

## **Phraseological patterns of English adjectives and nouns: with reference to the noun collocates of *new*, *good*, *old* and *high* in American English\***

**Hye-Kyung Lee**  
(Ajou University)

Lee, Hye-Kyung. 2014. **Phraseological patterns of English adjectives and nouns: with reference to the noun collocates of *new*, *good*, *old* and *high* in American English.** *Linguistic Research* 31(3), 541-567. This paper explores a corpus-driven approach to studying the phraseological patterns of adjectives and nouns using data and methods from corpus linguistics. The rationale behind this approach is the strong resonance between how the meanings of adjectives have been defined and a major claim proposed in corpus linguistics; the meaning of a lexical item is typically defined both by its inherent lexical content and its relation to any accompanying words. For the analysis, the four most frequent English adjectives and their most recurrent noun collocates were chosen from the Corpus of Contemporary American English (coca.byu.edu/coca). The data are *new/book*, *good/news*, *old/friend* and *high/levels*. The occurrences of the six different phraseological patterns of each pair were counted and the internal structures of each pattern were examined. Then, the concordance lines containing the two words were analyzed to determine how the words are related semantically and structurally. The analysis reveals that the noun phrase consisting of an adjective and its noun collocate is the canonical form throughout the four collocations. The analysis also finds that the four adjectives differ in terms of their phraseological patterning with the noun collocates. The adjectives *new* and *old* are rarely used as predicates when they are selected after *friend* and *book*, respectively. On the other hand, *good* and *high* are canonically employed to serve the predicative function when they follow *news* and *levels*, respectively. The two groups also differ from each other in the way they contrast with their antonyms. The adjectives *high* and *good* are frequently employed in a contrastive context with their antonyms *low* and *bad*, whereas the adjectives *new* and *old* are seldom employed in that way. (Ajou University)

**Keywords** Adjectives, COCA, Collocation, Phraseology

---

\* I would like to thank the anonymous reviewers of this journal for their constructive suggestions and corrections. However, all remaining errors are mine.

## 1. Introduction

The meanings of most adjectives have been defined with reference to their co-occurring words, especially nouns. At the same time, previous research has also examined adjectives in terms of the semantic properties of gradability and antonymy (e.g., Croft and Cruse, 2004; Cruse, 1986; Lyons, 1977; Kennedy and McNally, 2005; Lyons, 1977; Paradis and Willners, 2006). However, not much research has been conducted on how common adjectives actually interact with their closely related words. This paper investigates the interaction of adjectives with nouns using the data and methods available in corpus linguistics. Corpus linguistics is useful for analyzing the interaction because there is a strong resonance between how the meanings of adjectives have been defined and the claim in corpus linguistics that the definition of a lexical item derives from both its inherent lexical content and its accompanying words.

For the analysis, four of the most common adjectives and their most recurrent noun collocates were chosen from COCA ([coca.byu.edu/coca](http://corpus.byu.edu/coca)).<sup>1</sup> The chosen pairs are *new/book*, *good/news*, *old/friend* and *high/levels*. The occurrences of six different phraseological patterns of each pair (i.e., Adjective Noun, Adjective \* Noun, Adjective \* \* Noun, Noun Adjective, Noun \* Adjective, and Noun \* \* Adjective) were counted. Next, the internal structures containing the two words were analyzed by examining the relevant concordance lines. The results of the analysis identify the canonical form(s) of each pair and show that the four adjectives differ with respect to their phraseological patterning with noun collocates. The chosen adjectives also differ in the way they contrast with their so-called antonyms.

## 2. Previous studies on adjectives

Adjectives are typically studied with respect to their relationship with other parts of speech such as nouns or pronouns. This is largely because the meanings of most adjectives are rather elusive to define. For example, the adjective *big* is interpreted differently depending on the noun it modifies. In the phrases *big elephant* and *big*

---

<sup>1</sup> COCA stands for the Corpus of Contemporary American English, [corpus.byu.edu/coca](http://corpus.byu.edu/coca), which was developed by Marks Davies.

*ant*, however big an ant is, it cannot be bigger than the smallest elephant in the world.

In traditional grammars such as Quirk et al. (1985), adjectives have been classified in several categories according to their syntactic functions, the major two of which are attributive and predicative functions. An attributive adjective pre-modifies the head of a noun phrase, whereas a predicative adjective appears after the predicate of a sentence as a subject/object complement. The two functions are exemplified in (1) below.

- (1) Attributive: *new* information  
 Predicative: This information is *new*.

Adjectives have also been explored in terms of their semantic properties, such as their gradability. A vast majority of adjectives belong to the so-called gradable adjectives, which “map their arguments onto an open scale lacking end points” (Kamoen et al., 2011: 3140). Gradable adjectives have been investigated with respect to their antonyms (e.g., Croft and Cruse, 2004; Cruse, 1986; Kennedy and McNally, 2005; Lee, 2013; Lyons, 1977; Paradis and Willners, 2006). Some of the representative examples belonging to this category are *tall*, *short*, *high*, and *low*. The crucial diagnostic for gradable antonyms is that the negation of one gradable adjective does not equal its antonym. For example, being not tall does not necessarily mean being short. On the other hand, another group of adjectives are classified as so-called nongradable/bounded adjectives. These adjectives map their arguments onto (partially) bounded scales with endpoints. Representative examples include *full*, *empty*, *alive*, and *dead*. In traditional semantics literature, which focuses on the logical properties of lexical items, the negation of one nongradable adjective equals the affirmation of its counterpart. For example, being not alive is logically the same as being dead.

