A Study on the UES of English Collocation in Writing by Students at Thai Nguyen University

Vi Thi Trng, M.A.[1], Tran Thi Thao, M.A.[2]

[1,2]College of Education, Thai Nguyen University Thai Nguyen, Vietnam

ABSTRACT: The current research investigates the use of collocations in students' academic writings to obtain information about the popular types of collocations they use, the collocational errors, and the sources of errors. The design of the study is a qualitative research which employed document analysis as the instrument to collect data. 50 students were the population and the samples as well. The results show that students have a tendency of using Type 1 (Verb-Noun) and Type 2 (Adjective-Noun) collocations more than the other types. With regard to the collocational errors, it is noted that Type 1 and Type 2 are also the top types in which students make mistakes. Additionally, verbs and adjectives are the main parts that students mostly have problems with. On examining the sources of errors, the researcher found five causes including approximation, the ignorance of rule restriction, negative transfer, the use of synonyms, and false concept hypothesized. Among these error sources, negative transfer is the most important factor leading to students' collocational errors.

KEYWORDS: Academic writing, collocation, collocation errors, negative transfer

I.   INTRODUCTION
1.1.   Statement of the research problem
Writing has always been a challenge for many English language learners. Vietnamese students of English language are not exception. According to the latest statistics on Vietnamese students' performance in the Academic IELTS test (International English Language Testing System) in 2019, the average writing score is just 5.6 out of 9.0, the lowest among four skills (Reading: 6.2, Listening: 6.1, Speaking: 5.7) (https://www.ielts.org). Students are facing a variety of obstacles in learning writing skill in general and academic writing in particular. Thus, teachers of English are put under a hard task of training students to achieve proficiency in academic writing.

Nunan (1989) claimed that writing is a complicated activity since it involves the cognitive ability requiring students to have control over various factors. Faddah (2011) added that because writing is a product of mind, it is a mental activity. Many researchers have carried out research to examine the factors that affect students' writing competence.

For undergraduate students, writing syllabuses mostly attempt to supply students with academic writing skill. The reason for this is that academic writing is essential for them to enter the academic world of research. It should be noted that academic writing is different from other kinds of writing such as personal writing or narrative writing in the single factor mentioned previously. Therefore, the criteria to mark students' academic essay are, of course, different too.

One of the criteria in marking any sorts of students' writing is lexical resources, or in other words vocabulary use. To be more specific, a piece of writing is evaluated in terms of vocabulary range, word choice, and collocation. Obviously, words are used in specific contexts; therefore, which words to use and when to use the words are of important consideration. What is more, words do not exist in isolation from other words, they usually co-occur with each other in a group or a chunk. In any languages, the vocabulary is comprised of single words and multi-word expressions. Collocations are among the multi-word expressions (Brashi, 2009). To put it another way, when a native speaker writes or speaks, there is a tendency of using fixed expressions which are unconsciously stored in his mind through the process of language acquisition. Collocations are defined as the lexical relationship which is said to be more arbitrary and arise from common usage than from rules (Benson et al., 1986). A learner of a foreign language in general and a learner of English as a foreign language in particular may find it really challenging to become a master of the target language because they can hardly create a collocation by themselves. In many cases, language learners are usually influenced by their mother tongues. Hence, they may try to invent a collocation by translating word by word from their own language. Unfortunately, these inventions are rarely equivalent to the intended meaning they are making effort to describe (Bahns & Eldaw, 1993; Hussein, 1990). It is undeniable that the more EFL/ESL students can utilize collocations in English, the closer they are to the advanced level of the language. As can be seen, learning collocations play a vital role in being a successful language learner. However, in some cases

students must be fully aware of the differences between collocations used in speaking or informal writing and those used in writing academic essays. While academic writing requires formal use of language, spoken language and informal writing are colloquial. Therefore, all the afore-mentioned reasons encourage the researchers to conduct a research paper entitled *"An investigation into the use of English collocations in writing essays by second year students at Thai Nguyen University"*

