English Language & Literature Teaching, Vol. 19, No. 3 Fall 2013

Lexical Bundles in Korean College Students' English Essays: A Corpus-based Comparative Study

Junghyun Kim
(Seoul Theological University)

Kim, Junghyun. (2013). Lexical bundles in Korean college students' English essays: A corpus-based comparative study. *Language & Literature Teaching, 19*(3), 157-179.

The current study examined Korean college students' use of lexical bundles and compared with the use of those by native students' in their English essay writing. Here, the study chose 4-word bundles for data analysis. Because of the relatively small size (258 essays with 63,475 words in the learner writing corpus and 784 essays with 326,320 words in the native writing corpus) the raw frequency with more than 3 was included in the original data analysis. Both the Korean EFL learner corpus and the native reference corpus had many context-dependent bundles; those bundles were removed during the refinement procedure. The results indicated that Korean learners underused the lexical bundles which were frequently used by their native counterparts. Besides, due to the lack of Korean learners' proficiency and proper literacy, their colloquial language habits of the spoken discourse affected their use of lexical bundles in the written discourse, such as many uses of pronoun "I" and contractions in their selected lexical bundles. Finally, Korean learners used different structures and functions of lexical bundles compared with the native counterparts; they predominantly used VP-based structures and stance expressions, whereas native writers did not use the serious occupancy of specific structures and functions of bundles in their writing. Suggestions and teaching implications were drawn for the compounding and comprehensive understanding for the Korean EFL context.
[lexical bundles/corpus/learner corpus/reference corpus/multi-word-units/corpus-based study]

I. INTRODUCTION

There has been a growing interest in the study of chunks, multi-word units, or lexical bundles. The general term of the current study, 'lexical bundle' was first introduced in the *Longman Grammar of Spoken and Written English* (Biber, Johansson, Leech, Conard,

교보문고 KYOBO Book Centre

& Finegan, 1999), which considered extended collocations such as *do you want me to, I said to him, going to be a*, and *I don't know what*. These sequences of words usually co-occur in a specific register. Unlike idioms, the meanings of many lexical bundles cannot be guessed from the parts of the words. Thus, lexical bundles are more common than idioms and mainly originated from natural discourse.

Traditionally, lexical bundles have been used with many nicknames such as *clusters, phrasicon, recurrent words, combinations* (Chen & Baker, 2010), *multi-word vocabulary* (O'Keefee, McCarthy, & Carter, 2007), *lexical phrases* (Nattinger & Decarrico, 1992), *prefabricated patterns* (Hakuta, 1974), *routine formulas* (Coulmas, 1979), and *n-grams* (Greeves & Warren, 2010).

Many studies on lexical bundles have proposed that chunks seemed to be a good indicator of native-speaker identity as well as one component of the fluency (e.g., O'Keefee, et al., 2007). These multi-word-units are important because the meaning of the individual words can be identified by the words grouped with them. As Firth (1935) indicated, we can perceive a word by the company it stands together with.

Early studies with regard to lexical bundles gave background information on their fundamental understanding and classification method by analyzing those within various registers (Biber, 2006; Biber & Barbieri, 2007; Biber, Conard, & Cortes, 2004; Biber et al., 1999). The latter studies compared the uses of lexical bundles between expertise writers and novice writers (Cortes, 2004, 2006). Even, in EFL contexts, most of studies on lexical bundles were about analyzing the target bundles from the native corpus, and then comparing those target bundles with ones employed by non-native users in their writing (Bal, 2010; Chen & Baker, 2010; Hesamoddion, Behzad, & Azar, 2012; Slazar, 2011). However, few studies have been attempted to compare the learner corpora and the native reference corpora in the Korea EFL context.

Most non-native speakers have used lexical bundles from a stock of prefabricated and semi-automatic word chunks, so they overuse some specific lexical bundles, or underuse the most frequent lexical bundles employed by fluent native speakers. By comparing the non-native learners' use of lexical bundles to the use of those by native speakers, it is also possible to reduce these gaps and remedy their deficiency through teaching those target bundles to the non-native speakers.

Primarily, the present study will investigate Korean college students' use of lexical bundles in their English writing, and will compare those structurally and functionally with the ones used by native college students. The current study will also focus on Korean EFL learners' unique writing patterns regarding their use of lexical bundles, and try to reduce the gaps or differences from the use by native speakers through analyzing their individual English essay writing.

교보문고 KYOBO Book Centre

The next section will provide the background knowledge on the studies of lexical bundles. Then the list of the most frequent lexical bundles in the Korean learner corpus and the native corpus will be explained. Next, the structural and functional comparison of the use of lexical bundles in the two corpora will be described. The final section will discuss the conclusion and the implications of the current study.

II. LITERATURE REVIEWS

1. Studies of Multi-word-units

Until recently, the single word was the basic denomination and main concern of vocabulary acquisition studies. The importance of multi-word was established in the 1960s (e.g., Halliday, 1966; Sinclair, 1966). To understand the basic concepts of multi-word units, it is necessary to remember two important scholars, Sinclair and Biber, who have opposite notions on them.

Sinclair (1991, 2004, cited in Greeves & Warren, 2010) said that the lexis is in the core of making languages and creating their meaning. He proposed the two principles, the *idiom principle* and the *open-choice principle*. The idiom principle is about the grammar of ready-made chunks, which means that people can use those ready-made chunks at any moment, so it is not necessary to assemble again. In open-choice principle, the meaning is created by a co-selection of words. Therefore, in Sinclair's theory, interactions among people are very important in order to make discourse in everyday conversation, and chunks are treated as evidence of the single word choice rather than assembled ones, piece by piece (O'Keeffe et al., 2007).

On the other hand, some researchers have begun to identify the structures, functions, and registers of multi-word units (Biber 2006; Biber & Barbieri, 2007: Biber et al., 1999; Biber et al., 2004; Cortes, 2004, 2006; Hyland, 2008). They purely focused on quantitative models of bundles unlike Sinclair's approach. In Biber and his colleagues' view, chunks are treated as a mark of native speakers (Prodromou, 2005).

Most recently, researchers have an interest in the various types of fixed expressions in written production of native speakers (e.g., Granger, 1998; Salazar, 2011). Many of these studies discussed the relations between chunks and pragmatics, which showed the reasons of why many word strings are so common in pragmatics rather than in syntactic or semantics (O'Keefee et al., 2007).

