



A Corpus-Based Study on Vietnamese EFL Students' Use of English Lexical Collocations in Academic Writing

Hong Thi Tuyet Vo
Can Tho University

This descriptive corpus-based study aimed to investigate EFL students' use of English lexical collocations in their academic writing. The study included 120 freshmen majoring in English Studies at a state university in the southern region of Vietnam. Corpus-based tools, specifically Sketch Engine and the Corpus of Contemporary American English, were primarily employed to categorize and assess the appropriateness of collocations expressed in the students' essays. Additionally, in order to delve deeper into research-related matters, a Likert scale survey embedded with two open-ended questions was administered to the participants. The findings obtained from both quantitative and qualitative analyses of the collected data indicated that the students most frequently utilized adjective + noun collocations in their academic essays. However, the types of collocations that seemed to pose the most challenges for their usage were noun + noun, adjective + noun, and verb + adverb collocations. This was largely attributed to their limited vocabulary size, lack of collocational knowledge and learning strategies, and L1 negative transfer, which resulted in their misuse and incorrect selection of English collocations. Based on the research outcomes, important pedagogical implications have been further discussed for effective teaching and learning of English collocations to benefit EFL students' academic writing.

Nghiên cứu mô tả dựa trên ngữ liệu này nhằm khảo sát việc sử dụng cụm từ kết hợp tiếng Anh trong bài viết học thuật của sinh viên học tiếng Anh như một ngoại ngữ. Đối tượng của nghiên cứu gồm 120 sinh viên năm nhất chuyên ngành Ngôn ngữ Anh tại một trường đại học công lập ở khu vực phía Nam, Việt Nam. Công cụ phân tích dữ liệu bao gồm Sketch Engine và Corpus of Contemporary American English, được sử dụng chủ yếu để phân loại và đánh giá mức độ phù hợp của các kết hợp từ được biểu đạt trong bài luận của sinh viên. Ngoài ra, để nghiên cứu sâu hơn các yếu tố có liên quan, một khảo sát dựa trên thang đo Likert kèm với hai câu hỏi mở đã được thực hiện với đối tượng nghiên cứu. Kết quả thu được từ việc phân tích định lượng và định tính của dữ liệu cho thấy sinh viên sử dụng các kết hợp tính từ + danh từ thường xuyên nhất trong bài luận của họ. Tuy nhiên, loại cụm từ kết hợp dường như gây nhiều khó khăn nhất cho sinh viên là danh từ + danh từ, tính từ + danh từ, và động từ + trạng từ. Nguyên nhân chính của việc này được cho là do vốn từ vựng của sinh viên còn hạn chế, thiếu kiến thức và chiến lược học kết hợp từ, và ảnh hưởng của tiếng mẹ đẻ, dẫn đến việc sử dụng sai và chọn lựa không đúng các kết hợp từ tiếng Anh. Dựa trên kết quả nghiên cứu, một số đề xuất ứng dụng sư phạm quan trọng đã được thảo luận thêm để việc dạy và học các cụm từ kết hợp ngày càng có hiệu quả hơn góp phần nâng cao năng lực viết học thuật cho sinh viên học tiếng Anh như một ngoại ngữ.

Keywords: Lexical collocation, academic writing, corpus-based study, Vietnamese EFL students

Introduction

Recently, significant attention has been devoted to the study of collocations in both ESL and EFL contexts. This heightened focus arises from the paramount importance of collocations in showcasing learners' language proficiency and enriching their lexical resources, key to attaining native-like language



output (Dokchandra, 2019; McCarthy & O'Dell, 2008; Men, 2018; Saito, 2020; Wood, 2002). Despite the increasing globalization and demand for English proficiency, many English language learners, including Vietnamese EFL students, inevitably face substantial challenges in achieving native-like fluency and precision in their academic writing, particularly in the mastery of lexical collocations (Demir, 2017; Duong & Nguyen, 2021; Granger & Bestgen, 2014; Kuo, 2009; Mongkolchai, 2008; Trng & Thao, 2021). Lexical collocations refer to the co-occurrence of two or more content words, namely verbs, nouns, adjectives, and adverbs without incorporating any prepositions, infinitives, or relative clauses such as *bright future*, *data analysis*, or *incredibly difficult* (Benson et al., 1997, 2010; McCarthy & O'Dell, 2005; McIntosh et al., 2009). They play a pivotal role in shaping the quality, fluency, and naturalness of the English language learners' written discourse whereas any inappropriate use or wrong choices of a certain collocational pattern may impede comprehension, affect the overall coherence of written texts, and hinder academic success (Begagić, 2014; Boonraksa & Naisena, 2022; Granger & Bestgen, 2014; Shitu, 2015; Thongvitit & Thumawongsa, 2017). Hence, meticulous investigations into the patterns and tendencies of lexical collocation usage by English language learners in general and by Vietnamese EFL students in particular would have profound implications for language teaching and curriculum development.