1.2.   Aims of the study

Firstly, the researchers aim at identifying the common types of lexical collocations used in students' essays. Common errors in students' use of lexical collocations are also of the researcher's interest. Additionally, the researcher desires to figure out the sources of collocational errors. Finally, it is essential to inform the teachers about the research results in order to raise their awareness of incorporating collocations in their teaching writing as well as take some countermeasures to prevent students' collocational errors. In brief, to achieve these aims above, this paper is conducted to answer the following research questions:

1.2.1.   *What are the common errors students make in using lexical collocations in writing an essay?*

1.2.2.   *What are the sources of those collocational errors?*

1.3.   Significance of the study

The researchers expect that the results of the study could provide some certain implications for English teachers, especially those who are teaching writing skills. By being informed about students' limited use in some certain types of lexical collocations, teachers can draw some plans to expand students' vocabulary bank. What is more, when teachers are conscious of the errors frequently made by students, they could propose some solutions to tackle the situation. Finally, this study also contributes a small part as a reliable source for researchers who share interest in the same topic.

II.   LITERATURE REVIEW

2.1.   Definition of collocations

The term "collocations" has been used widely by linguists since its first presence in 1957 by J. R. Firth who defined collocations as "an abstraction at an syntagmatic level" (Firth, 1957). From then on, most linguistic scholars share the convergent opinion that a collocation is the "co-occurrence of words" or word combination.

According to McCarthy & O'Dell (2005), "a collocation is a pair or group of words that are often used together" and "they are difficult to guess". It means that the choice of which word can go with another one is extremely arbitrary. It is, of course, very natural to native speakers who use the language effortlessly; however, this troublesome aspect of the language may trigger some tension for EFL students. For example, the two words "fast" and "quick" are semantically synonymous; therefore, it is a tendency for a student to use these two words interchangeably in many cases such as " a quick car" (instead of "a fast car"), or "a fast meal" (instead of "a quick meal").

Likewise, Palmer (1993, p.4) stated that collocations are strings of words that must be learned as "an integral whole or independent entity, rather than by the process of piecing together their component parts." In other words, due to the arbitrary nature of collocations, there is a high chance to make mistake when combining words merely by basing on their individual meanings.

Woodlard (2000) also reckoned that collocations refer to the likelihood of two or more words to appear together rather than the random chance. Lewis (2008) concurred that collocational phenomenon is observable when certain words keep company with other words in natural text with greater than random frequency. Obviously, both Woodlard (2000) and Lewis (2008) put an emphasis on the natural and unpredictable essence of collocations which is a definite obstacle for learners of English.

In short, throughout the consistent literature, it can be concluded that collocations are the chunks of words that frequently occur together in natural speeches and texts, and the selection of component words in a combination is unpredictable.

2.2.   Characteristics of collocations

Although there has been a consensus regarding the definition of collocations as mentioned above, one may get confused since it is rather ambiguous when comparing collocations with other linguistic phenomenon such as idioms (*better late than never, hit the sack)*, discourse markers (*to begin with, in addition*), and figurative expressions (*as happy as a clam, as busy as a bee*), which also demonstrate the feature of frequent or fixed co-occurrence. Thus, attempts have been made to figure out the characteristics of collocations to distinguish them from other similar categories of English language.

Sinclair (1991) stated that both spoken and written English are based on two principles: the Open Principle (OP) and the Idiom Principle (IP). Collocations lie between the two ends of that continuum. To be more specific, OP is closely linked to the Chomskyan theory of Universal Grammar which emphasizes the capability of constructing sentences in natural language in accordance with grammatical rules and semantic rules. Meanwhile, IP explains the use of socio-lexical conventional combinations. To illustrate, it is grammatically and semantically acceptable to say "crowded traffic"; however, they may sound weird to a native speaker because they are not natural partners. Instead, native speakers use "heavy traffic" with the same meaning. Thus, word combinations can vary from weak company (the closest to OP) to strong and fixed company (the most akin to IP). To put it another way, we can make a distinction on one hand between free combinations and collocations and on the other hand between collocations and idioms. Free combinations are typical of OP when language users make a phrase which is grammatically and semantically sanctioned. For

example, "buy a book" is a free combination because "buy" can be replaced by "sell" or "select", "book" can also be replaced by "ticket", "pen", etc. Thus, the meaning of the phrase actually depends on what the speaker means to convey. Idioms, on the other hand, are fixed expressions with frozen meanings. While some collocations share the same feature of fixedness, they are mostly understood by calculating the meanings of the component words together. It is said that idioms are a special case of collocations (Duan & Qin, 2012). However, idioms are not the subject of interest in this paper.

