

Original Research Papers

The Use of Corpus and a Collocation Framework in the Comparison of the English Tests of Two Major College Entrance Examinations in Taiwan

ABSTRACT

Aims & study design: This study primarily investigated the English tests of two major college entrance examinations in Taiwan (i.e., the General Scholastic Ability Test and the Advanced Subjects Test) in terms of collocations via the use of corpus and a collocation framework proposed by Benson, Benson, and Ilson (1986 & 1997).

Methodology: Two corpora based on the GSAT and the AST (from 2002 to 2011) were built for analysis.

Results & conclusion: The results obtained through a program called AntConc demonstrated the following similarities. In relation to the most frequent lexical collocations derived from the content words that were among the 200 most frequently used words in both exams, *is not* (L7) ranked the 1st, and L3 was the most frequently occurring pattern among the seven types put forward by Benson et al. (1986 & 1997). As for the most frequent grammatical collocations generalized from the same content words in the two tests, *according to* (G5) and *such as* (G5) were among the top three.

Keywords: GSAT, AST, corpus, collocations

1. INTRODUCTION

1.1 The Definition of Collocation

An explanation of collocation has been outlined clearly by many researchers. In 1957, Firth first introduced the term of collocation explicitly. He stated that “you shall know a word by the company it keeps” (p. 181). The word “collocation” started receiving more attention after Firth coined the word and outlined its significance. In addition, Baker (1992) and Benson, Benson, and Ilson (1997) extended Firth’s definition by asserting that a collocation is a lexical unit which consists of a string of two or more than two words from different parts of speech. For example, *take a risk* is recognizable as a collocation, whereas *get a risk* should be deemed as a

misunderstanding of a unit of words. When you are doing something that is a risk, we normally use the verb of “take” not “get” to collocate with a risk. In other words, collocations do not only concentrate on a single word but also involve a focus on word associations.

Sinclair (1991) further adopted a frequency-based approach to collocations which started with Mr. Firth, and he proposed that a “collocation is the occurrence of two or more words within a short space of each other in a text” (p. 170). Moreover, he stated that words do not usually appear individually to make meaningful associations but must be associated with other words to be meaningful. Nation (2001) considered collocations as “items which frequently occur together and have some degree of semantic unpredictability” (p. 317). Furthermore, Lewis (1997) provided a typical definition of collocation. He asserted that collocations are associated words which naturally appear more frequently than randomly and they occur together although not all words that co-occur are collocations.

From the definitions above, it can be concluded that collocations are comprised of a cluster of two or more words which co-occur more frequently than word clusters which are infrequent.

1.2 Categories of Collocations

There are two representative categories of collocations that have emerged from among the many theories of researchers and linguists in regard to collocations. These come from Cowie (1981) and Benson, Benson, and Ilson (1986 & 1997). Cowie (1981) distinguished word combinations according to four types: free combinations (open collocations), restricted collocations, figurative idioms, and pure idioms (see Table 1).

Table 1. Four Types of Combinations (as cited in Kung, 2012, p. 34 and Nesselhauf, 2005, p. 14–15)

Free combinations (e.g., drink tea):
- The restriction on substitution can be specified on semantic grounds.
- All elements of the word combination are used in a literal sense.
Restricted collocations (e.g., perform a task):
- Some substitution is possible, but there are arbitrary limitations on substitution.
- At least one element has a non-literal meaning, and at least one element is used in its literal sense; the whole combination is transparent.
Figurative idioms (e.g., do a U-turn, in the sense of “completely change one’s policy or behavior”):
- Substitution of the elements is seldom possible.
- The combination has a figurative meaning but preserves a current literal interpretation.
Pure idioms (e.g., blow the gaff):

- 73 - Substitution of the elements is impossible.
- 74 - The combination has a figurative meaning and does not preserve a current literal
- 75 interpretation.

76 =====

77

78 In addition to the four types of combinations, there is another category of collocation which
 79 has been proposed by Benson et al. (1986 & 1997) who stated that “students must learn how
 80 words combine or “collocate” with each other. In any language, certain words regularly combine
 81 with certain other words or grammatical constructions. These recurrent, semi-fixed
 82 combinations, or collocations, can be divided into two groups: lexical collocations and
 83 grammatical collocations” (p. ix).

84 Based on Benson et al. (1986 & 1997), lexical collocations are comprised of adjectives,
 85 adverbs, nouns and verbs, not including infinitives, clauses, and prepositions. As Table 2
 86 illustrates, there are seven types of lexical collocations, including verb + noun or pronoun or
 87 prepositional phrase (L1), verb + noun (L2), adjective + noun (L3), noun + verb (L4), noun₁ of
 88 noun₂ (L5), adverb + adjective (L6), verb + adverb (L7). A difference between verbs in the L1
 89 and the L2 occurs in that verbs in the L1 denote creation and/or activation and the verbs in the
 90 L2 denote eradication and/or nullification.