Indeed, a multitude of studies have been conducted so far, concentrating on examining students' difficulties with their utilization of English collocations. Most of these explorations have yielded relatively similar outcomes, which revealed that both ESL and EFL learners commonly encountered great challenges in employing lexical and grammatical collocations (Boonraksa & Naisena, 2022; Mahmoud, 2005; Namvar, 2012; Shitu, 2015; Thongvitit & Thumawongsa, 2017; Umair et al., 2023). As indicated by a set of prior studies, students' problems in utilizing English collocations mostly stemmed from negative transfer of their native languages (Begagić, 2014; Boonraksa & Naisena, 2022; Fan, 2009; Farghal & Al-Hamly, 2007; Phoocharoensil, 2011, 2014; Yamashita & Jiang, 2010). Another predominant cause of students' incorrect use or selection of English collocations was their lack of proper collocational knowledge (Boonraksa & Naisena, 2022; Dokchandra, 2019; Duong & Nguyen, 2021; Fan, 2009; Namvar, 2012; Shitu, 2015; Yumanee & Phoocharoensil, 2013). Furthermore, a collection of previous research's findings confirmed that the overuse of synonyms was one of the underlying factors contributing to students' inappropriate use of English collocations (Boonyasquan, 2009; Kuo, 2009; Mongkolchai, 2008; Phoocharoensil, 2011). Apart from these fundamental causes, there were other elements affecting students' use of English collocations, including limited vocabulary size, paraphrasing strategies, low language proficiency, lack of explicit instruction, fossilization, overgeneralization, pedagogy, cultural differences, and learning materials (Ang et al., 2011; Begagić, 2014; Boonraksa & Naisena, 2022; Khoram & Mirsalari, 2020; Namvar, 2012; Tanihardjo, 2017).

In spite of numerous investigations into various aspects of English collocations, several gaps still exist in a series of previous studies. One significant limitation is that conventional explorations heavily relied on multiple-choice and gap-fill tests (Begagić, 2014; Dokchandra, 2019; Mongkolchai, 2008; Namvar, 2012; Yumanee & Phoocharoensil, 2013), which might not have accurately represented the students' real performance in collocations. In contrast, recent studies have adopted corpora and concordance tools; nevertheless, few have concentrated on specific lexical collocation categories in EFL learners' academic essays, particularly with a dearth of corpus-based research in the Vietnamese context. Another shortcoming of previous examinations is their failure to clearly identify the types of collocations that EFL university students with an intermediate level of English proficiency are capable of using in their academic writing, as well as which lexical collocation patterns would pose the most challenges for the learners' practical usage. One more gap in a body of prior research is their lack of explicit identification of the determining factors contributing to the incorrect use of English collocations by intermediate EFL students in their academic compositions. Moreover, those investigations did not pinpoint how intermediate EFL students could enhance their collocational competence to utilize English collocations more effectively in their compositions. Accordingly, the present research endeavors to bridge these gaps to gain deeper insights into how Vietnamese intermediate EFL students actually incorporate English lexical collocations into their

academic writing. To accomplish this, the study will focus on analyzing the students' written work using corpus-based tools, coupled with the use of a questionnaire to address the following research questions:

1. What kinds of English lexical collocations do Vietnamese intermediate EFL students frequently use in their academic essays?
2. What difficulties do these students have when using English lexical collocations in their academic essays?
3. What are the causes of the students' difficulties in utilizing English lexical collocations in their academic essays?

Literature Review

Definition and Classification of English Collocations

In English and various other languages, there exists a fundamental linguistic construct where combinations of words frequently occur together due to inherent syntactic, semantic, and pragmatic relationships among them (Shitu, 2015). This prominent linguistic phenomenon was initially coined as collocation by Firth (1957), signifying the co-occurrence of words through abstraction. Later, a more comprehensive definition of collocation has been largely agreed upon by a majority of researchers and scholars (Benson et al., 1997, 2010; Gao et al., 2019; Lewis, 2000; Mayor, 2009; McCarthy & O'Dell, 2005; McIntosh et al., 2009; Tanihardjo, 2017), regarding collocations as pairs or groups of words that frequently co-occur in a manner that sounds appropriate and natural in the target language (e.g., bitter cold, job interview, actively participate, incredibly important). Remarkably, Sinclair (1991) and Gries (2003) defined collocation as a recurrent combination of words that appear together more often than expected by chance, particularly concerning their individual frequencies within a specific corpus. In the context of this research, the two essential dimensions of a collocation would be considered, which encompasses its suitability in lexico-grammatical forms with semantic transparency and its prevalence in the communicative practice of the target language, grounded in a certain corpus.