This logic-oriented classification has been challenged by recent studies on the grounds that it mainly draws on the logical properties of the item and the researchers’ introspection (e.g., Jones 2002; Lee, 2013; Paradis and Willners, 2006). More recent studies instead rely on data produced by or elicited from ordinary language users. Paradis and Willners (2006) revealed in their experimental study that speakers tend to employ some of the bounded adjectives as unbounded/gradable. In

their experiments with Swedish speakers, they found that representative bounded adjectives differ from each other in the degree to which they are perceived to be bounded by ordinary speakers. Working with corpus data, Lee (2013) argued that some of the much-discussed gradable antonyms (e.g. *big/small* and *tall/short*) hardly qualify as antonyms as far as common collocates are concerned because these two pairs do not share collocates. These recent studies reveal a gap between theory-oriented research and data-based empirical research. They also back up the claim that the meanings of lexical items not only come from their inherent lexical contents but also from their co-occurrences with neighboring lexical items.

The current study investigates the interaction between common English adjectives and their frequent noun collocates. The motivation for this investigation is that the meanings of most adjectives are highly sensitive to the accompanying nouns.

### 3. Previous studies on collocations

One of the main issues in corpus linguistics involves recurrent word combinations, which have been described in many different ways: collocations (e.g., Sinclair, 1987), lexical bundles (e.g., Ädel and Erman, 2012; Biber et al., 2004; Csomay, 2012; Hyland, 2008), word clusters (e.g., Carter and McCarthy), or n-grams/concgrams (Cheng et al., 2006, 2008; Cheng & Leung, 2012). Among these, lexical bundles, word clusters and n-grams refer to the co-selection of *contiguous* words (e.g., *current assets*), while collocations and concgrams can refer to recurrent non-contiguous words (e.g., *current tax assets*) as well as contiguous words. In the present study, the term *collocations* will be employed to refer to the recurrent co-selections of words that are either contiguous or non-contiguous. Some collocations constitute phraseological units, which “coincide with traditional grammatical units (e.g., verb+noun collocations, such as *make a contribution*)” (Ädel and Erman, 2012: 82).

Collocations are crucial in language use in general because they provide language users with “a number of semi-preconstructed phrases that constitute single choices” (Sinclair, 1991: 110). That is, the meaning language users choose to convey is not individual words but the combinations of words, i.e., collocations. It is frequently the case that acquiring collocations is a crucial part of becoming a member of a given

discourse (Wray, 2006). At the same time, studies on collocations in various registers or genres have shown that language users engaged in different registers have different repertoires of collocations. (e.g., Biber et al., 2004; Biber, 2009; Cortes, 2004; Csomay, 2012; Hyland, 2008; Liu, 2012 and references therein). For example, Biber et al. (2004) investigated the use of lexical bundles in two registers, classroom teaching and textbooks, and found that the two registers employ different types of collocations, which are then argued to be a “unique linguistic construct” (Biber et al. (2004: 371). Ädel and Erman (2012) showed that non-native speakers of English possess a limited number of lexical bundles compared to native speakers of English. These findings suggest the crucial role of acquiring collocations in language use and language learning.

Noticing that studies on lexical bundles have not paid enough attention on recurrent noncontiguous words, Cheng and colleagues proposed the potential contributions of congrams, which encompass both contiguous and noncontiguous recurrent word combinations (e.g., Cheng et al., 2006, 2008; Cheng and Leung 2012). The main motivation for examining congrams was their “significance for better understanding of language use and meaning” (Cheng and Leung, 2012: 617), because they help identify all the possible patterns of word co-occurrence. In Cheng et al. (2008), the two-word collocation *role/play* was investigated to identify its collocational patterns alongside its respective frequencies. Given the frequency of each pattern, the canonical forms of the collocation were identified and their meanings were defined. The two identified canonical forms of the pair were the verb phrase *play \* \* role*, as in *he plays an important role*, and the noun phrase *role play*. The researchers also classified two types of variations: constituency variations and positional variations. The first type depends on the presence of any intervening words between the two words, while the second type is concerned with the occurrence order of the two words. Table 1 shows the constituency variations of the first positional variation, *play/role*, whereas Table 2 shows those of the second positional variation, *role/play*.

Table 1. Constituency variations of *play/role* (Cheng et al., 2008: 245)

Constituency Variants	Frequency
<b>PLAY * * ROLE</b>	<b>100</b>
PLAY * ROLE	22
PLAY * * * ROLE	19
PLAY * * * * ROLE	10
PLAY * * * * * * ROLE	1

Table 2. Constituency variations of *play/role* (Cheng et al., 2008: 246)

Constituency Variants	Frequency
<b>ROLE PLAY</b>	<b>40</b>
ROLE * PLAY	32
ROLE (2-6words) PLAY	32

The research by Cheng and colleagues on concgrams is an inspiring springboard for exploring collocations and suggests important implications in the field of language teaching and applied linguistics. Nevertheless, because the main focus of their research is to demonstrate the potential utility of a software program, the selection of data is not fully elaborated.

Thanks to the recent development of corpus tools, it is now possible to handle huge corpora of more than 100 million words and to perform searches in a more efficient way. One of these tools is corpus.byu.edu (Davies, 2013), which offers various querying and sorting functions. The number of the words contained in this corpus is much larger than those used in previous research, thus the queries can provide more comprehensive results on language use. For its improved search function and larger data size, the tool corpus.buy.edu is employed in the current study.