91

92 **Table 2. Lexical Collocations Structure (adopted from Benson, Benson, & Ilson, 1986 &**
 93 **1997)**

Type	Structure	Example
L1	verb + noun / pronoun (or prepositional phrase)	reach a verdict
L2	verb + noun	reject an appeal
L3	adjective + noun	strong tea
L4	noun + verb	blood circulates
L5	noun ₁ of noun ₂	a pack of dogs
L6	adverb + adjective	hopelessly addicted
L7	verb + adverb	affect deeply

94

95 In contrast to lexical collocations, grammatical collocations not only contain a dominant

word such as a noun, adjective, and verb, but also contain a preposition or grammatical structure such as infinitives and clauses. Grammatical collocations are delineated from G1 to G8 as listed in Table 3, and G8 has been further divided into 19 English verb patterns (see Table 4).

Table 3. Grammatical Collocations Structure (adopted from Benson et al., 1986 & 1997)

Type	Structure	Example
G1	noun + preposition (excluding noun + of, noun + by, noun + concerning; regarding; in regard to; with regard to)	<i>apathy towards</i>
G2	noun + to + infinitive	They made an <i>attempt to do</i> it.
G3	noun + that-clause	He took an <i>oath that</i> he would do his duty.
G4	preposition + noun	<i>by accident</i>
G5	adjective + preposition	They were <i>angry at</i> everyone.
G6	adjective + to + infinitive	She is <i>ready to go</i> .
G7	adjective + that-clause	It was <i>nice that</i> he was able to come home for the holidays.
G8	19 English verb patterns	see Table 4

Table 4. Nineteen English Verb Patterns (adopted from Benson et al., 1986 & 1997)

Type	Verb patterns	Example
A	S V O to O (or) S V O O	He sent the book to his brother. He sent him the book.
B	S V O to O	They mentioned the book to her.
C	S V O for O (or) S V O O	She bought a shirt for her husband. She bought him a shirt.

D	S V prep. O (*free combination and verb + by or with are excluded)	The dog barked at the car. *They came by train. *We cut bread with a knife.
E	S V to infinitive	He decided to come.
F	S V infinitive	We must work.
G	S V Ving	She quit smoking.
H	S V O to infinitive	We forced them to leave.
I	S V O infinitive	They saw her drive up to the house.
J	S V O Ving	He kept me waiting for two hours.
K	S V possessive Ving	I cannot imagine their stealing apples.
L	S V (O) that-clause	He denied that he had taken the money.
M	S V O to be C	We consider her to be a competent engineer.
N	S V O C	She dyed her hair red.
O	S V O O	The police fined him fifty dollars.
P	S V (O) adverbial	He carried himself with dignity.
Q	S V (O) wh-word	He wants what I want.
R	S (it) V O to infinitive S (it) V O that-clause	It behooves you to study more. It surprised me that our offer was rejected.
S	S V C (adjective or noun)	She became an engineer. She was enthusiastic.

104

105 The similarities between lexical collocations and grammatical collocations can be seen in
106 that both are comprised of adjectives, verbs, and nouns. In terms of differences, only adverbs
107 appear in lexical collocations, and prepositions, infinitives and clauses occur in grammatical
108 collocations. In addition, the disadvantage of the two categories is that they do not concentrate
109 on the collocations which are associated with two function words such as “as if”. Fortunately,
110 the advantage of the two classifications by Benson et al. (1986 & 1997) is that complete
111 collocations not only provide a framework for content words (lexical collocations) but also
112 provide collocates with function words (grammatical collocations). Therefore, this advantage
113 was adopted in the current research. The 19 English verb patterns of the G8 type were
114 excluded in this study because they have been designed to analyze verb sentence patterns, not
115 word associations.