When it comes to categorizing English collocations, there have been a wide array of collocational terms derived from the perspectives of various linguistic scholars for classification. These include strong-weak collocations (Hill, 2000; Lewis, 1997, 2000; McCarthy & O'Dell, 2008, 2017), open-restricted collocations (Mahmoud, 2005), and fixed collocations or idiomatic expressions (McCarthy & O'Dell, 2008, 2017; Wei, 1999). Although these categorizations provided valuable insights into multiple facets of collocations, their shortcomings lie in the fact that collocations were sorted at the two extremes of a continuous spectrum, leaving those in between overlooked. Therefore, from the lexico-grammatical and phraseological perspective, a more concrete and comprehensive model for classifying collocations emerged (Hoey, 2005; Howarth, 1996; Sinclair, 1991; Stubbs, 2007) and has been widely adapted in diverse studies (Demir, 2017; Dokchandra, 2019; Kuo, 2009; Shitu, 2015; Thongvitit & Thumawongsa, 2017; Trng & Thao, 2021). The model specifically developed by Benson et al. (1997, 2010) systematically grouped collocations into two primary categories comprising grammatical collocations and lexical collocations. The former involves a combination of a significant content word (i.e., verb, noun, adjective) and a preposition, infinitive or clause. On the contrary, the latter typically contains two or more content words without combining with any prepositions, infinitives, or relative clauses. Within this model, English lexical collocations are further divided into seven subtypes.

- adjective + noun (e.g., bright future, strong coffee)
- adverb + adjective (e.g., extremely difficult, remarkably successful)
- adverb + verb (e.g., carefully consider, swiftly respond)
- noun + noun (e.g., traffic jam, data analysis)

- noun + verb (e.g., time flies, research indicates)
- verb + adverb (e.g., investigate thoroughly, communicate effectively)
- verb + noun (e.g., provide evidence, implement strategies)

Importance and Challenges of Using English Collocations

In general, there is a scholarly consensus on three prominent roles of collocations. Firstly, since collocational knowledge is fundamental, encompassing nearly three-fourths of language input (Hill, 2000), mastering collocations contributes substantially to learners' vocabulary enrichment by promoting their acquisition of word pairings frequently occurring in native discourses (Nattinger & DeCarrico, 1992). In essence, learners who grasp the intricacies of collocations not only expand their lexicon but also develop a more intuitive understanding of how words interact in authentic usage (Begagić, 2014; Dokchandra, 2019; Duong & Nguyen, 2021; McCarthy & O'Dell, 2005; Siengsanoh, 2021). The second crucial role of collocations lies in facilitating natural and accurate language use, essential for native-like production (Dokchandra, 2019; McCarthy & O'Dell, 2008; Men, 2018; Namvar, 2012). In other words, a proper use of collocations assists students in steering clear of lexical errors and selecting appropriate words congruent with the intended register, which promotes their expertise in academic writing, adhering to standardized examination criteria, namely IELTS (Duong & Nguyen, 2021; Saito, 2020; Wood, 2002). Lastly, Sinclair (1991) claimed that proficiency in collocations enhances learners' cultural and pragmatic competence, enabling them to navigate varied communicative settings with greater ease and accuracy.

Despite the critical roles of collocations, language learners in general and English-major students in particular, often face considerable challenges in utilizing collocations, clearly evident in their academic writing (Farghal & Al-Hamly, 2007; Granger & Bestgen, 2014; Kuo, 2009; Mongkolchai, 2008; Phoocharoensil, 2011, 2014; Shitu, 2015). Up to now, a host of studies have been carried out in both ESL and EFL contexts, focusing on students' problems with their use of English collocations. Although the majority of previous investigations into students' utilization of English lexical and grammatical collocations adopted a similar classification model outlined by Benson et al. (1997, 2010) and Lewis (2000), their specific findings exhibited a certain degree of diversity. More precisely, Thongvitt and Thumawongsa (2017) discovered that advanced Thai EFL learners predominantly employed noun + preposition and adjective + noun collocations, yet they encountered the most difficulties with noun + preposition and verb + noun collocations spanning both grammatical and lexical categories. Mahmoud's (2005) research revealed that verb + noun and noun + verb collocations manifested the most errors in the Arabian EFL students' essays. Shitu (2015) found the highest number of verb + noun collocation errors, followed by noun + noun and adjective + noun patterns, with noun + verb collocations occurring the least in the essays of advanced Nigerian ESL learners. This finding somewhat aligns with other studies (Kuo, 2009; Umair et al., 2023), which demonstrated that Taiwanese EFL students and Pakistani EFL undergraduate students made more mistakes with verb + noun collocations than with adjective + noun category. On the contrary, Namvar (2012) identified adjective + noun and noun + verb collocations as the most difficult patterns for Iranian postgraduate students, while verb + noun collocations were the easiest for them. Phoocharoensil's (2011) study illustrated that Thai EFL low-proficiency learners had the most problems with verb + noun collocations, followed by noun + verb and adjective + noun ones, with verb + adverb collocations being the least troublesome; conversely, Boonraksa and Naisena's (2022) investigation showed that adverb + adjective collocational errors were the most frequent among Thai EFL students of all levels.

