

Fostering Lexis Awareness and Autonomy by Corpus-based Data-Driven Learning*

Heyoung Kim (Chung-Ang University)

Soojin Chun (Chung-Ang University)

Kim, Heyoung & Chun, Sooin. (2008). Fostering lexis awareness and autonomy by corpus-based data-driven learning. *English Teaching*, 63(2), 213-235.

This study investigates the effectiveness of corpus-based Data-Driven Learning (corpus DDL) in college English class in terms of enhancing vocabulary awareness/usability and autonomy in developing lexical knowledge. For this purpose, 48 college students were trained in corpus DDL approach for two weeks. Before and after instruction, students were tested with the same vocabulary and grammar questions and were also asked about their attitude toward vocabulary learning and its methods with a survey questionnaire. As a result, first of all, corpus DDL approach significantly enhanced college students' awareness of both vocabulary and grammar, especially of various types of collocations. In addition, corpus DDL approach fostered the college students' language learning autonomy. Students showed a significant increase in vocabulary learning control and confidence. This study suggests two pedagogical implications for English educators. First, in learning vocabulary and grammar, it is sometimes more important for teachers to provide good resources and tools to students and to teach how to use them effectively than just delivering linguistic knowledge to the students. Second, teachers should provide more various word chunks, different types of collocations.

I. INTRODUCTION

Vocabulary is often considered as the content of the language. Rich and appropriate word use is a crucial factor in successful oral and written communication. However, vocabulary has been introduced and studied mostly for the receptive purposes in Korean college English classes (to recognize the word meaning for reading and listening). This causes a lack of fluency and naturalness in producing language. More seriously, college English curriculum seldom provides any opportunity to develop *autonomy* (independent learnability) in building a lexical knowledge. Most students don't know how they can find

* This research was supported by the Chung-Ang University Research Grants in 2008.

and confirm appropriate words (or expressions) to specific communication contexts. Thus, in many cases, college graduates still feel challenged in preparing themselves for novel language tasks (e.g., writing emails in English or attending a business meeting with foreigners) even after taking many English courses.

Corpus-based Data Driven Learning (Corpus DDL) has been noted as an effective instructional approach to develop vocabulary knowledge (Chan & Liou, 2005; Hadley, 1997; Johns, 1991; Ko, 2005; Lee, 1999; Ma, 1993; Stevens, 1995; Whistle, 1999) and language learning autonomy or independent learning (Cobb, 1999; St. John, 2001; Sun & Wang, 2003; Thurstun & Candlin, 1998). First, corpus is real language data, so the corpus DDL approach gives authenticity of the learning context. Second, analyzing corpus by using a concordancing program, *concordancers*, gives L2 learners a strong analytical power. Third, corpus DDL helps L2 learners raise their target language awareness. Finally, increased exposure to self-accessible authentic texts through concordancing enhances learners' motivation and independence.

Thus, this study intends to investigate the effectiveness of corpus DDL in college English class in terms of enhancing vocabulary awareness/usability and autonomy in developing lexical knowledge. For this purpose, this study addresses two research questions with more specific sub-questions for each, namely:

1. Does the corpus DDL approach (corpus DDL) enhance the college students' awareness of vocabulary and grammar?
 - 1) Does the corpus DDL help them recognize and use appropriate lexis (single words and collocations) in the context?
 - 2) Does the corpus DDL help them raise consciousness of grammar (grammatical errors)?
2. Does the corpus DDL approach foster the college students' language learning autonomy?
 - 1) Does the corpus DDL increase language learning control?
 - 2) Does the corpus DDL increase language learning confidence?

The lesson types and findings of this study will give concrete information to English educators regarding corpus-based DDL as a new approach of teaching vocabulary and of developing autonomy in L2 learning in Korea.

II. THE BACKGROUND OF THIS STUDY

1. Lexical Approach (LA)

Lewis (1993, 2000, 2002) criticizes that vocabulary and grammar are among the most

misunderstood in language teaching. He disagrees on the traditional belief that language consists of grammar and vocabulary in the relation of ‘frame and slot.’ Rather, word chunks “lexis” in his term, such as words, collocations, semi fixed or fixed expressions are the most basic elements of language. Thus, learning lexis helps us handle highly probable events fluently and effortlessly while grammar helps us to use new word combination in relatively unusual situations. He tries to support his position with the following reasons. First, language is very arbitrary, and grammar cannot explain this with well-defined rules. Second, vocabulary is not just words, lexis including word chunks, so grammar is hardly distinct from vocabulary. In the lexical approach, Lewis describes lexis in the following four categories (Lewis, 2002).

- *Words*: traditional term of vocabulary, a single word (e.g., open, certainly), polywords (e.g., bread and butter, by the way) are included.
- *Collocations*: “certain words that co-occur in natural text with greater than random frequency (Lewis 2002, p. 8). There is a degree of partnership: strong (e.g., catch a cold), less strong (e.g., to make an effect).
- *Fixed Expressions*: social greetings (e.g., good morning), politeness phrases (e.g., No, thank you), Idioms (e.g., Hang on, Not too bad)
- *Semi-Fixed Expressions*: fixed expressions that permits minimal variations, spoken sentences with a simple slot or particular fillers (e.g., could you pass ... please?, I haven’t seen you + time expression with for or since). Sentence head (e.g., what was interesting/surprising/ annoying was....)

Another noticeable feature in LA is that receptive skills should be emphasized before productive skills in grammar practice to ‘raise language (grammar) awareness’. In most typical grammar instruction, learners are asked to produce correct sentences based on target grammar rules by transforming, matching or filling in the blank. However, in LA, “Can you see?” is more important than “Can you do?” in grammar practice. Students are encouraged to observe language patterns accurately, and perceive similarities and differences within target language data (target corpus data). Free practice is the last stage of learning (Lewis, 1993, 2000).