116

117 **1.3 The Importance of Collocations**

118 Researchers claim that collocation acquisition is an important element for foreigners to be
119 native-like in language use (Aston, 1995; Fillmore, 1979; Pawley & Syder, 1983). A few
120 research studies have further found that learners with sufficient collocation acquisition can
121 obtain productive and accurate language proficiency and perform with native-like fluency

(Benson, Benson, & Ilson, 1999; Chang, Chang, Chen, & Liou, 2008; Nation, 2001; Schmitt, 2000). More specifically, Shin and Nation (2008) have asserted that learning collocations is a way to increase a person's language fluency. Unfortunately, other researchers have indicated that a large number of EFL learners have insufficient collocation competence (Bahns & Eldaw, 1993; Farghal & Obiedat, 1995; Wu, 1996). For example, Tseng (2002) researched 94 participants who were senior high school students at National Yang Mei Senior High School in Taiwan. Forty-seven students out of the 94 students were the experimental group and were taught explicit collocation knowledge for 12 weeks. Meanwhile, the other 47 students were the control group and did not receive special collocation instruction. The results showed that the 94 students' collocation knowledge was not entirely proficient and they could not acquire collocation knowledge using a heuristic method. However, the collocation instruction attracted the experimental students' awareness successfully and improved their collocation acquisition. In addition to the studies above, Carter and McCarthy (1988) asserted that "collocations teach students expectations about which sorts of language can follow from what has preceded. Students will not have to go about reconstructing the language each time they want to say something but instead can use these collocations as pre-packaged building blocks" (p. 75). Likewise, Woolard (2000) has stated that the "collocation has emerged as an important category of lexical patterning and it is fast becoming an established unit of description in language teaching courses and materials" (p. 28). To make a brief conclusion in regard to the discussion above, awareness of collocations is obviously one vital means to make students fluent in listening, speaking, reading, and writing of the target language, and teachers can help to guide students to broaden their language knowledge by incorporating them (i.e., collocations) into teaching materials. Therefore, investigating collocations becomes such an important thing. Its importance and the aforementioned advantage of the two classifications by Benson et al. (1986 & 1997) have also motivated the researchers of this study to attempt to analyze the English tests of two major college entrance examinations (i.e., the General Scholastic Ability Test and the Advanced Subjects Test) in Taiwan. Hopefully, the results could benefit the teachers who are assisting their students in entering a good university and the students who are studying or will study for both tests. The following are the research questions that this project sought to address.

1. What are the most frequent collocations with respect to the content words retrieved from the 200 most frequent words in the English tests of the General Scholastic Ability Test (GSAT) and the Advance Subjects Test (AST) respectively?
2. Based on the framework of Benson et al. (1986 & 1997), what patterns do these collocations belong to?

2. MATERIAL AND METHODS

2.1 Instrument

The researchers of this project employed a corpus-based approach with the AntConc (3.2.4w) concordance program as the instrument of the study. Basically, a concordance tool, a concordance plot, a file view, clusters (N-Grams), collocates, a word list, and a keyword list were the major tools installed in the program. The concordance tool displays the results of a searched target word in a KWIC (Keyword in context). This tool shows that the target word or phrase appears in a corpus. The concordance plot primarily demonstrates the results plotted in a "barcode" bar where a given target word is shown as it exactly occurs in a text. The file view tool exhibits the entire text of the KWIC and the concordance plot. The clusters reveals two or more words or phrases accompanying the word that is sought. The collocates presents which words co-occur with the searched word. The word list shows all the words in an ordered list based on their frequency. The keyword list is used to compare two corpora. One corpus functions as a reference corpus that contains a large number of word lists to provide background information; the other is a small corpus that is to be investigated and analyzed. The intention of having the keyword list is to explore words in the smaller corpus to compare it with the other reference corpus for such features as word frequency or stylistic characteristics. From the introduction of the main function of each of the tools above, the word list is to demonstrate word frequencies, and the clusters is to display which words accompany target words. Therefore, the word list aided in discovering the most frequent words from the collected English tests of the GSAT and AST and the clusters to further reveal collocations in the current research.

2.2 Corpora Collection

The AST was first held back in 2002 and the GSAT around the same time. Hence, the data of the GSAT in this research study was gathered from 2002 to 2011 with 28,722 tokens, and the AST covered the same period as well with 32,544 tokens. In its first 10 years, 20 examination sessions based on the GSAT and AST were scheduled to be held once separately every year. However, the AST held one more examination than the GSAT as a result of a disruption by a typhoon in 2004. For that reason, a total of 21 English examination papers were collected and made into analyzable electronic files respectively by year. The corpus of the GSAT comprised 10 files, and the corpus of the AST contained 11 files that were constructed for linguistic analysis.

In reference to the components of the two corpora from the 21 examination papers, there were non-multiple choice questions and multiple choice questions in both tests. The non-multiple choice questions in the tests were in Mandarin, so they were excluded from this study. Accordingly, the current research focused on multiple choice questions to discover frequency words, content words (word forms), and collocations of the corpora. In the multiple choice question section, there were both a number of similarities and a number of differences between the English section of the GSAT and the English section of the AST. With respect to

the similarities, both of them were composed of a vocabulary test, a cloze test, a WYXT¹ test, and a reading comprehension test. The vocabulary test, the cloze test, and the WYXT test not only evaluated students on lexical words, grammatical words, and transitions, but also tested them on collocations based on various sentence structures and meanings. As regards the reading comprehension test, it assessed the examinees' reading ability through a few articles.