#### 4. Materials and procedure

The main aim of this paper is to examine how common English adjectives interact with their accompanying nouns. For that purpose, four adjectives whose frequencies belong to the top eight in COCA were chosen: *new*, *good*, *old*, and *high*. Their ranks among the adjectives and frequencies in the corpus are presented in Table 3.

Table 3. Adjectives used in the current study

Adjectives	Ranks	Frequencies
<i>new</i>	2	492585
<i>good</i>	3	409902
<i>old</i>	7	206743
<i>high</i>	8	202614

The top eight adjectives in COCA are as follows in decreasing order: *other*, *new*, *good*, *American*, *great*, *big*, *old*, and *high*. The adjectives *other* and *American* are excluded because they do not quite fit into the category of lexical adjectives. The two adjectives *great* and *big* are also excluded because their most frequent uses are in proper nouns (e.g., *Big Bang*, *Great Britain*) or (semi)-idiomatic expressions (e.g., *big difference*, *great deal*).

Next, the noun collocates of the chosen adjectives were searched within the span of three words to the left and right. The nouns with the highest frequencies were chosen for the analysis. When choosing the nouns, those mainly employed for proper nouns or compounds were excluded.<sup>2</sup> The list of noun collocates for the chosen adjectives are provided in Table 4.

Table 4. Noun collocates for the adjectives

Adjectives	new	good	old	high
Noun Collocates	book	news	friend	levels

The next step was to examine the collocational patterns of each pair. For each pair, six different patterns were searched: ADJECTIVE NOUN, ADJECTIVE \* NOUN, ADJECTIVE \* \* NOUN, NOUN ADJECTIVE, NOUN \* ADJECTIVE and NOUN \* \* ADJECTIVE. The first three are the constituency variations of the positional variation ADJECTIVE/NOUN (adjective-first variation henceforth), whereas the latter three are those of the variation NOUN/ADJECTIVE (noun-first variation). The frequencies of each pattern were counted and the internal components and structures were analyzed. When the frequencies and internal structure were examined,

<sup>2</sup> For example, in the case of *new*, the most frequent nouns were *city*, *times*, *years*, *world* and *book* in decreasing order. These nouns were mostly employed for proper nouns or in the compounds *New York City*, *New York Times*, *New Year* and *New World* respectively. Hence, the common noun *book* was selected for the current analysis.

it was determined whether the co-occurrences are meaningfully associated (following Cheng et al., 2008). For example, co-occurrences of the two words in different clauses or in proper nouns were excluded from the analysis, as illustrated in (2).

- (2) a. into entertainment time or a chance to learn something **new**.  
**AUDIOBOOK**...
- b. In thee name of creating Californicated' literature, **New York book** editors have blurred the line...

## 5. Results and Discussion

### 5.1 *New/book* collocation

There are 6,487 occurrences of *new/book* within the span of three words to the left and right. Out of all these occurrences, 5,953 cases are found to be semantically associated. The breakdown of the occurrences is presented in Table 5.

Table 5. Breakdown of the *new/book* collocation

<b>new/book</b>	<b>Occurrences (%)</b>	<b>book/new</b>	<b>Occurrences (%)</b>
new book	5654 (95.0)	book new	0 (0)
new * book	135 (2.3)	book * new	12 (0.2)
new * * book	109 (1.8)	book * * new	43 (0.7)

There are 5,654 occurrences of the noun phrase *new book*. This noun phrase is usually preceded by adjectives, most of which possess positive semantic prosody (see Sinclair (1991)). The most frequent adjectives used with *new book* include *brand*, *controversial*, *terrific*, *provocative*, *explosive*, *wonderful*, *extraordinary*, *excellent*, *stunning*, and *sensational*. Hence new books are generally valued as positive in American English. On the other hand, the noun phrase is followed either by a verb such as *is*, *claims*, *reveals*, and *tells* or by the preposition *for*. It is also modified by participles such as *called*, *coming*, *entitled*, and *written*.

In the corpus, 135 occurrences of the *new \* book* sequence are found. The expressions whose occurrences are greater than three are presented in Table 6.



Table 6. Expressions of the *new* \* *book* sequence

Expressions	Occurrences	Occurrences of * + book
new tell-all book	7	69
new picture book	7	290
new comic book	5	811
new Pray Book	4	177
new coffee-table book	4	89
new address book	4	366
new non-fiction book	4	102
new phone book	4	726
new photo book	3	33
new companion book	3	60
new hit book	3	18
new history book	3	185

The numbers in the third column specify the occurrences of the combinations of the intervening word and the noun *book*. The numbers are given to see whether the combination of the middle word and *book* is also meaningful. Overall, the combinations prove to be statistically meaningful; the Mutual Information (MI) scores are over 2 except for the expressions *photo book* (0.6), and *hit book* (-0.13).<sup>3</sup> As pointed out by one anonymous referee, most of the combinations of the intervening word the noun *book* can be regarded as compound nouns. Hence the three-word sequence *new* \* *book* is mainly employed in such a way where the adjective *new* modifies the combination of an intervening word and the noun *book*.

There are 109 occurrences of the *new* \* \* *book* sequence. The expressions containing two intervening words with more than three occurrences are listed in Table 7.

