*영어어문교육, 15*권 **3**호 **2009**년 가을

A Corpus-based Lexical Analysis of the Speech Texts: A Collocational Approach*

Nahk Bohk Kim
(Korea Nazarene University)

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

Recently speech texts have been increasingly used for English education because of their various advantages as language teaching and learning materials. The purpose of this paper is to analyze speech texts in a corpus-based lexical approach, and suggest some productive methods which utilize English speaking or writing as the main resource for the course, along with introducing the actual classroom adaptations. First, this study shows that a speech corpus has some unique features such as different selections of pronouns, nouns, and lexical chunks in comparison to a general corpus. Next, from a collocational perspective, the study demonstrates that the speech corpus consists of a wide variety of collocations and lexical chunks which a number of linguists describe (Lewis, 1997; McCarthy, 1990; Willis, 1990). In other words, the speech corpus suggests that speech texts not only have considerable lexical potential that could be exploited to facilitate chunk-learning, but also that learners are not very likely to unlock this potential autonomously. Based on this result, teachers can develop a learners' corpus and use it by chunking the speech text. This new approach of adapting speech samples as important materials for college students' speaking or writing ability should be implemented as shown in samplers. Finally, to foster learner's productive skills more communicatively, a few practical suggestions are made such as chunking and windowing chunks of speech and presentation, and the pedagogical implications are discussed.

[corpus-based lexical analysis/collocation/lexical approach/chunking/windowing]

I. INTRODUCTION

In recent years, authentic materials such as dramas, newspapers, news, Youtube videos,

* This work was supported by the Research Grant of Korea Nazarene University in 2009.

UCCs (User Created Contents) and even LCCs (Learner Created Contents) have been utilized to develop English communication skills as the main source of a given course in college. In addition to materials, speech texts are widespread in the spotlight of Korean learners of English, influenced by the convincing attraction of Obama's speech in the process of the Presidential election last year. Historically, speech is a powerful means of self-expression (Nancy, 2008), a way of showing one's self to the world with a professional knowledge from a variety of concerns. Thus far, speech texts, especially the presidents' speeches from the States have also been used as a communicative tool, in many cases for reading or listening materials, to improve English proficiency. However, now there is a need to develop a new course framework to meet students' needs for a new vehicle of English learning in a productive skill-focused society.

With the emergence of the real-time global communication era, a number of students want to make convincing and exceptional impressions on their communication activities, like actors, or stars on a stage in front of people. This matches the current emphasis on speaking and writing. Actually, more emphasis is now placed on the productive skills like speaking and writing over the receptive skills like reading and listening. In light of this, many are also looking for appropriate materials to improve their productive skills. To do this, one's attention should be drawn to a lexical approach in which language consists not of traditional grammar and vocabulary but often of multi-word prefabricated chunks. Many researchers and EFL educators have pointed out that students improved their communicative skills through this lexical approach (Kang & Kim, 2008; Kim, 2003, 2005, 2007; Kim & Chun, 2008; Lewis, 1997; Y. Kim., 2005).

To do this, it is meaningful to examine written texts of different genres from a lexical perspective, focusing on a collocational approach since a lexical view provides a general strategy and procedure for effective translation of the target language. This can also meet English-learning students' requirements which include wanting to encode and decode the texts they encounter. Until now we have analyzed the texts of textbooks used in middle and high schools in order to identify the lexical characteristics in terms of collocational analysis (Kim, 2004; Kwon, 2002, 2004). Textbooks full of collocational chunks are usually set up for students in an ordered and systematic manner. Also, it is not difficult to identify useful lexical items from many popular EFL coursebook texts which deal with daily conversations. Hill (2000) stresses that one needs to compare fiction, financial reports, newspaper articles, and typical EFL texts with each other. He also concludes that well-chosen coursebook texts are full of collocational expressions. As McCarthy (1990) ascertained, languages are full of strong collocational pairs and so it is interesting to examine the speech text. Teachers should be asking students to notice and highlight useful lexical items and encouraging them to store, retrieve, and recycle them as fixed or semi-fixed expressions in speech acts. From the classroom point of view, speech texts can be

beneficial and recyclable ways for students to improve their speech and public speaking skills.

Hence, this paper aims to examine the given speeches through corpus-based lexical analysis and discuss some general strategies for making our classroom approach more lexical. There is a strong need to introduce a detailed description of how to present more collocational work to different kinds of classes. The research questions are the following:

(1) What are the high frequency words used in each speech chosen?
(2) What are the lexical characteristics of the speech texts in terms of collocational approaches?
(3) What is an effective method for utilizing authentic materials such as speech texts in speech or presentation?

II. THEORECTICAL BACKGROUND

1. Corpus Analysis and the Lexical Approach (LA)

What is a corpus, a word on the tip of many tongues? It literally means, as defined by the *Longman Dictionary of Language Teaching & Applied Linguistics(LDLTAL)* "a collection of naturally occurring samples of language which have been collected and collated for easy access by researchers and material developers who want to know how words and other linguistic items are actually used" (2002, p. 126). O'Keeffe, McCarthy and Carter (2007) defined it more as "a corpus is a collection of texts, written or spoken, which is stored on a computer" (p. 1). In a more concrete and numerical perspective, Biber, Conrad and Reppen (2000) describe the essential characteristics of corpus-based analysis in the following four domains (p. 4).