2. Data-Driven Learning (DDL) and Autonomy in L2 Learning

Since the advent of Information and Communication Technology (ICT), DDL has been attended as a new way of learning in language education (Chan & Liou, 2005; Hadley, 1997; Johns, 1991; Ko, 2005; Lee, 1999; Ma, 1993; Stevens, 1995; Whistle, 1999). DDL is an inductive method of learning language with authentic language samples (e.g., corpus

data), sometimes by using software programs (e.g., concordancer). Johns (1991) writes that the “language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data --hence the term ‘data-driven learning (DDL)’ to describe the approach”(p. 2). This is a good example of a grammar-consciousness raising method because it is language learners themselves who explore and discover the language patterns by accessing real-life language data. Bastone (1995) sees strength of DDL also in L2 grammar teaching. According to him, there is a critical gap between the product and process grammar teaching approach, and DDL can bridge this gap by helping learners to move along the pedagogic continuum from product to process (p. 99).

The idea of DDL is also much similar to that of language learner autonomy. Benson (2001) recently identifies the meaning of learning control (e.g., autonomy) within the framework of three levels: learning management, cognitive processes, and learning content. Two major practices to foster learning control in Benson’s six approaches (2001) are *Resource-based* and *Technology-based* approaches, both of which focus on independent use of learning sources. Resources in these approaches mean self-accessible learning materials, authentic materials, activities and technology, a part of which resembles Johns’ DDL approach. Sheerin (1991) defines it as “a way of describing materials that are designed and organized in such a way that students can select and work on tasks on their own” (recited from Benson, 2001, p.113). Technology-based approach can help learners’ exploratory learning and their control over selecting materials and strategies of interpretation.

Thus, teaching language with the DDL approach can help L2 learners with not only effective grammar and vocabulary learning, but also with the development of autonomy. Based on these approaches, students can foster their learning control and responsibility including learning management (e.g., utilizing concordancers, cognitive process (e.g., defining rules, finding language patterns), and language content (e.g., lexical knowledge) data) .

3. Concordance and Concordancing Instruction Methods

1) Concordance

A concordance is “a list of occurrences of a specified word or phrase along with a piece of the passages where they occur from a large body of texts, or corpus” (de Szendeffy, 2004 p. 1). As Figure 1 illustrates, concordance is sentence chunks center-aligned with a keyword. The concordance data provides varied linguistic information such as word frequency, word uses, grammatical patterns, collocated words etc. From many studies (Cobb, 1999; Flowerdew, 1993; Johns, 1988; Lamy, 2007), the values of concordancing in

language teaching can be summarized as follows. First, corpus is real language data and a concordancing approach gives authenticity to the learning context. Second, analyzing corpus by using concordancing program, *concordancers* has become much more user-friendly and given a more analytical power. Third, reviewing concordances helps L2 learners raise their target language awareness. Finally, increased exposure to self-accessible authentic texts through concordancing enhances motivation and learner independence.

2) Concordancing Activities

To use the corpus as a valuable resource for language learning and teaching, users need to be fully trained for the functions of concordancers. Concordancers can successfully function as DDL tools only after sufficient practice. For this purpose, there have been introduced many concordancing activities which are designed for independent learning or group tasks. The following activities and games have been most commonly introduced and used for corpus-DDL practices in the literature (de Szendeffy, 2004; Kim, 2001; Lamy & Klarskov Mortensen, 2007; Ma, 1993; Thurstun & Candlin, 1998).

- Concordancing exercises:
 - Finding affixes of a key word using wildcards (e.g., *)
 - Finding compound words of a key word using wildcards
 - Finding the fitting word to a key word given using tags
 - Finding missing words in a work chunk or a sentence using tags
 - Filling a common word fitting in the blanks in concordance lines given
- Concordancing tasks:
 - Making a list of possible collocation with a key word
 - Unscrambling collocations or phrasal expressions from a word bank
 - Correcting mis-collocations in a paragraph

The literature has also offered examples of online exercises and games as classroom activities. By using the activities appropriately in class, students can be effectively trained for autonomous DDL.

3) Terms

KWIC is defined as “Key Word in Context.” *KWIC* is a target language form (e.g., ‘support’ in Figure 1) normally aligned at the center of concordance lines. In concordance a *KWIC* index is formed by sorting and aligning the words to allow each word to be

searchable alphabetically in the index, e.g.,

A corpus can be classified in two by its size: *large corpus* and *small corpus*. A large corpus consists of millions of words which are selected and sorted out by the builders; for example, BNC (British National Corpus), MICASE (Michigan Corpus of Academic Spoken Corpus), ANC (American National Corpus).¹ Concordancing can be possible from an individual genre in the corpus or from a whole corpus as one source.

On the other hand, a small corpus consists of small numbers of words. It can be defined as the specificity of sectioning the sources of text rather than the number of words which consists of a corpus. For instance, “Web Concordance by Virtual Language Center (VLC)”² can be classified a small corpus because of the specificity of classifying and the independency of concordancing each source of text (See Figure 1).

While a large corpus offers sufficient concordance lines from the enormous text, a small corpus shows limited numbers of text. However, for pedagogical purposes, a small corpus in which the users can select the genre or find the limited numbers of text from this selection can be more useful. Small-corpora-concordancing (SCC) has also been known as classroom concordancing (CC) or pedagogical concordancing, which is beneficial for ESL (Johns, 1988; Ma, 1993; Stevens, 1991; Tribble, 1990).

