A Corpus-based Analysis of Collocations in Tenth Grade High School English Textbooks*

Hye-Young Choi (Hanyang University)
Yuah V. Chon (Hanyang University)

Increased corpus-based language studies have resulted in research on individual words in textbooks, but these teaching materials have not been extensively analyzed on the type and frequency of collocations. Previous research has been limited to studying certain parts of textbooks or specific types of collocations. By analyzing 16 Common English I High School textbooks of English (1,441,402 running words), 852 lexical collocations were extracted with the 41 most frequent content words used as nodes. For detailed analysis, 50 high-frequency lexical collocations were compared to collocations in the 2k Graded Reader native-speaker corpus for their frequencies and usages. Collocations that are closely related to learners' real life and interest (e.g., *volunteer work, good grade, fast food*) were found while the number of collocations combined with adverbs was relatively small comprising only 41 out of 852 collocations (less than 5%). The comparison of the fifty most frequent collocations in the textbook with the 2k Graded Reader native-speaker corpus produced considerable differences. The results suggest that a variety of high-frequency collocations from native-speaker corpus need to be given thought in the process of materials development. The study suggests contrastive instruction for teaching collocations to L2 learners.

I. INTRODUCTION

Knowledge of multiword units and the learning problems associated with them have been recognized for some time (Granger, 1998; Lewis, 2000; Moon, 1997; Palmer, 1933; Pawley & Syder, 1983; Schmitt, 2004; Sinclair, 1991; Wray, 2002). Different terms have been used to express multiword units, which have also been referred to as prefabricated patterns, preassembled units, formulaic sequences, chunks, ready-made utterances, and so forth (Foster, 2001; Howarth, 1998; Nattinger & DeCarrico, 1992; Wray, 1999). More importantly, they are viewed as a necessary component of second language (L2) lexical competence in addition to the knowledge of single words. Many have documented on how knowledge of multiword units contributes to the quality and fluency of spoken and written language (e.g., Altenberg & Granger, 2001; Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Laufer & Girsai, 2008; Laufer & Waldman, 2011; Lee, Shin, & Chon, 2009) where mastery of the idiomatic dimension of language is seen to make the learner come across as proficient and fluent. In fact, Hill (1999) has noticed how learners may often be at a disadvantage when they do not know four or five most important collocations that is crucial to expressing their intended messages.

In accordance with consensus on the importance of collocational knowledge, research has been conducted on native-speaker corpus (Biber, Conrad & Cortes, 2004; Grant & Nation, 2006; Liu, 2003; Shin, 2007; Shin & Nation, 2008) or learner corpus (Altenberg & Granger, 2001; Laufer & Waldman, 2011; Nesselhauf, 2003; Yamashita & Jiang, 2010). Within the Korean context, research has been conducted for single word items (Kim & Suh, 2006; Kwon, 2002, 2004; Shin & Chon, 2011; Yoon, 2009) or specific parts of speech (E.-J. Lee, 2010), but we are still limited in the number of studies that have been conducted on collocations of English textbooks, while some exceptions are Jung (2009), J.-K. Lee (2009), Y. M. Lee (2005) and Moon (2009). The studies are also limited in analyzing the range and types of textbooks, or focus on restricted types of collocations. Based on these findings, the researchers felt it was propitious to take a comprehensive approach to analyzing collocations by compiling the corpus of 16 high school textbooks at the 10 grade level, which resulted in the analysis of 1,441,402 running words. Considering that the main source of language input for teaching and learning in the EFL context is the textbook, it was indubitably valid to conduct an analysis of collocations in the textbooks by also comparing them to those in the native-speaker corpus for authenticity and usefulness. The findings are expected to inform researchers, practitioners, and materials writers on the types of collocations that need to be considered when future

materials are used as sources of language input for L2 learners.

II. BACKGROUND

1. Defining Collocations

Different researchers have used different definitions to define collocations so that there is not one simple and precise definition. Nation (2001) seems to include under the term 'collocations' most multiword units, including idioms (e.g., *kick the bucket)* and fixed expressions (e.g., *to and fro, leap year*). Boers et al. (2006) explain collocations as one type of multiword expression, other types being fillers (e.g., *sort of*), functional expressions (e.g., *excuse me*), idioms (e.g., *back to square one*), proverbs (e.g., *let's make hay while the sun shines*), and standardized phrases (e.g., *There' a growing body of evidence that*) (p. 246).

Schmitt and Carter (2004) distinguish between collocations and formulas. Collocations have been described as multiword units that consist primarily of open class items, such as *heavy traffic*, *strong tea*, and *kill time*, which may share certain characteristics with formulas. That is, both types of multiword units may be stored holistically as single lexical units, at least in the case of highly frequent, and probably short collocations. However, they distinguish collocations from formulaic sequences (e.g., *at first, in front of, on the other hand*) to be different in some important ways. First, collocations are looser combinations of words than formulaic sequences, in the sense that a component word in a collocation may collocate frequently with many other words to form other collocations. For example, 'heavy' not only collocates with 'traffic' but also with 'stone' and 'smoker', while 'in front of' makes a unit in which the constituent words cannot be replaced by another.

In the study on the high-frequency collocations of spoken and written English, Shin (2007) generally defines a collocation as a group of two or more words that occur frequently together where a collocation is made up of two parts—a pivot word (also called a 'node' which is the focal word in the collocation and its collocate(s), the word or words accompanying the pivot word. As such, *next week* is a collocation made up of the two collocates next and week, one of which is the pivot word. Nesselhauf (2003) distinguishes between free combinations (e.g., *want a car*), collocations (e.g., *take a picture*), and idioms (e.g., *sweeten the pill*) to classify verb-noun collocations. Examples of collocations are take a picture/photograph in which the noun used is unrestricted; however, the sense of the verb is restricted, so that the verb in the sense in which it is used can only be

combined with certain nouns. Following this criterion, *take a film/movie* would not be acceptable. To define collocations in our study, we used Nesselhauf's definition of collocations, but we did not restrict these to those of verb + noun combinations, but extracted multiwords when the pivot words of the collocations were content words. However, as Nesselhauf (ibid) has admitted, the classification among the multi-word items has a number of limitations, which may make their exact delimitation impossible. To operationalize a clear coding scheme for our analysis, we selected only content words as pivot words for analysis of collocations, so that all function words and proper nouns were excluded when selecting the pivot words.

2. Corpus-based Studies on Collocations

The previous research on corpus-based studies of collocations provide a great deal of information on how the use of collocations is problematic for L2 learners, regardless of the instruction they have received in L2. Bahns and Eldaw (1993) found that the number of collocation errors was twice as high as the number of errors involving single lexical items in the translation tasks they assigned. Barfield (2007) found that Japanese learners of English reported better knowledge of single nouns and verbs than they did of combinations of the same nouns and verbs into collocations. In comparison to native-speakers, learners have also been found to overuse some collocations; for example, collocations constructed with core verbs (*be, have, make, etc.*) or particular amplifiers (*very, completely, highly, strongly*), whereas they do not use other nativelike collocations (Altenberg & Granger, 2001; De Cock, Granger, Leech, & McEnery, 1998; Granger, 1998). Hasselgren (1994) argued that inappropriate collocations resulted from overdependence on the familiar structures that learners learned early, used widely, and with which they felt comfortable. She referred to these as 'lexical teddy bears.'