(1) It is empirical, analyzing the actual patterns of use in natural texts;
(2) it utilizes a large and principled collection of natural texts, known as a "corpus," as the basis for analysis;
(3) it makes extensive use of computers for analysis, using both automatic and interactive techniques;
(4) it depends on both quantitative and qualitative analytical techniques.

To sum this up, a corpus is a principled collection of texts available for qualitative and quantitative analysis. Unlike the generative grammar perspective, which focuses on identifying the structural units and classes of a language as a tool for linguistic analysis,

corpus analysis is moving from the field of applied linguistics into the real classroom where it can be used for practical pedagogic purposes. Essentially, it is emphasizing the actual language used in naturally occurring texts.

Thus far, from structure-based analysis, many previous intuition-driven studies of chunks mostly focused on fixed and semi-fixed expressions which are structurally well-formed and semantically idiomatic. However, thanks to the development of computer technology, corpus-driven research paradigm, a much wider range of lexical sequences displaying complex structural and functional characteristics, even discourse or conversational analysis, can be tackled.

What is the lexical approach? The lexical approach (LA) to second language teaching has received interest in recent years from a number of researchers and linguists as a viable alternative to traditional grammar-based approaches. This approach concentrates on developing learners' proficiency with lexis, or words and word combinations such as collocations, fixed or semi-fixed expressions. It is based on the idea that an important part of language acquisition is the ability to comprehend and produce lexical phrases as unanalyzed wholes or ‘chunks,’ and that these chunks become the raw data by which learners perceive patterns of language traditionally thought of as grammar (Lewis, 1993, p. 95). Instruction focuses on relatively fixed expressions that occur frequently in spoken language such as, “I'm sorry,” “I didn't mean to make you jump,” or “That will never happen to me,” rather than on originally created sentences (Lewis, 1997, p. 212). The LA makes a distinction between vocabulary—traditionally understood as a stock of individual words with fixed meanings—and lexis, which includes not only the single words but also the word combinations that are stored in one’s mental lexicon. Furthermore, the LA advocates argue that language consists of meaningful chunks that, when combined, produce continuous coherent text and only a minority of spoken sentences are entirely novel creations.

2. Chunks and Chunking

In using authentic materials for learning English, more attention needs to be given to lexical chunks which are regularly occurring strings of two or more words that seem to represent unitary meaning or function, behaving almost as one unit. Many chunks are as frequent as or more frequent than the single-word items which appear in the core vocabulary (O’Keeffe et al., 2007). As Schmitt and Carter (2004) point out, it is worth noting that “chunks may not necessarily be acquired in an ‘all-or-nothing’ manner” (p. 4). Thus, acquiring chunks takes some time and a lot of effort, just like other components of English such as grammatical structures and phonological features. Learning the chunks and the appropriate use of chunks may take place over time after a number of exposures and

language interactions. Because languages consist of prefabricated chunks of different kinds, to greater or lesser degree, the chunks, large or small, are being acquired simultaneously and collaboratively (Lewis, 2007). Ellis (2003) underlines that chunks that repeated across learning experiences are also better remembered. Therefore, one can "chunk together chunks", increasing the clusters of chunks gradually. In fact, short chunks are linked together to form longer chains of meaning.

Here, the concept of collocation, which is the central idea of the lexical chunks, is briefly handled. Collocation is also included in the term lexical chunk, but is referred to separately from time to time, so it is defined as a pair of lexical content words commonly found together. Following this definition, Table 1 shows a collocation that includes several types of word combinations such as noun+noun, adjective+noun, verb+noun, adverb+adjective, and even multiple collocations consisting of more than one collocational cluster.

In essence, collocation is a marriage contract between words, and some words are more firmly married to each other than others (McCarthy, 1990). However, a phrase like 'up until now' is not a collocation because it combines a lexical content word and a grammar function word. Identifying chunks and collocations is often a question of intuition, unless you have access to a corpus. Therefore, there are a few specific examples of words that group together: words (including polywords like 'by the way'), collocations, lexical chunks (fixed, or semi-fixed expressions) and idioms which are lexical phrases where the meaning of the whole phrase may not be comprehensible even if you know the meaning of the individual words[1]. Here are some examples:

TABLE 1

Types of the Lexical Chunks

Lexical Chunks (that are not collocations)	Lexical Chunks (that are collocations)
• up until now (fixed expression)	• withdraw an offer (V+N)
• nice to see you (semi-fixed expression)	• crushing defeat (A+N)
• by the way (polywords)	• blizzards rage (N+V)
• if I were you (sentence frame)	• a sense of humor (N+N)
• hit the nail on the head (idiom)	• deeply absorbed (Adv+A)
• my point is that…(sentence builder)	• appreciated sincerely (V+Adv)
• long time no see (institutionalized utterance; fixed expression)	• notoriously hot and humid weather conditions (Adv+A, A+N, N+N; multiple collocation)

[1] For further information, you can see the 'comparisons between collocations and idioms' from Kim's article (2008, pp. 35-36).

A chunk is a lexical concept. There are four major categories of lexical chunks: words or polywords like ‘certainly’ and ‘back and forth’, collocations or word partnerships like ‘make a mistake’ and ‘powerful engine’, fixed expressions like ‘long time no see’, and ‘face the music’, and semi-fixed expressions like ‘nice to see you’, and ‘my opinion is that…’.