Causes behind Students' Difficulties with English Collocations

As reported in a number of previous studies, students' difficulties in utilizing English collocations mostly originated from two major causes, including internal factors (i.e., language transfer, collocational knowledge, learners' cognitive and metacognitive abilities, overgeneralization, language proficiency), and external influences (i.e., pedagogy, materials, teaching and learning strategies). With respect to the external

factors, instructional approaches, learning materials, pedagogical methods, and learning tactics greatly contribute to students' struggles with English collocations (Begagić, 2014; Boonraksa & Naisena, 2022; Yumanee & Phoocharoensil, 2013). In particular, Begagić (2014) who examined the receptive and productive knowledge of lexical collocations among the first and fourth-year English-majored students in Zenica attributed the students' poor performance to their L1 negative transfer and a lack of explicit instruction of collocations in classrooms. Kuo (2009) and Tanihardjo (2017) further supported the significant impact of various strategies, specifically approximation and literal translation, on the misuse of collocations among Taiwanese EFL learners and Indonesian EFL students. Meanwhile, Phoocharoensil (2011, 2014), and Boonraksa and Naisena (2022) emphasized learning tactics, namely synonyms and transliteration as contributors to learners' inappropriate use of English collocations.

Internally, language transfer or L1 negative transfer, has been recognized as a leading cause inhibiting learners from using English collocations appropriately (Begagić, 2014; Boonraksa & Naisena, 2022; Duong & Nguyen, 2021; Phoocharoensil, 2011; Shitu, 2015; Yumanee & Phoocharoensil, 2013). Specifically, Fan (2009) found that Hong Kong ESL students' misuse of English collocations was adversely influenced by their mother language transfer. Likewise, Yamashita and Jiang (2010) corroborated that Japanese learners' acquisition of English collocations was affected by their native language interference. Aside from language transfer, a lack of collocational knowledge could hinder learners from using English collocations correctly (Boonraksa & Naisena, 2022; Fan, 2009; Thongvitit & Thumawongsa, 2017). Trng and Thao (2021) noted the challenge for L2 learners in mastering collocations due to the intrinsic characteristics of the target language's collocations. On top of that, Duong and Nguyen (2021) accentuated students' collocational knowledge, vocabulary size, and low language proficiency as underlying causes of their collocational problems. While Namvar's (2012) research stressed the strong correlation between students' collocational knowledge and their overall proficiency, Khoram and Mirsalari (2020) underlined the significance of collocational knowledge and suggested more attention to learning materials and instructional modes. These findings are in accord with Dokchandra's (2019) research outcomes which confirmed that lacking collocational knowledge was a widespread problem triggering the learners' failure in acquiring and producing the target language.

In brief, numerous prior investigations have indicated that the primary reasons behind students' misuse of English collocations result from a lack of collocational knowledge and L1 negative transfer, alongside various additional factors. These include a limited vocabulary size, low language proficiency, reliance on synonym strategies, overgeneralization, instructional practices, and the quality of learning materials. Therefore, comprehending the elements that govern and influence students' use of English collocations is of utmost importance as it enables careful consideration of necessary measures to address any associated problems, thereby maximizing students' opportunities to enhance their collocational competence.

Corpus-Based Studies on English Collocations

Over the past two decades, corpus-based studies have emerged as a pivotal methodological approach to investigating specific aspects of students' utilization of English multi-word sequences. Corpus-based studies offer a robust framework for delving into the intricacies of how students authentically use recurrent word combinations, fostering a deeper understanding of collocational patterns in diverse contexts (Choi, 2019; Gablasova et al., 2017; González Fernández & Schmitt, 2015; Granger & Bestgen, 2014; Ishikawa, 2024; Paquot & Granger, 2012).

Following Stubbs (2007) who advocated for applying concordance analysis to identify recurrent word patterns and their co-occurrences, providing insights into collocation frequency and distribution, recent researchers have employed various tools to deepen into different facets of English word combinations. Granger and Bestgen (2014) deployed the Constituent Likelihood Automatic Word-tagging System (CLAWS) 7, associated with the referencing corpus, British National Corpus (BNC), to compare the use of collocations in academic essays written by intermediate and advanced learners of English from 3 different mother tongue backgrounds (i.e., French, German, and Spanish). The results uncovered that the