In terms of the differences between the GSAT and AST corpus, the sentence matching in the GSAT corpus which appeared in 2002 and 2003 only required that the test takers smoothly combine two sentences into one complete sentence. This part was also included in the corpus analysis in this research. The PZJG² test in the AST corpus was used to test if the participants had comprehension capability and could smoothly organize every sentence to make an entire passage. Moreover, the GSAT corpus covered approximately 4,500 words, and its average article length for the cloze test, WYXT test, and reading comprehension test was 150 to 250 words. In contrast, the AST corpus extended the vocabulary size from 4,500 to 7,000 words, and its average article length for the cloze test, WYXT test, PZJG test, and reading comprehension test was 200 to 300 words. In brief, the English exam of the AST was harder than that of the GSAT due to its increased amount of vocabulary, length of the article, and difficulty level of sentence structure.

2.3 Data Analysis

This study was conducted in five steps. These steps were implemented by using AntConc. First, all data of the GSAT and AST were collected and made in MS-Word documents (.doc). Then the MS-Word documents were converted into plain text files (.txt) respectively by year with AntConc 3.2.4w. Third, from the plain text files, AntConc generated a word frequency list for each corpus via the tool of word list, and the results showed the most frequent words in an ordered list for each corpus. Right after that, based on the word frequency list, four comparative sub-lists of nouns, adjectives, adverbs, and verbs from the 200 most frequent words of each corpus were created manually. Finally, the tool of clusters was used to show the most frequently occurring collocations in relation to the content words.

3. RESULTS AND DISCUSSION

3.1 The Most Frequent Collocations Generalized from the 134 Content Words in the GSAT Corpus

In an attempt to analyze the GSAT corpus, the current study initially sought the most frequent 200 words via the word list tool of AntConc program. A total of 134 content words,

¹ WYXT is an abbreviation of Wen Yi Xuan Tian (文意選填) and is a part of the GSAT. This part of the test requires test takers to read a passage and then the test takers have to select the most appropriate word/phrase for each blank in the passage.

² PZJG is an abbreviation of Pian Zhang Jie Gou (篇章結構) and is a part of the AST. This part of the test requires test takers to read a passage and then the test takers have to select a correct sentence for each blank in the passage.

which were among the top 200 frequently used words in the GSAT corpus, were then sorted out and grouped into four categories manually based on their parts of speech in the sentence. Afterwards, the content words were used as search terms to discover their collocations. From the results generated by the clusters tool of AntConc as shown in Table 5, L3 and L7 were the most frequently occurring lexical collocation patterns among the seven types. As can be seen from the table, *is not* appeared in 26 occurrences in L7 collocations. *Following statements* and *green tea* occurred 12 times and 11 times respectively, both in L3 collocations. The L7 collocation, *are not*, also took place 11 times.

More specifically, *is not* is an L7 collocation which is comprised of a verb and an adverb. It was mostly used in declarative sentences in the corpus. However, “NOT” was particularly demonstrated in uppercase in questions once in a while to emphasize to test takers that among the answer options of the question, there existed one that was not mentioned in the text. For instance, on the 2006 exam, this declarative sentence can be found: “However, stress is not always bad.” (p. 4) For question number 49 from the 2003 exam, “NOT” was emphasized in uppercase: “According to the passage, which of the following is NOT used in health spas?” (p. 6) Another frequent collocation that consists of an adjective and a noun (Type L3) was *following statements*. It was always used in questions in the tests. An example of this comes from test item number 48 on the 2006 exam: “Which of the following statements can be inferred from the passage?” (p. 5) Furthermore, test question number 41 from the 2011 exam asked, “Which of the following statements is true according to the passage?” (p. 3) Although *green tea* is also a frequent L3 collocation, it mostly appeared in the declarative sentences in the 2006 and 2008 tests because *green tea* was one of the topics in the passages of the tests. Examples of its use in the texts are, “Green tea leaves are heated in steam.” (2006, p. 5) and “The secret power of green tea is its richness in a powerful anti-oxidant.” (2008, p. 2) In terms of the *are not* (a verb + an adverb) collocation, it was presented in the declarative sentences for the plural subject in the exams. For example, “And in the real world there are not always happy endings.” (2002, p. 4) and “90% of the bottles used are not recycled and lie for ages in landfills.” (2010, p. 3) Therefore, from the data analyzed, *is not*, *green tea*, and *are not* were used frequently in declarative sentences in the corpus of the GSAT. *Following statements* was always employed in questions.