Nesselhauf (2003) analyzed collocations of learner corpus at the advanced German-speaking learners of English on the use of verb-noun collocations in free written production, which as a result produced categorization on the types of mistakes and seeing the influence of learners' L1 on the production of collocations. In another study on learners' production of collocations, Laufer and Waldman (2011) investigated learners' production of collocations by comparing to the corpus of young native-speakers (i.e., LOCNESS corpus). The data revealed that the learners even at different proficiency levels produced far fewer collocations than native-speakers, that the number of collocations increased only at the advanced level, and that errors, particularly interlingual ones, continued to persist even at advanced levels of proficiency. Alternberg and Granger (2001)

report on EFL learners' use of high-frequency verbs, in particular the use of the verb 'make.' Learner data was compared with native-speaker data. Results showed that EFL learners, even at high proficiency levels, had great difficulty with a high-frequency verb such as 'make.' In comparison, there have been studies examining collocations in native-speaker corpus (Biber, Conrad & Cortes, 2004; Grant & Nation, 2006; Shin, 2007; Shin & Nation, 2008). One example is Shin's (2007) study where he examines the high-frequency collocations of spoken and written English by utilizing a 10,000,000 token spoken corpus from the British National Corpus, and a 10,000,000 token written corpus, including the Australian Corpus of English (ACE), the Brown corpus, the Lancaster-Oslo/Bergen (LOB) corpus, the Freiburg-Brown (FROWN), Freiburg-LOB (FLOB) corpora and more to find the most frequent, grammatically well-formed English collocations. All in all, although the types of studies conducted on the learner corpus provide valuable insights on learners' problems regarding collocations in comparison to those used by native-speakers, the type of studies are limited in informing researchers and practitioners in the range of collocations that are taught through their learning materials (i.e., textbooks). The next section elaborates on this area to investigate studies that have been conducted in relation to collocations that are presented in textbooks.

3. Collocations and Corpus of Textbooks

There has been some research devoted to analyzing single word items in learning materials, particularly for textbooks based on the national curriculum (Kim & Suh, 2006; Kwon, 2002, 2004; Shin & Chon, 2011; Yoon, 2009), but to the researchers' knowledge, there are scarcity of studies conducted in particular on how collocations are presented in these textbooks. The few studies have been conducted by N.-B. Kim (2004), J.-K. Lee (2009), and Moon (2009). N.-B. Kim identifies the characteristics of collocations and the extent of their uses in high school 10th grade textbooks. The textbook was seen to offer a variety of 'adjective + noun' and 'noun + noun' collocations, but not 'verb + noun' and 'adverb + verb' collocations. Moon (2009) analyzed 35 textbooks of the 7th National Curriculum by examining collocations that included nouns in the reading passages of the materials. When the results were compared to native-speaker corpus (British National Corpus, Longman Lancaster Corpus) for evaluation on frequency and naturalness, some common collocations were found across both corpora, but noticeable differences were found in the frequency of the collocations.

While alerting us to the problem-orientedness of the Basic Word List of the 2007 National Curriculum, J.-K. Lee (2009) investigated the collocability of the word list in the

revised national curriculum, in particular for 'verb + noun' combinations. She claims that it is mainly on medium-strength collocations (e.g., *strong coffee*) that learners are expected to have most difficulties, but that the textbooks do not provide training on collocations for these types. She alerts us to how there should be not just a basic word list, but also a collocation list to supplement the use of textbooks. As demonstrated so far, the limited number of research conducted on the presentation of collocations in textbooks for Korean high school learners compelled us to conduct the present study to extract a comprehensive collocation list, and analyze the high-frequency collocations. The research questions of the present study are:

RQ1: What are the high-frequency collocations that can be compiled from High School English I or tenth grade textbooks of the Common National Curriculum?

RQ2: What are the features of the high-frequency lexical collocations?

RQ3: How do the 50 high-frequency textbook collocations in particular differ (in terms of frequency and use) with those in the native-speaker corpus?

III. METHODS

1. Textbook and Native-Speaker Corpora

For the analysis of collocations, High School English I or 10th grade high school textbooks of the Common National Curriculum (Revised 7th National Curriculum) from 16 different publishers was compiled for analysis. Our choice of the textbook originated from seeing that the tenth grade is the last year of the Common National Curriculum, which most learners are required to study. In total, the textbooks yielded 1,441,402 running words, which included the corpora of workbooks available to accompany the textbooks.

Once the textbooks were analyzed for the collocations, there was need to find a native-speaker corpus valid for comparative analysis. The choice of corpus that seemed valid for comparison was the Graded Corpus (consisting of 920,000 running words) at the 2nd 1,000 word level (2k) whose corpus could be accessed at *Compleat Lexical Tutor* (http://www.lextutor.ca/concordancers/concord_e.html). According to the citation notes, the Graded corpus is 'formed of hundreds of graded readers, scanned and digitized over 10 years, and the overall Vocabulary profile of the 2,000 word families is equal to 95% of the running words overall (when not counting the proper nouns).' That is, the corpus is

largely composed of items that serve the need for pedagogical concordancing, those lexical items or other patterns that learners need to become familiar with. Considering that the textbook corpus has also been controlled in terms of the vocabulary that can be presented (i.e., the number of word families permitted for teaching is within 1,810 words for High School English I; Shin & Chon, 2011), comparison with the Graded Corpus was deemed suitable.

2. Instrument of Analysis

For the analysis of the textbook corpus, Wordsmith Tools 4.0 (Scott, 2004) was utilized to retrieve the collocations. The Wordsmith Tools is an integrated suite of programs for looking at how words behave in texts. For our analysis, the 'WordList function' and 'Concord function' was used. The WordList function allows the researcher to see a list of all words in a text set out in frequency order. Once the pivot words of interest were selected via the WordList function, a concordancer, the Concord function provided us with chances to see any word or phrase in context. That is, the concordancer allowed us to see the type of words that accompany the pivot words.

3. Procedure for Data Analysis

Since the focus of the present study was to produce a list of high-frequency collocations, the Wordlist function was first used to sequence the textbook corpus in the order of word frequency. The next step of the analysis involved selecting the pivot words that would make up the collocations. Since in the present study we were interested in meaningful units of words, we selected content words, such as, nouns, verbs, and adjectives as the pivot words. For instance, prepositions, pronouns, determiners, conjunctions, and auxiliary verbs were excluded from the list of pivot words (Shin, 2007). Proper nouns were also eliminated for generalizability. Once the pivot words were selected, each of the pivot words was searched in Concord function with horizons set to 2 for both left and right sides and a cutoff point of 4 to obtain output on the collocates of the pivot words.