intermediate learners tended to overuse high frequency collocations while utilizing a smaller proportion of lower frequency but strongly associated collocations within the five categories of word pairs (i.e., premodifier noun sequences, noun + noun, adjective + noun, adverb + adjective, and all bigrams). Kim and Lee (2021) applied AntConc 3.5.7, the corpus analysis toolkit, to explore the structural and functional characteristics of lexical bundles or multi-word sequences in medical journal abstracts performed by native speakers of English and Korean non-native speakers of English. The findings revealed that Korean writers of English relied heavily on verb phrase-based and text-oriented bundles, whereas native English speakers prominently employed noun phrase-based and research-oriented bundles. Likewise, Hosseinpur et al. (2023) implemented the latest software AntConc 3.5.9 to detect frequent 4-word lexical bundles used by L1 and L2 authors in economic research articles. The results highlighted similar features in the structural and functional use of lexical bundles by both groups of writers with prepositional phrase bundles more frequently occurring than verb phrase and dependent clause bundles. In a similar manner, nonetheless, Yuliawati et al. (2021) extracted lexical bundles from English and Indonesian research articles on literature and linguistics using Sketch Engine (Kilgarriff et al., 2014). The outcomes demonstrated that prepositional-based bundles were predominant in the English corpus, followed by noun-based bundles. In contrast, the Indonesian corpus was primarily composed of noun-based bundles, with verb-based multi-word sequences being the second most common.

By employing the concordance program WordSmith Tools 3.0 to extract collocations in Malaysian ESL learners' essays, along with referencing the Oxford Collocations Dictionary (2009) and the BNC, Ang et al. (2011) found that among all seven types of collocational errors adapted from Nesselhauf's (2003) framework, preposition-related collocations were the most prominent, which was mainly caused by the learners' ignorance of rule restrictions and their L1 transliteration. Thongvitit and Thumawongsa (2017) relying on AntConc 3.4.3 and TagAnt 1.2.0, coupled with the Oxford Collocations Dictionary for Students of English (McIntosh et al., 2009) discovered that noun + preposition and adjective + noun pairings were the most frequently employed grammatical and lexical collocations in the research article abstracts written by Thai advanced EFL writers. However, noun + preposition and verb + noun combinations were identified as the most commonly misused collocations among the writers. Trinant and Yodkamlue (2019) also leaned to the AntConc to extract keywords, and the BNC as a reference corpus to explore lexical collocations used in Suranaree nursing research articles. The findings illustrated that the dominant type of collocation was noun + noun, followed by adjective + noun and noun + verb collocations.

In short, corpus-based studies on students' use of English collocations form a critical component of the literature through the employment of sophisticated tools, and rigorous analytical frameworks to unravel the complexities of collocational usage patterns. Nevertheless, several constraints persist in the previously reviewed studies, underscoring the need for more targeted research in the domain of lexical collocations. First and foremost, traditional investigations primarily depended on multiple-choice and fill-in-the-blank tests (Begagić, 2014; Dokchandra, 2019; Mongkolchai, 2008; Namvar, 2012; Yumanee & Phoocharoensil, 2013), potentially failing to accurately reflect the learners' actual performance of collocations in academic writing. Conversely, while most recent studies have utilized corpora and concordance tools to strengthen research outcomes, few have paid attention to distinct categories of lexical collocations within L2 learners' compositions with a notable absence of corpus-based research in the Vietnamese context. Moreover, most of the previous corpus-based studies in various EFL settings have focused on examining the use of lexical bundles or collocations in advanced groups, either with or without a comparison to native speakers of English while the lower-level group, namely intermediate EFL learners, appears to have been underresearched. Another gap arises from the fact that the investigation of which types of English collocations remain challenging for intermediate EFL learners, what factors influence their usage of collocations in academic writing, and how to enhance their collocational competence is still lacking. Thus, conducting a comprehensive analysis to address such issues will provide valuable insights to inform educators and curriculum developers in designing more effective instructional strategies aimed at promoting the collocational competence of intermediate EFL learners, ultimately contributing to their linguistic proficiency and academic success. In an attempt to bridge these gaps, the present research

examines how Vietnamese EFL students, at an intermediate level of English proficiency, incorporate English lexical collocations into their academic essays. This investigation is grounded in the lexical collocation framework proposed by Benson et al. (2010) and utilizes corpus tools, particularly Sketch Engine and the Corpus of Contemporary American English (COCA), to validate and triangulate the results.

Method

Setting and Sampling

This descriptive corpus-based study was conducted with 120 consenting English-majored freshmen from four distinct writing classes, at a state university in the southern region of Vietnam during the academic year 2022-2023. All the EFL participants, comprising 92 females and 28 males, were aged between 18 and 20 years at the time of data collection. The research population was selected through the convenient and purposive sampling method, encompassing all freshmen enrolled in the four-year English Studies program. All of them had been formally studying English for over seven years since their secondary school, including approximately one academic year at the university level. Based on their first-year learning outcomes at university, especially with their IELTS writing scores ranging around 4.0-5.0, this population strongly supports the scope and objectives of the current study, which focuses on intermediate EFL students attending *Writing 2* course for English academic essays. The textbook utilized in their writing course was *Mindset for IELTS Level 1* (Crosthwaite et al., 2017), which is appropriate for intermediate EFL learners.