4. Previous Studies on Corpus-Based Instruction

Although the invention of concordancers has given great contribution to developing corpus linguistics or analyzing texts in English materials, few teachers have introduced concordances to their students in English class. There are not many studies that investigate pedagogic values in this case.

¹ Examples of large corpus:

- BNC <http://corpus.byu.edu/bnc/>
- ANC <http://www.americannationalcorpus.org/>
- MICASE <http://quod.lib.umich.edu/m/micase/>

² Examples of small corpus:

- VLC Web concordancer <http://www.edict.com.hk/concordance/WWWConcappE.htm>

FIGURE 1
Concordance for KWIC ‘support’ in students’ writing corpus data

Web Concordancer is now searching corpus academicwriting.txt for support		
Concordances for support = 19		
1	will examine the arguments which	support and against of turning Kai Tak into
2	have the responsibility to give	support and backup to those designers. The
3	essay, I examine arguments which	support commercial area and those which sup
4	their company. In my opinion, I	support environmental consideration should
5	ow found in CBD can no longer to	support further economic growth. In order t
6	stion, a land is needed to	support infrastructure for urgent road
7	essay, I examine arguments which	support introducing new rules and those whi
8	a lot of commercial building to	support its further development of commerci
9	at time. The second argument in	support of recreational redevelopment is th
10	rt commercial area and those which	support recreational area to be the futur
11	essay, I examine arguments, which	support recreational purpose and commerci
12	Airport. The first argument to	support redevelopment focuses on the sugges
13	essay, I examine arguments which	support redevelopment for recreational purp
14	and commercial center, they will	support the airport's workforce and associa
15	ad opened last year. In order to	support the airport, Tung Chung and Tai Ho
16	ssay, I examine arguments, which	support the commercial, or those support th
17	area, the residents will agree and	support the government to use the site fo
18	h support the commercial, or those	support the recreational use. I then go
19	larly, there are several points to	support the side of environmental consid

Most pedagogical corpus research focuses on vocabulary and grammar. These studies mostly investigate the effectiveness of using concordancers for learning words, collocations, or grammatical structures. St. John (2001) used a parallel corpus and concordancer to teach grammar to beginners in German class and reported that concordancing is beneficial for beginners and motivating both FL learners and teachers. What he provided to his students were self-awareness tasks for the target grammar. Students were asked to find the answers to the question of each task by using concordance software and relying on their intelligence for language analysis. Even beginning students were able to find correct answers. Thurstun and Candlin (1998) also confirm the value of concordancing in the development of teaching materials focusing on vocabulary and grammar. The participants (undefined students) were asked to complete exercises with a concordance program in their working project, and they mostly found all the exercises helpful in giving insights into collocations and grammatical structures. Sun and Wang (2003) experimented with the effectiveness of teaching collocations to high school students in Taiwan with two different treatment groups, an inductive and a deductive group. As a result, the inductive group which studied collocations by using concordancers improved significantly better than deductive group who studied given rules and patterns. Chan and Liou (2005) also reported similar results with college students in the concordancer learning setting. Some concordance studies attend to the potential power of writings. Yoon (2005) notes the importance of corpus use in learning collocation and writing skills.

However, there are several weaknesses in the previous concordance studies. First of all, most corpus-based educational studies focus only on the short-term improvement of *the learning content*, such as vocabulary, grammar knowledge, and *motivation*. New tools can easily motivate students and additional reference will better facilitate learners. However, more important thing is the students' awareness of lexis and the ability to independently manage their learning with new vocabulary and grammar patterns. There isn't much evidence from the early research that DDL enhances learners' autonomy, independence, or vocabulary or grammar learning skills. Second, only a limited number of corpus studies has been conducted in Korean educational context. Quite a lot of corpus studies (Han, 1999; Hong & Oh, 2008; Jung & Sur, 2007; Kim & Ahn, 2005; Ko 2005; Kim, 2001; Lee, 1999; Lee, 2004; Lee, 2003; Yoon, 2005) have been conducted in this area for over the past few decades in Korea, but there are few findings (Hong & Oh, 2008; Kim & Ahn, 2005; Ko, 2005; Yoon, 2005) from actual classroom experiments in the concordancing learning setting. Furthermore, none of these studies deal with lexis awareness or autonomy development through DDL. Thus, this research identifies the need of this study from the literature review as follows. Unlike the previous findings, this study focuses on the improvement of 1) students' lexis awareness and vocabulary learning skills by using concordancers and 2) autonomy in learning vocabulary including learning control and confidence.

III. THE STUDY

1. Subjects

The subjects in this study were 48 college students, 37 female and 11 male, who registered for a major course offered by the English Education department at a university in Korea. Three students out of 51 were excluded due to missing tests or survey reports. Of the remainder, 36 were English education majors and 12 were 'other' education or language majors. They ranged from sophomores to seniors. Their English proficiency was from intermediate to advanced according to their self-report or their TOEIC scores. Students were not selected or graded by language level or by previous language experience, and no control or comparison groups were used.