Now, two concepts regarding chunks and chunking are briefly discussed. Newell (1990) defines these two terms this way “a chunk is a unit of memory organization, formed by bringing together a set of already formed elements (which, themselves, may be chunks) in memory and welding them together into a larger unit. Chunking implies the ability to build up such structures recursively, thus leading to a hierarchical organization of memory. Chunking appears to be a ubiquitous feature of human memory” (p. 7). To be more specific, “lexical chunk” is a umbrella term which includes all the other terms that combine words together. So we define a lexical chunk as any pair or group of words commonly interlocked together, or next of kin to each other.

In the same vein of chunks, chunking is the key to comprehensibility, to making yourself understood in speech, and from a language teaching point of view, successfully turning input into intake. Therefore, chunking is central to effective communication and efficient acquisition (Lewis, 2007, p. 58). In studying English, students have to grasp the meaning of the sentence or passage in a chunking manner. By processing each chunk, students can read the given sentence in an ordered and thought-or-breath group fashion. In order to substantially increase communicative competence, learners of English should acquire checking ability to chunk from a small breath unit to a larger unit that can also be produced in the same way.

III. METHOD

1. Subjects: Speech Texts

The subjects in this study are ten speech texts which the incumbent president Obama delivered up through 2009. The reason for choosing his speeches is that he became increasingly popular from a little known local Senator to a national celebrity after his 2004 Democratic National Convention Keynote Speech. Non-native speakers of English have studied the texts of the presidents’ speeches, because the speeches are usually made by professional speechwriters, and the contents as well as the sentences used include many informational messages and pedagogical usages. Also, they provide a contextual understanding of not only many different vocabularies and rhetorical expressions but also

sentence structures and other lexical characteristics within their contents.

Among the Obama speeches, some are somewhat special occasion speeches, while others are informative or persuasive speeches. He used either logos, ethos, pathos, poetic or the narrative approach when delivering to people from all walks of life, races, and religions. Many parts of the speeches reflect his life, philosophy and vision for the future as a world leading politician. He is a good speaker and gifted orator. These are sufficient reasons to study the lexical chunks he used in his speeches. He chose easy but laboriously-selected words. As the intention can be inferred from the key lexical chunks, a wide variety of lexical chunks can be recycled, as well as a number of basic sentence structures from each speech. Here are the speeches chosen in a chronological order for this study.

TABLE 2

Texts of Obama Speeches

No.	Speech Titles	Dates and Other Information
1	The Audacity of Hope	July 27, 2004: Democratic National Convention Keynote Speech
2	Politics of Change	Feb. 10, 2007: Campaign Speech
3	A More Perfect Union	Mar.18, 2008: Speech in Constitution Center Philadelphia
4	A World that Stands as One	July 24, 2008: Speech to the Citizens of Berlin
5	The American Promise	Aug. 28, 2008: Democratic Convention Speech
6	A 21st Century Education	Sep. 09, 2008: Education Speech
7	Tonight is Your Answer	Nov. 04, 2008: Election Night Victory Speech
8	We Seek a New Way Forward.	Jan. 20, 2009: Inaugural Address
9	Speech to Congress	Feb. 24, 2009: Address to a Joint Session of Congress
10	A New Beginning	June 04, 2009: Speech in Cairo, Egypt

2. Data Collection Procedures

As mentioned earlier, this research is to analyze speech texts and find pedagogical implications after identifying the characteristics of lexical chunks, and focusing on collocational analysis. In order to get core data collections, first, a number of internet websites, chiefly *google.com* and *BarackObama.com*, were searched for this study. Some websites included only manuscripts; some included manuscripts and videos. Among his speeches, there have been some topics chosen including education. He delivered a wide range of speeches, based on his numerous life experiences, before and after becoming president of the USA. His speeches cover a variety of topics and concerns that are relevant

to current society. After initially choosing thirteen main speeches to be used in this study, three of them were dropped because of their length and lack of balance in genres. The final ten speeches are sorted in a text file for the analysis of this study.

The next necessary step was to create a corpora (a plural form of a corpus) for analyzing and extracting some necessary items from these texts. In this vein, building a pedagogic corpora that students can use for a communicative purpose should be developed in the process of data collection. Therefore, this corpus was built according to each speech, in total ten of each corpus. Later arranged in order by the whole characteristics of the lexical chunks of Obama speeches, the total corpus was built after putting ten of each corpus into one large corpus.

3. Data Analysis

To verify whether speech texts include pedagogic materials and implications for teaching when they are used in class, this research uses the corpus analysis program NLPtools (NLP: Natural Language Processing), which includes various functions such as frequency count, collocation/KWIC(Key Word in Context), English tagger, grammar usage, and so on. First of all, by using the function of frequency count, high frequency words are listed from each corpora, and then from the total corpus. Also to identify the overall feature of the words used in speech, one must check the Type, Token, then TTR Ratio (Type/Token Ratio) which represent the basic characteristics of the number of running words in each speech. Next, so as to see what type of lexical collocation each speech corpus has, each corpus is analyzed from the collocational point of view. Finally, each speech corpus is analyzed to identify the collocational usages and perspectives of the most frequent words used in speech texts: *have* and *do* in verbs; *new* in an adjective, and *world* in a noun.

