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

Maosheng Hung[1*] and Yi-Hsien Chin[1]

[1]*Department of Applied English, Ming Chuan University, 5 De-Ming Rd., Gwei-Shan Township, Taoyuan 333, Taiwan, ROC.*

Authors' contributions

This work was carried out in collaboration between both authors. Author MH designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author YHC managed the analyses of the study and the literature searches. Both authors read and approved the final manuscript.

Original Research Article

ABSTRACT

Aims and 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 [3,8].
Methodology: Two corpora based on the GSAT and the AST (from 2002 to 2011) were built for analysis.
Results and 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. [3,8]. 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.

_____

Corresponding author: E-mail: allergy86@yahoo.com.tw;

1. INTRODUCTION

1.1 The Definition of Collocation

An explanation of collocation has been outlined clearly by many researchers. In 1957, Firth [1] 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 [2] and Benson, Benson, and Ilson [3] 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. However, not all words that co-occur are collocations. For example, *take a risk* is recognized as a collocation, whereas *get a risk* should be deemed as a misunderstanding of a unit of words. When people are doing something that is a risk, they normally use "take" (not "get") to collocate with a risk. Therefore, collocations not only concentrate on a single word but also involve a focus on word associations.

Sinclair [4] 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 [5] considered collocations as "items which frequently occur together and have some degree of semantic unpredictability" (p. 317). Furthermore, Lewis [6] 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

Two representative categories of collocations have emerged among the many theories of researchers and linguists in regard to collocations. These come from Cowie [7] and Benson, Benson, and Ilson [3,8]. Cowie [7] distinguished word combinations according to four types: free combinations (open collocations), restricted collocations, figurative idioms, and pure idioms (see Table 1).

In addition to the four types of combinations, there is another category of collocation which has been proposed by Benson et al. [3,8] who stated that "students must learn how words combine or "collocate" with each other. In any language, certain words regularly combine with certain other words or grammatical constructions. These recurrent semi-fixed combinations or collocations can be divided into two groups: lexical collocations and grammatical collocations" (p. ix).

Table 1. Four types of combinations (as cited in Kung [9], p. 34 and Nesselhauf [10], 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):
- Substitution of the elements is impossible.
- The combination has a figurative meaning and does not preserve a current literal interpretation.

Based on Benson et al. [3,8], lexical collocations are primarily comprised of adjectives, adverbs, nouns, and verbs. They do not include infinitives, clauses, and prepositions. As Table 2 illustrates, there are seven types of lexical collocations, including verb + noun or pronoun or prepositional phrase (L1), verb + noun (L2), adjective + noun (L3), noun + verb (L4), noun$_1$ of noun$_2$ (L5), adverb + adjective (L6), verb + adverb (L7). A difference between verbs in the L1 and the L2 lies in that verbs in the L1 denote creation and/or activation but verbs in the L2 denote eradication and/or nullification.

In contrast to lexical collocations, grammatical collocations not only contain a dominant word such as a noun, adjective, and verb, but also involve 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).

Lexical collocations and grammatical collocations share a similar characteristic: both are comprised of adjectives, verbs, and nouns. In terms of their differences, only adverbs appear in lexical collocations. Prepositions, infinitive, and clauses occur in grammatical collocations. In addition, a disadvantage of the two categories is that they do not concentrate on the collocations which are associated with two function words such as "as if". Fortunately, an advantage of the two classifications by Benson et al. [3,8] is that complete collocations not only provide a framework for content words (lexical collocations) but also provide collocates with function words (grammatical collocations). Therefore, this advantage was adopted in the current research. The 19 English verb patterns of the G8 type were excluded in this study because they have been designed to analyze verb sentence patterns, not word associations.

1.3 The Importance of Collocations

Researchers claim that collocation acquisition is an important element for foreign language (FL) leaners to use the FL like first language (L1) speakers (Aston [11]; Chen [12]; Fillmore [13]; Pawley & Syder [14]). A few research studies have further found that learners with sufficient collocation acquisition can obtain productive and accurate language proficiency and perform with L1 fluency (Benson, Benson, & Ilson [15]; Chang, Chang, Chen, & Liou [16]; Nation [5]; Schmitt [17]). More specifically, Shin and Nation [18] have asserted that

Table 2. Lexical collocations structure (adopted from Benson, Benson, & Ilson [3,8])

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

Table 3. Grammatical collocations structure (adopted from Benson et al. [3,8])