교보문고 KYOBO Book Centre

2. Studies of Lexical Bundles

As discussed above, lexical bundles are referred to as recurrent sequence of words, so Biber and his associates have adopted corpus-based research for the most common multi-word-units in various spoken and written registers. Biber, Johansson, Leech, Conrad, and Finegan (1999) attempted to describe all syntactic phenomena of English based on corpus analysis with various interpretations of quantitative results using a number of spoken and written registers (e.g., transcribed conversations, fictions, news, and academic prose). Studies regarding lexical bundles have been officially investigated in one of the chapters of this study. Biber, Conard, and Cortes (2004) investigated the use of lexical bundles in two different university registers: classroom teaching and text book. They lay great emphasis on the structural and functional differences of lexical bundles between spoken and written registers. For example, classroom teaching used a wide variety of NP/PP-based referential bundles. They finally supported the idea that high frequencies of patterns of the specific bundles are not accidental, nor are they explanatory. They insisted that lexical bundles are not complete grammatical structures nor are they idiomatic, but they act as a basic unit of discourse. Biber (2006) conducted another famous study related to lexical bundles in one of the chapters in *University Language* using the same corpus (i.e., T2k-SWL) as other studies. He also used the similar functional taxonomies of lexical bundles: 1) stance bundles, 2) discourse organizer, and 3) referential bundles. They concluded that stance bundles and referential bundles were common in both classroom teaching and textbook, but discourse organizing bundles were the most common in classroom teaching. They treated lexical bundles as a basic linguistic construct with important functions in university register. Finally, Biber and Barbieri (2007) extended the various registers such as instructional registers (e.g., classroom teaching and textbook), management registers (e.g., class management talk, written course management), institutional registers (e.g., service encounters and institutional writing), and student-student academic interaction (e.g., study group). The researchers concluded that lexical bundles were more prevalent in non-academic university registers than in the core instructional registers. Furthermore, lexical bundles were very common in written course management (e.g., course syllabi), while other previous research showed more bundles in speech than in writing.

Cortes (2004) analyzed lexical bundles of two different types of corpora: published authors and students in history and biology. First, the study discussed the target bundles of professionals. Then, she analyzed the use of target bundles in students' writing. The students rarely used these target bundles in writing, so the researcher said that it would be necessary to conduct further research on instructor's perceptions of the use of lexical bundles. The influence of Cortes' (2004) study lead Cortes (2006) to examine the effects

교보문고 KYOBO Book Centre

of teaching lexical bundles of a group of university students in a writing-intensive history class. Five 20-minute micro lessons at five different times were carried out during the semester, such as filling in the blank, multiple choice, and correcting inappropriate use. There was no difference between the before and after instruction analysis on the use of lexical bundles, but there was an increase in the students' awareness of the interest in these expressions. Hyland (2008) analyzed the most frequent 3-word, 4-word, and 5 word bundles in a 3.5 million word corpus of research articles, doctoral dissertations, and Master's theses in four disciplines on their frequencies and preferred uses. The noun phrase with of-phrase fragment was the most common structure in general. He used the different functional classification of lexical bundles: research-oriented bundles, text-oriented bundles, and participant-oriented bundles. Using this classification, the researcher found that some functional categories were strongly connected to the structural patterns; 'noun phrases+ of structures' were dominant in research-oriented bundles, prepositional phrase patterns were common in text-oriented bundles, and *anticipatory it* usually happened in participation oriented bundles.

Studies in EFL contexts are consequential because these mainly compared the lexical bundles of native speakers' target bundles and non-native speakers' use of them, and they adopted more sophisticated or better methodology for data analysis (Bal, 2010; Chen & Baker, 2010; Hesamoddion, et al., 2012; Slazar, 2011). Bal (2010) investigated the use of lexical bundles in research articles written in English by Turkish scholars in six different academic disciplines. Four-word lexical bundles appearing at least twenty times in one million words were identified and analyzed structurally and functionally based on the taxonomies of Bibier and his colleagues. Overall, ninety-nine frequent lexical bundles were found in TSRAC (Turkish Scholars Research Articles Corpus). Chen and Baker (2010) collected the lexical bundles from one corpus of published academic texts and two corpora of students' academic writing. Here, three broad structural categories were identified: NP-based, PP-based, and VP-based. Then the researcher removed some context-dependent and overlapping bundles before analyzing specific lexical bundles. L2 students used the smallest range of lexical bundles, while published academic writers used the widest range of lexical bundles. L2 students underused some high frequency expressions in published expressions, while they overused certain expressions (e.g., *all over the world*) which native academic writers rarely used. Salazar's (2011) analyzed three- to six-word lexical bundles from a 1.3 million word samples from the Health Science Corpus (HSC), which is a collection of published articles in biology and biochemistry. This dissertation adopted taxonomies of Biber et al. (1999) and Hyland (2008), and confirmed how native expert writers employed these recurrent word strings coherently and convincingly according to the convention of the scientific genres, and compared these target bundles of the lexical

교보문고 KYOBO Book Centre

bundles to the lexical bundles written by Spanish non-native writers. At the end of study, extensive lists of the target bundles were provided. Finally, Hesamoddion, Behzad, and Azar (2012) interestingly analyzed four-word bundles target bundles from a corpus of 200 research article introductions written by published authors in the area of Applied Linguistics, and compared these target bundles to a similar corpus by Iranian, non-native writers. Iranian authors used more 4-word lexical bundles in their writing compared with their native speaker counterparts.

In Korea, some studies also attempted to analyze the chunks in Korean high school English textbooks (e.g., N. Kim, 2004) and in native speakers' speech texts (e.g., N. Kim, 2009). However, M. Huh's (2008) study is worth considering because it could be the starting point for investigating the learners' written data of the lexical bundles. She analyzed four-word bundles in 96 essays of 28,400 words with the same functional classification of Biber and his colleagues. The result of this study can be criticized and questioned because it only showed 15 types of lexical bundles with 27 tokens, and did not compare use of bundles with the native reference corpus.

In conclusion, some earlier studies gave a basic understanding and classification method on identifying lexical bundles, and these studies analyzed and compared the lexical bundles in various spoken and written registers (Biber, 2006; Biber & Barbieri, 2007; Biber et al., 1999; Biber et al., 2004). Latter studies differentiated the use of lexical bundles between expert writers and novice writers (e.g., Cortes, 2004, 2006). These trends continued even in EFL contexts (Bal, 2010; Chen & Baker, 2010; Hesamoddion et al., 2012; Slazar, 2011). However, little study on lexical bundles attempted to compare native reference corpus to Korean English learner corpus in Korea. Therefore, the current study will focus on comparing the use of lexical bundles employed by Korean EFL leaners to those used by native writers with more specific and concrete methodology for the data analysis.