IV. RESULT AND DISCUSSION

1. Results

1) Frequency Analysis

The basic analysis of the corpora shows what kind of high frequency word each corpus has. Table 3 shows the most frequent items, a mixed list in a ten-million-word corpus made up of the five-million-word CANCODE spoken corpus and a five-million-word general written corpus sample from the Cambridge International Corpus (CIC) (O'Keeffe et al., 2007). There are a few differences between spoken and written, reflected in the high rank

of *I* and *you* in the spoken data, along with *yeah, er*, and *oh* representing items of high frequency in conversational speech. However, the written list shows a greater prevalence of third-person references, prepositions and conjunctions largely representing 'the world out there' (O'Keeffe et al., 2007, p. 33).

TABLE 3

Most Frequency Words: Written+Spoken Corpus(CIC), Written Corpus, and Speech Corpus

frq	CIC	Written Corpus	*Speech corpus*	frq	CIC	Written Corpus	*Speech corpus*
1	the	the	the	26	as	from	with
2	and	to	and	27	at	not	they
3	to	and	to	28	we	they	all
4	a	of	of	29	her	by	us
5	of	a	that	30	had	this	or
6	I	in	a	31	not	are	America
7	you	was	we	32	no	were	their
8	it	it	in	33	what	all	people
9	in	I	our	34	this	him	more
10	that	he	is	35	like	up	from
11	was	that	I	36	all	an	by
12	yeah	she	for	37	mm	said	my
13	he	for	this	38	er	there	has
14	is	on	it	39	there	one	so
15	on	her	will	40	do	been	what
16	for	you	have	41	his	would	one
17	but	is	you	42	well	out	do
18	she	with	are	43	one	so	there
19	they	his	but	44	just	their	new
20	have	had	not	45	if	what	world
21	with	as	can	46	are	when	when
22	be	at	on	47	oh	we	was
23	It's	but	be	48	right	if	must
24	so	be	as	49	or	me	know
25	know	have	who	50	from	my	at

Meanwhile, another feature from Table 3 can be seen. The speech corpus uses different words: *we, us*, and *our* instead of *I* and *my*; auxiliary *will* and *can* that did not appear in

other corpora. This is attributable to the concept that speech has a unique characteristic in its convincing popularity and future-oriented presentation to attract people's attention.

In addition to frequency analysis, Type/Token Ratio (TTR) was analyzed to identify how diverse words are used. Type means the number of different word forms, and Token means the number of running words. Table 4 shows the Type, Token, and Type/Token Ratio of the words that appeared in these speech texts. The total number of words used in a given speech is more or less 1,000 words in word tokens. It will vary according to the length of the speech. In fact, the total number of words is not as many as expected. In other words, it is not difficult for students to speak and write English by using the famous speech texts. Regarding the TTR, the written texts for this study are generally higher than spoken words. TTR here is very high compared to the BNC (British National Corpus) sampler of spoken words (TTR: 2.69). However, Table 4 reveals that some speech texts (No. 1, 2, 8 and others) do not have big gaps compared to those of the BNC. That means that a characteristic of Obama speeches is using repetitive and common words that are familiar and popular so that everybody can understand them well. Also he repeats what he says to ensure much common ground has been properly transmitted to the audience in order to communicate sympathetically and empathically.

TABLE 4

Types, Tokens, and Type/Token Ratio of the Words in 10 Speech Texts

	1	2	3	4	5	6	7	8	9	10	Total
Word Types	2336	2824	5015	3029	4681	5197	2019	2423	6311	5947	38926
Word Tokens	793	874	1377	876	1235	1096	679	893	1509	1455	4724
Type/ TokenRatio	2.95	3.23	3.64	3.46	3.79	4.74	3.11	2.71	4.18	4.09	8.24

* TTR formula: (Types+Tokens) divided by100

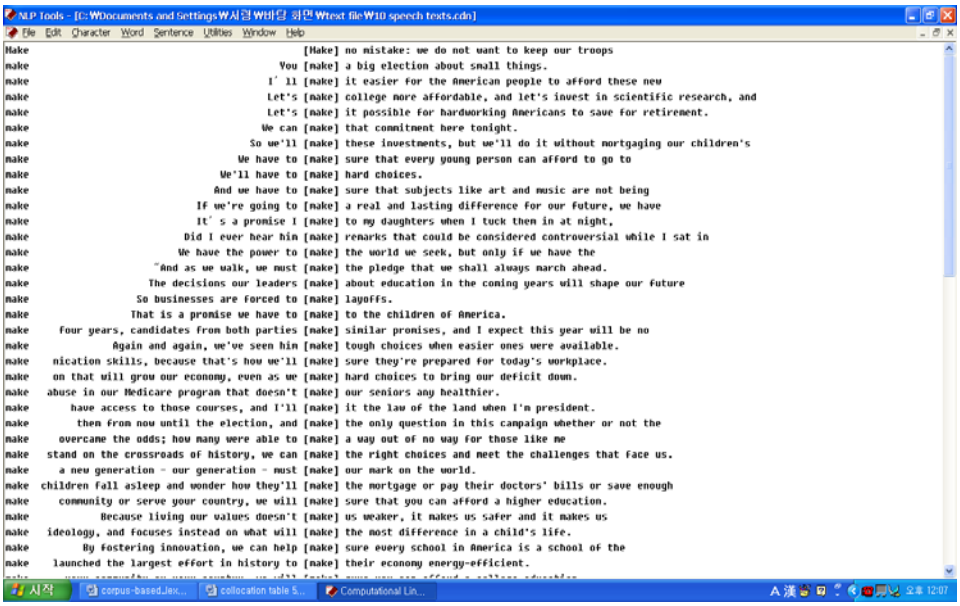
2) Collocational Analysis

As a major referential framework sets out to describe and discuss the structures and functions of lexical chunks, the lexicalized collocation shows the characteristics of the vocabulary. Kim (2008) points out that the lexical chunks such as collocations have emerged as an important category of lexical patterning, but they have not yet become an established unit of description in language teaching courses and materials. However, through this analysis, one can find out the speaker's preferred words and word partnerships.