Throughout the literature, the following characteristics of collocations are agreed upon:

2.2.1. *Prefabrication*

The most notable characteristics of collocations agreed by many researchers is the prefabrication of a collocational phrase (Howarth, 1998a; Hill, 2000; Pawley & Syder, 1993 as cited in Seretan, 2011). According to Tode (2013), prefabrication has been usually understood in contrast to generation. While generation refers to the production of language based on rules which enable speakers to construct the language creatively, prefabricated language consists of meaningful units which are stored, memorized, and retrieved as wholes in performance (Tode, 2013). Simply put, prefabricated phrases or prefabs refer to the conventionally ready-made vocabulary stored in the brain of native speakers. Thus, collocations are word combinations that are natural and available for use, which promotes the fluency of speeches.

2.2.2. *Non or limited substitutability*

Stulpinaitė, Horbačauskienė & Kasperavičienė (2016) claimed that the meaning of a collocation could be lost if a word of a collocation is replaced with another word. This property demonstrates that the components of collocation can be substituted neither syntactically nor semantically. Nesselhauf (2005) also concurred that lexical substitution in the case of collocation is restricted. Kurosaki (2012) used another term with the same meaning as substitutability which is combinability. This property is applied even in the case when a substitute word is the same part of speech or has a closely similar meaning – a synonym, e.g., *strong tea* vs. *powerful tea*. Even though *strong* and *powerful* are synonymous, they cannot replace each other in this context.

2.2.3. *Non or limited modification*

This feature describes "the syntagmatic effect that there is only a limited number of collocations that can be modified because other lexical items cannot be inserted within a phrase" (Stulpinaitė, Horbačauskienė & Kasperavičienė, 2016). As in the example: kick the bucket vs. kick the large bucket, this idiomatic expression cannot be modified by adding an additional lexical item based on grammatical rules since the phrase itself loses its collocational meaning and is unnatural for native speakers.

2.2.4. *Transparency*

Transparency is generally deemed to mean whether the elements of the combination and the combination itself have a literal or a non-literal meaning (Nesselhauf, 2005). This property is usually taken as a criteria to distinguish between collations and idioms. As mentioned before, idioms are perceived as a special case of collocations where the meaning of an idiom cannot be conveyed through its constituents. That is to say, the meaning of idioms are opaque while the meaning of collocations can be more transparent. Take the idiom "under the weather" and the collocation "make a decision" as a case in point, it is evident that the meaning of "under the weather" cannot be counted on the component words, but the meaning of "make a decision" is clear to a non-native speaker.

2.3.   Classifications of collocations

Based on the aforementioned approaches to collocations, there are different terms and different ways to categorize collocations. However, it can be concluded that the classifications of collocations fall into three main streams: Degree of fixedness; Upward and Downward collocations; Grammatical and Lexical collocations.

2.3.1. *Degree of fixedness*

Many linguistics scholars chose degree of fixedness as a criteria to classify collocations. There are, however, varying terms created to name different kinds of collocations.

Lewis (2000) classified collocations as follows:

Strong collocations: are collocations that have very limited number of collocates and most collocates are fixed such as "rancid butter" or "rancid oil".

1. Weak collocations: refers to collocations which have a wide range of collocates, for example, many things can be described as "good" or "bad".
2. Medium-strength collocations: words that go together more frequently than weak collocations such as "hold a meeting", "carry out a study".