Table 5. Lexical Collocation Patterns Generalized from the 134 Content Words in the GSAT Corpus

Type	Structure	Examples	Frequency
L1	verb + noun / pronoun (or prepositional phrase)	take care	2
L2	verb + noun	none	0
L3	adjective + noun	following statements	12

		green tea	11
		many people	10
		red ant	10
		high school	10
		casual clothes	9
		bottled water	8
		every day	7
		second language	7
		old man	6
		severe pain	5
		million years	5
		folk medicine	5
		North America	5
		some people	5
		ancient civilizations	5
		other hand	4
		red clothing	4
		long time	4
		best title	4
		same part	4
		each year	4
		real world	3
		recent study	3
		first time	3
		different language	3
		both groups	3
		young children	2
		enough food	2
		good health	2
		giant tree	2
		new chapter	2
		last night	2
L4	noun + verb	stress is	6
		proms were	5
		people feel	4
		researchers found	4
		employers said	4

		girls are	4
		animals are	3
		leaves need	3
		boys were	3
		fireflies use	2
L5	noun ₁ of noun ₂	none	0
L6	adverb + adjective	so many	5
		very much	4
		most famous	3
		how much	3
		too much	3
L7	verb + adverb	is not	26
		are not	11
		is more	9
		was not	8
		is also	5
		is always	4
		is very	4
		is just	4
		is why	3
		is now	3
		find out	3
		cut down	3
		came back	3
		are often	2
		live longer	2
		heard only	2
		go off	2

265 Remark: Freq. represents Frequency

266

267 In contrast with lexical collocations that are primarily composed of adjectives, adverbs,
 268 nouns, and verbs, grammatical collocations not only contain a dominant word such as a noun,
 269 adjective, and verb³, but also involve a preposition or grammatical structure such as infinitives

³ As stated in the Introduction section, the 19 English verb patterns of the G8 type were excluded in this study since they have been designed to analyze verb sentence patterns as opposed to word associations.

270 and clauses. According to the outcome produced by the cluster tool of Antconc as
 271 demonstrated in Table 6, G5 was the most frequently occurring grammatical collocation
 272 pattern among the seven types with regard to the 134 content words in the GSAT corpus. In
 273 particular, the elements of the G5 collocations, *according to*, *true about*, and *such as*,
 274 co-occurred with each other more than 10 times.

275 *According to* is an adjective + a preposition type of G5 collocation which appeared 37
 276 times in the corpus, and this collocation was mainly used at the beginning of questions and
 277 declarative sentences. Two examples from the past exams can be illustrated. Test item 45 in
 278 the 2011 exam asked, "According to the passage, which of the following statements is true
 279 about Kamba and Maasai people?"(p. 4) Moreover, a declarative sentence in the 2006 exam
 280 expressed, "According to a recent study, almost three-quarters of American high school juniors
 281 said they felt stress at least once a week, some almost daily. Fewer than half of Japanese and
 282 Taiwanese eleventh graders reported feeling stress that often."(p. 4) In addition, *true about*, a
 283 G5 collocation comprised of an adjective + a preposition, occurred 13 times, and it was always
 284 used in the middle of interrogative sentences. For example, question 47 of the exam held in
 285 2007 asked, "Which of the following is true about public television stations?" (p. 5) Question 51
 286 of the 2011 exam queried, "What is true about Deborah Duffey? " (p. 5) Finally, *such as*, which
 287 also consists of an adjective + a preposition, appeared 11 times, and it was mainly used to
 288 introduce examples in declarative sentences. An example can be found in the 2005 test:
 289 "Small birds such as baby quails are fair game to the expanding colony." (p. 6) Another
 290 example was in the 2006 test, "Also, astronauts conduct experiments on the shuttle to learn
 291 how space conditions, such as microgravity, affect humans, animals, plants, and insects." (p. 5)
 292 In sum, from the data analyzed, *according to*, *true about*, and *such as* were the most frequently
 293 used collocations among the investigated collocations in relation to the 134 content words, and
 294 they appeared frequently in questions (i.e., *according to* and *true about*) and in declarative
 295 sentences (i.e., *according to* and *such as*) respectively in the corpus of the GSAT.