<sup>3</sup> “The Mutual Information score expresses the extent to which observed frequency of co-occurrence differs from what we would expect (statistically speaking). In statistically pure terms this is a measure of the strength of association between words *x* and *y*. In a given finite corpus MI is calculated on the basis of the number of times you observed the pair together versus the number of times you saw the pair separately.” (<http://wordbanks.harpercollins.co.uk/Docs/Help/statistics.html>)

Table 7. Expressions of the *new \* \* book* sequence

Expressions	Occurrences	Occurrences of * * + book
new children's book	20	720
new Bob Woodward book	5	14
new Harry Potter book	4	25

In these expressions, the adjective *new* also has the role of modifying the combination of the intervening words and the noun *book*. The figures in the last column show the frequencies of the two intervening words and the noun *book*. Their combinations are all proved statistically significant. Even though there are two intervening words, they can be regarded as one chunk in terms of meaning; *children's*, *Bob Woodward*, and *Harry Potter* refer to one entity. Thus, the two sequences *new \* book* and *new \* \* book* have a similar semantic function; the adjective *new* modifies the combination of the following expressions, most of which are compound nouns.

Let us move to a discussion of the positional variation *book/new*. There are not semantically associated occurrences of the *book new* sequence. The 12 occurrences of the *book \* new* sequence can be classified into two types: *book PREPOSITION new* (five tokens) and *book VERB new* (seven tokens). The examples of the first type are provided in (3).

- (3) a. poet Gary Snyder has well spent the time since his last **book of new** poetry was published. . .  
 b. Gina Maranto is at work on a **book about new** reproductive technologies.

This type is employed to refer to a book of/on/about/with a new entity, with the adjective *new* modifying the immediately following noun. The second type is used as a part of sentence, in which the noun *book* belongs to the subject and the adjective *new* to the predicate, usually modifying the following noun. Some examples of this type are given in (4).

- (4) a. This **book offers new** ideas for small business owners, including how to compete in a service economy

- b. Writing a **book was new** to him, but he spent a lot of time writing his monologues

The 43 occurrences of the *book \* \* new* sequence can be classified into two broad categories. The first type is *book PREPOSITION a/the/any/POSSESSIVE new* (15 tokens), as illustrated in (5).

- (5) a. He's at work on a **book about the new** business models developing around three-dimensional printing. . .  
 b. You should plan to take the instruction **book on any new equipment**.

The other type is *book VERB a/the/not. . . new* sequence (28 tokens). Similar to the *book \* new* sequence, in this type the noun *book* belongs to the subject and the adjective *new* belongs to the predicate respectively, as in (6).

- (6) a. This nicely written and carefully researched **book offers exciting new** insights on the Revolutionary War.  
 b. In that regard, the **book mirrors the new** age on the slopes.

Only two occurrences of this sequence reveal the predicative function the adjective *new*, as shown in (7).

- (7) a. ...on the sixtieth as on the ninth reading, the **book is as new**.  
 b. Now, explain this. This **book is not new?** HEATHER MILLS MCCARTNEY, AUTHOR. . .

The examination of the patterns of the *book/new* variation reveals that this positional variation is used for two constructions: *book on/about. . . new entity* and *book VERB new entity*. The adjective *new* is rarely used for the predicative function in this positional variation. Given the data the *new/book* collocation appears to have two canonical forms, which are the noun phrase *new book* and the form *book PREPOSITION/VERB new*.

## 5.2 *Good/news* collocation

There are 10,323 occurrences of *good/news* within the span of three words to the left and right, among which 8,688 semantically associated cases are found. The breakdown of the occurrences is presented in Table 8.

Table 8: Breakdown of the *good/news* collocation

<b>good/news</b>	<b>Occurrences (%)</b>	<b>news/good</b>	<b>Occurrences (%)</b>
good news	8271 (95.2)	news good	2 (0)
good * news	73 (0.8)	news * good	130 (1.5)
good * * news	78 (0.9)	news * * good	134 (1.5)

The noun phrase *good news* is preceded by the quantifier *some* (447 tokens) or modified by the intensifier *very* (143 tokens). There are 15 cases that contain the preceding adverb *necessarily*. Interestingly, among these 15 cases, 14 examples are couched in negative contexts, as shown in concordance lines (8).

- (8) a. It's difficult to say why this **wasn't necessarily good news**.  
 b. Either way, its **not necessarily good news** for the Republican party.

The examination of the collocation patterns of the adverb *necessarily* reveals that the adverb is chiefly employed in negative contexts, since the top two leftmost collocates of the adverb are *not* and its contracted form *n't*.

As opposed to the noun phrase *new book*, the noun phrase *good news* is seldom preceded by a modifying adjective; the word *shocking* is the only adjective included in the top ten leftmost collocates.<sup>4</sup> The most frequent rightmost collocate is the verb *is* (1,595), which is followed by the prepositions *for* (1,346 tokens) and *about* (212 tokens). Thus, *good news* tends to be mentioned with reference to the target recipient. The other top ten rightmost collocates are quite heterogeneous.

In the data, 73 occurrences of the *good \* news* sequence are found. The sequences with more than three occurrences are presented in Table 9.

<sup>4</sup> To investigate whether the difference comes from the nature of nouns or verbs, the leftmost collocates of the four words were searched. The results revealed that the leftmost collocates of the individual words are quite different from those of the NPs, *new book* and *good news*. This supports that the NPs, *new book* and *good news* function as lexical bundles/chunks.

Table 9: Expressions of the *good \* news* sequence

Expressions	Occurrences	Occurrences of * + news
good economic news	42	390
good earning news	3	27
good inflation news	3	7

Most of the expressions specify the type of good news such as *economic, earnings, inflation, financial, macroeconomic* and *television news*. The numbers in the third column specify the occurrences of the combination of the intervening word and the noun *news*. Overall, the combination was statistically meaningful because the MI score is over 2. Thus, the sequence *good \* news* is mainly employed to convey that certain type of news is good.