Hill (2000) also provided a similar category of collocations:

1. Unique collocations: refer to collocations which are fixed and irreplaceable by any other words such as "to foot the bill".
2. Strong collocations: these collocations are strong or very strong but not unique. To be more specific, strong collocations have few other possible collocates. Some examples of this are "moved to tears" or "reduced to tears".
3. Weak collocations: consists of word combinations which can be easily guessed such as "a white shirt", "a red shirt", or "a green shirt".
4. Medium-strength collocations: these collocations are of the same meaning as suggested by Lewis (2000).

As can be seen, these two ways of classifying collocations are similar in that they both examine collocations on a continual spectrum ranging from weak to strong and fixed. However, Hill made a distinction between strong and unique collocations while Lewis included the term unique collocations into the definition of strong collocations.

Another categorization is offered by Mahmoud (2005) in which there are only two types of collocations: Open and restricted collocations. Open collocations refers to words that can cluster with a variety of other words. This definition suggests a similarity with the category of weak collocations provided by Lewis (2000) and Hill (2000). The other one, restricted collocations, shares the same characteristics with strong and unique collocations in that it refers to fixed word combinations or like idioms such as "rain cats and dogs". Thus, the limitation of this classification is it only sees collocations as the two extremes of a continual spectrum, but leaves some others which are in between behind.

2.3.2. *Upward and downward collocations*

Sinclair (1991) also offered a different classification of collocations: the upward collocations and the downward ones. The former consists of words which often collocate with the other words more frequently used in English than they are themselves. Most of them are prepositions, adverbs, conjunctions and pronouns. Take the word "back" as an example, "back" can collocate with at, down, from, into, on, all of which are more frequent words than "back". The latter includes words which habitually collocate with other words which are less frequent than they are themselves. For example, the words "arrive", "bring" can combine with "back", but they are less frequent in the corpus than "back". By categorizing collocations into upward and downward groups, Sinclair (1991) figured out a systematic difference between them: "Upward collocations, of course, is the weaker pattern in statistical term, and the words tend to be elements of grammatical frames, or super-ordinates. Downward collocations by contrast give us a semantic analysis of a word." While the notion of upward and downward collocations would definitely contribute a large part in corpus-based linguistics, it is somewhat challenging in the teaching context because it is hard to know which words are more or less frequent than other words.

2.3.3. *Grammatical and lexical collocations*

According to Benson, Benson and Ilson (1997) as cited in Begagic (2014), collocations can be of two types: grammatical and lexical collocations. Grammatical collocations are clusters in which a content word such as a noun, a verb, and an adjective frequently co-occurs with a grammatical item, usually a preposition. Grammatical collocations basically consist of eight types as presented below in.

III. RESEARCH METHODOLOGY

3.1. Research design: qualitative research design

Being aware of the aims and characteristics of the current research, the researcher employed qualitative research design with content analysis of textual materials as the primary tool. According to Crossman (2019), qualitative research is a type of social science research that gathers non-numerical data, which targets interpreting meaning from these data to understand a social phenomenon. It allows the researchers to investigate the meanings people attribute to their behaviors, actions, and interactions with others. Furthermore, as opposed to quantitative research which focuses on macro level trends and phenomena, qualitative research mainly focuses on micro level ones (Crossman, 2019).

The subject for investigation in the study is the use of lexical collocations in written materials; therefore, the data are not numerical. To identify the types of lexical collocations, the common errors, and the appropriateness of these collocations with respect to academic register, the researcher needs to adopt a theoretical framework which helps to decode the meaning from the students' behaviors manifested in the data. Thus, there is no more suitable method than qualitative one. Another thing is within the time and human resource limit, the researcher could not carry out the research on a large scale. Hence, the qualitative research design is selected.

3.2. Data collection instrument

In order to collect the data, the researcher used writing samples as the main instrument. It is a suitable instrument in this research because it tailored to the aim of the researcher which is to examine the use of collocations in academic writing.

3.3. Data collection method and procedure

In the first phase, the researcher selected two writing topics for opinion essay which is one of the academic genres the students had learnt. After that, on the due date of final examination all the students were required to choose one topic and finish the writing within 60 minutes. No cell phones and documents were allowed to be brought into the room. In the third phase, the researcher collected all the essays and scanned them to store the documents in the personal computer. The underlying reason for making the soft copies of the exam papers was that the hard copies must be submitted to the department as the university regulations indicate.