Although the researchers used horizons set to 2 for both left and right sides of the pivot words (i.e., we permitted collocations that constituted a pivot word + 2 collocates to the left and right), some exceptions had to be permitted when meaningful collocations appeared beyond the number of intended collocates. For instance, when the word 'have' was used as a pivot word, the words that could be coded within the word span of
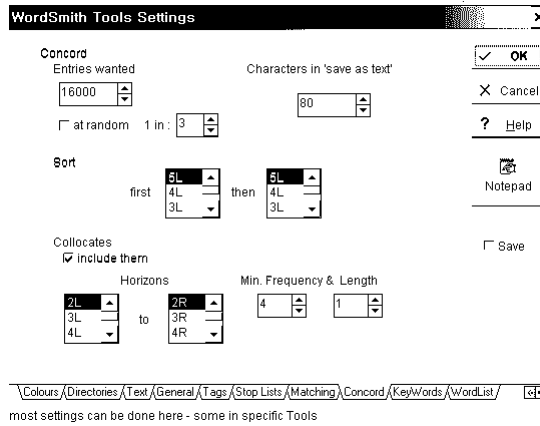
[Figure 1] Settings of Horizons for Collocates

collocates were *have time*, *have more time*, and *have enough time*. However, when meaningful collocations were found, such as in *have a good time* or *have a great time*, we extended the number of collocates to 3 words. As such, the coding of collocations required fine manual analysis of the pivot words and the accompanying collocates to identify meaningful collocational units.

As a way to validate the mutiwords as collocations, we also used the *Oxford Collocation Dictionary for students of English in CD-ROM* (McIntosh, Francis, & Poole, 2009). That is, we coded words as collocations only when they appeared in the reference material (Such methods of coding collocations have been adopted by Barfield, 2007; Laufer & Waldman, 2011; Nesselhauf, 2003; Wang & Shaw, 2008). Also, auxiliary verbs (e.g., *will listen, would listen*) were excluded since they did not fit within our classification of collocations (i.e., which are neither 'free combinations' or 'idioms'). After the list of collocations were produced, there was a process of lemmatization (i.e., the process of grouping together the different inflected forms of a word so they can be analysed as a single item). That is, all different inflected forms (e.g., *get old: get old, getting old, get older, getting old*) were categorized under a lemma (i.e., a word considered as its citation form together with all the inflected forms).

To produce the final list of collocations, we utilized 41 high-frequency words (i.e., content words). The collocations were also compared with the 2nd 1,000 word (2k) Graded Corpus (Refer back to *1. Textbook and Native-Speaker Corpora* for description of the corpus). As done with the textbook corpus, a cutoff point of 4 was used to search for the pivot words in the 2k corpus, and the number of collocations was also calculated in the same way by counting the lemma words.

IV. RESULTS AND DISCUSSION

1. Preliminary Analysis

In order to extract the high-frequency collocations of High School English I textbooks of the Common National Curriculum, there was first need to retrieve the pivot words for the analysis of collocations. The textbook corpus from the 16 publishers produced 145 different word types. However, since we were interested in using only the content words as the pivot words (nodes), the elimination of functions words (i.e., prepositions, pronouns,

[Table 1] High-frequency Words of High School English I

Rank	High-frequency words	Rank	High-frequency words
1	DO* (11,068)	25	WRITE (2,061)
2	HAVE* (8,525)	26	USE (2,017)
3	LIKE (4,689)	27	TALK (1,981)
4	THINK (4,539)	28	DAY (1,964)
5	PEOPLE (4,519)	29	LOOK (1,937)
6	LISTEN(4,010)	30	FIRST (1,854)
7	GOOD (3,618)	31	WORK (1,846)
8	READ (3,495)	32	GOING (1,834)
9	TIME (3,410)	33	WORLD (1,743)
10	GO (3,174)	34	SURE (1,722)
11	DID* (3,029)	35	RIGHT (1,683)
12	MAKE (2,991)	36	GREAT (1,680)
13	GET (2,730)	37	READING (1,628)
14	SCHOOL (2,567)	38	LESSON (1,447)
15	HAS* (2,529)	39	STEP (1,436)
16	KNOW (2,511)	40	WAY (1,434)
17	WANT (2,469)	41	MAN (1,402)
18	ENGLISH* (2,374)	42	FOOD (1,377)
19	TAKE (2,319)	43	NEED (1,369)
20	HELP (2,231)	44	LIFE (1,361)
21	HAD* (2,215)	45	CHECK (1,317)
22	SEE (2,144)	46	USING (1,299)
23	NEW (2,076)	47	WRITING (1,253)
24	WORDS (2,068)		

Note: ( ) indicate frequencies; words marked with * were eliminated from the list of pivot words

conjunctions, modal verbs) and proper nouns produced 47 words. Further elimination of grammatical words (i.e., *do, did*, *have, has, had that have*), and the proper noun English produced 41 pivot words as seen in Table 1.

2. Textbook Collocations

This section answers the first research question: What are the high-frequency collocations that can be compiled from High School English I textbooks of the Common National Curriculum?

In this study, we were ultimately interested listing the high-frequency collocations of the tenth grade high school textbooks rather than the single word items. The search with the pivot words at Concord function to extract concordance lines produced 852 high-frequency collocations. In the process, idioms (e.g., *get started*), and free combinations (e.g., *great people*) were eliminated from the analysis. As mentioned previously, the collocations that were valid for coding were those that fell between the categories of free combinations and idioms. Flexibility was permitted to some extent, but we excluded collocations that included grammatical functions. With lemmatization in the final stage of analysis, collocations such as *want to see*, *wants to see*, *wanted to see* were all included under the category of want to see. Table 2 presents 50 of those collocations. Due to space limitation, only 300 of the 852 collocations are presented in the Appendix.

As seen in Table 2, it can be seen that words, such as, *high school*, *school student*, *middle school*, *go to school*, *make friend*, *good friend*, *new friend*, and *volunteer work* were associated with learners' immediate lives. Collocations such as *use a word*, use expressions, and key words seem to have been listed due to how they often appear in the textbooks for stating directions in exercises or activities. There were other collocations including expressions of time (e.g., *have time, free time, take time*) being closely related to how the learners manage time. The list, however, as useful as they may be for their daily lives, indicates that the collocations may not be so effective for expanding the learners' academic vocabulary knowledge, or to prepare for the CSAT (College Scholastic Ability Test) which requires vocabulary at the 4,200 word level (i.e., a bit beyond the 4th 1,000 word level) (Joo, 2008). Although the list can be considered useful for learners' coverage of high-frequency collocations, such as at 1,000 to 2,000 word levels, extra training of collocations of those including low frequency words may separately be needed.

[Table 2] 50 High-frequency Collocations

Rank	Collocations	Rank	Collocations
1	high school (555)	26	take a picture (102)
2	volunteer work (344)	27	good job (99)
3	good idea (317)	28	use computer (99)
4	have time (271)	29	use expressions (98)
5	make sure (232)	30	sound great (96)
6	long time (226)	31	see a movie (95)
7	get ready (204)	32	good friend (91)
8	do the work (200)	33	read newspaper (89)
9	read a book (181)	34	take order (89)
10	take care (159)	35	real life (89)
11	school student (147)	36	first time (87)
12	fast food (144)	37	poor people (83)
13	use a word (142)	38	need help (81)
14	next time (141)	39	go to see (81)
15	old man (129)	40	work together (81)
16	one day (127)	41	new friend (80)
17	eat food (124)	42	good luck (74)
18	spend time (123)	43	take a look (73)
19	look good (118)	44	take time (72)
20	want to see (117)	45	go to school (71)
21	very good (113)	46	go to bed (69)
22	make friends (108)	47	let sb see (69)
23	free time (106)	48	good time (67)
24	middle school (106)	49	get sleep (67)
25	good point (103)	50	key words (66)

Note: The numbers in parentheses indicate frequency of collocations

3. Features of Collocations

In accordance with our second research question, this section tries to illustrate the features of the collocations. There was first an examination of the pivot words and the

collocates. This was followed by a semantic analysis of the collocations, and a categorization of the collocations according to their forms.