296
 297 **Table 6. Grammatical Collocation Patterns Generalized from the 134 Content Words in**
 298 **the GSAT Corpus**

Type	Structure	Example	Frequency
G1	noun + preposition (excluding noun + of, noun + by, noun + concerning; regarding; in regard to; with regard to)	area for	2
G2	noun + to + infinitive	children to break	1
		children to do	1
		way to measure	1

		way to live	1
		way to fight	1
		men to cut	1
		men to feel	1
G3	noun + that-clause	none	0
G4	preposition + noun	of ice	8
		in place	5
		in life	5
		of time	4
		on earth	4
		of magic	3
		in love	3
		of trees	3
		to elephants	3
		in space	2
		of problems	2
		at work	2
G5	adjective + preposition	according to	37
		true about	13
		such as	11
		all over	8
		better than	5
G6	adjective + to + infinitive	none	0
G7	adjective + that-clause	none	0

299 Remark: Freq. represents Frequency

300

301 **3.2 The Most Frequent Collocations Generalized from the 132 Content Words in** 302 **the AST Corpus**

303 In an attempt to analyze the AST corpus, the present project first sought the most frequent
304 200 words via the word list tool of AntConc program. A total of 132 content words were then
305 manually identified from the 200 most frequently occurring words, and they were used as
306 search terms to discover collocations via the clusters tool of AntConc. According to Table 7,
307 the results of this study showed that in terms of the 132 content words, the top three most
308 frequent lexical collocations were *is not*, *are not*, and *young people*. *Is not* appeared 22 times,
309 and *are not* occurred 14 times. These collocations include a verb followed by an adverb as in
310 type L7 collocations. *Young people* was found 10 times and consists of an adjective + a noun
311 as in an L3 type.

312 *Is not* was mostly used in declarative sentences. However, "NOT" was displayed in

uppercase in questions once in a while to emphasize to test takers that among the answer options of the question, there existed one that was not mentioned in the text. For example, on the 2010 exam, this declarative sentence can be found: "He is not only a novelist, poet and playwright, but also a renowned painter and sculptor." (p. 6) For question number 47 on the exam held in 2006, "NOT" was shown in uppercase: "Which of the following is NOT a reason for the Japanese to use more packaging?" (p. 5) Regarding *are not*, it occurred frequently in declarative sentences. The only difference between the two collocations is that *is* was used with the third person singular subject; whereas, *are* was used with the second person singular or plural subject. Two examples of *are not* in the corpus of the AST can be provided: "If you do, you are not alone." (2010, p. 3) and "PhD graduates in Austria are not encouraged to work outside university." (2011, p. 7). As can be seen in the table below, another most frequently occurring collocation was *young people* which was mostly used to signify a generation in declarative sentences. It occurred in sentences like, "Slang is most commonly used among young people today." (2010, p. 2) and "The number of young people is shrinking." (2011, p. 6) To conclude, *is not*, *are not*, and *young people* were all mainly used in declarative sentences. However, *is not* was sometimes utilized in questions. When it was, *not* was often demonstrated in uppercase in these questions to inform test takers that one of the answer options was not mentioned in the text and they had to find it out to score.

331

332 **Table 7. Lexical Collocation Patterns Generalized from the 132 Content Words in the**
333 **AST Corpus**

Type	Structure	Examples	Freq.
L1	verb + noun / pronoun (or prepositional phrase)	take care	3
L2	verb + noun	none	0
L3	adjective + noun	young people	10
		ethnic groups	8
		following statements	8
		other hand	7
		second paragraph	6
		some people	6
		high school	6
		every year	5
		each year	5
		many people	5
		presidential US	5
		animal experiments	5
		old building	4

		great Britain	4
		gray wolf	4
		same time	4
		long hair	3
		white man	3
		free newspaper	3
		rough diamonds	3
		new technology	3
		last year	3
		recent years	3
		copyrighted music	2
		all countries	2
		fresh water	2
		no idea	2
		first film	2
		American prospect	2
		best title	2
L4	noun + verb	people have	4
		hair has	3
		animals used	3
		countries are	3
		child is	3
		family had	2
		policeman did	2
		museum is	2
		problems are	2
		computer is	2
		parents are	2
		statistics show	2
L5	noun ₁ of noun ₂	none	0
L6	adverb + adjective	so far	3
L7	verb + adverb	is not	22
		are not	14
		was not	9
		are more	8
		is also	7

	are still	7
	is most	6
	are now	6
	is why	5
	is very	4
	is much	4
	were not	4
	come out	3
	are only	3
	are often	3
	are just	2
	is always	2
	was later	2
	find out	2

334 Remark: Freq. represents Frequency

335

336 With respect to the most frequent grammatical collocations, the following Table 8
 337 indicates that concerning the 132 content words in the AST corpus, the top three were
 338 *according to*, *such as*, and *for example*. *According to* appeared 26 times, and *such as*
 339 occurred 23 times. They were both G5 collocations which include an adjective and a
 340 preposition. Another collocation pattern *for example* which was a G4 collocation appeared 19
 341 times and is comprised of a preposition and a noun.