There are 82 occurrences of the *good \* \* news* sequence. The expressions containing two intervening words with more than three occurrences are listed in Table 10.

Table 10. Expressions of the *good \* \* news* sequence

Expressions	Occurrences
good and bad news	35
good or bad news	13
good piece of news	3
good pieces of news	3

This sequence is mainly employed to present *good news* in contrast to *bad news*, as shown in Table 10 (50 tokens). Since *news* is an uncountable noun in English, the expressions *piece(s)/bit of* appear between *good* and *news*.

Let us turn to the discussion of the positional variation. The sequence *news good* occurs just twice, as shown in (9).

- (9) a. Her pulse raced faster than a hummingbird’s wings. “ Is the **news good?**” she said...
- b. One of the most important targets was the famous VOA slogan “ we bring you **news good** or bad. . .

Among the 130 occurrences of the *news \* good* sequence, 105 occurrences displayed the variants of the proposition, NEWS BE GOOD (81%). The variants and their respective occurrences are presented in Table 11. The concordance lines containing these sequences are presented in (10).

Table 11. Variants of the proposition NEWS BE GOOD in the *news \* good* sequence

Expressions	Occurrences
news is good	74
news was good	29
news being good	1
news be good	1
Total	105

- (10) a. ... the Dow is now up more than eleven percent. But not all economic **news is good**
- b. These were calls that she dreaded, but this time the **news was good**.  
Amazingly so
- c. “This is an example of bad **news being good** news,” says Dr. Vincent DeVita Jr
- d. has been holding her breath, waiting for her post-surgery lab report.  
Will the **news be good**?

This tendency is more pronounced in the *news \* \* good* sequence. Among the 134 occurrences, the variants of the proposition, NEWS BE GOOD, account for 98 cases (73%). Some of the variants are presented in Table 12 in decreasing order of frequencies.

Table 12. Variants of the proposition NEWS BE GOOD in the *news \* \* good* sequence

Expressions	Occurrences
news is not good	21
news was n't good	13
news was not good	13
news is n't good	13
news is all good	3
news has been good	2
news may be good	2
news is also good	2
news was pretty good	2
news was so good	2
news will be good	2

It could be concluded that the positional variation *news/good* with one or two intervening elements is mostly employed to express the proposition NEWS BE (NOT) GOOD. As predicted, the sequence *news \* good* is employed to convey a positive proposition, whereas the sequence *news \* \* good* is mainly for its negative counterpart. Unlike the adjective *new*, the adjective *good* is favored in the predicate position as well as in the modifier position when used with the noun *news*. There are two canonical forms for the *good/news* collocation: the noun phrase *good news* and the proposition NEWS BE GOOD.

### 5.3 *Old/friend* collocation

There are 2,762 occurrences of *old* and *friend* within the span of three words to the left and right, among which 2,701 cases are semantically associated. The breakdown of the occurrences is presented in Table 13.

Table 13. Breakdown of the *old/friend* collocation

<b>old/friend</b>	<b>Occurrences (%)</b>	<b>friend/old</b>	<b>Occurrences (%)</b>
old friend	2422 (89.7)	friend old	1 (0)
old * friend	173 (6.4)	friend * old	10 (0.4)
old * * friend	75 (2.8)	friend * * old	20 (0.7)

The noun phrase *old friend* is mainly preceded either by possessive pronouns such as *his*, *my*, and *our* or by adjectives such as *dear*, *reliable* and *trusted*. As for the rightmost collocates, the preposition *from* is the most frequent one, and the relative pronouns *who* and *whom* are the second and third most frequent collocates. Interestingly, male proper names constitute the remaining top ten rightmost collocates, including *George*, *David*, *Bob*, *Bill*, *Jerry* and *Ed*. These data show that in American English an old friend must be predominantly associated with male friends.

There are 173 occurrences of the *old \* friend* sequence with 76 types. The *old \* friend* sequences with more than three occurrences are presented in Table 14.<sup>5</sup>

Table 14. Expressions of the *old \* friend* sequence

Expressions	Occurrences	Occurrences of * friend
old family friend	32	760
old college friend	28	132
old school friend	18	153
old childhood friend	5	360
old army friend	4	13
old best friend	4	6236
old political friend	3	14
old high-school friend	3	14

The most frequently used intervening words include *family*, *college* and *school*. This finding reveals that *family friends*, and *college/school friends* are regarded as old friends most prominently in American English. Most of the intervening words and the word *friend* can be collocated, as revealed by the frequency numbers in the third column in Table 14.

Except for the word *family*, the intervening words are classified according to their functions, which are presented in Table 15.

<sup>5</sup> The sequences that have MI score of less than 1 are *army friend*, and *political friend*.



Table 15. Functional classification of the *old \* friend* sequence

Types	Occurrences
old ACADEMIC INSTITUTION friend	51
old NATION/ORIGIN friend	12
old PERSON'S NAME friend	7
old ADJECTIVE friend	30

Expressions such as *old (high-)school friend*, *old college friend*, and *old Stanford friend* belong to the first category. Related to this category is the use of words that specify when or where the friends met. Expressions such as *old army/navy friend*, and *old boyhood friend* are instances of this type. For the second category, words such as *Polish*, *Arkansas* and *Cuban* are employed. As for the person's name category, names such as *Eastwood*, *Beatty*, *Bush* and *Rudolph* are found. Another prominent construction is the *old ADJECTIVE friend* sequence, which includes adjectives such as *best*, *political*, *familiar*, *dear*, and *hick*.