1) Pivot Words and Collocates

In order to see which of the pivot words were most commonly used, we collated them by the common pivot words and arranged the collocates according to their frequency. Table 3 presents the first 10 high-frequency pivot words and their collocates. In the process, not only the function words *do* and *have*, but other function words (i.e., *get, give, keep, look, make, put, take*) were additionally eliminated due to them being de-lexicalized words (Lewis, 1997). As mentioned previously, we were interested in pivot words that have meaning even on its own (i.e., content words).

As seen in Table 3, the pivot words combined with collocates most frequently in the order of GOOD (78 collocates), GREAT (49), TIME (45), USE (40), FOOD (36), NEW (35), LIFE (30), SCHOOL (28), READ (27), and WORLD (26). In fact, when these pivot words were searched in the Basic Word List of the 2009 Revised National Curriculum, the pivot words were found to be those in the list of the 736 words recommended for learning in elementary schools.

Some of the most frequently mentioned collocations can be listed as being *good idea*, *look good*; *great time*, *sound great*; *have time*, *time management*; *use (a) computer*, *usually use*; *fast food*, *food group*; *new friend*; *real life*, *life story*; *high school*, *school life*; *read (a) book*, *just read*; *globalized world*, and *world cup*. At this point, we may need to question whether these collocations are fit to be introduced at the high school level. At an exploratory level, when we examined the corpus of middle school textbooks from grades 7-9, we found that a large number of them had already been introduced. As such, we claim that the presentation of the collocations may have worked for consolidation rather than in expanding the learners' collocational knowledge.

[Table 3] 10 High-frequency Pivot Words and Collocates

Rank	Pivot Word	Collocates
1	GOOD	(Total of 78 collocates) GOOD ___: idea, point, job, friend, luck, time, way, grade, habit, deed, place, choice, news, student, listener, things, writer, example, advice, care, cook, shape, chance, movie, book, player, reader, impression, sense, food, manners, etiquette, management, speaker, day, plan, qualities, singer, figure, fortune, health, jokes, people, presentation, score, site, marks, acting, boy, education, friendship, leaders, neighbors, question, result, start, story, athlete, attitude, communication skills, condition, dancer, diplomat, fight, memories, mood, person, relationships, selection, teacher ___ GOOD: look, very, really, pretty, sound, feel, taste, especially
2	GREAT	(Total of 49 collocates) GREAT ___: time, idea, job, thing, chance, importance, opportunity, success, musician, music, place, advance, help, hero, love, book, news, experience, music, teacher, difference, genius, achievement, curiosity, fun, length, friends, variety, painter, beauty, dancer, day, effort, influence, joy, mind, museum, poetry, writing, scientist, sense, singer, talent great to see, a great number of ___ GREAT: sound, really, look, feel
3	TIME	(Total of 45 collocates) ___ TIME: have, long, next, spend, free, first, take, good, great, this, hard, save, last, make, waste, every, little, each, spare, need, quiet, short, third, wonderful, find, use, devote, computer, departure, invest, leisure, set aside, that amount of time, period of time, many times, couple of times, all time favorite TIME ___: travel, management, passes, flies, period, goes by, limit
4	USE	(Total of 40 collocates) USE ___: word, computer, expression, information, cell phone, internet, transportation, products, phone, language, fuel, spray, phrase, lighting, energy, tactic, technology, creativity, shampoo, trick, code, machine, strategy, time, vocabulary, chemicals, dictionary, card, pencil, materials, sound, tip, web get used to, make use of, let sb use, easy to use, be allowed to use ___ USE: usually, illegal
5	FOOD	(Total of 36 collocates) ___ FOOD: fast, eat, junk, cook, like, favorite, traditional, healthy, enjoy, spicy, enough, good, buy, order, delicious, fatty, leftover, nutritious, right, serve, have, processed, GM, love, make, provide, real give sb food FOOD ___: group, package, source, company, contains, come, court, taste

Rank	Pivot Word	Collocates
6	NEW	(Total of 35 collocates)
		NEW ___: friend, car, idea, invention, way, year, technology, book, kind, neighbor, record, generation, job, wave, fashion, method, toy, year's eve, rule, biography, source, form, look, research, world, recipe, buildings, clothes, design, direction, novel, skills, style, system
7	LIFE	(Total of 30 collocates)
		___ LIFE: real, school, everyday, daily, human, animal, early, save, live, change, enjoy, healthy, rural, spend, city, devote, long, modern, whole, entire, private, battery, busy, happy, later, normal, student, future, hard
		LIFE ___: story
8	SCHOOL	(Total of 28 collocates)
		___ SCHOOL: high, middle, elementary, medical, enter, start, law, open, local
		go to school
		SCHOOL ___: student, life, uniform, year, days, team, trip, newspaper, bus, band, building, rules, teacher, counselor, lunch, work, boy, friend
9	READ	(Total of 27 collocates)
		READ ___: book, newspaper, article, story, novel, map, words, passage, letter, sentence, instruction, magazine, text, poem, paper, essay, manual, paragraph, biography, label, message, work, aloud, blog, statement
		learn to read
		___ READ: just
10	WORLD	(Total of 26 collocates)
		___ WORLD: globalized, real, old, change, whole, natural, digital, outside, imagine, save, modern, small, new, travel, create, explore, entire, see
		live in world
		WORLD ___: cup, famous, record, war Ⅰ, cultures, championships, record

In the textbook, we also found other collocations related to health, such as, *good health, good figure, good shape, healthy life,* and *healthy people*, which indicate that the textbook writers were also concerned about the well-being of the high school learners in accordance with the National Curriculum (i.e., topics about habits, health, activities, etc. is proposed as one of the 'Subject Matters' to be included in textbooks).

In the corpus, TIME was the third ranking pivot word for having the most number of collocates. When the usages of the collocations were examined, it was apparent that they were substantially related to efficient time management. Related to the topic were *spend*

time, free time, save time, time management, make time, schedule time, and *waste time,* and the specific examples from the corpus are illustrated as follows. The context in which the collocations are used is didactic in tone, trying to alert students on the importance of time:

Examples with *time:*
- I don't think doing homework as a group is helpful for students. There is a greater chance that time will be wasted. For example, they can **spend time** chatting with friends instead of studying.
- Volunteers spend their **free time** doing things to make their community a good place to live in.
- You'd better take the subway. You can **save time** and help reduce air pollution.
- How you **use** this **time** is the key to **time management**. Before you schedule time for watching TV or playing computer games make sure you **schedule time** for your important activities.