342 In addition, *according to* was mostly used before questions and declarative sentences in
 343 the corpus. For instance, in exam item 50 in the 2004 exam, the question was found:
 344 "According to the passage, which of the following statements is true about Robot Aibo?" (p. 5),
 345 and the researchers of the current study detected in the 2010 exam text: "According to Kopi
 346 Lowak suppliers, palm civets eat the fruit whole." (p. 5). As for the collocation *such as*, it was
 347 used in the middle or at the end of declarative sentences. The following sentence on the 2008
 348 exam provided an example: "The events in which they compete include both team sports, such
 349 as baseball, basketball, and soccer, and individual sports, such as golf, tennis, and bowling."
 350 (p. 3). In a 2011 text, another example was found: "Alternatives to animal testing, such as
 351 using human tissue or computer models, are now strongly recommended." (p. 5) With
 352 reference to *for example*, it was mainly employed before declarative sentences or sometimes
 353 as one of the test answer options. This can be seen in the 2009 exam text: "For example,
 354 many schools offer community service activities for students to join." (p. 5) and in the 2002
 355 exam choice (p. 2):

356 _____, paintings of historical characters and stories of everyday life became extremely

357 popular.

358 16. (A) meanwhile (B) besides (C) in fact (D) for example

359 Based on the findings above, *according to* frequently appeared in questions and declarative

360 sentences, *such as* was mostly employed in declarative sentences, and *for example* was

361 mainly used in declarative sentences or sometimes as one of the answer options.

362

363 **Table 8. Grammatical Collocation Patterns Generalized from the 132 Content Words in**
 364 **the AST Corpus**

Type	Structure	Examples	Freq.
G1	noun + preposition (excluding noun + of, noun + by, noun + concerning; regarding; in regard to; with regard to)	women in	4
		books on	3
G2	noun + to + infinitive	way to make	2
		way to stay	1
		way to show	1
		way to deliver	1
		way to compete	1
		way to enjoy	1
		way to solve	1
		time to make	1
		time to see	1
		time to get	1
		time to keep	1
		time to learn	1
		time to go	1
		children to take	2
		children to understand	1
		children to read	1
		students to learn	2
		students to read	1
		students to join	1
		students to download	1
G3	noun + that-clause	...those new computer words that seem to be popping up everywhere...	1

		They are words that often make us smile.	1
		in favor of words that are simple, fresh and playful	1
		Many words that were found in Shakespeare	1
G4	preposition + noun	for example	19
		in order	10
		in Taiwan	9
		in place	5
		in research	3
		in life	3
		in language	3
		of bag	3
		of art	2
		of animal	2
		of painting	2
		at home	2
		at work	2
		in world	2
G5	adjective + preposition	according to	26
		such as	23
		true about	5
		different from	4
		good at	3
		important in	3
G6	adjective + to + infinitive	none	0
G7	adjective + that-clause	none	0

365 Remark: Freq. represents Frequency

366

367 4. CONCLUSION

368 Although the GSAT and AST are two various major college entrance examinations in
 369 Taiwan and are held in different months every year, from the results of this study, they exhibit
 370 quite a few similarities in terms of the content of their English tests. In relation to the main
 371 focus of the current research (i.e., the most frequent lexical and grammatical collocations
 372 generalized from the content words that were among the 200 most frequently used words in
 373 both exams), it was found that *is not*, *following statements*, *green tea*, and *are not* were the

most frequently occurring lexical collocations in the corpus of the GSAT and in the AST corpus, they were *is not*, *are not*, and *young people*. Therefore, *is not* and *are not* were both among the top most frequent lexical collocations in the two exams, and they both belonged to L7 patterns based on the framework of Benson et al. (1986 & 1997). Furthermore, in the corpus of the GSAT and AST, *is not* and *are not* were often utilized in declarative sentences. However, "NOT" was once in a while displayed in uppercase in questions to emphasize to test takers that among the answer options of the question, there existed one that was not mentioned in the text. With respect to the most frequent grammatical collocations derived from the same content words, *according to*, *true about*, and *such as* were the top three most frequent grammatical collocations in the GSAT, and *according to*, *such as*, and *for example* were the top three commonly used grammatical collocations in the AST. Hence, *according to* and *such as* were both among the top three most frequent grammatical collocations in the two exams, and they both belonged to G5 patterns based on the framework of Benson et al. (1986 & 1997). Moreover, it was discovered that *according to* primarily existed in questions and declarative sentences in both the GSAT and AST and *such as* mainly occurred in declarative sentences in the two tests.