The study made two hypotheses: high frequency patterns are not accidental, but they also are not also explanatory themselves (Bibier, 2006), and the lexical bundles used in Korean EFL college students' English essays will show different aspects and patterns compared with that of native speakers. Based on these hypotheses, the present study attempts to answer the following:

1) What are the most frequent lexical bundles in Korean college students' English essays, and what are their unique characteristics of them compared with those in native speakers' English essays?

2) What are the structural and the functional differences in the use of lexical bundles in the Korean English learner corpus and the native reference corpus?

교보문고 KYOBO Book Centre

III. METHOD

1. Corpora

The data used in this study was drawn from two corpora, the Korean English Learner Corpus (KELC) and the LOCNESS[1]. The KELC[2] is a collection of essays written by Korean college students who took a college English course as a requirement at two different universities, aged from 19 to 25: Seoul National University (100 samples, collected in 2007) and Sookmyung Women's University (158 samples, collected in 2010 and 2011). Most of the essays were argumentative with some being descriptive and narrative essays. Both timed (42 samples) and un-timed essays were included in the current study. The general language proficiencies of the participants in the KELC were intermediate (i.e., moderate mid) or upper-intermediate (i.e., moderate high) based on MATE (Multimedia Assisted Test of English). Table 1 shows the number of essays and words of the two corpora.

TABLE 1
The Two Corpora

	KELC	LONCESS
Number of Essays	258	784
Number of Words	63,475	326,320

2. Data Analysis Procedure

Usually, lexical bundles are frequent combinations of three or more words (Hyland, 2008; Slazar, 2011), but most studies on lexical bundles have analyzed 4-word because of convenience and overlapping. Here, the study also choose only 4-word bundles for data analysis. Hyland (2008) supported the merits of using 4-word bundles because "4-word bundles are more common than 5-word bundles and show clear ranges and

[1] "The LOCNESS, which were compiled as a reference corpus in the ICLE (International Corpus of Learner English) project, consists of argumentative (and some literary) essays written by British and American University students" (cited in Oh, 2009, p.57). All the participants in the LOCNESS are fully English native speakers, aged from 17 to 48. Most essays in the LOCNESS were argumentative or persuasive with various topics, such as euthanasia, yoga, nuclear power, surrogate mother hood, sex equality, teenagers, freedom of press, abortion, and etc.

[2] Topics of the KELC were announcers' appearing on the entertainment programs, smoking at the public places, having part time jobs, using vending machines, animal testing, private education, wearing school uniforms, euthanasia, abortion, and etc.

교보문고 KYOBO Book Centre

functions compared to 3-word bundles". The analysis tool of bundles was undertaken with AntConc (version: 3.21) with the n-gram analysis (see Anthony, 2004). The actual frequency cut-off used to identify lexical bundles in the current study is seemingly arbitrary. At first, because both the learner and the reference corpus are small, the raw frequency with more than 3 was included in the original data analysis. Then, to ensure comparability, only parts of each corpus were selected for investigation; the bundles with more than 6 raw frequency were selected in the LOCNESS, and the bundles with more than 3 raw frequency in the KELC were chosen for data analysis. Finally, both corpora have quite the number of context-dependent bundles, so these bundles were removed after refinement (Chen & Baker, 2010; Salazar, 2011)[3]. Table 2 shows the number of tokens and types after the data refinement.

TABLE 2

Number of Bundles Before and After the Refinement

	Before Refinement		After Refinement	
	Types	Token	Types	Token
KELC	639	3,299	137	638
LOCNESS	1,857	8,281	181	1,752

3. Classification of Lexical Bundles

Tables 3 and 4 briefly show the classification of the structures and functions of the lexical bundles of the current study. The specific definitions and examples of taxonomies will also be explained in the next section.

TABLE 3

Structural Classifications of Lexical Bundles (Biber et al., 1999)

	Types of LB	Example of LB
NP-based	NP+ of-phrase fragment	*the end of the*
	NP with other post-modifier fragment	*the way in which*
	Other NP	*the past few years*
VP-based	Passive+ PP fragment	*I have been drunken,*
	Other verbal fragment	*will be able to*
PP-based	Prepositional phrase + of	*as a result of*
	Other prepositional phrase (fragment)	*at the same time*

[3] Three types of bundles were removed during the refinement procedure: 1) topic specific bundles (e.g., a *part time job*, and *important for college students*, 2) context indicating bundles (e.g., *in the U.K.* and *in the United States)*, and 3) fragment of two bundles (e.g., *other hand some people*). Here, apostrophes used for the contractions and the possessive forms were counted as one token of frequency for the data analysis.

교보문고 KYOBO Book Centre

TABLE 4

Functional Classifications of Lexical Bundles (Biber et al., 2004)

	Types of LB	Example of LB
Stance Expressions	Epistemic S	*I think it is*
	Attitudinal modality S	*it is important to*
Discourse Organizers	Topic introduction/focus	*I would like to*
	Topic elaboration/clarification	*as well as the*
Referential Expressions	Identification	*is one of the*
	Imprecision	*or something like that*
	Specification of attributes	*a lot of the*
	Time/place/text reference	*at the same time*

IV. RESULTS AND DISCUSSION

The analyses in the following section are based on the recurrent word associates retrieved and refined (see Table 2). In this section, the list of the most frequent lexical bundles in the two corpora, individual essay samples from the learner corpus, and the structural and functional comparison of lexical bundles in the two corpora are made.

1. Lexical Bundles in the Learner Corpus (KELC) and the Native Corpus (LOCNESS)

The first research question was about analyzing the most frequent lexical bundles from the Korean college students' English essays and those from the American or British college students' English essays. Then, the first research question also had to do with discovering Korean learners' unique or ill-formed characteristics of the use of lexical bundles from the individual essay samples.

Table 5 presents the raw frequencies and the normalized frequencies (per 100,000) of the occurrence of 4-word bundles in each corpus. Three major results can be discussed from this.