Research Instruments

In order to answer the first and second research questions, all the final exam essays of the 120 participants were collected with their consent. Regarding the final exam, all the students of the Writing 2 course were required to complete two writing tasks within 60 minutes, following the format similar to that of the IELTS writing test. This paper-based exam took place in the classroom with all forms of documents and phone use being strictly prohibited. In light of the scope of this study, only the participants' argumentative essays of the IELTS academic writing task 2 were referred to as the written artifacts of the investigation.

Sketch Engine was then employed to group content words of all the participants' collected essays into 7 categories (i.e., adjective + noun, adverb + adjective, adverb + verb, noun + noun, noun + verb, verb + adverb, and verb + noun). This qualitative categorization was grounded upon Benson et al.'s (2010) framework for classifying English lexical collocations. Sketch Engine is indeed a web-based Corpus Query System that not only allows users to organize and analyze their texts lexically and grammatically, but also provides access to a variety of common corpora, including the BNC, the British Academic Written English Corpus (BAWE), English Web 2020 (enTenTen20), just to name a few.

Furthermore, to qualitatively assess the appropriateness of lexical collocations in the participants' essays after identifying 7 groups of content words using Sketch Engine, the academic written subcorpus of COCA was deployed, along with the Collocations Dictionary for Students of English (McIntosh et al., 2009) and the BBI Dictionary of English Word Combinations (Benson et al., 2010), to confirm whether the students used English lexical collocations correctly or incorrectly in their writing. According to Davies (2010), the COCA is a sizable, publicly accessible corpus containing over one billion words of texts. It includes nearly 25 million words each year from 1990 to 2019 and covers a wide range of genres, including spoken language, fiction, popular magazines, newspapers, academic texts, blogs, and other web contents. This corpus tool supplies abundant information not only for individual words, such as collocates, topics, genres, related words, and websites, but also for phrases and strings (e.g., frequency, clusters, concordance lines).

To address the final research question, a 5-point Likert scale questionnaire consisting of 15 items and two open-ended questions adapted from Duong and Nguyen's (2021) study, was administered to the participants. This questionnaire aimed to explore the causes of the students' difficulties in using English

collocations and gather their suggestions for enhancing collocational competence. The design of all questionnaire items was rooted in the research aims outlined in the literature review, aligning with the theoretical frameworks that underpin this investigation. All the Likert scale items from the collected questionnaire were analyzed quantitatively using SPSS 26.0, while the students' responses to the two open-ended questions underwent qualitative analysis to obtain deeper understanding of the factors contributing to the learners' difficulties in utilizing English collocations.

Research Procedure

After all the participants' written essays were collected from their final exam, they were transformed into electronic files, which were then processed using Sketch Engine. Firstly, the Sketch Engine was used to generate a foundational analysis of the students' essays involving metrics such as word count, tokens, and sentence count. Thereafter, the data went through a comprehensive analysis focusing on N-grams, keywords, and concordance lines.

During the process of qualitatively analyzing the students' essays using the Sketch Engine for the N-grams pertaining to parts of speech, seven groups (i.e., adverb + adjective, adjective + noun, adverb + verb, noun + noun, noun + verb, verb + adverb, and verb + noun) were identified. The occurrences were alphabetically sorted through the key words with each category representing a type of collocation derived from the framework developed by Benson et al. (2010) for the classification of lexical collocations. Lastly, the results were exported into Excel files to facilitate a more in-depth examination of the appropriateness of each collocation occurrence. Afterwards, collocations within each category were manually classified as either appropriate or inappropriate based on two criteria: 1) matches in two reference dictionaries, namely the Collocations Dictionary for Students of English (McIntosh et al., 2009) and the BBI Dictionary of English Word Combinations (Benson et al., 2010); and 2) occurrences within the academic written subcorpus of COCA. In essence, considering the focus of this study on the students' usage of collocations in academic writing, collocations that lacked matches in the aforementioned dictionaries and had no instances of academic usage in the COCA were deemed inappropriate. The number of appropriate and inappropriate occurrences within each type of lexical collocation was quantified and used to calculate percentages to address the first and second research questions.

As for the questionnaire employed to tackle the last research question, to ensure its reliability and validity, a pilot survey was conducted on 21 students from the same writing course, representing the target population of this study before the questionnaire was administered to all the participants for the full data collection. The quantitative analysis of the collected pilot testing data, computed by the SPSS version 26.0, resulted in a Cronbach's alpha of .878. This value indicated that the instrument was significantly reliable for use in the entire data collection process of the research. One week after the pilot survey, the main questionnaire was directly distributed to all of the study participants in their classrooms with clear guidance, which required no more than 15 minutes to finish. Due to some students being absent on the data collection date, only 103 responses out of the total population of 120 participants were included for quantitative analysis. After the data collection, all Likert scale questionnaire items were subjected to quantitative analysis using the SPSS program to address the research question. Plus, responses from the participants to the two open-ended questions were qualitatively analyzed to gain additional insights into the reasons behind the students' challenges in using English collocations in their academic essays. These responses also offered further suggestions to guide instructors and curriculum developers in devising productive classroom activities that boost EFL learners' mastery of English collocations.