Type	Structure	Example
G1	noun + preposition (excluding noun + of, noun + by, noun + concerning; regarding; in regard to; with regard to)	*apathy towards*
G2	noun + to + infinitive	They made an *attempt to do* it.
G3	noun + that-clause	He took an *oath that* he would do his duty.
G4	preposition + noun	*by accident*
G5	adjective + preposition	They were *angry at* everyone.
G6	adjective + to + infinitive	She is *ready to* go.
G7	adjective + that-clause	It was *nice that* 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. [3,8])

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	The dog barked at the car.
	(*free combination and verb + by or with	*They came by train.
	are excluded)	*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	It behooves you to study more.
	S (it) V O that-clause	It surprised me that our offer was rejected.
S	S V C (adjective or noun)	She became an engineer.
		She was enthusiastic.

learning collocations is a way to increase a person's language fluency. Unfortunately, other researchers have indicated that a large number of English as a foreign language (EFL) learners have insufficient collocation competence (Bahns & Eldaw [19]; Farghal & Obiedat [20]; Wu [21]). For example, Tseng [22] 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 [23] 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 [24] 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. [3,8] 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. MATERIALS 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 (Hung & Chin [25]). 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 (Hung & Chin [25]).

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 WYXT[1] 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 (Hung & Chin [25]).

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

---

[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.*

corpus analysis in this research. The PZJG[2] 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 (Hung & Chin, [25]).

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, 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,

[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.*

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* (both in L3 collocations) occurred 12 times and 11 times respectively. 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*, g*reen tea,* and *are not* were used frequently in declarative sentences in the corpus of the GSAT. *Following statements* was always employed in questions.

In contrast with lexical collocations that are primarily composed of adjectives, adverbs,

nouns, and verbs, grammatical collocations not only contain a dominant word such as a noun, adjective, and verb [3], but also involve a preposition or grammatical structure such as infinitives and clauses. According to the outcome produced by the cluster tool of Antconc as demonstrated in Table 6, G5 was the most frequently occurring grammatical collocation pattern among the seven types with regard to the 134 content words in the GSAT corpus. In particular, the elements of the G5 collocations, *according to*, *true about*, and *such as*, co-occurred with each other more than 10 times.

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

---

[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.*

to and *true about*) and in declarative sentences (i.e., *according to* and *such as*) respectively in the corpus of the GSAT.

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

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

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.

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

In addition, *according to* was mostly used before questions and declarative sentences in the corpus. For instance, in exam item 50 in the 2004 exam, the question was found: "According to the passage, which of the following statements is true about Robot Aibo?" (p. 5), and the researchers of the current study detected in the 2010 exam text: "According to Kopi Lowak suppliers, palm civets eat the fruit whole."(p. 5). As for the collocation *such as,* it was used in the middle or at the end of declarative sentences.

The following sentence on the 2008 exam provided an example: "The events in which they compete include both team sports, such as baseball, basketball, and soccer, and individual sports, such as golf, tennis, and bowling." (p. 3). In a 2011 text, another example was found: "Alternatives to animal testing, such as using human tissue or computer models, are now strongly recommended." (p. 5) With reference to *for example,* it was mainly employed before declarative sentences or sometimes as one of the test answer options. This can be seen in the 2009 exam text: "For example, many schools offer community service activities for students to join." (p. 5) and in the 2002 exam choice (p. 2):

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

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

Based on the findings above, *according to* frequently appeared in questions and declarative sentences, *such as* was mostly employed in declarative sentences, and *for example* was mainly used in declarative sentences or sometimes as one of the answer options.

Table 5. Lexical collocation patterns generalized from the 134 content words in the GSAT Corpus

Type	Structure	Examples	Frequency
L3	adjective + noun	**following statements**	**12**
		green tea	**11**
		many people	10
		red ant	10
		high school	10
L7	verb + adverb	**is not**	**26**
		are not	**11**
		go off	2

Remark: Freq. represents Frequency

Table 6. Grammatical collocation patterns generalized from the 134 content words in the GSAT Corpus

Type	Structure	Example	Frequency
G5	adjective + preposition	**according to**	**37**
		true about	**13**
		such as	**11**
		all over	8
		better than	5

Remark: Freq. represents Frequency

Table 7. Lexical collocation patterns generalized from the 132 content words in the AST Corpus

Type	Structure	Examples	Freq.
L3	adjective + noun	**young people**	**10**
		ethnic groups	8
		following statements	8
		other hand	7
L7	verb + adverb	**is not**	**22**
		are not	**14**
		was not	9
		are more	8

Remark: Freq. represents Frequency

Table 8. Grammatical collocation patterns generalized from the 132 content words in the AST Corpus

Type	Structure	Examples	Freq.
G4	preposition + noun	**for example**	**19**
		in order	10
		in Taiwan	9
		in place	5
G5	adjective + preposition	**according to**	**26**
		such as	**23**
		true about	5
		different from	4
		good at	3
		important in	3

Remark: Freq. represents Frequency

4. CONCLUSION

Although the GSAT and AST are two various major college entrance examinations in Taiwan and are held in different months every year, from the results of this study, they exhibit quite a few similarities in terms of the content of their English tests. In relation to the main focus of the current research (i.e., the most frequent lexical and grammatical collocations generalized from the content words that were among the 200 most frequently used words in 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; 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. [3,8]. 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. [3,8]. 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 [26] 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 [27] found that senior high school students usually overlook 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 receive 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 when they are teaching vocabulary in sentences.