TABLE 5

The Most Frequent Lexical Bundles in Two Corpora

KELC	RF	NF	LOCNESS	RF	NF
it is important for	24	37.8	the end of the	66	20.2
I think it is	19	29.9	*on the other hand*	47	14.4
I think that it	15	23.6	at the end of	42	12.9
For these reasons I	13	20.5	*is one of the*	30	9.2

교보문고 KYOBO Book Centre

is one of the	11	17.3	one of the most	30	9.2
On the other hand	11	17.3	as a result of	25	7.7
are a lot of	10	15.8	the beginning of the	23	7.0
for a long time	10	15.8	have the right to	20	6.1
So I think that	8	12.6	in the case of	20	6.1
there are a lot	8	12.6	the rest of the	19	5.8
to go to the	8	12.6	*at the same time*	*18*	*5.5*
don't have to	7	11.0	a great deal of	17	5.2
don't want to	7	11.0	due to the fact	17	5.2
have the right to	7	11.0	the only way to	17	5.2
I don't think	7	11.0	at the beginning of	16	4.9
the most important thing	7	11.0	is for the best	15	4.6
at the same time	6	9.5	not be able to	15	4.6
When I was a	6	9.5	to a certain extent	15	4.6
In my opinions I	6	9.5	the role of the	14	4.3
it is good for	6	9.5	to be able to	14	4.3

RF= Raw Frequency, NF=Normalized Frequency (per 100,000 words)

First, the most frequent lexical bundles in these two corpora did not overlap much. The Korean college students generally did not use the frequent lexical bundles often employed by native speakers; only three types of lexical bundles were found in both the KELC and the LOCNESS on the top 20 frequently used lexical bundles in the two corpora: *is one of the*, *on the other hand,* and *at the same time*. In addition, the Korean learners overused these three types of bundles compared with their native counter parts (*is one of the*: KELC-NF 17.3, LOCNESS-NF 9.3, *on the other hand*: KELC-NF 17.3, LOCNESS- NF 9.2, and *at the same time*- KELC-NF 9.5, LOCNESS-NF 5.5). Another point to consider is that although the two groups used these three lexical bundles frequently in their essay writing, their ways or rules of composition might be different in the choice and the use of them in the sentences. For instance, the two groups used *on the other hand* in the different positions of the sentences; Korean EFL writers used it 10 times at the beginning of the sentences and used it only once in the middle of the sentences, whereas native writers wrote it 20 times in the initial position and used it 17 times in the middle of the sentences.

Second, the Korean learners used the pronoun "I" excessively in their uses of lexical bundles in their essay writing. In Table 5, 7 types of lexical bundles in the KELC were related to the pronoun "I": *I think it is, I think that it, For these reasons I, So I think that, I don't think, When I was a,* and *In my opinion I*. It is possible to say that Korean English learners used the pronoun "I" a lot because they might lack proficiency, literacy, or

교보문고 KYOBO Book Centre

variations to select other lexical bundles. Therefore, the colloquial language usage habits from the spoken discourse in daily life might have an influence on their use of lexical bundles in the written discourse. The word list of the KELC and the LOCNESS can also confirm this result. In Table 6 with regard to the word list of the two corpora, the pronoun "I" in the KELC is in the third position with 1,394 raw frequency and 2,196.1 normalized frequency, while there is no position of the pronoun "I" in the LOCNESS (i.e., native corpus).

TABLE 6

Word Lists of the Two Corpora

	KELC	RF	NF	LOCNESS	RF	NF
1	The	2,030	3,198.1	The	18,856	5,778.4
2	To	1,721	2,711.3	Of	10,697	3,278.1
3	**I**	**1,394**	**2,196.1**	To	10,659	3,266.4
4	A	1,387	2,185.1	And	8,256	2,530.0
5	Of	1,372	2,161.5	A	6,547	2,006.3
6	Is	1,298	2,044.9	Is	6,272	1,922.0
7	And	1,279	2,015.0	In	5,616	1,721.0
8	In	1,029	1,621.1	That	4,879	1,495.2
9	That	842	1,326.5	Be	3,192	978.2
10	For	651	1,025.6	For	2,961	907.4

RF= Raw Frequency, NF=Normalized Frequency (per 100,000 words)

Last, the Korean English language learners used more contractions in their use of lexical bundles than their native counterparts. Two types of frequent lexical bundles in KELC were related to contractions in Table 5: *don't have to* and *don't want to* while native writers did not use any contractions. Also, from the whole lists of lexical bundles used for data analysis, 15 types of lexical bundles were using contractions whereas the native writers used contractions in three types of lexical bundles (see Appendix). Like the above, it might have been the possibility that the Korean English learners' spoken language usage habits affected this result.

Aforementioned, the Korean EFL learners' excessive use of pronoun "I" and contraction in their use of lexical bundles can be also found in their individual essay writing. These expressions are rarely used in the essay samples written by the native speakers. The following essay sample shows the use of undesirable lexical bundles, *I'm telling you*, three times.

교보문고 KYOBO Book Centre

I will describe me. By the way, because I'm normal I have head, upper body and lower body. My height is 170 cm and my weight is secret. I'm telling you about me under essay into three parts.

First, I'm telling you about my head. my head is normal, so I have hair, eyes, nose, and mouse etc. My hair is bobbed hair and my hair color is black but black isn't perfect black. And my eyes is normal and my pupil is smaller than person. So, I have to wear contact lenses. Well, my nose is normal but I heard my nose is beautiful. And my mouse is bigger and lip is thick. So I don't like my mouse.

Second, I'm telling you about my upper body. I have broad shoulder so I'm worried because I can't wear any clothes. And I have a small bust. I don't have confidence. If I lost weight, my small bust is smaller than after. And I have too much fat in my midsection. I want to lose some of my belly fat. And my hand is smaller than person who have in the same height with me.

Finally, I'm telling you about my lower body. I have thick thigh. These days skinny jeans is trend. So I don't have confidence. Skinny jeans make me come out my thigt. And I have thinner calf than person who has weight in the same weight with me. And I have thin ankles. I have bigger feet than person. I think normal women have small feet size about 235mm to 240mm. But my feet size is 250mm. So, I have big feet.

In conclusion, I'm nomal.

(Essay Sample1)

The student above wrote an essay regarding describing herself. Many inapplicable expressions in well-written essays were here: excessive use of the pronouns "I" (26 times), contractions (6 times), two spelling errors ("thigt" and "nomal"), improper position of coordinate conjunctions such as *so* and *and* at the beginning of the sentences (8 times), and capitalization. The student did not distinguish the spoken language and the written language appropriately; so many expressions frequently found in spoken discourse were in their English essay writing. It is expected that this student used different lexical bundles such as *when it comes to* (13 times in the LOCNESS) or *I would like to* (12 times in the LOCNESS), if he or she had chances to practice repeatedly to acquire these expressions through the classroom teaching.