In the corpus, 75 occurrences of the *old \* \* friend* sequence are found. The expressions containing two intervening words with more than three occurrences are listed in Table 16.

Table 16: Expressions of the *old \* \* friend* sequence

Expressions	Occurrences
old high school friend	11
old and dear friend	9
old, old friend	7
old and trusted friend	4
old man's friend	4
old ally and friend	3

The most frequent four-word sequence is *old high school friend*. Again, this finding proves that school friends are typically mentioned as old friends. The *old \* \* friend* sequence can be classified into several major categories, which are presented in Table 17.

Table 17: Classification of the *old* \* \* *friend* sequence

Types	Occurrences
old AND ADJECTIVE friend	35
old NOUN AND friend	9
others	31

The most recurrent adjectives include *dear* and *trusted*. Other adjectives used in this construction are *close*, *valued*, and *drunken*. When the second type is employed, as in (11), the whole sequence refers to the same person who plays two roles simultaneously.

- (11) ... she is joined by Dr. Suhaila Siddiqi, an **old ally and friend**, in speaking up for women.

In (11), the person called Dr. Suhaila Siddiqi must be regarded as both an old ally and an old friend at the same time.

The positional variation *friend old* is employed just once, as shown in (12).

- (12) I tried not to remember overhearing Banks ask Grace when I first arrived, “ Is your **friend old** enough to drive yet?”

Out of the 10 occurrences of the sequence *friend* \* *old*, five cases contain the expression *friend and old*, which are illustrated in (13).

- (13) China agreed that its closest **friend and old** communist ally North Korea should abandon its nuclear weapons program.

Here the noun phrase containing *friend* and the noun phrase containing *old* are juxtaposed and refer to the same entity. The other five occurrences are quite heterogeneous.

The 20 occurrences of the *friend* \* \* *old* sequence are mainly employed in two constructions: FRIEND FROM ARTICLE/POSSESSIVE OLD NOUN (11 tokens) and FRIEND BE (ARTICLE/ADVERB) OLD (5 tokens). Typical examples of each construction are provided in (14) and (15).

- (14) a. A **friend from the old neighborhood** was due over for dinner.  
 b. I saw a **friend from my old neighborhood** and told him about the bees.
- (15) a. Her parents had died; her only real **friend was an old** parrot, with green feathers.  
 b. “Your **friend is very old**, right?” Rachel nodded.

The positional variation *friend/old* with one or two intervening elements appears to be seldom used to express the proposition, FRIEND BE OLD. The identified canonical form of this collocation is the noun phrase *old friend*.

#### 5.4 *High/levels* collocation

There are 4,247 occurrences of *high* and *levels* within the span of three words to the left and right. Among them, 4171 cases are semantically associated. The breakdown of the occurrences is presented in Table 18.

Table 18. Breakdown of the *high/levels* collocation

<b>high/levels</b>	<b>Occurrences (%)</b>	<b>levels/high</b>	<b>Occurrences (%)</b>
high levels	3144 (75.4)	levels high	47 (1.1)
high * levels	646 (15.5)	levels * high	108 (2.6)
high ** levels	129 (3.1)	levels ** high	97 (2.3)

Out of the 3,144 occurrences of *high levels*, 2,772 cases (88%) are followed by the preposition *of*, forming the phraseological unit, *high levels of*. The examination of expressions after this phraseological unit reveals that the words presented in Table 19 usually follow the phrase, *high levels of*.

Table 19: Words immediately following the phrase *high levels of*

<b>Words</b>	<b>Occurrences</b>	<b>MI</b>
stress	63	7.60
lead	38	5.60
identification	32	7.94
unemployment	32	7.87
anxiety	30	7.38
social	25	3.74
mercury	24	8.27
parental	24	7.95
poverty	24	6.86
education	21	3.68

Except for *social* and *parental*, the words listed in Table 19 are nouns. Thus, *high levels* are usually mentioned in reference to materials such as either *lead* or *mercury* or abstract notions such as *stress*, *unemployment* or *poverty*. This finding can be compared with the words used in the sequence, *high \* levels*. As for the leftmost collocates of *high levels*, the words are classified either as modifying adverbs such as *very*, *relatively* or *extremely*, or verbs such as *report(ed)*, *contain*, and *found*. This indicates that the noun phrase *high levels* is modified by adverbs or used as the object of those verbs.

In the corpus, 646 occurrences of the sequence *high \* levels* are found. The intervening words whose occurrences are greater than three are presented in Table 20.

Table 20. Words used in the *high \* levels* sequence

Occurrences	Words
70	school
38	cholesterol
18	lead
14	energy/ ozone
13	stress
10	mercury/ blood
9	education/ radon
8	radiation/ water/ cortisol
7	debt/ enough/ skill
6	poverty/ triglyceride/ homocysteine/ immigration/ insulin/ activity
5	light/ testosterone/ nitrate
4	noise/ socioeconomic/ hemoglobin/ estrogen/ glucose/ academic/ CRP/ circulating/ blood-cholesterol/ blood-sugar
3	copper/bacteria/ exposure/ force/ dosage/ hematocrit/ intensity/ irradiance/ metal/ nutrient/ LDL/ PCB/ production

The two constructions, *high levels of* and *high \* levels*, are semantically synonymous and therefore it is predicted that the same expressions can appear either after the first construction or as the intervening word in the second construction. A comparison of the words in Table 19 and those in Table 20 reveals that at least two of the words, *stress* and *lead*, are commonly employed in these two constructions.