2. Materials

1) Vocabulary Test Sheet

A vocabulary and grammar exercise sheet (Appendix) was used for both the pre-test and the post-test. It comprised 40 questions in eight parts, based on the target lexis patterns in Table 1:

TABLE 1
The Organization of Vocabulary & Grammar Test Sheet

Test section	Types of word chunks	Types of questions	Numbers of questions
Part A	adjective + noun	matching	8
Part B	adjective + noun	odd-one-out	4
Part C	verb + noun	odd-one-out	5
Part D	verb + noun (in context)	odd-one-out	5
Part E	adverb + verb (participle)	fill-in-the blanks	5
Part F	verb + preposition + noun	true or false	4
Part G	verb + preposition + noun	fill-in-the blanks	8
Part H	Mixed	error correction	5

The types of collocation were chosen based on Lewis' and Hill's suggestion (*LPT Dictionary of Selected Collocations*, 1997; *Teaching Collocation*, 2000). Among these, "combinations with nouns that could strongly suggest the partner words" (Lewis, 2002, p.28) were mainly covered in Part A, Part B, Part C, and Part D. Collocations related to adverbs, which comprise adverb + verb, verb + adverb, adverb + adjective, were covered by one type (adverb + participle form of verb) in E. Part F, G, and H were designed to focus on grammatical structure, with bigger word chunks than in the other parts. They required knowledge outside of the collocative combination of two words. In F and G, appropriate choices of prepositions were made regarding the nouns that followed each preposition. Error corrections in H (mixed types of collocation) required an understanding of the context of the story.

The questions were written after referring to *The LTP Dictionary of Selected Collocations* (Hill & Lewis, 1999), and *Oxford Collocations Dictionary* (Oxford University Press, 2002), whereas the types of questions were adopted from *English Collocations in Use* (McCarthy & O'Dell, 2005), *Key Words for Fluency* (Woolard, 2005), and *VLC edict*.

2) Student's Attitude Questionnaire Sheet

The survey was also conducted to find out each student's attitude towards 1) the vocabulary test, 2) vocabulary and grammar learning in general, 3) problem solving methods, and 4) other affective aspects like confidence or emotional difficulties in language learning. Questions regarding autonomy (language learning management, etc.) were created based on "Questionnaire on responsible attitudes (Scharle & Szabo, 2000)." It consisted of 24 items about vocabulary and grammar. Twenty one items were rated on a five-point Likert scale (1 = absolutely no, 2 = no, 3 = not sure, 4 = yes, and 5 = absolutely yes). The other three questions were answered by students providing percentages, verbal descriptions, and ranks.

3. Procedures

The research was carried out in three stages (see Table 2): 1) the pre-test (take-home) and survey (in-class) at the beginning of the semester, 2) four lessons in two consecutive weeks for corpus DDL practice in the middle of the course, and 3) the post-test (take-home) and survey (in-class) at the end of the course. The same test and the same questionnaire sheets were used before and after the treatment (corpus DDL instruction).

The pre-test sheets were distributed in the second week, and students were asked to return them a week later. No directions and instructions were given for the assignment. The survey was done in class on the same day the test sheets were collected. Its purpose was to find out how successfully the students could solve the lexis problems themselves. A take-home test was considered as a better way to examine *the degree of students' autonomy* in dealing with lexis (words and grammars), which includes their appropriate method use as well as accuracy of their lexis knowledge.

A month later, corpus DDL instruction was given in a computer lab for four lessons (six class hours, two consecutive weeks), to teach students how to use corpus in online dictionaries and online concordancers, and to practice with concordancing activities (see p.5). In the first lesson, the instructor introduced an online thesaurus dictionary,³ a visual dictionary,⁴ BNC,⁵ JustTheWord,⁶ VLC Web concordancer,⁷ and Google,⁸ for a general understanding of corpus and concordance. The second lesson was concordancing

³ The FreeDictionary <http://www.thefreedictionary.com/>

⁴ Thinkmap Visual Thesaurus <http://www.visualthesaurus.com/>

⁵ British National Corpus <http://corpus.byu.edu/bnc/>

⁶ JustTheWord <http://193.133.140.102/JustTheWord/>

⁷ VLC Web Concordancer <http://www.edict.com.hk/concordance/default.htm>

⁸ Google <http://www.google.com>

practice. The students learned how to utilize tags for POS (part of speech) and wildcards in the concordancers. In the third lesson, the instructor introduced web-based corpus exercises.⁹ Students learned about the types of corpus exercises and practiced the questions individually or in groups. In the last lesson, the instructor offered them a group activity with a flash game that included the various collocation activities mentioned on page 5. The game was designed for the group competition. When each group reached the final stage of the game, the instructor added each group's scores and found the winning group. The entire activity was designed to enable students to fully understand the concordance functions covered in their lessons.

Finally, a post-test was distributed approximately one month after the last lesson. Students were again asked to take it home and complete the assignment within one week. The same test sheet was used for the post test because the researchers wanted to check changes in methods the students had employed for the same problems and improvement of their scores in each section.

There might be some repetition effect on the results of the second test, but the effect doesn't seem significantly affect the major finding of this study in the following reasons. First, the answers were not given to the students, so they could not confirm their test results. Second, the test interval was long enough (10 weeks) for students to be hardly able to recall the questions. Third, the instructor did not deal with any of the target words or grammatical elements during her lessons.

TABLE 2
The Procedures of Treatment

Week	Stages	Treatments
2	<i>The pre-test & pre-survey</i>	<i>Take-home assignment with the test sheet, and an in-class survey after collecting the pre-test sheet</i>
6	<i>Lessons in a computer lab</i>	<i>The 1st lesson: Introduction of corpus, concordancers, and other online tools</i> <i>The 2nd lesson: Practice for utilizing the functions of concordancers</i> <i>The 3rd lesson: Practice with web-based corpus exercises</i> <i>The 4th lesson: A group activity for concordancing exercises using a flash game</i>
12	<i>The post-test & post-survey</i>	<i>Take-home assignment with the identical test sheet, and an identical in-class survey after collecting the post-test sheet</i>

⁹ UGRU UAEU Concordance <http://www.ugru.uaeu.ac.ae/concordance/inde.g.,html>
 Virtual DDL Library http://www.eisu2.bham.ac.uk/johnstf/ddl_lib.htm
 Usti nad Labem DDL Workshop http://www.eisu2.bham.ac.uk/johnstf/unl_ddl.htm

V. RESULTS

1. Awareness of Vocabulary and Grammar

1) Appropriate Use of Collocations

Students obtained significantly more correct answers on the second test than the first. The average score of the pre-test was 25.04 (out of 40.00) while that of the post-test was 32.20. The mean difference was 7.16. As the Table 3 shows, the student achievement significantly improves on the post vocabulary and grammar test ($t=8.65$, $df=43$, $p<.01$).