Another unusual example of use of lexical bundles' use employed by the non-native learners is in their repeated use of ill-formed or asyntactic lexical bundles in their writing. The following essay sample shows of the use of ill-formed expression, *is need thing for*, three times.

교보문고 KYOBO Book Centre

Recently, the Seoul Metropolitian Company consider setting the place for women. But people have different opinion. Some people say that this is need thing for women but other people say that this is not fair for men.
I think that setting the place for women is need thing for women.

First, men and women have different biological body structure.
For example, women can give own milk for their baby, but men cannot this. Now when women give own milk their baby, they have to go another place.
But after the place for women setted, women don't have to go anouther place to give milk for their baby.
Second, men can also suggest their opinion about setting the place for men to the Seoul Metropolitian Company.
Then the company will consider about men's opinion.

For these reasons, I think that setting the place for women is need thing for women.
And people have to think about the diffrence about men and women.

(Essay Sample2)

The essay above is an argumentative type introducing female only spaces in the Seoul Metropolitan system. The student wrote this essay without the aid of an English dictionary within the limited allotted time (i.e., 30 minutes). Probably, *is need thing for* is the improper expression of *is necessary for~*. Some ill-formed expressions include spelling errors (e.g., *Metropolitian, diffrence*), starting a paragraph without an indentation, and using two conjunctions successively (e.g., *But after the place~*). It is suggested that the students could reduce the chances of writing these lexical bundles by learning good examples of lexical bundles from the native speakers as well as proper grammar and literacy in a classroom setting.

To sum up, the Korean EFL students did not use the frequent lexical bundles employed by the native college students. Their spoken language usage habit and lack of literacy might have affected their use of lexical bundles in their writing, such as excessive use of the pronoun "I" and contractions. These undesirable expressions were also found in their individual essay samples. It is expected that the Korean learners can correct these lexical bundles by practicing the good expressions from the lexical bundles used by the native writers during the class. The next section will describe the structural and functional difference between the Korean English learners and the native writers.

교보문고 KYOBO Book Centre

2. The Structural and Functional Comparisons of Using Lexical Bundles in Two Corpora

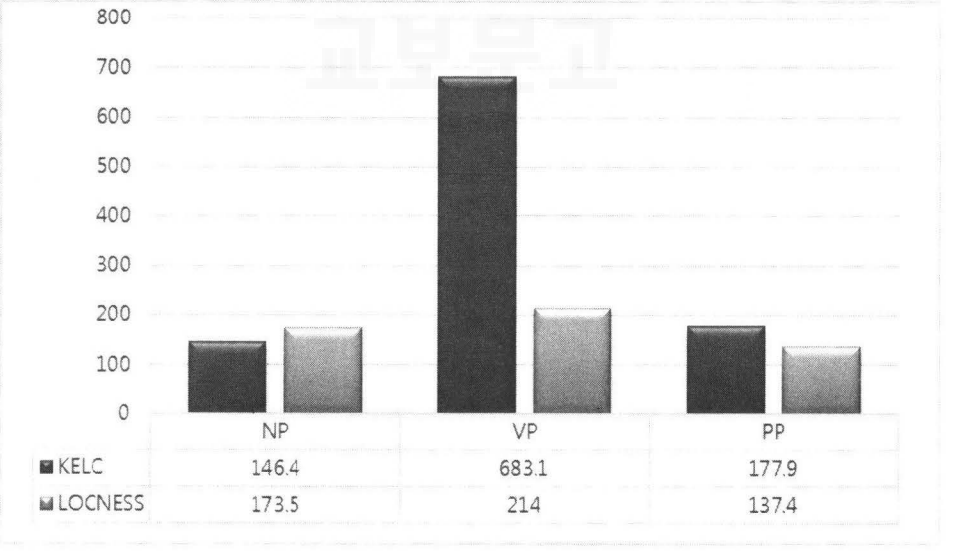
1) Structures

The structural classification of lexical bundles in the current study utilized the taxonomies by Biber et al. (1999). Three broad structural categories were distinguished here: 'NP-based', 'VP-based', and 'PP-based' (refer to Table 3). NP-based bundles include any noun phrase with post-modifier fragments, such as *the end of the,* or *the way in which.* VP-based bundles stand for any word combinations with a verb component, such as *will be able to,* or *should be allowed to*. PP-based bundles include those starting with a preposition plus a noun-phrase fragment, such as *at the same time* or *as a result of.*

As shown in Figure 1, the use of VP-based bundles differs between these two writing groups. The Korean EFL learners predominantly used VP-based bundles compared with other types of bundles (NP-based bundles: NF 146.4, VP-based bundles: NF 683.1, PP-based bundles: NF 177.9, normalized per 100,000 words). On the other hand, native writers used VP-based bundles the most, but did not show the serious occupancy of the specific types of bundles (NP-based bundles: NF 173.3, VP-based bundles: NF 214, PP-based bundles: NF 137.4, normalized per 100,000 words).