In the corpus, *cell phone* was another frequently occurring multi-word item. While *cell phone* appeared 353 times in the corpus, the most commonly appearing collocations were *use cell phone* (55), *use phone* (18), *make a phone call* (6), and *buy a cell phone* (6). Except the instance of *make a phone call*, the rest of the collocations were referring to the use of cell phones. Following is an extended detailed example:

Examples with *cell phone:*
A: Do you agree with the idea that students should be allowed to use **cell phones** at school?
B: I'm not happy with that idea.
A: Why?
B: Because my friends sometimes send text messages during class.

As seen in the above example, the context was related to how students should not be allowed to use cell phones not only in schools but also in public places (e.g., theaters, libraries) and while driving. In relation to the recent high-tech instruments, *use the internet* (48), *use computer* (99), *computer time* (4), *use the web* (4), *read blogs* (4) could also be found. The use of *laptop* (20) was also found in the corpus with *have a laptop*

computer (3), and *buy a laptop computer* (6). The following are the examples:

Examples with *laptop computer:*
W: How do you like your new **laptop computer?**
M: It isn't as fast as my old computer.
W: Really? It good you can carry it around though.

W: Yesterday, I went to buy a new **laptop computer** with my dad. There were so many kinds to choose from at the store. But, we finally found a good one. It was cheap, but it looked as fancy as my sister's new computer. I'm really happy to have it.

In the same vein, Yoon (2009) has also found the use of these high-tech words, such as, *cell phone, text message, blog, mobile, cyber* to be a common feature of the middle school textbook corpus. All in all, the inclusion of modern and futuristic word items or collocations can be considered attempts of textbook writers to present most recent trends and concepts.

2) Different Forms of Collocations

In this section, there is the analysis of the different forms of collocations that have been found in the textbooks corpus. Collocations that included adverbs (i.e., 'adverb + adjective', 'adverb + verb', 'verb + adverb') were largest in number, and they also exceeded those that came in the combinations of 'adjective + noun', 'verb + adjective', 'noun + noun', 'verb + noun', and 'verb + verb.' Of the 852 collocations, there were 41 (4.81%) of them that included adverbs where 'really' ranked the highest. This is in accordance with how N.-B. Kim (2004) also found similar results in his analysis of high school textbooks, finding *really* the most common adverbial collocate. The collocations that included the use of adverbs are presented in the order of frequency as follows in Table 4, where the most frequently appearing adverbial collocation was very good.

For more detailed analysis, the high-frequency adverbial collocations (i.e., *really, together, just, well, especially*) were also listed according to their collocates whenever appearing at least twice in the list of adverbial collocations, as presented in Table 5.

When the adverbs were examined, collocations including *really* appeared the most number of times. Adverbial collocations including *really* (288) appeared more than three times over those including *together* (91), and this may indicate that really has been

overused in comparison to other adverbs. In fact, the range of adverbs that have appeared in the textbooks is relatively limited. In comparison, the type of adverbs used in native-speaker corpus have been *maybe, probably, apparently, of course, definitely* (Karkkainen, 2003), and *probably, maybe, of course, certainly, definitely,* and *perhaps* (Biber, Johansson, Leech, Conrad, & Finegan, 1999) researched all in the Corpus of Spoken American English. In relation to the types of collocations, Park (2003) has found Korean learners to have the most problems with collocations in the form of 'adverbs + adjectives.' As such, this type of finding would need to be considered in the development of future materials for Korean learners.

[Table 4] Adverbial Collocations of Textbooks (adv. + adj., adv. + verb, verb + adv.)

Rank	Collocations	Freq.	Rank	Collocations	Freq.
1	very good	113	22	not so sure	12
2	work together	81	23	talk loudly	12
3	really want	66	24	know well	11
4	really good	65	25	like better	11
5	really like	62	26	see clearly	11
6	pretty good	60	27	go together	10
7	work hard	55	28	always want	9
8	really great	40	29	especially like	9
9	really need	25	30	especially good	7
10	like best	24	31	think again	7
11	just read	19	32	think twice	7
12	like ___ a lot	19	33	know exactly	6
13	listen carefully	19	34	quite right	6
14	go well	18	35	write so well	6
15	usually use	18	36	just right	5
16	just need	17	37	read aloud	5
17	really think	16	38	look closely	4
18	really help	14	39	never want	4
19	work well	13	40	no longer need	4
20	just want	12	41	particularly like	4
21	like sb/sth very much	12			

[Table 5] Frequent Adverbs and Collocates

Frequent adverbs	Collocates	Freq.
REALLY	want(66), good(65), like(62), great(40), need(25), think(16), help(14)	288
TOGETHER	work(81), go(10)	91
JUST	read(19), need(17), want(12), right(5)	53
WELL	go(18), work(13), know(11), write(6)	48
ESPECIALLY	like(9), good(7)	16

4. Native-speaker Collocations and Textbook Collocations

This section answers the third research question on how the 50 high-frequency collocations differ (in terms of frequency and use) with native-speaker collocations. In order to compare the 50 high-frequency collocations from the corpus of 1,441,402 words, the 2k Graded Corpus consisting of 920,000 words was used (see Table 6). Since there was some difference in the sizes of the two corpora, the frequency of the collocations was calculated per 100,000 words for normalization.

[Table 6] 50 High-frequency Collocations in Textbook and Native-speaker Corpus

Rank	Collocations	Textbooks		Native-speaker corpus	
		Freq.	Rate (per 100,000)	Freq.	Rate (per 100,000)
1	▸ high school	555	[39]	28	[3]
2	▸ volunteer work	344	[24]	0	[0]
3	▸ good idea	317	[22]	52	[6]
4	have time	271	[19]	147	[16]
5	make sure	232	[16]	107	[12]
6	long time	226	[16]	175	[19]
7	▸ get ready	204	[14]	25	[3]
8	do the work	200	[14]	117	[13]
9	read book	181	[13]	40	[4]
10	take care	159	[11]	100	[11]

Rank	Collocations	Textbooks		Native-speaker corpus	
		Freq.	Rate (per 100,000)	Freq.	Rate (per 100,000)
11	▶ school student	147	[10]	1	[0]
12	▶ fast food	144	[10]	3	[0]
13	use a word	142	[10]	27	[3]
14	next time	141	[10]	37	[4]
15	▶ old man	129	[9]	196	[21]
16	▶ one day	127	[9]	184	[20]
17	eat food	124	[9]	24	[3]
18	spend time	123	[9]	126	[14]
19	look good	118	[8]	15	[2]
20	want to see	117	[8]	114	[12]
21	very good	113	[8]	143	[16]
22	make friends	108	[7]	31	[3]
23	free time	106	[7]	8	[1]
24	middle school	106	[7]	5	[1]
25	good point	103	[7]	2	[0]
26	take a picture	102	[7]	19	[2]
27	good job	99	[7]	16	[2]
28	use computer	99	[7]	4	[0]
29	use expressions	98	[7]	3	[0]
30	sounds great	96	[7]	1	[0]
31	see a movie	95	[7]	3	[0]
32	good friend	91	[6]	38	[4]
33	read newspaper	89	[6]	23	[3]
34	take order	89	[6]	0	[0]
35	real life	89	[6]	7	[1]
36	▶ first time	87	[6]	150	[16]
37	poor people	83	[6]	18	[2]
38	need help	81	[6]	42	[5]
39	go to see	81	[6]	33	[4]
40	work together	81	[6]	19	[2]