3.4. Data analysis method and procedure

The researcher employed document analysis as the sole method of the study. Since the raw data show no numerical information, to make sense of the data the researcher developed coding categories. With regards to the first research question about the types of lexical collocations used in the essays, the initial step was finding out all the lexical collocations. This was followed by placing

these lexical collocations into the appropriate coding categories which are based on the classification of Benson, Benson & Ilson (1986).

IV. RESULTS AND DISCUSSION

4.1. *Research question 1: What are the common errors students make in using lexical collocations in writing an essay?*

On analysing students' essays, the researcher has calculated 187 collocations of 6 types in total. Of all the types, Verb- Noun collocations accounted for 40.6%, therefore, ranked first in the list. Adjective-Noun was also preferred by students with 36.9%, which is quite similar to the former. Meanwhile, Verb- Adverb collocations contributed nearly three times less than the percentage of Verb- Noun, only 13.5%. The other three types of collocations were less selected in the writing samples, especially both Noun- Verb and Quantity-Noun collocations were used the least with just above two percent.

Table 1. Frequency of use in the writing samples

No.	Types of collocations	Frequency	Percent	Total
1.	Verb + Noun	76	40.6	
2.	Adjective + Noun	69	36.9	
3.	Noun + Verb	4	2.1	
4.	Quantity + Noun	4	2.1	**187**
5.	Adverb + Adjective	9	4.8	
6.	Verb + Adverb	25	13.5	

It is also noticeable that almost all writing samples share the same pattern of collocational frequency which is presented in Table 6, as can be seen from Table 7. Within each essay, Verb-Noun and Adjective-Noun were most used, except for essay No. 31 and No. 47 where Type 6 (Verb- Adverb) was utilized more than Type 1 and Type 2. Moreover, attention must be paid to the enormous disparity of the total number of collocations among 50 essays. While some students were well aware of using collocations such as No.1, No.2, and No.10 with 17, 13, and 14 collocations used respectively, many other students merely expressed their ideas in long sentences with few collocational expressions. For example, student No. 20 and No. 37 only had 2 collocations in their writings. Additionally, it is of great consideration that compared to the length of an essay which is about 250 words, generally the collocations found in each piece of writing are quite small in number. Specifically, 17 is the highest number of collocational expressions identified in one essay.

4.2. *Research question 2: what are the sources of those collocational errors?*

In order to understand why students made such mistakes in writing, the erroneous collocations were categorized in light of the classification provided by Liu (1999b). The incorrect collocations could be attributed to: False concept hypothesized, the use of synonym, ignorance of rule restrictions, negative transfer, and approximation. Negative Transfer, the proportion of which is about 49% of the total errors. As can be seen, the majority of negative transfer errors pertain to Type 1 (19 out of 27 errors). The other Types contributing to this category include Type 2 (5 out of 27 errors) and Type 3 (2 out of 27 errors). It is evident that the mother tongue has a tremendous impact on the production of L2 language. Students tend to invent a collocation based on their knowledge of L1 language to compensate for their lack of collocational knowledge. The following table shows some examples of collocational errors caused by Negative Transfer. False concept hypothesized errors are caused by students' inability to differentiate the meaning of de-lexical verbs such as "do" and "make", "have" and "get", "take" and "bring". These words may confuse students because they carry little meaning of their own when they are used with particular nouns. In fact, the meanings of these structures are mostly connected to the nouns, not the verbs. Delexical structures are very common in English, and the mastery of these will ensure the students' fluency of the language. Table 7 shows that the majority of false concept hypothesized errors belong to Type 1 (Verb-Noun) with 7 errors. As mentioned before, students encounter problems with some similar meaning verbs; therefore, it is comprehensible why Type 1 is the most troublesome for them.