FIGURE 1

The Structures of the Lexical Bundles in Two Corpora



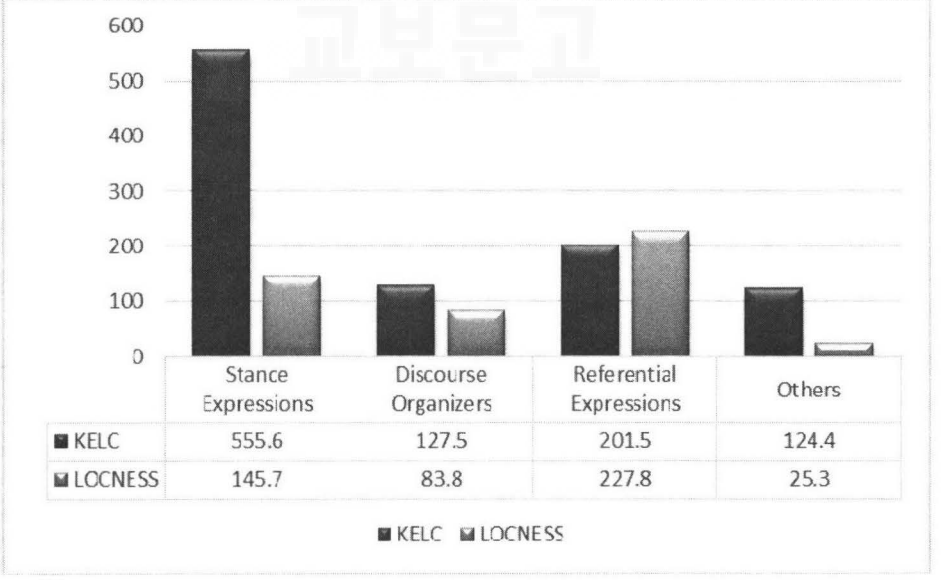
교보문고 KYOBO Book Centre

2) Functions

The functional categorization here adopted the taxonomies by Biber and his colleagues (Biber, 2006; Biber & Barbieri, 2007; Biber, et al., 2004). Three core functional categories were distinguished: stance expressions, discourse organizers, and referential expressions. Stance bundles express attitudes for assessment of certainty that frame other positions, such as epistemic stance bundles (e.g., *I think it is*) and attitudinal or modality stance bundles (e.g., *I want you to*). Discourse organizers means the reflections of the relationship between prior and the subsequent discourse like topic introduction or focus bundles (e.g., *I would like to*) and topic elaboration or clarification bundles (e.g., *on the other hand*). Referential bundles make direct reference to physical or abstract certainty, such as identification (e.g., *was one of the*), imprecision (e.g., *or something like that*), specification of attributes (e.g., *a lot of the*), time, place, or text reference (e.g., *at the end of*).

As seen in Figure 2, the Korean non-native learners used stance bundles excessively, so stance expressions were the most common lexical bundles. (KELC-stance expressions: NF 555.6, discourse organizers: NF 127.5, referential expressions: 201.5, normalized per 100,000 words). On the other hand, referential expressions were the most common bundles in the LOCNESS (stance expressions: NF 145.7, discourse organizers: NF 83.8, referential expressions: NF 227.8, normalized per 100,000 words).

FIGURE 2



교보문고 KYOBO Book Centre

In sum, the Korean English learners and the native writer used the lexical bundles differently regarding the structural and functional use of lexical bundles. Koreans predominantly used VP-based bundles and stance expressions in their use of lexical bundles whereas native writer used VP-based bundles and referential expressions the most, but did not show the serious occupancy of using specific types of lexical bundles in their essay writing.

The present study is valuable in that it provided an understanding of the unique characteristics regarding the uses of the lexical bundles employed by the Korea EFL learners such as the use of the pronoun "I" and the excessive use of VP-based bundles and stance expressions through the comparison of bundles used by the native writers. The study also found the gaps or the differences between these two groups in writing, and suggested the necessity for formal instructions of the EFL learners to include good example from the native writers.

Another value of the current study can be found in the refinement process of the data analysis. Previously mentioned in method section, many context-dependent bundles were removed during the process of refinement; those bundles which blurred the original intentions and purposes of the study were excluded. During this process, many redundant bundles such as *a part time job* (raw frequency: 133), *have a part time job* (raw frequency: 51), and *in the United States* (raw frequency: 49) were eliminated. Finally, the results of the study became clear and meaningful because lexical bundles which were commonly found in the previous studies were extracted from this refinement procedure.

VI. CONCLUSION and IMPLICATION

The current study explored the Korean EFL learners' use of lexical bundles in their essay writing by comparing these bundles with the native writers' use of lexical bundles. The study also investigated the uses of lexical bundles from individual writing samples by Korean English learners and found their use of ill-form expressions and differences compared to native writers. Korean learners did not use the frequent lexical bundles employed by the native writers much. Their lack of literacy and spoken language usage habits affected their use of lexical bundles, so they used the pronoun "I" and contractions excessively in their English essays. With regard to the structural and the functional comparisons of the use of lexical bundles, Korean EFL learners used the lexical bundles differently compared to their native counterparts; Koreans used VP-based bundles and stance expressions excessively in their use of lexical bundles.

The current study offer major points to consider in terms of the pedagogical implications. One of the findings from the frequent list of lexical bundles and the essay

교보문고 KYOBO Book Centre

samples is that exposure to the uses of lexical bundles employed by the native writers is necessary in a formal classroom. However, only exposure to those target bundles is not enough because simple exposure to the use of lexical bundles in reading texts does not directly change the students' use of lexical bundles in writing (Cortes, 2006). Therefore, the systemic and continuous exposure to the good uses of lexical bundles as well as correcting the use of ill-formed expressions and providing proper grammar and activities are essential in a class curriculum. By doing this, Koreans English learners' excessive uses of specific types of bundles in VP-based bundles and stance expressions will be ameliorated.

There are also limitations of the current study. Among the limitations, the relatively small size of corpus is the first thing to consider (63,475 words in the KELC and 326,320 words in the LOCESS); because of the corpora size, the cut off scores here were also arbitrary. This corpus size is much smaller than other studies in the uses of lexical bundles in academic settings (e.g., Biber and his colleagues' T2k-SWL corpus with 2,737,200 words, Hyland's (2008) with 3,400,400 words), although there were some studies with the smaller size of the corpus (e.g., M. Huh's (2008) with 28,400 words). However, due to the refinement procedure of the lexical bundles, aforementioned, the corpus size is not a big deal in this regard. Biber (1990) said that only 1,000 words of the data could produce reliable results. Tribble (1977) also supported the small corpus size indicating that if a register were very specialized, a smaller corpus would be adequate to provide meaningful results.

Finally, based on the implications and the limitations previously mentioned, the future suggested studies are as follows. First, researchers can investigate similar types of research with a larger size of the corpus to make the results more meaningful and great implications. Second, based on the target bundles from the native reference corpus, the future researchers can explore the effects of training on the use of lexical bundles.

The present study will contribute to the essay writing classes by making Korean EFL leaners' aware of good uses of lexical bundles by comparing the uses of those with native writers. It is hoped that this paper will contribute to our understanding of the characteristics of lexical bundles for language teaching and learning.

REFERENCES

Anthony, L. (2004). AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit. *An Interactive Workshop on Language e-Learning (IWLeL 2004)*, 7-13.