Table 5 shows the collocates with the collocator ‘make’ such as ‘make mistakes’, ‘make a big election’, ‘make investments’, ‘make a difference’, ‘make remarks, ‘make pledges’, ‘make hard/tough choices’, ‘make promises’, and so on. The verb ‘make’, one of the delexical verbs[2], has a large number of collocates that can be directly used in the classroom.

TABLE 5

Collocation Pairings with the Verb ‘Make’



One can see the collocates with the noun ‘world’ and the adjective ‘new’ from Table 6 and 7. In terms of noun-collocates, Table 6 shows that the noun ‘world’ is generally combined with propositional phrases such as ‘of the world’, ‘around the word’, ‘in the/our world’, ‘for the world’, ‘on the world’, and ‘to the world’. Considering speech texts, one can also see the verb-collocates such as ‘light the world’, ‘share the world’, ‘change the world’, ‘threaten the world’, and so on. When introducing the noun ‘world’, one should teach a verb with a preposition and a verb with a noun.

From Table 7, one can see adjective-collocates such as ‘new beginning’, ‘new century’, ‘new energy’, ‘new vision’, ‘new age’, ‘new partnerships’, ‘new effort’, ‘new

[2] O’Keeffe et al. (2007) describe this category as including extremely high-frequency verbs such as *do, make, take and get* in their collocations with nouns, propositional phrases and particles. They are termed ‘delexical’ because of their low lexical content and the fact that their meanings in context are conditioned by the word they co-occur with (e.g. compare *to make a mistake* with *to make progress* or *to make it* [*to a place*] (pp. 37-38).

TABLE 6

Collocation Pairings with the Noun 'World'

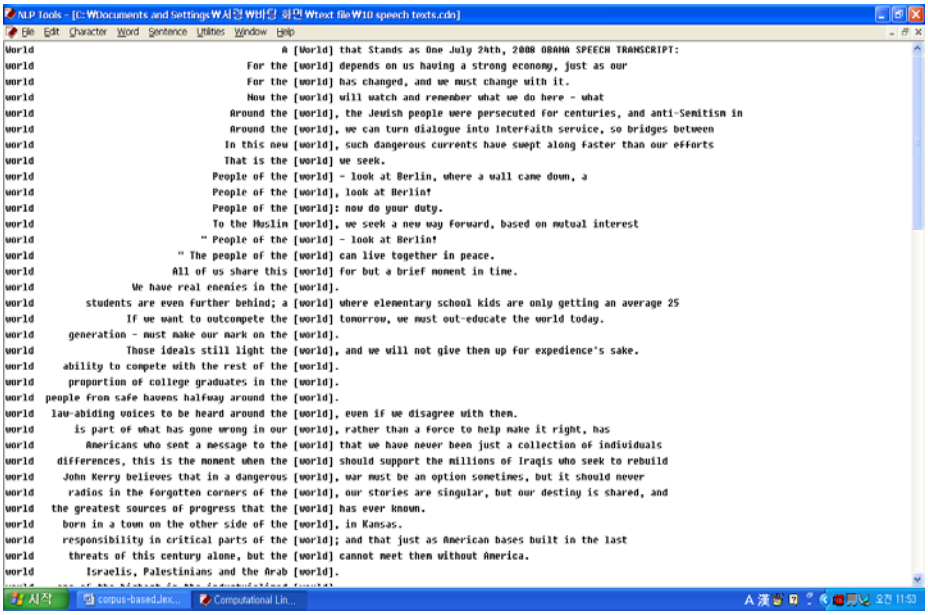
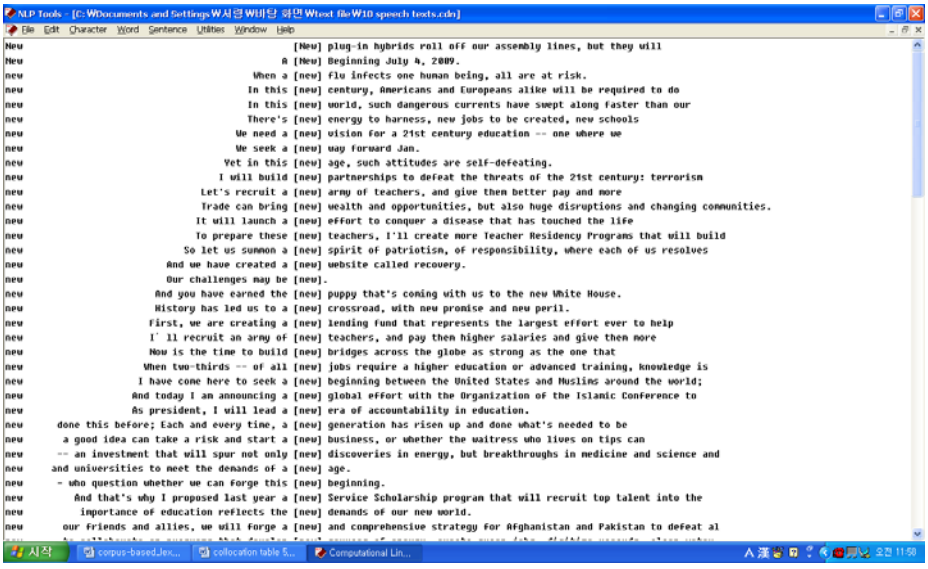