Among the words used in the *high \* levels* sequence, the most frequently used word is *school*, forming a compound *high school*. The other intervening words employed in the sequence *high \* levels* can be classified into two broad categories. The first category includes abstract nouns such as energy, stress, education, and radiation, whereas the other category contains nouns referring to materials such as cholesterol, lead, ozone and mercury. All the words presented in Table 20 can be classified as belonging to either of these two categories. Thus, it could be concluded that *high levels* are mentioned with respect to either abstract notions or materials.

The sequence *high \* \* levels* has 129 tokens and those with more than three occurrences are presented in Table 21.

Table 21. Expressions of the *high \* \* levels* sequence

Expressions	Occurrences
high and low levels	17
high blood sugar levels	11
high or low levels	9
high blood cholesterol levels	7
high versus low levels	4
high sound pressure levels	3
high blood lead levels	3
high job satisfaction levels	3
high carbon dioxide levels	3

The intervening words used in this sequence can be classified into several categories. Noticeably, the word *low* is frequently juxtaposed with *high* as in *high and low levels*, *high or low levels*, or *high versus low levels*.<sup>6</sup> These constructions account for 32 tokens (25%). Similar to the sequence *high \* levels*, the words appearing between *high* and *levels* in this four-word sequence specify either materials (e.g., *blood sugar/cholesterol* and *carbon dioxide/monoxide*) or abstract notions (e.g., *sound pressure* and *job satisfaction*). As a minor category, there are constructions in which the word *high* is juxtaposed with another adjective, forming phrases such as *high and stable/sustained levels*.

Let us turn to a discussion of the positional variation *levels/high*. Overall, 47 tokens of the two-word sequence *levels high* are found. Interestingly, out of 47 tokens, 14 cases (30%) are followed by the adverb *enough*, as shown in the concordance lines (16).

- (16) a. Almost 90 percent of primary school children tested had **levels high enough to impair their developmental and learning** abilities...
- b. That would protect against droughts and also keep lake **levels high enough** for recreational boating

The other constructions where this two-word sequence is used are presented in (17).

<sup>6</sup> There are also *high to low levels* and *high or low-skill levels*, which appear just once.

- (17) a. Consistent eating will also **keep** your protein **levels high**, helping you build muscle.  
 b. huts of brick and stone and wood, some of them two or three **levels high**, sagging roofs of shingle or slate.

In (17a), the word *high* is used as an object complement of the verb *keep*, forming a phraseological unit, *keep something levels high*. On the other hand, in (17b), the two-word sequence *levels high* forms a (quasi) measure expression similar to *feet high*.

In the corpus, 108 occurrences of the *level \* high* sequence are found. The expressions whose occurrences are greater than three are presented in Table 22.

Table 22. Expressions of the *levels \* high* sequence

Expressions	Occurrences
levels are high	30
levels as high	20
levels were high	13
levels and high	8
levels of high	7
levels remained high	3

It is noteworthy that the proposition LEVELS COPULAR VERB HIGH accounts for 54 cases (50%). Some of the concordance lines containing this proposition are given in (18).

- (18) a. When carbon dioxide **levels are high**, it's impossible to light a flame.  
 b. She had kidney malfunctions, liver malfunctions, her hormone **levels were high**..  
 c. The early risers' **levels remained high** all day. If you are trying to keep your cortisol normal, do not rise too early.

All occurrences of the second most frequent sequence *levels as high* are followed by the word *as*, forming a phraseological unit *levels as high as*, as shown in concordance lines in (19).

- (19) a. In 1998, chemical analysis of the water supply well at the View-Master factory in Beaverton, Oregon, found TCE at **levels as high as** 1,600 parts per billion...
- b. Nationwide, the company has placed stores in neighborhoods with income **levels as hig has** \$100,000, Covert said.

In the corpus, 97 occurrences of the four-word sequence *levels \* \* high* are found. The expressions whose occurrences are greater than 3 are presented in Table 23.

Table 23. Expressions of the *levels \* \* high* sequence

Expressions	Occurrences
Levels are too high	10
Levels are so high	7
Levels are very high	5
Levels were too high	5
Levels were so high	4
Levels are extremely high	3

The dominance of the variants of the proposition LEVELS COPULA VERB HIGH is pronounced, as illustrated in Table 23. Indeed out of 97 occurrences, 80 (82%) cases are variants of that proposition. Some of the concordance lines are provided in (20).

- (20) a. if you wonder if the noise **levels are too high** for you, there's a website called Yelp and there's another one called Open Table.
- b. Our view is that if **levels are so high** on our property that it requires condemning, then condemn it!
- c. The water **levels are very high**. But all of the vital safety systems at those plants are -- are being protected.

The two sequences *levels \* high* and *levels \* \* high* are used to express the proposition LEVELS COPULA VERB HIGH in various forms. The canonical forms for the *high/levels* collocation are the NP *high levels* and the proposition LEVELS COPULA VERB HIGH.