Bal, B. (2010). Analysis of Four-word Lexical Bundles in Publish Research Articles

교보문고 KYOBO Book Centre

Written by Turkish Scholars. *Alesl Theses. Paper 2*. Retrieved April 1, 2013. from: http://digitalarchive.gsu.edu/alesl_theses/2

Biber. D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing, 5*, 257-269.

Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam and Philadelphia: John Benjamins.133-175.

Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purpose, 26*, 263-286.

Biber, D., Conard, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conard, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson.

Chen, Y. H. & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397-423.

Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education, 17*, 391-406.

Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae, *Journal of Pragmatics, 3*, 239-266.

Firth, J. R. (1935). The technique of semantics. *Transactions of the Philological Society*. 36-72.

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

Greaves, C. & Warren, M. (2010). What can a corpus tell us about multi-word units? In A. O'Keefee & M. McCarthy (Eds.) *The Routledge Handbook of Corpus Linguistics*. Routledge. 212-225.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24*, 287-298.

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Baszell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.) *In Memory of J. R. Firth*. London: Longman, 148-162.

Hesamoddion, S. A, Behzad, G., & Azar, H. F. (2012). Analyzing research article introductions by Iranian and native English-speaking authors of Applied Linguistics. *International Journal of Research Studies in Language Learning, 20*, 1-11.

Huh, Myung-Hye. (2008). "*I think it is*": Lexical Bundles in EFL Writing. *Korean*

교보문고 KYOBO Book Centre

Journal of Applied Linguistics, 24(3), 129-146.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21.

Kim, Nahk Bohk. (2004). A collocational analysis of Korean high school English textbooks and suggestions for collocation instruction. *English Language & Literature Teaching, 10*(2), 41-66.

Kim, Nahk Bohk. (2009). A Corpus-based lexical analysis of the speech texts: A collocational approach. *English Language & Literature Teaching. 15*(3), 151-170.

Nattinger, J & DeCarrico, J. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Oh, Sun-Young (2009). Korean college students' use of English demonstratives in Argumentative essays. *English Teaching, 64*(1), 51-78.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press. 58-79.

Prodromou, L. (2005). *"You see, it's sort of tricky for the L2-user": The puzzle of idiomaticity in English as a lingua franca.* Unpublished Doctoral dissertation, University of Nottingham, UK.

Salazar, D. J. L. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing.* Unpublished Doctoral Dissertation, University of Barcelona. 1-304.

Sinclair, J. (1966). Beginning the Study of Lexis, In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.) *In Memory of J. R. Firth*. London: Longmans.

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

Sinclair, J. (2004). Meaning in the framework of corpus linguistics. *Lexicographica, 20,* 20-32.

Tribble, C. (1997). Improving corpora for ELT: Quick and dirty ways of developing corpora for language teaching, In B Lewsandowska-Tomaszczyk and J Melia (Eds.) *Proceedings of the First International Conference on Practical Applications in Language Corpora*. Lodz: Lodz University Press, 106-107; Retrieved March 16, 2004. from: www.cribble.co.uk

APPENDIX

Lexical Bundles in Frequency Order

KELC			LOCNESS		
	RF	NF		RF	NF
it is important for	24	37.8	the end of the	66	20.2
I think it is	19	29.9	on the other hand	47	14.4
I think that it	15	23.6	at the end of	42	12.9

교보문고 KYOBO Book Centre

For these reasons I	13	20.5	is one of the	30	9.2
is one of the	11	17.3	one of the most	30	9.2
On the other hand	11	17.3	as a result of	25	7.7
are a lot of	10	15.8	the beginning of the	23	7.0
for a long time	10	15.8	have the right to	20	6.1
So I think that	8	12.6	in the case of	20	6.1
there are a lot	8	12.6	the rest of the	19	5.8
to go to the	8	12.6	at the same time	18	5.5
don't have to	7	11.0	a great deal of	17	5.2
don't want to	7	11.0	due to the fact	17	5.2
have a right to	7	11.0	the only way to	17	5.2
I don't think	7	11.0	at the beginning of	16	4.9
the most important thing	7	11.0	is for the best	15	4.6
a lot of people	6	9.5	not be able to	15	4.6
but I think that	6	9.5	to a certain extent	15	4.6
don't need to	6	9.5	the role of the	14	4.3
In my opinion I	6	9.5	to be able to	14	4.3
it is good for	6	9.5	will be able to	14	4.3
it is hard to	6	9.5	would be able to	14	4.3
Some people say that	6	9.5	One of the main	14	4.3
these reasons I think	6	9.5	that it is a	13	4.0
think it is not	6	9.5	the majority of the	13	4.0
to get a job	6	9.5	the root of all	13	4.0
When I was a	6	9.5	the start of the	13	4.0
It is not a	6	9.5	there would be a	13	4.0
at the same time	5	7.9	when it comes to	13	4.0
be the portion of	5	7.9	would have to be	13	4.0
for the following reasons	5	7.9	a result of the	12	3.7
have a lot of	5	7.9	for the good of	12	3.7
have enough time to	5	7.9	I would like to	12	3.7
have the right to	5	7.9	in the long run	12	3.7
I agree with the	5	7.9	that it is not	12	3.7
I'm in the	5	7.9	the best of all	12	3.7
if he or she	5	7.9	all over the world	11	3.4
is the most important	5	7.9	for the best and	11	3.4
new employees to take	5	7.9	has the right to	11	3.4
of well educated female	5	7.9	in the form of	11	3.4
other people's health	5	7.9	it should not be	11	3.4
There are so many	5	7.9	should be able to	11	3.4
they are going to	5	7.9	the amount of money	11	3.4
a good idea for	4	6.3	the introduction of the	11	3.4
a strong impression on	4	6.3	the way in which	11	3.4
agree the idea that	4	6.3	a member of the	10	3.1
all over the world	4	6.3	a part of the	10	3.1
and so on I	4	6.3	both sides of the	10	3.1
As a result we	4	6.3	changed people's lives	10	3.1
At that time he	4	6.3	do not want to	10	3.1
female who wants to	4	6.3	in the field of	10	3.1
First of all I	4	6.3	is shown to be	10	3.1
for students to work	4	6.3	it is possible to	10	3.1
for the women in	4	6.3	of all possible worlds	10	3.1
have a bad effect	4	6.3	the idea of the	10	3.1
have negative effects on	4	6.3	the power of the	10	3.1