TABLE 7

Collocation Pairings with the Adjective 'New'



spirit', 'new world', 'new website', 'new leading', 'new business', and so on. Here

another important thing is that only the adjective ‘new’ seems adjective+noun collocations by combining with a noun; they are seen only as free combinations. However, in a real classroom, students can not create as many collocator ‘nouns’ as this. In other words, even free combinations are meaningful in a collocational approach. Therefore, an expanded perspective can occur in the verb+noun collocations (strictly speaking, verb+adjective +noun collocations) such as ‘need a new vision’ ‘seek a new way’, ‘build new partnerships’, ‘launch a new effort’, ‘summon a new spirit of patriotism’, ‘create a new website’, ‘lead a new era’, and so on.

To discuss more, in regards to lexical chunks, according to O’Keeffe et al. (2007), two-word and three-word chunks comprise more than 90 percent of all 2-wd to 6-wd chunks. Among them, 2-wd chunks account for more than 60 percent. Table 8 shows the top 20 3-wd chunks from five million words of mixed written CIC data (O’Keeffe et al., 2007, p. 68) and the equivalent number of 3-wd chunks extracting from speech corpus for this study.

TABLE 8

Top 20 Three-word Chunks (Written)

frq	CIC	*Speech Corpus*	frq	CIC	*Speech Corpus*
1	one of the	one of the	11	it would be	many of the(these)
2	out of the	part of the	12	in front of	a(the) number of
3	it was a	out of the	13	it was the	most of us
4	there was a	be able to	14	some of the	some of the
5	the end of	around the world	15	I don’t know	a set(couple) of
6	a lot of	it is that	16	on to the	that is why(the)
7	there was no	people of the	17	part of the	now is the
8	as well as	there is a	18	be able to	be here to
9	end of the	all of the	19	the rest of	the end of
10	to be a	be willing to	20	the first time	in the face

This study sets out to describe and discuss the structures and functions of lexical chunks in the Speech Corpus, with a view to characterize the collocational features of sophisticated speech texts. It has been found that speech texts frequently use many recurrent sequences largely associated with the making of propositions, in particular ‘of’, while using far fewer chunks (and using them much less frequently), which are basically associated with pragmatic functions. As seen in Table 8, more than half of the top 20 three-word chunks is about noun+of+noun collocations. Additionally, there are a significant number of chunks found in speech contexts different from those in CIC chunks.

2. Discussion

Essentially, as seen from my experience, Korean learners of English prefer using memory-based language to rule-based language. Pawley and Syder (1983) and Nattinger and DeCarrico (1992) claimed language users carry in their memories a vast store of pre-fabricated language consisting of multi-word lexical items or chunks. In reality, much of the language encountered in daily life may derive directly from memory and not be created from an analytic mechanism. The learning and use of this prefabricated language can be conducive to a useful and attractive communication strategy to exploit, but when the pressure of real-time communication is a factor, a lexical mode of communication will predominate (Skehan, 1998). In a corpus linguistic focus on the raw data, it is any collection of texts that people can actually produce in forms of not only prose, newspaper, and so on, but also poetry and drama, word lists, dictionaries, and so on.

FIGURE 1

Collocation Map for *Have* (Thornbury, 2002. p. 120)