TABLE 3
T-test of Difference between Pre- and Post- Vocabulary-Grammar Test ($\alpha = .01$)

Test	N	M	SD	MD	t	df	Sig.
Pre	47	24.70	4.764				
Post	45	32.18	4.402	7.159	8.650	43	0.000*

Based on the result of each test section in Table 4, the students show greater improvement on collocation matching than grammar. Students achieved a better score on every question, but they solved the questions particularly about nouns + adjectives collocations (Part B) and verbs + nouns collocation (Part C). It is probably because the instructor focused more on vocabulary (e.g., words and collocations) since this course was designed for learning and teaching vocabulary.

TABLE 4
Pre-and Post- test Difference of Each Question

Part	Test	N	M	SD	MD	t	df	Sig.
A	pre	47	6.77	1.463	0.89	3.908	44	.000*
	Post	46	7.70	1.030				
B	Pre	47	1.77	1.068	1.44	6.809	44	.000*
	Post	46	3.24	0.923				
C	Pre	47	3.23	1.272	1.02	4.619	44	.000*
	Post	46	4.35	0.924				
D	Pre	47	2.47	1.039	1.60	7.472	44	.000*
	Post	46	4.11	0.994				
E	Pre	47	2.89	1.289	1.16	4.735	44	.000*
	Post	46	4.07	1.181				
F	Pre	47	3.28	0.772	-0.02	-.158	44	.875
	Post	46	3.28	0.779				
G	Pre	47	2.91	0.974	0.33	2.187	44	.034
	Post	46	3.30	0.511				
H	Pre	47	1.34	0.984	0.80	4.161	44	.000*
	post	46	2.17	1.039				

2) Grammar Consciousness-raising

As Table 4 indicates, students mostly achieved a better score on the second grammar test (question F, G, and H). The only question on which there is no improvement after receiving instructions was F. It is assumed that students had already had quite good previous knowledge about this target grammatical point (verbs + prepositions) such as “discuss/discuss about” from the first test (they got average score 3.28/4.0), so the students usually relied on their knowledge without further investigation even on the second test.

In particular, on question H, most students show significantly better performance in finding errors on paragraph writing. In addition, their second answers show that their lexis awareness has been remarkably raised. In Figure 2, student A’s answer is very different between the two test sheets. For example, student A got 25 out of 40 at first, but 33 on the post test. On the first test, she found three grammatical errors (wider-> wide, living-> live, makes-> goes on). She initially paid attention to only morphological errors based on the sentence structure. However, on the post-test, her grammar awareness was extended to other lexical knowledge, such as word combination, or appropriateness of word use. Furthermore, her answer was significantly more accurate and thorough than before. As Figure 2 shown, she generated more than one possible answers (e.g., great-> important, significant, suggests-> offers, gives) by using web concordancers. It probably implies that she started thinking about replaceable word combinations.

FIGURE 2

Differences of Student A's Answer to Question H Between Pre- and Post Tests

<Pre-test Result: 25/40>

will be a great addition to their collection. Bob recently got solo after five years with the folk band Blue Mountain. He is proud of the musical heritage of his native Kentucky.

Track 3 and 7 feature his old friend Wiz Carter on guitar. With this CD Coppin says he hopes to capture a wider audience for folk music. His excellent living performance at the recent Lockwood Folk Festival suggests he has a good chance of succeeding. He makes a tour next month. Don't miss him.

Handwritten notes: "wide" under "wider", "live" under "living", "goes on" under "makes".

<Post-test Result: 33/40>

will be a great addition to their collection. Bob recently got solo after five years with the folk band Blue Mountain. He is proud of the musical heritage of his native Kentucky.

Track 3 and 7 feature his old friend Wiz Carter on guitar. With this CD Coppin says he hopes to capture a wider audience for folk music. His excellent living performance at the recent Lockwood Folk Festival suggests he has a good chance of succeeding. He makes a tour next month. Don't miss him.

Handwritten notes: "important, significant" under "great", "has been solo" under "got solo", "live performance" under "living performance", "offers/gives" under "makes".

① music lovers → lovers.

② a great addition → Important, significant.

③ got solo → has been solo.

④ living performance → live.

⑤ suggests a good chance → offers, gives.