Rank	Collocations	Textbooks		Native-speaker corpus	
		Freq.	Rate (per 100,000)	Freq.	Rate (per 100,000)
41	new friend	80	[6]	12	[1]
42	good luck	74	[5]	26	[3]
43	take a look	73	[5]	17	[2]
44	take time	72	[5]	38	[4]
45	go to school	71	[5]	15	[2]
46	go to bed	69	[5]	97	[11]
47	let sb see	69	[5]	35	[4]
48	good time	67	[5]	22	[2]
49	get sleep	67	[5]	5	[1]
50	key words	66	[5]	0	[0]

Note: ▸ indicates collocations that differ by more than 10 between the two corpora

In the comparison, collocations that differed by more than 10 between the two corpora were *high school, volunteer work, good idea, get ready, school student, fast food, old man, one day,* and *first time.* Of these words, *old man*, *one day*, *first time* were used comparatively more in the native-speaker corpus, but this seems to have occurred due to the discourse feature of the graded readers that we chose as the reference corpus. All in all, the differences show how collocations, such as, *high school, volunteer work, good idea, get ready, school student, fast food,* and *see a movie* that appeared in the textbooks may be lacking authenticity for representing common collocations of English. In the following, we compared some of these collocations to the native-speaker collocations for the differences in usages. We will pay separate attention to *volunteer work*, *fast food*, and *sound great.*

Although *volunteer work* was the second ranking collocation in the textbook corpus, the specific collocation could not be found in the reference corpus. The concordance output with *volunteer work* indicated that the term had been used to encourage volunteering in schools. The following are some of the example sentences:

Examples with *volunteer work*:
- I get a lot out of doing **volunteer work**. Helping others is like helping myself.

- If everyone did some kind of **volunteer work** the world would be a better place to live in.

As such, *volunteer work* was further used to discuss how work should be conducted at 'Habitat for Humanity', nursing homes, orphanages, hospitals, food bank, and work camp overseas. Through volunteer work, students are advised to learn to appreciate the virtues of cooperation and act as members of the society while also valuing humanity, altruism, and ethics. However, the frequency indicates that volunteer work appeared excessively in the textbook due to being a part of their school lives. In fact, *volunteer work* could not be found in the BNC Spoken Corpus (1 million words), BNC Written Corpus (1 million words), or the Brown Corpus (1 million words) when searched at http://www.lextutor.ca/concordancers/concord_e.html As such, illustration with *volunteer work* indicates that it has hardly been used in native-speaker corpus, and that we may need to consider the learning benefits of collocations when including them in learning materials, which are the main sources of input.

At an exploratory level, when we searched the 2k corpus to see how *work* was collocated with other words, we found *hard work* (28 times), *best work* (6), *extra work* (4), *important work* (4), *detective work* (3), and *greatest work* (3). For collocations similar to volunteer work, we found *humanitarian work* and *voluntary work* respectively appearing once in the reference corpus. The following illustrates the examples from the Native-speaker corpus:

Examples with *work*:
- It is able to deploy highly trained, well-equipped troops to distant places, whether it be to fight wars, keep peace, provide training or do *humanitarian* **work**, and the population at home likes to see them do it.
- I have some *voluntary* **work** lined up with a community project in Liverpool.

Another collocation worthy of comparison was *fast food*. In the textbook corpus, *fast food* appeared 144 times (10 by normalization). On the other hand, the particular collocation was found only 3 times (0 by normalization) in the 2k Graded Corpus. See examples below:

Examples with *fast food* from 2k Graded Corpus:
1. There had been killings at **fast-food** restaurants and playgrounds, too.

2. All year long I have to cook ***fast-food****:* pizzas and burgers and chips.

3. While our evaluation indicated that McDonald's is a solid company, it also told us that the venerable ***fast-food*** chain probably would not continue to grow at 15 percent annually.

As mentioned previously, while the textbook was warning the students about the harmfulness of fast food and recommending the students to eat healthier food, the collocations have different functions in the 2k Graded Corpus. The exemplary sentences show that *fast food* in sentences 1 and 3 have been used as modifiers such as in *fast-food* restaurants and *fast-food* chain. In comparison, it is only in sentence 2 that it is used as a collocation as in the textbook, however, while not carrying any didactic tone. Also, while *junk food* appeared 51 times in the textbook corpus (ranked as 79 on the list of collocations), it did not appear at all in the 2k Graded Corpus. This can be attributed to how the textbooks are emphasizing the harmful effects of junk food. The following are some examples from the textbook with *junk food* in the concordancer.

---

33 B Well the secret lies in changing I quit drinking sodas and eating **junk food.** I jog and play badminton after school. A I'm going to try tho

35 g in the park. And on the way home from school, I almost bought some **junk food**. C If I really want to change myself, I've got to do something

39 tritious diet. Scientists say **junk food** can stress cause. Try to avoid junk food and keep your body in shape. Being in shape is one of the bes

40 do you think I should do? Sam: Your problem is that you eat too much **junk food**. And you don't exercise enough. Jongsu: Yes that right.

[Figure 2] Concordance Output for *Junk Food*

In fact, when we also searched the 2k Graded Corpus for the collocations related to *food,* we obtained collocations in the order of *eat food* (22 times), *have food* (15), *cook food* (7), *good food* (7), *enough food* (7), *buy food* (6), *serve food* (5), *plate of food* (5), *prepare food* (4), *real food* (4), *fast food* (3), *fine food* (3), *fresh food* (3), and *excellent food* (2), which demonstrates that we may need to include a larger range of collocations that are related to *food* in the textbook. Also, the Brown corpus (1 million words), produced food-related collocations, such as, *adequate food, best-looking food, cafeteria food, refrigerated food*, and *swallow food*. The comparative searches suggest that textbook writers may want to become more data-driven when developing materials by searching

authentic sources of English for checking on frequency information, degrees of usefulness, and the range of collocations that needs to be acquired by students at particular levels.

Sound(s) great was another collocation that differed conspicuously in terms of frequency between the corpora of the textbook and the 2k Graded Corpus. While it was found 96 times (7 by normalization) in the textbook, it was found only once (0 by normalization) in the latter, which shows that the collocation was far often used in the textbook. The examples that could be found according to the different types of corpora are as follows:

Examples with *sound great*:
[Textbook corpus]
W: Tomorrow we're visiting the World Cup Stadium.
M: **Sounds great**. I've always wanted to see it.

[2k Graded Reader Corpus]
001. before the heat gets too bad. Listen···?' 'Sorry, John. **SOUNDS GREAT**, but I must go!' said Hu and she quickly put the phone down.

Considering that *sounds great* is likely to be more prominent in spoken discourse but that the 2k Graded Corpus is closer to being a type of written discourse, we also tried finding the collocation in other corpora. When the search was conducted on two different types of corpora (i.e., BNC Spoken Corpus, and Brown Corpus), a hit could be found only once in the Brown Corpus. On the other hand, when the collocation *sounds good* was searched instead, the collocation could be found 50 times in the textbook, 3 times in the 2k Graded Corpus, 10 times in the BNC Spoken Corpus, and 3 times in the Brown Corpus. As such, the results show how *sounds good* rather than *sounds great* is a more frequently used collocation among native-speakers.