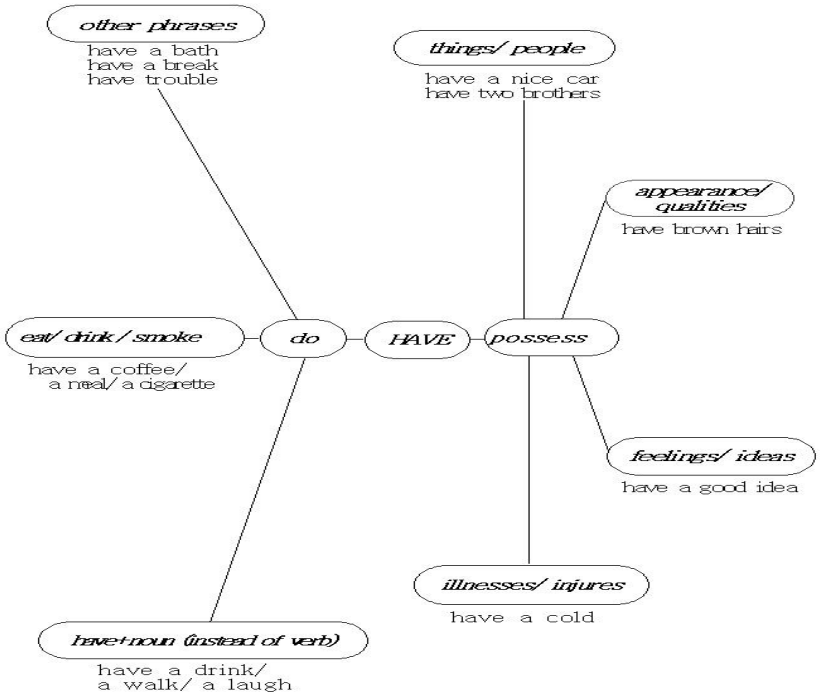


Figure 1 exemplifies a collocation map for *have*, which shows its range of collocations organized into meaning categories (Thornbury, 2002. p. 120). Through this mapping activity, students who are taught and noticed lexical chunks can either create their own maps or add to an existing map, using collocation dictionaries and collocation notebooks.

Next, as a main teaching framework, a corpus syllabus is very useful. Willis (1990) and Lewis (1997) suggested a lexically-oriented syllabus, which focuses on the lexical entry in order to make learners aware of the natural language through their conscious-raising of the chunks. Regarding the lexical syllabus, Kim (2008) suggested that a well-organized notebook is essential for all students in order to encourage them to be aware of the principle of collocations. Through using the notebook, lexical chunks need to be recycled to be learned through a variety of exposures.

Another method for using the corpora is chunking and windowing. As aforementioned, chunking is interlocking each small chunk into an extended, larger, meaningful unit to deliver an idea or opinion on the specific subject or topic. Chunking, eventually consisting of ranges from simple individual words to a variety of complicated phrases or items, recursively builds up sophisticated systematic or syntagmatic organization in the working memory system. Take a look at the following section of *speech corpus*:

> "That's why we'll have to **set priorities**. We'll have to **make hard choices**. And although government will **play a crucial role** in **bringing about the changes** we need, **more money and programs** alone will not get us where we need to go. Each of us, **in our own lives**, will have to **accept responsibility** - for **instilling an ethic of achievement** in our children, for **adapting to a more competitive economy**, for **strengthening our communities**, and **sharing some measure of sacrifice**." (excerpt from a text of Barack Obama's announcement for President; February 10, 2007)

From the above text, we can find a wide range of lexical chunks, including V+N collocations, A+N collocations, and Prep.+N collocations such as 'set priority', 'make choices', 'play a crucial role in –ing', 'bring about the changes', 'in our own lives', 'accept responsibility', 'instill an ethnic of achievement', 'adapt to a more competitive economy', 'strengthen out communities', and 'share some measure of sacrifice'. As part of making our classroom activity more lexical, we should introduce the concept of collocations and the extended lexical chunks. That is why we pay close attention to the speech texts that consist of very polished and sophisticated sentences and messages.

Last but not least, to improve productive skills such as speech or presentation, more attention should be placed on the concept 'windowing of attention'. It is like opening windows, one by one, to ventilate a place with the air. Communication is basically like a

ping-pong game, in giving and taking an object interactively. In order to deliver meanings more communicatively by using words, things can be described using chunks. Therefore, a chunk is equal to a window. Look at the following two sentences:

1) Some middle school students/ spend endless hours/ shouldering a huge academic burden/ at various cram schools/ after their regular school classes/ to prepare to get into specialized high schools/ such as foreign language schools,/ science schools,/ and independent private high schools.//

(2) We are now studying an English Speaking Course/ on the fifth floor of the Goryo building/ at Hangook University/ at the front of the campus/ located in Cheonan/ in Chungnam Province/ in Korea.//

Of course, there are extreme examples of 7-9 chunk sentences. Most English sentences may be 4 or 5 chunks. Anyway, as seen above, the two sentences are comprised of multiple collocations and lexical chunks, dividing by just using slashing. When doing a speech or presentation, one opens their attention 'windows' one by one like opening each window of the classroom from left to right or vice versa in an ordered and hierarchical fashion. While speaking, the learners of English should open their attention like opening each window one by one through the process of windowing. This approach like building chunk blocks can be helpful to enhance learners' speaking ability as well as writing ability. In other words, the chunking and windowing approach plays a vital part in improving productive skills through *speech corpus* as seen in the above samplers. In addition to productive purposes, speech texts can be made into a corpora appropriate for use with lower-level learners' reading and listening. That is called Data-Driven Learning (DDL). Here a speech corpus for DDL can be used to make this approach more accessible to multi-level classroom students. Generally, the simplified nature of the corpora may limit the learner's exposure to lexical chunks, which are fundamental to the acquisition of natural and fluent language. More suitable supplementary materials can be added by using collocation dictionaries and lexical notebooks. Despite some lack of variety, it is argued that the scale and type of lexical chunks are sufficient to provide input that reflects authentic language, suggesting that speech corpora may offer an acceptable balance of accessibility and authenticity.

V. CONCLUSION

Based on the results, this study concludes with the proposition that as the core component, collocations or word combinations must be integrated and taught in order to

improve the productive proficiency of Korean learners of English. Knowledge about collocations is of great importance to language and production (Kim, 2003). Chunk-based English learning plays an important role in many English language processing tasks such as information retrieval and text mining. Therefore, there is also a need to explore how the word collocates with or is used in conjunction with other words. The present study identifies how lexical chunks are produced in written contexts. Furthermore, it suggests using the speeches of native speakers to more effectively and productively use collocation, word groups, or chunks in the EFL classroom, which could significantly improve comprehension, acquisition, and production.