Results and Discussion

Types of English Lexical Collocations Frequently Used in Students' Academic Writing

Regarding the first research question about the kinds of English lexical collocations commonly utilized in academic essays by Vietnamese intermediate EFL students, the descriptive statistics generated from the analysis of the participants' actual writing are presented in Table 1.

TABLE 1
Occurrences of English Lexical Collocations in Students' Academic Essays

Types of collocations	Nt	P (%)	Typical examples
1. Adjective + noun	1324	42.25	advanced devices, bad effects
2. Adverb + adjective	165	5.26	completely wrong, extremely crucial
3. Adverb + verb	96	3.06	absolutely agree, directly affect
4. Noun + verb	663	21.16	companies produce, lives change
5. Noun + noun	286	9.13	communication skills, computer screen
6. Verb + adverb	79	2.52	work effectively, increase significantly
7. Verb + noun	521	16.62	play sports, reduce stress
Total	3134	100	

Note. Nt = number of collocational occurrences; P (%) = percentage

It is obvious from Table 1 that all seven types of English lexical collocations were expressed in the students' academic essays. To be more specific, the analysis of the students' essays revealed that the most frequently used collocations were adjective + noun collocations (42.25%), followed by noun + verb collocations (21.16%), and verb + noun collocations (16.62%). On the contrary, the least common types of English lexical collocations were verb + adverb collocations (2.52%), and adverb + verb collocations (3.06%). Additionally, noun + noun collocations (9.13%) and adverb + adjective collocations (5.26%) represented a relatively low percentage of usage, indicating that these two types of English lexical collocations were still not easy for the students to incorporate into their writing.

These findings are more in line with the previous study outcomes (Demir, 2017; Granger & Bestgen, 2014; Siengsanoh, 2021; Thongvitit & Thumawongsa, 2017; Trinant & Yodkamlue, 2019), which disclosed that complex noun phrases, particularly adjective + noun collocations, were common in academic writing. As Taguchi et al. (2013) asserted, phrasal complexity, often involving sequences of noun phrases, is prevalent in academic written prose as it allows writers to present information succinctly. In the same vein, Hyland (2008), Biber et al. (2011), and Crossley (2020) argued that complex noun phrase constituents serve as essential components of quality academic discourse, which requires explicit instruction in both lexical cohesion and semantic precision. Thus, it is imperative to provide additional support for L2 learners in general, and for Vietnamese intermediate EFL students in particular, to enhance their proficiency in constructing complex noun phrases, specifically through the usage of adjective + noun collocations in academic writing. This can be achieved by implementing targeted instructional strategies that stress the significance of mastering this aspect of phrasal complexity. More focused instruction and practice on phrasal complexity would ultimately benefit L2 writing lessons, empowering learners to navigate the intricacies of scholarly discourse with confidence and proficiency (Taguchi et al., 2013).

Students' Difficulties with Using English Lexical Collocations in Academic Writing

In terms of the challenges that Vietnamese intermediate EFL students might encounter when using English lexical collocations in their academic papers, the statistical analysis of collocational errors discovered in the participants' writing has been calculated and documented in Table 2.

TABLE 2
Occurrences of English Collocational Errors in Students' Academic Essays

Types of collocational errors	Ne	P ₁ (%)	P ₂ (%)	Typical examples
1. Adjective + noun	435	32.85	43.81	*affordable time, *controllable mind, *multifarious people, *profound presence
2. Adverb + adjective	23	13.94	2.32	*easy sick, *gradually familiar, *more worse, *more bad, *quite unfound
3. Adverb + verb	12	12.50	1.21	*easily balance, *equally separate, *increasingly diagnose, *relevantly make
4. Noun + verb	234	35.29	23.56	*opinion said, *people founded, *survey say, *idea show, *weight risen
5. Noun + noun	110	38.46	11.08	*desktop lifestyle, *drawback aspects, *modernization world, *trouble effect
6. Verb + adverb	28	35.44	2.82	*affect badly, *impact effectively, *increase strongly, *see cleanly
7. Verb + noun	151	28.98	15.21	*brings failures, *do cooks, *make decrease, *reduce intensity, *rise health
Total	993	31.68	100	

Note. Ne = number of collocational errors; P₁ = percentage of collocational errors of each type over the total collocational occurrences of each type; P₂ = percentage of collocational errors of each type over the total errors of all types