As Lin (2012) suggested, students should receive direct collocation training as early as possible since the scores of students who have acquired collocation knowledge have been higher than those who have never heard of collocations. Moreover, collocation knowledge improves students' reading and writing competence according to a significant positive relationship found to have occurred between the Joint College Entrance Examination and collocation tests. In addition, Huang (2009) found that senior high school students usually overlook the collocation patterns in English sentences, indicating the need for further training. Hence, the results of this study can serve as guidance for students to pay more attention to collocation knowledge in order to acquire the highest score in reading and writing on the English tests of these two vital examinations (i.e., the GSAT and AST). Also, the data derived from this project can be used to develop teaching materials so that language instructors can bring up the collocation patterns at the same time as they are teaching vocabulary in sentences.

The findings of this study can be expected to be a reference material for further research. The current project collected data from 2002 to 2011 because the AST was initiated in 2002. Prospective researchers could involve an investigation continuously of the following yearly examinations as this study concentrates only on the first decade. Furthermore, the first step in the present research was a focus on the top 200 frequent words, so if further studies can expand the frequency size, perhaps more findings will be revealed. In addition, G8 grammatical collocations involve 19 verb patterns and were omitted in the current research study because this study only focused on collocations not verb sentence patterns. A final suggestion for further research is to focus on the 19 verb patterns in the GSAT and AST. The

results from such a study of 19 verb sentence patterns should reveal in depth what kind of verb sentence patterns appear in the GSAT and AST in order to benefit senior high school students who want to enroll in a prestigious college or university through the GSAT and AST.

REFERENCES

1. Aston, G. (1995). Corpora in language pedagogy: Matching theory and practice. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H.G. Widdowson* (pp. 257-270). Oxford: Oxford University Press.
2. Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21,101-114.
3. Baker, M. (1992). *In other words*. London: Routledge.
4. Benson, M., Benson, E., & Ilson, R. (1986). *The BBI dictionary of English word combinations: A guide to word combinations*. Amsterdam, Netherlands: John Benjamins.
5. Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. Amsterdam, Netherlands: John BenJamins.
6. Benson, M., Benson, E., & Ilson, R. (1999). *The BBI dictionary of English word combinations (Revised ed.)*. Taipei: Bookman.
7. Carter, R., & McCarthy, M. (1988). *Vocabulary and language teaching*. New York: Longman.
8. Chang, Y. C., Chang, J. S., Chen, H. J., & Liou, H. C. (2008). An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. *Computer Assisted Language Learning*, 20, 283-299.
9. Cowie, A. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2, 223-235.
10. Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *IRAL*, 33, 315-333.
11. Fillmore, J. C. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. Wang (Eds.), *Individual differences in language ability and behavior* (pp. 85-101). New York: Academic Press.
12. Firth, J. R. (1957). *Papers in linguistics*. London: Oxford University Press.
13. Huang, C. M. (2009). *A study of the effects of collocation instruction on Taiwanese EFL senior high school students' learning of grammatical collocations*. Master's thesis, National Taiwan Normal University, Taiwan.
14. Kung, L. C. (2012). *The relationship between vocational high school students' productive vocabulary size and written collocation errors*. Master's thesis, National Kaohsiung Normal University, Taiwan.
15. Lewis, M. (1997). *Implementing the lexical approach: Putting theories into practice*. London: Language Teaching Publications.

- 450 16. Lin, K. S. (2012). *A study of Taiwanese EFL senior high students' collocation competence*
451 *and collocation learning strategy use*. Master's thesis, National Kaohsiung Normal
452 University, Taiwan.
- 453 17. Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge
454 University Press.
- 455 18. Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- 456 19. Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection
457 and nativelike fluency. In J. C. Richards, & R.W. Schmidt (Eds.), *Language and*
458 *communication* (pp. 191-226). New York: Longman.
- 459 20. Schmitt, N. (2000). *Vocabulary in language teaching*. New York: Cambridge University
460 Press.
- 461 21. Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in
462 spoken English. *ELT Journal*, 62, 339-348.
- 463 22. Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- 464 23. Tseng, F. P. (2002). *A study of the effects of collocation instruction on the collocational*
465 *competence of senior high school in Taiwan*. (Master's thesis). National Taiwan Normal
466 University, Taipei, Taiwan.
- 467 24. Woolard, G. (2000). Collocation-encouraging learner independence. In M. Lewis (Ed.),
468 *Teaching collocation: Further developments in the lexical approach* (pp. 28-46). London:
469 Language Teaching Publications.
- 470 25. Wu, W. S. (1996). Lexical collocations: One way to make passive vocabulary active. *The*
471 *Proceedings of the 11th Conference of English Teaching and Learning in the Republic of*
472 *China*, (pp. 461-480). Taipei: Crane.
- 473
- 474