison Wesley London & New York. Longman.

Gui, S. & Yang H. (2002). *Chinese learner English corpus (CLEC).* Shanghai: Shanghai Foreign Language Education Pres.

Halliday, M. A. K. (1993). Towards a language-based theory of learning. *Linguistics and Education,* 5(2), 93-116.

Hasselgard, H. (2001). *Corpora and their use in research and teaching.* Retrieved from http://www.folk.uio.no/hhasselg/UV-corpus.htm. (accessed 20.06.2008).

Hoey, M. (2000). A world beyond collocation: new perspectives on vocabulary teaching. In M. Lewis (ed.) *Teaching Collocations,* (pp. 224-245). Hove: Language Teaching Publications.

Hoey, M. (2003). *Lexical priming and the properties of text.* Retrieved from www.monabaker.com/tsresources/LexicalPrimingandthePropertiesofText.htm. (accessed 05.04.2008).

Hunston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Pres.

Johns, T. (1990) From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data- driven Learning, *CALL Austria,* 10, 14-34.

Johns, T. (1991). Should you be persuaded: two samples of data-driven learning materials' in T. Johns & P. King (eds.) *Classroom Concordancing ELR Journal,* 4, 1-16.

Johns, T. (1994). Contexts: the Background, Development, and Trialling of a Concordance-based CALL Program. *TACT Conference*, Lancaster University.

Johns, T. (1997) Contexts: the Background, Development and Trialling of a Concordance-based CALL Program. In Wichmann, A., Fligelstone, S., McEnery, T. and G. Knowles (eds). *Teaching and language corpora* (100-115), London and New York: Longman.

Kasapoğlu-Akyol, P. (2010). Using educational technology tools to improve language and communication skills of ESL students, *Novitas-ROYAL*, 4(2), 225-241.

Kennedy, G. (2003). 'Amplifier collocations in the British National Corpus: implications for English language teaching'. *TESOL Quarterly, 37(3),* 467-487.

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology, 5(3),* 77-90. Retrieved from http://llt.msu.edu/vol5num3/kennedymiceli/. (accessed 02.11.2008).

Kenning, M. (2000). Concordancing and comprehension: preliminary observations on using concordance output to predict pitfalls. *ReCALL,* 12(2), 157–169.

Kita, K., & Ogata, H. (1997). 'Collocations in language learning: corpus-based automatic compilation of collocations and bilingual collocation concordancer'. *Computer Assisted Language Learning,* 10(3), 229-238.

Koç, G. (2006). *Developing Collocational Awareness*. Unpublished MA Thesis. Bilkent University. Ankara.

Kurtul, K. (1999). *Bağlamda Sözcük Öğretimi.* Unpublished MA Thesis. Ankara University. Ankara.

McAlpine, J., & Myles, J. (2003). Capturing phraseology in an online dictionary for advanced users of English as a second language: a response to user needs. *System,* 31(1), 71–84.

McEnery, A., & Wilson, A. (2001). *Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics,* 24(2), 223-42.

Richard, J.C., Platt, J. & Platt, H. (1992). *Dictionary of language teaching & applied linguistics.* Essex: Longman.

Shei, C. C., & Pain, H. (2000). An ESL writer's collocation aid. *Computer Assisted Language Learning,* 13, 167-182.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Sinclair, J. (1997). 'Corpus Evidence in Language Description' in Wichmann, Fligelstone, McEnery & Knowles (eds.), *Teaching and Language Corpora.* (pp. 27-39). London and New York. Longman.

Sokmen, A. (1997). Current trends in teaching second language vocabulary**.** In N. Schmitt and M. McCarthy. (eds.), *Vocabulary: Description, Acquisition and Pedagogy.* (237-257). Cambridge: Cambridge University Press.

Stevens, V. (1995). Concordancing with language Learners: Why? When? What? *CAELL Journal,* 6(2), 2-10.

Supatranont, P. (2005). *A Comparison of the Effects of the Concordance-Based and the Conventional Teaching Methods on Engineering Students' English Vocabulary Learning.* Unpublished Doctoral Dissertation. Graduate School Chulalongkorn University.

Sun, H. C., & Wang, C. Y. (2003). Concordancers in the EFL Classroom: Cognitive Approaches and Collocation Difficulty. *Computer Assisted Language Learning,* 16(1), 83–94.

Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System,* 29(1), 91-102.

Tribble, C., & Johns, G. (1990). *Concordances in the classroom.* London: Longman.

Willis, D. (1990). *The Lexical Syllabus: A New Approach to Language Teaching*. London: HarperCollins.