V. CONCLUSION

The present corpus-based study was conducted on analyzing 1,441,402 running words when corpus was compiled from 16 high school English I textbooks. As a result, 852 collocations were extracted by analyzing 41 most frequent content words. Analysis of the

pivot words and collocates yielded collocations closely related to learners' real life and interest, for example, *volunteer work, good grade, get grades, use cell phone, use computer, use the internet.* Collocations that are related to health such as *fast food,* as well as *junk food, healthy food, spicy food,* and *good food* were also found with high frequencies. Although the learning and teaching of the high-frequency collocations are related to the learners' daily life in general, the range of collocations do not suffice to prepare the students to communicate about various facets of their lives. Also, we suspect that the textbooks would be able to meet the needs of the learners who are preparing for the CSAT, which requires vocabulary at the 4,200 word level (Joo, 2008). Even when the collocation list is glanced through, which illustrates collocations such as *need help*, *go to see*, *work together*, *let sb see*, and *get sleep*, it can be seen that they may not be suffice in terms of coverge of low frequency words.

The comparison of the fifty most frequent collocations in the textbook corpus with the 2k Graded Corpus produced considerable differences. First, between the textbook corpus and native-speaker corpus, there was more than a difference of 10 frequencies (per 100,000 words), such as for *volunteer work, good idea, get ready,* and *fast food.* It may be concluded from the findings of the study that high-frequency collocations in the textbooks to not correspond to the collocations that commonly appear in the native-speaker corpus. However, we interpret this to have occurred in some parts due to how there is difference in the type of topics dealt with between the high school textbooks and the 2k Graded Reader corpus. The textbooks focus on the high school learners' daily school lives (e.g., volunteer work) while the 2k graded readers may deal with a larger range of topics.

On the basis of the above results, the researcher suggests some pedagogical implications. First, it would be essential for high-frequency collocations of native-speaker corpus to be given some consideration in the process of developing materials for L2 learners, who rely mainly on textbooks for language input. Even when intake occurs among learners by being exposed to these textbooks, the outcome may not be as productive when they occur for collocations that are not often used in native-speaker discourse. For this purpose, in accordance with what J.-K. Lee (2009) has mentioned, it may be necessary to include not just a basic word list of the National Curriculum of English, but also a basic collocation list so that it can be incorporated in textbook development. As suggested by Bernardini (2004), materials would be useful for the purpose of instruction when able to include words or multi-word items through which learners can be presented with comparative views of learner and native-speaker corpus.

Through this process, the learners are likely to notice gaps in their linguistic knowledge, which is a prerequisite for language learning.

Last but not least, we suggest 'contrastive instruction' as an effective way to teach collocations to L2 learners. As seen in the study conducted by Laufer and Girsai (2008), incorporating contrastive analysis and translation activities, where learners were led to understand the similarities and differences between their L1 and L2 in terms of individual words, made significant differences in recalling collocations. As such, we learn from the study that one way to make a foreign language feature noticeable or salient in the input is to enhance it by providing contrastive association with the corresponding L1 item. The explanation is that lexical items, such as collocations, will be acquired more efficiently when items become salient in the input and when learners receive information about the particular difficulties resulting from L1-L2 differences. Also, when trying to teach collocations, teachers should try to have learners aware of the interlingual differences which may result in errors. For instance, among Korean learners of English, we can occasionally see instances of *see an exam* rather than *take an exam,* which can be overcome by contrastive instruction. On this matter, (Nesselhauf, 2003) has also made suggestions on how it is not sufficient to merely teach the lexical elements that go together, but that it is necessary to teach entire combinations including prepositions, articles, etc. (e.g., knowing that *pass* can combine with *judgement* is less useful, than knowing that it is *pass judgement on* and not *pass one's judgement* or *pass judgements* or *pass judgements about*). To conclude, the researchers propose that there should be more consciousness-raising activities to draw the learners' attention to dealing with different types of multi-word items (e.g., idioms, phrasal verbs) in their process of language learning and use.

REFERENCES

Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics, 22*, 173-195.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System, 21*, 101-114.

Barfield, A. (2007). *An exploration of second language collocation knowledge and development.* Unpublished doctoral dissertation. University of Wales, Swansea.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future

developments. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam and Philadelphia: John Benjamins.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at . . .: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*, 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English.* London and New York: Longman.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*, 245-261.

De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67-79). London: Longman.

Foster, P. (2001). Rules and routines: A consideration of their role in task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swan (Eds.), *Researching pedagogical tasks: Second language learning, teaching and assessment* (pp. 75-97). London: Pearson.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.

Grant, L., & Nation, I. S. P. (2006). How many idioms are there in English? *International Journal of Applied Linguistics, 151*, 1-14.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*, 237-258.

Hill, J. (1999). Collocational competence. *English Teaching Professional, 11*, 3-7.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*, 24-44.

Joo, H. W. (2008). *A corpus-based analysis of vocabulary in the BEWL and the CSAT.* Unpublished master's thesis, Korea University, Seoul.

Jung, Y. (2009). A corpus-based study on the use of adverb synonyms 'Entirely, totally, and wholly'. *Linguistic Research, 26*, 85-101.

Karkkainen, E. (2003). *Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think.* Amsterdam: John Benjamins.

Kim, N.-B. (2004). A collocational analysis of Korean high school English textbooks and

suggestions for collocation instruction. *English Language & Literature Teaching, 10*(3), 41–66.

Kim, Y.-M., & Suh, J. (2006). Usage analysis of vocabulary in Korean high school English textbooks using multiple corpora. *English Language & Literature Teaching, 12*(4), 139–157.

Kwon, I.-S. (2002). A corpus-based lexical analysis of middle school English textbooks. *English Teaching, 57*(4), 409–444.

Kwon, I.-S. (2004). A corpus-based lexical analysis of middle school English textbooks of the 6th and the 7th National Curriculum. *Foreign Languages Education, 11*(1), 211–251.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29,* 694–716.

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61,* 647–672.

Lee, E.-J. (2010). The use of stance adverbials in the Korean EFL college students' writing corpus. *Foreign Languages Education, 17*(3), 347–366.

Lee, J.-K. (2009). Analysis of collocability of word list in the revised national curriculum. *Studies in Modern Grammar, 58,* 249–271.

Lee, M., Shin, D., & Chon, Y. V. (2009). Online corpus consultation in L2 writing for in-service teachers of English. *English Teaching, 64*(2), 233–254.

Lee. Y. M. (2005). *A Corpus-based analysis of <verb + NP> collocations in high school English textbooks*. Unpublished master's thesis, Hanyang University, Seoul.

Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice*. Hove: Language Teaching Publications.

Lewis, M. (2000). *Teaching collocation: Further developments in the lexical approach*. Hove: Language Teaching Publications.

Liu, D. (2003). The most frequently used spoken American English idioms: A corpus analysis and its implications. *TESOL Quarterly, 37,* 671–700.

Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 40–63). Cambridge: Cambridge University Press.