The speaker's basic argument is that being able to put together strings of language such as idiomatic expressions and common collocations is more important than grammar for the development of language fluency. If good proficiency in L2 entails the acquisition not only of many single words but of many lexical chunks as well, it must then be asked how all this additional lexis is to be committed to long-term memory in the limited time available in non-extensive classroom-based language courses. A significant fraction of conventionalized lexical chunks occupies much of English (Sinclair, Jones, & Daley, 2004). Evidence was noted that speech corpus is relatively common in lexical chunks in English. Because of this evidence, some of the lexical chunks will be autonomously noticed after being used quite often, like 'make a mistake'. This study identifies the characteristics and actualization of the collocations in speech corpus. In conclusion, the speech corpus suggests that speech texts have not only considerable lexical potential that could be exploited to foster chunks-based learning, but also that learners are not very likely to unravel this interlocked potential autonomously.

In the age of globalization, commanding good communication skills in English as an international language is a fundamental quality of all learners of English. These days, nurturing world leaders with communicative skills, particularly productive skills, is an unavoidable mission that universities have continuously pursued since these skills most certainly determine the global competitiveness of any region or country. Consequently, in realization of its vision of producing global leaders, this new approach does not happen in a moment, but reflects on-going high-quality speech or presentation achievements which attempt to inform English education towards language comprehension and its production. Chunk-focused approaches appear in two dimensions which put existing practice categories in the realm of education and newly added context categories for activating productive skills. These approaches are also shown in the academic attempts that suggest the level and scope of speech text in accordance with both classroom activity and the real world of language use.

REFERENCES

Biber, D., Conrad. S., & Reppen. R. (2000). *Corpus linguistics*. Cambridge: Cambridge University press.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Hong (Eds.), *The handbook of second language acquisition* (pp. 63-103), Oxford: Blackwell.

Harmer, J. (2007). *How to teach English* (new. Ed.). Harlow, England: Longman.

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47-69). London: Language Teaching Publications.

Kang, J., & Kim, H. (2008). An analysis on the practical usage of the word *look*. *English Language & Literature Teaching, 14*(4), 25-49.

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

Kim, N. (2003). An investigation into the collocational competence of Korean high school EFL learners. *English Teaching, 58*(4), 225-248.

Kim, N. (2004). An collocational analysis of Korean high school English textbooks and suggestions for collocation instruction. *English Language & Literature Teaching, 10*(3), 41-66.

Kim, N. (2005). A general survey of English collocations and research on the collocation teaching methods. *Foreign Languages Education, 12*(2), 141-165.

Kim, N. (2007). Effects of collocation-based vocabulary instruction on improving English reading ability for high school learners. *English Language & Literature Teaching, 13*(3), 157-176.

Kim, N. (2008). Teaching in chunks: Facilitating English proficiency. *Modern English Education, 9*(1), 30-51.

Kim, Y. (2005). The effects of collocational competence on college students' English proficiency and writing abilities. *English Language & Literature Teaching, 11*(4), 189-208.

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

Kwon, I. (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.

Lewis, M. (1993). The lexical approach: The state of ELT and a way forward. Hove, England: Language Teaching Publications.

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

McCarthy, M. (1990). *Vocabulary*. Oxford: Oxford University Press.

Nancy, G. H. (2008). *Public speaking in American English*. Boston, MA: Pearson Education, Inc.

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

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harard University Press.

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

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

Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching & applied linguistics*. Harlow, London: Person Longman.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. In N. Schmitt. (Ed.) *Formulaic sequences* (pp. 55-71). Amsterdam: John Benjamins.

Sinclair, J. M., Jones, S., & Daley, R. (2004). *English collocational studies: The OSTI report*. London: Continuum.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Thornbury, S. (2002). *How to teach vocabulary*. Harlow, London: Pearson Longman.

Willis, D. (1990). *The lexical syllabus: A new approach to language learning*. London: Collins ELT.

Websites:

www.americanrhetoric.com/speeches/convention2004/barackobama2004dnc.htm

www.barackobama.com/2007/02/10/remarks_of_senator_barack_obam_11.php

www.huffingtonpost.com/2008/03/18/obama-race-speech-read-h_n_92077.html

http://my.barackobama.com/page/community/post/obamaroadblog/gGxyd4

www.barackobama.com/2008/08/28/remarks_of_senator_barack_obam_108.php

www.edwize.org/obamas-education-speech

www.barackobama.com/2008/11/04/remarks_of_presidentelect_bara.php

www.whitehouse.gov/blog/inaugural-address/

www.cbsnews.com/stories/2009/02/24/politics/main4826494.shtml

www.whitehouse.gov/the_press_office/Remarks-by-the-President-at-Cairo-University-6-04-09/

Examples in: English
Applicable Languages: English
Applicable Levels: Secondary and Tertiary

Nahk-Bohk Kim
Korea Nazarene University, English Department
456, Ssangyoung-dong, Seobuk-gu, Cheonan city,
Chungnam, 331-946, Korea
Tel: 041-570-1514
Email: knb2030@kornu.ac.kr

Received in July, 2009
Reviewed in August, 2009
Revised version received in September, 2009