## 6. General discussion and conclusion

As far as the four collocations discussed in this paper are concerned, the NPs consisting of an adjective and a noun comprise the vast majority of the occurrences: *new book* (95%), *good news* (95.2%), *old friend* (89.7%), and *high levels* (75.4%). Thus, an NP consisting of an adjective and a noun is the canonical form. Among the four collocations, the NP *high levels* has the lowest occurrence percentage. The reason for this result must be that in this NP, a certain type of information is lacking semantically. In other words, *high levels* tend to be mentioned with respect to something. In contrast, the other NPs are self-contained semantically. This claim is partially supported by the fact that the *high \* levels* sequence has the highest occurrence percentage compared to the other ADJECTIVE \* NOUN sequences: *new \* book* (2.3%), *good \* news* (0.8), *old \* friend* (6.4) and *high \* levels* (15.5%). In terms of the function/meaning, the two-word sequence ADJ \* NOUN and the three-word sequence ADJ \* \* NOUN share similarities throughout the four collocations. That is, in these two sequences, the adjective modifies the combination of the intervening words and the noun. As the percentage figures reveal, the two-word sequences of each pair outnumber their three-word counterparts.

Throughout the four collocations, the noun-first sequences have dramatically lower occurrences than the adjective-first sequences: *book/new* (0.9%), *news/good* (3%), *friend/old* (1.1%) and *levels/high* (6%). This proves that the four adjectives serve the attributive function much more frequently than the predicative function when accompanied by the relevant nouns. Another noticeable point concerning the noun-first sequences is that the four collocations are divided into two types. That is, the two/three-word sequences, *book \* (\*) new* and *friend \* (\*) old*, are seldom employed in such a way that the adjective plays the predicative function. On the other hand, the adjectives in the two/three-word sequences *news \* (\*) good* and *levels \* (\*) high* are predominantly used to serve the predicative function. That is, the adjective *good* and *high* mainly occur as the predicates of the nouns *news* and *levels*, respectively, in this positional variation. Thus, the four adjectives differ in terms of the extent to which they display the predicative functions.

As is discussed in Section 2, the meanings of adjectives are largely subject to accompanying parts-of-speech, mostly nouns. Moreover, in the traditional semantics literature, most adjectives are considered to have two major functions: attributive and

predicative. This study showed that the most common English adjectives are differentiated in terms of their predicative functions when they co-occur with their closely related nouns. Furthermore, the adjectives behave differently in the way they interact with their antonyms. The two adjectives *high* and *good* are easily juxtaposed with their antonyms *low* and *bad* in four-word adjective-first sequences, while the other two adjectives are not. The findings in this paper hope to suggest a corpus-driven approach to investigating the interaction between adjectives and frequently co-occurring nouns. In addition, these findings can be applied to language teaching and teaching material production, so that language learners can benefit by being exposed to authentic language data.

### References

- Ädel, Annelie, and Britt Erman. 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes* 31(2): 81-92.
- Biber, Douglas. 2009. A corpus-driven approach to formulaic language in English multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14(3), 275-311.
- Biber, Douglas, Susan Conrad, and Viviana Cortes. 2004. *If you look at...: Lexical bundles in university teaching and textbook*. *Applied Linguistics* 25(3): 371-405.
- Carter, Ronald, and Michael McCarthy. 2006. *Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Cheng, Winnie, Chris Greaves, John Sinclair, John, and Warren Martin. 2008. Uncovering the extent of the phraseological tendency: Towards a systemic analysis of concgrams. *Applied Linguistics* 30(2), 236-252.
- Cheng, Winnie, Chris Greaves, and Martin Warren, Martin. 2006. From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics* 11(4), 411-433.
- Cheng, Winnie, and Maggie Leung. 2012. Exploring phraseological variations by concgramming: The realization of complete patterns of variations. *Linguistic Research* 29(3), 617-638.
- Cortes, Viviana. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes* 23(4), 397-423.
- Croft, William, and Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

- Cruse, Alan. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Csomay, Eniko. 2013. Lexical bundles in discourse structure: A corpus-based study of classroom. *Applied Linguistics* 34(3): 369-388.
- Hyland, Ken. 2008. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27(1), 4-21.
- Jones, Steven. 2002. *Antonymy: a Corpus-based Perspective*. Routledge, London.
- Kamoen Naomi, Bregje Holleman, Rick Nouwen, Ted Sanders, and Huub van den Bergh. 2011. Absolutely relative or relatively absolute? The linguistic behavior of gradable adjectives and degree modifiers. *Journal of Pragmatics* 43(12), 3139 - 3151.
- Kennedy, Christopher, and Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81(2), 345 - 381.
- Lee, Hye-Kyung. 2013. Antonymy and gradability: A corpus-based approach on English gradable antonyms. *Linguistic Research* 30(2), 335-354.
- Liu, Dilin. 2012. The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes* 31(1), 25-35.
- Lyons, John. 1977. *Semantics*. Cambridge University Press, Cambridge.
- Paradis, Carita, and Cariline Willners. 2006. Antonymy and negation - the boundedness hypothesis. *Journal of Pragmatics* 38(7), 1051 - 1080.
- Quirk, Randolph, Sidney Greenbaum, Sidney, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Sinclair, John McH. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Wray, Alison. 2006. Formulaic language. In Keith Brown (Ed.), *Encyclopedia of Language and Linguistics*, 590-597. Oxford: Elsevier.

**Hye-Kyung Lee**

Department of English

Ajou University

206 Worldcup-ro Yeongtong-gu Suwon-si Gyeonggi-do 443-749 Korea

E-mail: hklee@ajou.ac.kr

Received: 2014. 06. 23.

Revised: 2014. 09. 20.

Accepted: 2014. 09. 20.