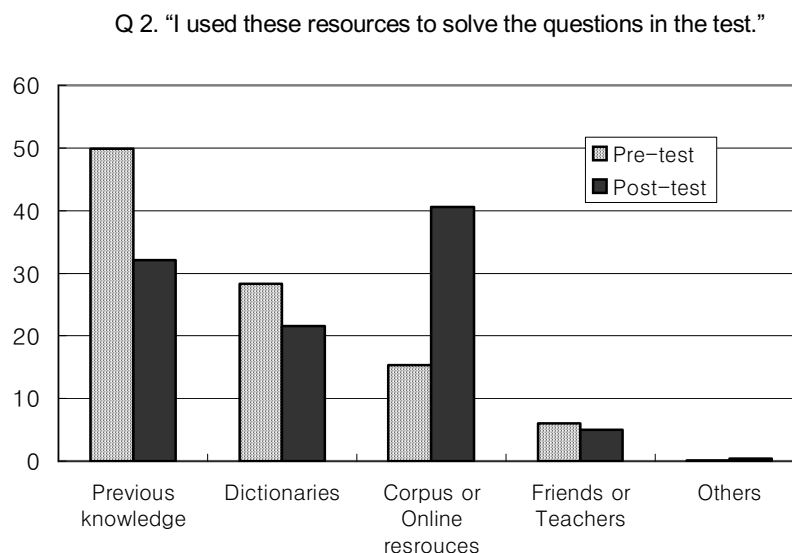
Thus, data analysis supports that corpus DDL approach such as concordancing activities help the participants raise their both lexis and grammar awareness. Students showed significant improvement especially word combination, such as adjectives+ nouns, verbs + nouns, and adverbs + adjectives. However, there is no remarkable change between the two tests asking such collocations as verbs + prepositions, or nouns + prepositions.

2. Autonomy in Learning Vocabulary

1) Vocabulary Learning Control (Usability)

The way students look for the answer in the test was changed considerably. First of all, as Figure 3 shows, when comparing the two tests, students used quite different resources when they tried to find the answers. At the beginning of the semester, students usually relied mostly on their previous knowledge first and/or a dictionary. Meanwhile, students reported that they used online resources the most and their previous knowledge next. These changes are all significant. 41% of the students used online resource on the second test, while only 15% of the students did so on the first one.

FIGURE 3
Differences of Reference Use for Vocabulary Test between Pre- and Post-Test



More importantly, students who used corpus or online resources (DDL group) on the

second test improved a lot more, compared with students who used other methods, such as previous knowledge or dictionary (No-DDL group) ($t=2.514$, $df=30$, $p<.05$). DDL group showed a significantly greater difference (9.47) between the post- and the pre-test scores than No-DDL group (5.12) in Table 5. Thus, there seems to be close relationship between types of reference use and achievement on the second test.

TABLE 5
Difference of Improvement between Different Reference Groups ($\alpha = .05$)

Groups	N	M	SD	MD	t	df	Sig.
DDL	15	9.47	5.303	4.35	2.514	30	0.018*
No-DDL	17	5.12	4.484				

Another noticeable aspect is that students are significantly more satisfied with their methods of solving questions after instruction (See Table 6). They more strongly feel the usefulness of the resources they chose. Therefore, this suggests that students can now better control their learning with vocabulary and grammar.

TABLE 6
Students' Attitude Survey for the Increase of Language Learning Control

Questions	Test	N	M	SD	t	df	Sig.
3. The resources were helpful to solve the questions.	Pre-	42	3.48	0.943	3.279	41	0.002*
	Post-	42	4.05	0.825			

2) Language Learning Confidence

Students regarded the second test as being much easier than the first one although they were the same tests, and there's no lesson directly for the target words or grammars. "Feeling easier" in solving the problems between pre- and post made a remarkably huge difference ($t=13.375$, $df=40$, $p<.01$).

More importantly, they were greatly more confident with their answers on the second test than the first. ($t=4.942$, $df=41$, $p<.01$). They were sure of their answers after the instruction. The t-test result of two survey questions shows that students' attitude changed significantly after treatment.

However, the students' confidence shown on Q1 and 4 did not yet seem transferred to confidence with their English vocabulary use in general or overcoming the feeling of a lack of vocabulary knowledge. More than half of the students still feel that their vocabulary is not good as their friends. There is no improvement in Q14 which is important for developing learner autonomy.

TABLE 7

Improvement of Students' Autonomy in Solving Vocabulary Problems

Questions	Test	N	M	SD	t	df	Sig.
1. The test was difficult.	Pre	41	4.10	0.539	-13.375	40	0.000
	Post	41	2.49	0.637			
4. I am sure of my answers.	Pre	42	2.69	0.563	4.942	41	0.000
	Post	42	3.36	0.692			
14. My vocabulary is not as good as my friends.	Pre	43	3.26	0.902	0.771	42	0.445
	Post	43	3.33	0.919			

* $\alpha = .01$

VI. DISCUSSION AND CONCLUSION

The purpose of this study is to investigate the effectiveness of corpus DDL in college English class in terms of enhancing vocabulary awareness/usability and autonomy in developing lexical knowledge. The findings of this study give a very positive signal on the effectiveness of corpus DDL.

First of all, the corpus DDL approach significantly fostered college students' awareness of both vocabulary and grammar. Students achieved significantly better scores on the post vocabulary assignment. This result can be supported by many previous studies (St. John, 2001; Thurstun & Candlin, 1998; Sun & Wang, 2003). It truly helped the students recognize and use appropriate lexis. Another interesting finding in this study is that their improvement was particularly distinctive in some collocation types such as verbs + nouns, adjectives + nouns, and adverbs + nouns. It is assumed that chunking these word combinations was unfamiliar to them, and they haven't learned or noticed them before. Corpus DDL (e.g., concordancing activities) in this class successfully helped the students raise their lexis awareness, especially of various types of collocations. Grammatical consciousness was also significantly improved in some part as regard to appropriate word uses, but not very much, when compared with lexis awareness. This is probably due to the fact that the students had already developed good grammatical knowledge for a long time, and the target grammar questions were not very new to them. However, they were noticeably better at finding errors in the paragraphs by considering lexical as well as morphological (or syntactic) information.