Moon, A.-N. (2009). Frequency distribution of noun collocations in Korean high school English textbooks: A corpus-based analysis. *The Journal of Studies in Language,*

24(4), 731–753.

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

Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*, 223–242.

McIntosh, C., Francis, B., & Poole, R. (Eds.). (2009). *Oxford collocations dictionary* (2nd ed.). New York: Oxford University Press.

Palmer, H. (1933). *Second interim report on English collocations.* Tokyo: Kaitakusha.

Park. S. H. (2003). *Lexical collocation use by Korean EFL college learners.* Unpublished master's thesis, Seoul National University, Seoul.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–225). London: Longman.

Schmitt, N. (2004). Formulaic sequences: Acquisition, processing, and use. Amsterdam: John Benjamins.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 1–22). Philadelphia: John Benjamins.

Scott, M. (2004). *Word Smith Tools. Version 4.0.* Oxford: Oxford University Press. Available online from http://whttp://www.lexically.net/wordsmith/version4/index.html

Shin, D. (2007). The high frequency collocations of spoken and written English. *English Teaching, 62*(1), 199–218.

Shin, D., & Nation, I. S. P. (2008). Beyond single words: the most frequent collocations in spoken English. *ELT Journal, 62*, 339–348.

Shin, D., & Chon, Y. V. (2011). A corpus–based analysis of curriculum–based elementary and secondary English textbooks. *Multimedia–Assisted Language Learning, 14*(1), 149–175.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Wang, Y., & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal, 32*, 201–232.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching, 32*, 213 - 231.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge: Cambridge University Press.

Yamashita J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*, 647-668.

Yoon, H. S. (2009). A corpus-based analysis of vocabulary in the revised middle school English 1 textbooks. *Modern English Education, 10*(2), 87-107.

APPENDIX

Rank	Collocations	Freq.
51	really want	66
52	make sb/sth look	65
53	really good	65
54	make a difference	64
55	use the information	64
56	first step	62
57	really like	62
58	get used to	61
59	project work	61
60	go shopping	60
61	pretty good	60
62	see a doctor	60
63	take part	59
64	time travel	59
65	get information	58
66	let sb know	58
67	good way	56
68	use cell phone	55
69	work hard	55
70	go home	54
71	read article	54
72	meet people	53
73	write letter	53
74	good grade	52
75	great time	52
76	this time	52
77	best way	51
78	hard time	51
79	junk food	51
80	have right	50
81	sounds good	50
82	save time	49
83	listen to music	48
84	use the internet	48
85	writing workshop	48
86	good habit	47
87	make an appointment	47
88	read story	47
89	hard work	46
90	make a complaint	46
91	school life	46
92	group work	45
93	write a report	45
94	feel good	44
95	get better	44
96	read novel	44
97	see things	44
98	everyday life	43
99	make a decision	42
100	school uniform	42
101	first prize	41
102	great idea	41
103	next day	41
104	go to college	40
105	really great	40
106	take a message	40
107	take a rest	40
108	take a walk	40
109	close reading	39
110	young people	39
111	same way	38
112	school year	38
113	take medicine	38
114	make mistakes	37
115	write answers	37
116	write the word	37
117	make use of	36
118	old people	36
119	take bus	36
120	take notes	36
121	cook food	35
122	like food	35
123	reading skill	35
124	take a shower	35
125	like music	34
126	look happy	34
127	make plan	34
128	time management	34
129	poor man	33
130	use public transportation	33
131	get grades	32
132	globalized world	32
133	local people	32
134	real world	32
135	young man	32

136	daily life	31	184	right hand	24
137	find a way	31	185	sick people	24
138	nice to see	31	186	some day	24
139	take a bath	31	187	take a class	24
140	take line	31	188	good student	23
141	homeless people	30	189	Let's talk	23
142	make a list	30	190	not need to worry	23
143	right amount	30	191	real people	23
144	take a taxi	30	192	take a trip	23
145	write an essay	30	193	traditional food	23
146	elementary school	29	194	work station	23
147	finish the work	29	195	world famous	23
148	go fishing	29	196	different way	22
149	good deed	29	197	look upset	22
150	good place	29	198	make a reservation	22
151	write the name	29	199	school days	22
152	amount of time	28	200	take a break	22
153	make sense	28	201	take a step	22
154	take turns	28	202	blind people	21
155	write the number	28	203	easy to use	21
156	human life	27	204	get a job	21
157	new car	27	205	good listener	21
158	get exercise	26	206	good things	21
159	good choice	26	207	look great	21
160	good news	26	208	school team	21
161	last time	26	209	get angry	20
162	look sad	26	210	good writer	20
163	make time	26	211	long way	20
164	memorize words	26	212	animal life	19
165	use products	26	213	change way	19
166	disabled people	25	214	get to know	19
167	great job	25	215	healthy food	19
168	many times	25	216	help (to) solve, help sb (to) solve	19
169	read map	25			
170	read words	25	217	just read	19
171	really need	25	218	like sb/sth a lot	19
172	take a nap	25	219	listen carefully	19
173	want to talk	25	220	make a living	19
174	waste time	25	221	make an effort	19
175	world cup	25	222	make sb happy	19
176	check progress	24	223	read passage	19
177	easy way	24	224	time passes	19
178	favorite food	24	225	wait to see	19
179	first thing	24	226	get nervous	18
180	like best	24	227	go well	18
181	look worried	24	228	good example	18
182	old world	24	229	great thing	18
183	own way	24	230	new idea	18

231	school trip	18	279	get news	14
232	thanksgiving day	18	280	good chance	14
233	the other day	18	281	good movie	14
234	use phone	18	282	great musician	14
235	usually use	18	283	little time	14
236	famous people	17	284	look tired	14
237	feel great	17	285	make a poster	14
238	good advice	17	286	make sth easy	14
239	good care	17	287	need information	14
240	just need	17	288	period of time	14
241	learn to read	17	289	read sentences	14
242	new invention	17	290	really help	14
243	new way	17	291	right person	14
244	ordinary people	17	292	save life	14
245	read letter	17	293	school band	14
246	school newspaper	17	294	see the sign	14
247	write a book	17	295	surprised to see	14
248	writing skills	17	296	take action	14
249	change the world	16	297	write a paragraph	14
250	happy to see	16	298	write words	14
251	new year	16	299	environmental work	13
252	really think	16	300	food group	13
253	stressed words	16			
254	whole world	16			
255	write a review	16			
256	ask for help	15			
257	common people	15			
258	every time	15			
259	feel this way	15			
260	get in touch with	15			
261	get married	15			
262	get rest	15			
263	good cook	15			
264	good shape	15			
265	great chance	15			
266	great importance	15			
267	great opportunity	15			
268	great success	15			
269	great to see	15			
270	know one's place	15			
271	new technology	15			
272	next step	15			
273	school bus	15			
274	write a journal	15			
275	early life	14			
276	elderly people	14			
277	enjoy food	14			
278	get hold of	14			

Author(s): Choi, Hye-Young (Hanyang University, 1st author);
            woomy@hotmail.com
            Chon, Yuah V. (Hanyang University, corresponding author);
            vylee52@hanyang.ac.kr