교보문고 KYOBO Book Centre

I agree the idea	4	6.3	a large number of	9	2.8
I am convinced that	4	6.3	An example of this	9	2.8
I couldn't do	4	6.3	in such a way	9	2.8
I have been drunken	4	6.3	of the most important	9	2.8
I'm telling you	4	6.3	should be allowed to	9	2.8
I was years old	4	6.3	should not be allowed	9	2.8
In conclusion I think	4	6.3	that there is no	9	2.8
In my case I	4	6.3	there has been a	9	2.8
Many people think that	4	6.3	there would be no	9	2.8
Of course there are	4	6.3	to have a child	9	2.8
people who do not	4	6.3	would not be able	9	2.8
Second reason is that	4	6.3	as well as the	8	2.5
should not be permitted	4	6.3	came to power in	8	2.5
Some people think that	4	6.3	can be seen as	8	2.5
strong impression on me	4	6.3	could be used to	8	2.5
that it is not	4	6.3	fact that it is	8	2.5
the first time I	4	6.3	I think that the	8	2.5
The second reason is	4	6.3	in today s society	8	2.5
There are many reasons	4	6.3	is not the only	8	2.5
There are three reasons	4	6.3	on the part of	8	2.5
There are two reasons	4	6.3	one of the greatest	8	2.5
There were many changes	4	6.3	that it is the	8	2.5
think that they are	4	6.3	that there is a	8	2.5
When I'm in	4	6.3	the head of the	8	2.5
a bad effect on	3	4.7	the idea of a	8	2.5
a chance to think	3	4.7	there will be no	8	2.5
a long time and	3	4.7	as a means of	7	2.1
a lot of foods	3	4.7	At the same time	7	2.1
a lot of students	3	4.7	be seen as a	7	2.1
a result of that	3	4.7	be seen to be	7	2.1
a source of my	3	4.7	become a part of	7	2.1
above reasons I think	3	4.7	can be seen to	7	2.1
agree that it is	3	4.7	can be used to	7	2.1
agree with the idea	3	4.7	did not want to	7	2.1
are already work in	3	4.7	do not have the	7	2.1
are going to school	3	4.7	don't have to	7	2.1
are not fit for	3	4.7	he or she is	7	2.1
As a result I	3	4.7	I believe that the	7	2.1
As a result of	3	4.7	I don't think	7	2.1
as the most important	3	4.7	I feel that the	7	2.1
because they have to	3	4.7	in a way that	7	2.1
both female and male	3	4.7	in order for the	7	2.1
but they don't	3	4.7	in order to make	7	2.1
can be a good	3	4.7	in order to prevent	7	2.1
chance to think about	3	4.7	In the case of	7	2.1
couldn't do anything	3	4.7	in the face of	7	2.1
didn't like to	3	4.7	in the light of	7	2.1
don't have enough	3	4.7	in the middle of	7	2.1
for the first time	3	4.7	in the number of	7	2.1
for their children's	3	4.7	is because of the	7	2.1
get money from their	3	4.7	is seen to be	7	2.1
has a lot of	3	4.7	It has also been	7	2.1
has some negative effects	3	4.7	it is clear that	7	2.1

교보문고 KYOBO Book Centre

have to appear in	3	4.7
he or she has	3	4.7
he or she wants	3	4.7
I agree that it	3	4.7
I agree that the	3	4.7
I didn't know	3	4.7
I don't have	3	4.7
I'm sure that	3	4.7
I think that the	3	4.7
I think that we	3	4.7
I want to say	3	4.7
If they don't	3	4.7
In my opinion it	3	4.7
in the center of	3	4.7
in various fields of	3	4.7
is a person who	3	4.7
is more harmful to	3	4.7
is need thing for	3	4.7
is not a good	3	4.7
is not a problem	3	4.7
is not necessary for	3	4.7
is the origin of	3	4.7
it is certain that	3	4.7
it is important to	3	4.7
it is not enough	3	4.7
it is not fair	3	4.7
it is not good	3	4.7
It is true that	3	4.7
It is important to	7	2.1
It is obvious that	7	2.1
of the opposite sex	7	2.1
rest of the world	7	2.1
that there should be	7	2.1
that there would be	7	2.1
that they do not	7	2.1
the case of the	7	2.1
the fact that it	7	2.1
the minds of the	7	2.1
the past few years	7	2.1
the question of whether	7	2.1
the rights of the	7	2.1
the size of the	7	2.1
This is a very	7	2.1
This would mean that	7	2.1
was one of the	7	2.1
will have to be	7	2.1
a lot of people	6	1.8
an integral part of	6	1.8
and in some cases	6	1.8
as a symbol of	6	1.8
as long as they	6	1.8
At the beginning of	6	1.8
By the end of	6	1.8
cars on the road	6	1.8
choose the sex of	6	1.8
do not have a	6	1.8
do not have to	6	1.8
does not mean that	6	1.8
does not seem to	6	1.8
don' t want to	6	1.8
for a long time	6	1.8
for the rest of	6	1.8
I have come to	6	1.8
in a number of	6	1.8
In many cases the	6	1.8
in the eyes of	6	1.8
in the hands of	6	1.8
in the long term	6	1.8
in the same way	6	1.8
is seen as a	6	1.8
is supposed to be	6	1.8
is the root of	6	1.8
It is difficult to	6	1.8
live life to the	6	1.8
more and more people	6	1.8
no use to the	6	1.8
not be allowed to	6	1.8
not being able to	6	1.8
not for the best	6	1.8
that it will be	6	1.8
that the only way	6	1.8

교보문고 KYOBO Book Centre

that they are not	6	1.8
the best way to	6	1.8
the case against the	6	1.8
the creation of the	6	1.8
the good of the	6	1.8
the middle of the	6	1.8
the nature of the	6	1.8
the needs of the	6	1.8
the number of people	6	1.8
the people in the	6	1.8
the rest of their	6	1.8
the use of a	6	1.8
there is no need	6	1.8
This is a good	6	1.8
This is because the	6	1.8
this is not the	6	1.8
to a loss of	6	1.8
to the conclusion that	6	1.8
value of human life	6	1.8

Examples in: English
Applicable Languages: English
Applicable Levels: College

Junghyun Kim
Seoul Theological University
52 489th St. Hoyhyun-ro Sosa-gu Bucheon-si
Gyeonggi-do, 422-742, S. Korea
Tel: 032, 340-9574
Mobile: 010-9085-9444
Email: selinainse@stu.ac.kr

Received in July, 2013
Reviewed in August, 2013
Revised version received in September, 2013

교보문고 KYOBO Book Centre