Second, overall, the corpus DDL approach also intensified the college students' language learning autonomy. First of all, the students showed a significant increase in vocabulary learning control. Students were better able to utilize effective tools and resources in solving problems. They mostly used online resources including web concordances with their previous knowledge to find the answers. They were greatly more satisfied with the outside resources they selected. They also tend to think that online

resources were significantly more useful than the resources that they had previously relied on. Furthermore, corpus DDL enhanced language learning confidence. The students felt much less challenged with the tasks and a lot more assured of their answers. Their attitude toward the vocabulary assignment changed significantly. This study provides strong evidence for Benson's (2001) resource-based approach. Students developed language learner autonomy with concordancing activities by improving language learning management and confidence.

The limitations of this study are as follows. First, the findings were yielded with a small sample size and with a short-term treatment. Forty students were definitely not sufficient and not randomly chosen either. A bigger sample size in a different context might yield different findings even with the same treatment. Also, the pre- and post-test did not cover more comprehensive lexical knowledge. The test validity was not thoroughly examined although the question types were carefully selected and all borrowed from DDL studies and books. In addition, it is also possible that there might be unexpected external effects during the period between the two tests.

However, this study is meaningful in terms of the following research and pedagogical facets. This research is a new attempt of using corpus DDL approach with lexis-based concordancing activities in the Korean context. This study intended to view the study effect of various types of collocations and grammar after systematically developed DDL instruction. Moreover, this study attends to developing vocabulary learning control (autonomy), which hasn't been explored in any previous studies. This study provides good evidence in the major aspects of learner autonomy such as vocabulary learning management and vocabulary learning confidence.

This study suggests some pedagogical concerns to English educators. First, in learning vocabulary and grammar, it is sometimes more important for teachers to provide good resources and tools to students and to teach how to use them effectively than just delivering their knowledge to the students. Second, teach more various word chunks, different types of collocations (e.g., verbs + nouns, adjectives + nouns, adverbs + adjectives) from what the students have learned from the textbooks (verbs + prepositions, nouns + prepositions).

It is hoped that there will be future studies that further explore the effectiveness of the corpus DDL approach with more concrete contexts and further examine the relationship between DDL and learning autonomy development in vocabulary and other language skill areas.

REFERENCES

- Batstone, R. (1995). Product and process: Grammar in the second language classroom. In M. Bygate, A. Tonkyn, & E. Williams, (Eds.), *Grammar and the language teacher* (pp. 224-236). London: Prentice Hall.
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow, UK: Pearson Education.
- Chan, T., & Liou, H. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer-Assisted Language Learning*, 18(3), 231-250.
- Cobb, T. (1999). Breath and Depth of lexical acquisition with hands-on concordancing. *Computer-Assisted Language Learning*, 12(4), 345-360.
- de Szendeffy, J. (2004). Vocabulary and usage activities with concordances. *Computer-Assisted Language Learning: A Practical Guide for Teachers*.
- Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, 21(3), 231-243.
- Hadley, G. (1997). Sensing the winds of change: An introduction to data-driven learning. In J. Field, A. Graham, & M. Peacock (Eds.), *Insights 2*, Witlestable: IATEFL. Retrieved January 6, 2008 from the World Wide Web <http://www.nuis.ac.jp/~hadley/publication/windofchange/windsofchange.htm>
- Han, I. (1999). Concordancing in linguistic research and language learning. *Studies in Language*, 15(2), 251-260.
- Hill, J., & Lewis, M. (1999). *LTP dictionary of selected collocations*. Hove, England: Language Teaching Publications.
- Johns, T. (1988). Whence and whither classroom concordancing? In T. Bongaerts, P. de Haan, S. Lobbe, & H. wekker (Eds.), *Computer applications in language learning* (pp. 9-33). Dordrecht: Foris.
- Johns, T. (1991). Should you be persuaded: Two examples of data driven learning. In T. Johns & P. King (Eds.), *Classroom concordancing*, *ELR Journal* 4 (pp.1-16). University of Birmingham.
- Jung, Y.-C., Sur, R.-I., & Kim, E. (2007). A corpus-based study of Korean college students' collocation ability of English synonymous verbs. *The New Korean Association of English Language and Literature*, 49(2), 175-914.
- Kim, Y.-K. (2001). Use of concordancing in ESP vocabulary teaching. *Journal of the Applied Linguistics Association of Korea*, 17(1), 1-19.
- Ko, M. H. (2005). Korean university student use of phrasal and one-word verbs and concordance-based learning. *Korean Journal of Applied Linguistics*, 21(2), 1-39.
- Lamy M-N., & Klarskov Mortensen H. J. (2007) Using concordance programs in the modern foreign languages classroom: Module 2.4. In Davies G. (ed.), *Information and Communications Technology for Language Teachers (ICT4LT)*, Slough,