The findings of this research can be viewed as a reference material for further studies. The researchers of this project collected data from 2002 to 2011 because the AST started in 2002. Prospective researchers could involve an exploration continuously of the following yearly examinations as the current study focuses only on the first 10 years. Moreover, the frequency size of the present research was set to the top 200 frequent words. Therefore, 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.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Firth JR. Papers in linguistics. London: Oxford University Press; 1957.
2. Baker M. In other words. London: Routledge; 1992.
3. Benson M, Benson E, Ilson R. The BBI dictionary of English word combinations. Amsterdam, Netherlands: John Ben Jamins; 1997.
4. Sinclair J. Corpus, concordance, collocation. Oxford: Oxford University Press; 1991.
5. Nation ISP. Learning vocabulary in another language. Cambridge: Cambridge University Press; 2001.
6. Lewis M. Implementing the lexical approach: Putting theories into practice. London: Language Teaching Publications; 1997.
7. Cowie A. The treatment of collocations and idioms in learners' dictionaries. Applied Linguistics. 1981;2:223-235.
8. Benson M, Benson E, Ilson R. The BBI dictionary of English word combinations: A guide to word combinations. Amsterdam, Netherlands: John Benjamins; 1986.
9. Kung LC. The relationship between vocational high school students' productive vocabulary size and written collocation errors. Master's thesis, National Kaohsiung Normal University, Taiwan; 2012.
10. Nesselhauf N. Collocations in a learner corpus. Amsterdam: John Benjamins; 2005.
11. Aston G. 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. Oxford: Oxford University Press. 1995;257-270.
12. Chen MH. A study of the english collocation competence of college students in Taiwan. Master's thesis, National Taiwan University of Science and Technology, Taiwan; 2008.
13. Fillmore JC. On fluency. In C. J. Fillmore, D. Kempler, & W. Wang (Eds.). Individual differences in language ability and behavior. New York: Academic Press. 1979;85-101.
14. Pawley A, Syder FH. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards, & R.W. Schimidt (Eds.), Language and communication. New York: Longman. 1983;191-226.
15. Benson M, Benson E, Ilson R. The BBI dictionary of English word combinations (Revised ed.). Taipei: Bookman; 1999.
16. Chang YC, Chang JS, Chen HJ, Liou HC. An automatic collocation writing assistant for Taiwanese EFL learners: A case of corpus-based NLP technology. Computer Assisted Language Learning. 2008;20: 283-299.

17. Schmitt N. Vocabulary in language teaching. New York: Cambridge University Press; 2000.

18. Shin D, Nation P. Beyond single words: The most frequent collocations in spoken English. ELT Journal. 2008; 62:339-348.

19. Bahns J, Eldaw M. Should we teach EFL students collocations? System. 1993; 21:101-114.

20. Farghal M, Obiedat H. Collocations: A neglected variable in EFL. IRAL. 1995;33: 315-333.

21. Wu WS. Lexical collocations: One way to make passive vocabulary active. The Proceedings of the 11th Conference of English Teaching and Learning in the Republic of China, Taipei: Crane. 1996; 461-480.

22. Tseng FP. A study of the effects of collocation instruction on the collocational competence of senior high school in Taiwan. (Master's thesis). National Taiwan Normal University, Taipei, Taiwan; 2002.

23. Carter R, McCarthy M. Vocabulary and language teaching. New York: Longman; 1988.

24. Woolard G. Collocation-encouraging learner independence. In M. Lewis (Ed.), Teaching collocation: Further developments in the lexical approach. London: Language Teaching Publications. 2000;28-46.

25. Hung MS, Chin YH. A corpus-based comparison of the English tests of two major college entrance examinations in Taiwan. Journal of National Huwei University of Science & Technology. 2016; 33(2):87-109.

26. Lin KS. A study of Taiwanese EFL senior high students' collocation competence and collocation learning strategy use. Master's thesis, National Kaohsiung Normal University, Taiwan; 2012.

27. Huang CM. 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; 2009.