V. CONCLUSION

With respect to the second research question "*What are the common errors students make in using lexical collocations in writing an essay?*" the data show that among 187 collocations found in the sample writings, 51 collocations were incorrectly used. The largest number of erroneous collocations arose from Type 1 (Verb-Noun), followed by Type 2 (Adjective-Noun). In fact, Type 1 and Type 2 are the most common types of collocations students used in their writings; therefore, they stand more chance of committing errors related to these two categories. What is more, although Noun- Verb collocations were not a popular type among students' essays, almost all of the collocations in this type were problematic. In other words, the ratio between the frequency of Noun-Verb collocations and its number of errors take the first place in comparison to the other types. It means students have the most difficulty in utilizing this type of collocations.

A Study on the UES of English Collocation in Writing by Students at Thai Nguyen University

In terms of the second research question *"What are the sources of those collocational errors?"*, the researcher investigated the erroneous collocations in light of the classification provided by Liu (1999b). The results show that students commit collocational errors due to five reasons which are false concept hypothesized, the use of synonym, ignorance of rule restrictions, negative transfer, and approximation. Specifically, negative transfer is the leading factor causing collocational errors with nearly half of the total errors originated from this source. The other three sources which are false concept hypothesized, the use of synonym, and approximation share quite the same proportion of collocational errors, but these proportions are much fewer than that of negative transfer. Most notably, students make mistakes when using many synonyms of "big" to collocate with different nouns, for example, "huge" and "vast" are two words that students incorrectly combined with some particular nouns. The last source of collocational errors is the ignorance of rule restriction, which contributed only a small number of mistakes

REFERENCES
1) Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1), 101-114.
2) Begagic, M. (2014). English language students' productive and receptive knowledge of collocations. *Explorations in English Language and Linguistics*, *2* (1), pp.46-67.
3) Benson, M., Benson, E., & Ilson, R. (1986). *The BBI combinatory dictionary of English*. Amsterdam and Philadelphia: John Benjamins.
4) Brashi, A. (2009). *Collocability as a problem in L2 production*. Reflections on English Language Teaching, 8 (*1*), pp.21-34.
5) Crossman, A. (2019). *An Overview of Qualitative Research Methods.* Retrieved 29th October 2019 from https://www.thoughtco.com/qualitative-research-methods-3026555
6) Duan, M. & Qin, X. (2012). Collocation in English Teaching and Learning. *Theory and Practice in Language Studies*, *2*(9), pp. 1890-1894. DOI: 10.4304/tpls.2.9.1890-1894.
7) Faddah, H. (2011). Difficulties in Academic Writing: From the Perspective of King Saud University Postgraduate Students. *English Language Teaching*, *5*(3), pp.123-130.
8) Firth, J. (1957). Modes of meaning. In J. Firth (Ed.), *Papers in linguistics.* Oxford: Oxford University Press.
9) Hill, J. (1999). Collocational competence. *English Teaching Professional*, *11*, pp. 3-6.
10) Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further development in the lexical approach* (pp.47-69). Oxford: Oxford University Press.
11) Hussein, R. (1990). Collocations: The missing link in vocabulary acquisition amongst English foreign learners. In J. Fisiak (ed.), *Papers and studies in contrastive linguistics*. The Polish–English contrastive project (Vol. 26, pp. 123-136). Pozan: Adam Mickiewicz University.
12) Lewis, M. (2008). *Teaching collections-Further development in the lexical approach*. New York: Heinle.
13) Lewis, M. (Ed.). (2000). *Teaching collocation: Further development in the lexical approach*. Oxford: Oxford University Press.
14) McCarthy, M. & O'Dell, F. (2005). *English Collocations in Use Intermediate*. Cambridge: Cambridge University press.
15) Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Armsterdam: J. Benjamins Publishing Company.
16) Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press, United Kingdom.
17) Palmer, H. E. (1993). *Second interim report on English collocations*. Tokyo: Kaitakusha.
18) Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
19) Stulpinaitė, M., Horbačauskienė, J. & Kasperavičienė, R. (2016). Issues in Translation of Linguistic Collocations. *Studies about languages*, *29*, pp. 4-37.
20) Tode, T. (2013). Chunking and prefabrication. The Routledge Encyclopedia of Second Language Acquisition, Peter Robinson (ed). New York: Routledge.