- Thames Valley University. Retrieved February 17, 2008 from the World Wide Web http://www.ict4lt.org/en/en_mod2-4.htm.
- Lee, S. (2003). "Pour boiling water?" or "Pour boiled water?": How to apply lexical approach and concordance analysis in secondary English education. *Foreign Languages Education*, 10(4), 87-100.
- Lee, S.-M. (2004). Learner corpora-based error analysis and its application in English classes. *Foreign Languages Education*, 11(3), 33-57.
- Lee, S.-R. (1999) Use of corpus: Data-driven learning. *English History*, 8(11), 203-220.
- Lewis, M. (1993). *Lexical approach*. Boston: Thomson & Heinle.
- Lewis, M. (2000). *Teaching collocation*. Boston: Thomson & Heinle.
- Lewis, M. (2002). *Implementing the lexical approach: Putting theory into practice*. Boston: Global ELT.
- Ma, B. K. C. (1993). Small-corpora concordancing in ESL teaching and learning. *Hong Kong Papers in Linguistics and Language Teaching* 16. (ERIC Document Reproduction Service No. ED 365119).
- McCarthy, M. & O'Dell, F. (1995). *English collocations in use*. Cambridge: Cambridge University Press.
- St. John, E. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning & Technology*, 5(3), 185-203.
- Stevens, V. (1991a). Concordance-based vocabulary exercises: A viable alternative to gap-fillers. *ELR Journal (New Series)* 4, 47-62.
- Stevens, V. (1991b). Classroom concordancing: Vocabulary materials derived from relevant, authentic text. *English for Special Purposes*, 10(1), 35-46.
- Stevens, V. (1995). Concordancing with language learners: Why? when? what?, *CAELL Journal*, 6(2), 2-10.
- Sun, Y., & Wang, L. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer-Assisted Language Learning*, 16(1), 83-94.
- Thursun, J., & Candlin, C. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, 17(3), 267-280.
- Tribble, C. (1990). Small scale corpora in ELT: An investigation into vocabulary use. *CAELL Journal*, 1(4), 13-17.
- Whistle, J. (1999). Concordancing with students using an 'off-the-Web' corpus. *ReCALL*, 11(2), 74-80.
- Woolard, G. C. (2005). *Keyword for fluency*. Hove, England: Language Teaching Publications.
- Yoon, H. (2005). L2 students' collocation learning through corpus use: An exploratory study. *Foreign Languages Education*, 12(3), 91-107.

APPENDIX

Vocabulary Test

A. 왼쪽의 형용사(수식어)와 어울리는 명사를 상자에서 찾아 빈칸 안에 쓰시오.

ex) heavy rain

- | | |
|-----------------|-------------------|
| 1. dense _____ | 2. equal _____ |
| 3. public _____ | 4. grave _____ |
| 5. prime _____ | 6. relevant _____ |
| 7. vivid _____ | 8. wild _____ |

candidate	doubts	fog	transportation
guess	imagination	information	opportunity

B. 밑줄 친 명사와 어울리지 않는 형용사에 X표 하시오.

9. The high / huge / enormous / wide cost of energy was a problem for consumers.
10. We took for trains, and the maximum / overall / total / all duration of the journey was 70 hours.
11. John made amazing / excellent / remarkable / surprising recoveries after strokes.
12. Regional managers occupy a crucial / essential / key / focusing role in developing a strategic framework.

C. 주어진 동사와 어울리지 않는 명사에 X표를 하시오.

- | | | | | | |
|-----------------|-------------------|-------------------|--------------------|------------------|--------------|
| 13. do | a. the dishes | b. 'Hamlet' | c. research | d. damage | e. a mess |
| 14. make | a. an appointment | b. an attempt | c. a degree | d. a guess | e. a mark |
| 15. have | a. action | b. cancer | c. a chat | d. an interest | e. a feeling |
| 16. take | a. a bite | b. a heart attack | c. his temperature | d. notes | e. a look |
| 17. give | a. a chance | b. a cry of pain | c. a lift | d. a priority to | e. a noise |

D. 밑줄 친 명사와 어울리지 않는 동사에 X표를 하시오.

18. A department store spokesman says that their new outlet will create / provide / produce more than 75 permanent jobs in the city.
19. You probably think carefully before deciding how to respond to / reply to / answer to a business letter.
20. The scientists failed to arrived at / decide / draw any firm conclusions from the study.

21. Someone came up with / presented / put forward the **suggestion** that we should have an auction.
22. He had to do two jobs to clear / pay off / pay up his **debts**.

E. 주어진 글자로 시작하는 부사를 빈칸에 써 넣으시오.

23. One of his legs was b_____ broken.
24. Consumers must be f_____ informed of the services available.
25. She was h_____ influenced by the writer.
26. The result can be v_____ presented in the form of a graph.
27. The cleaning liquid should not be d_____ applied to the surface.

F. 문법적으로 맞는 문장에는 O를, 틀린 문장에는 X를 괄호 안에 쓰시오.

28. Can we meet to **discuss about** the policy changed? ()
Can we meet to **discuss** the policy changed? ()
39. **Education in** the internet can save a lot of resources. ()
Education on the internet can save a lot of resources. ()
30. This figure could reflect their **dissatisfaction of** the lack of training. ()
This figure could reflect their **dissatisfaction with** the lack of training. ()
31. The USA made a formal **complaint on** that country's environmental policy. ()
The USA made a formal **complaint about** that country's environmental policy. ()

G. 밑줄 친 단어와 어울리는 전치사를 빈칸에 쓰시오.

32. Her entrance was **acquainted** _____ suitably dramatic music.
33. No one could **resist** _____ her beauty smile The English Linguistics Society Of Korea.
34. She **insisted** _____ inviting him.
35. I was **alarmed** _____ the latest crime statistics.

H. 다음 이야기에서 단어가 잘못 선택된 5곳을 찾아 올바른 단어로 바꾸시오.

For all folk music likers, Johnny Coppin's new CD, The long Harvest, published last week, will be a great addition to their collection. Bob recently got solo after five years with the folk band Blue Mountain. He is proud of the musical heritage of his native Kentucky. Track 3 and 7 feature his old friend Wiz Carter on guitar. With this CD Coppin says he

hopes to capture a wider audience for folk music. His excellent living performance at the recent Lockwood Folk Festival suggests he has a good chance of succeeding. He makes a tour next month. Don't miss him.

Applicable levels: university

Key words: corpus, DDL, concordance, concordancers, lexical approach, autonomy

Heyoung Kim & Sooin Chun

Department of English Education

Chung-Ang University

Seoul 156-756, Korea

Email: englishnet@cau.ac.kr; sooic@unitel.co.kr

Received in February, 2008.

Reviewed in April, 2008

Revised version received in May, 2008