As can be seen from both Table 1 and Table 2, concerning the students' difficulties with using English lexical collocations, the quantitative examination of the learners' English compositions uncovered that roughly 32% of the collocations occurring in the participants' writing were used incorrectly (Ne=993, Nt=3134). Specifically, among the seven types of English lexical collocations, the students made the most mistakes related to adjective + noun collocations (Ne=435, P₂=43.81%). This was followed by noun + verb collocations (Ne=234, P₂=23.56%), which accounted for almost half of the frequently discovered errors in the first category. As the three least commonly used types of collocations in the students' academic compositions, verb + adverb collocations (Ne=28, P₂=2.82%), adverb + adjective collocations (Ne=23, P₂=2.32%), and adverb + verb collocations (Ne=12, P₂=1.21%) witnessed the fewest mistakes committed by the participants.

Most remarkably, when looking closely at the proportion of collocational mistakes made over the total number of the collocational occurrences within each of the seven types of English lexical collocations, the students seemed to face the most challenges in utilizing noun + noun collocations (Ne=110, Nt=286, P₁=38.46%). Despite being the least frequently used category of lexical collocations in the students' academic essays, mistakes committed with verb + adverb collocations accounted for more than a third of all collocational tokens of this type (Ne=28, Nt=79, P₁=35.44%). As the two most frequently utilized categories of collocations among the seven types, adjective + noun and noun + verb collocations made up approximately one third of mistakes occurring within each of these two types, (Ne=435, Nt=1324, P₁=32.85%), (Ne=234, Nt=663, P₁=35.29%), respectively. Being the third commonly used collocational type of collocations in the students' English compositions, verb + noun collocations experienced more than a quarter of mistakes committed within this category (Ne=151, Nt=521, P₁=28.98%).

By and large, among the seven types of English lexical collocations, Vietnamese intermediate EFL students generally made their most mistakes relevant to adjective + noun collocations, followed by noun + verb collocations, with their fewest errors pertaining to adverb + adjective, and adverb + verb collocations. However, the types of English lexical collocations which appeared to be the most challenging for the Vietnamese intermediate EFL students' usage were noun + noun, adjective + noun, verb + adverb, and noun + verb collocations because the percentage of collocational errors within each of these categories was over 31.68%, which was the average of all the seven types. Such research results corroborate Begagić's (2014) findings, which demonstrated that verb + adverb collocations posed the greatest challenges for both first and fourth-year English-majored students in Zenica. In effect, these research results are more in accordance with Namvar's (2012) study outcomes which disclosed that adjective + noun, and noun + verb

collocations were the most problematic categories of English lexical collocations for Iranian postgraduate students. Likewise, the results from this research are also congruent with previous studies (Kuo, 2009; Phoocharoensil, 2011; Umair et al., 2023), showing that adjective + noun collocations were among the most frequent errors for Taiwanese EFL students, Thai EFL learners, and Pakistani EFL students, respectively. Nonetheless, the findings partially diverge from earlier investigations (Farghal & Al-Hamly, 2007; Mahmoud, 2005; Thongvitit & Thumawongsa, 2017), which discovered that verb + noun collocations were among the most misused English lexical collocations by Kuwaiti EFL learners, Arabian EFL undergraduate students, and Thai advanced EFL writers.

Based on the research findings, it is evident that Vietnamese intermediate EFL learners face considerable challenges in using adjective + noun, noun-noun, noun + verb, and verb + adverb collocations in their academic essays. Thus, these categories should receive special treatment from writing instructors, which calls for more effective teaching activities and more focused learning materials. To put it another way, it is vital to prioritize raising the students' awareness of the challenging collocational patterns through offering explicit collocation-focused instruction associated with guided practices and providing students with sufficient exposure to authentic use of the target language's collocations (Granger & Bestgen, 2014; Siengsanoh, 2021). By integrating corpus-based resources which emphasize typical collocations and their contextual usage, it is believed that students will be able to develop their collocational competence suitable for academic writing (Hyland, 2008; Lee et al., 2019).

Causes of Students' Difficulties in Using English Lexical Collocations

In reference to the reasons behind the students' challenges in utilizing English lexical collocations in their academic writing, the descriptive analysis of the participants' perspectives via the questionnaire highlighted several crucial factors contributing to their improper use of English lexical collocations. These quantitative findings have been summarized in Table 3 and Figure 1.

TABLE 3
Causes of Students' Difficulties in Using English Lexical Collocations in Academic Writing

	Mean	Standard Deviation
Limited vocabulary size	4.24	.734
Low English proficiency	4.17	.729
Lack of appropriate collocational knowledge	4.02	.598
Unawareness of the importance of collocations	3.93	.744
Lack of appropriate learning strategies	4.13	.645
Pedagogy and learning materials	3.98	.779
L1 negative transfer	4.06	.790
Overuse of synonym strategy	4.02	.727

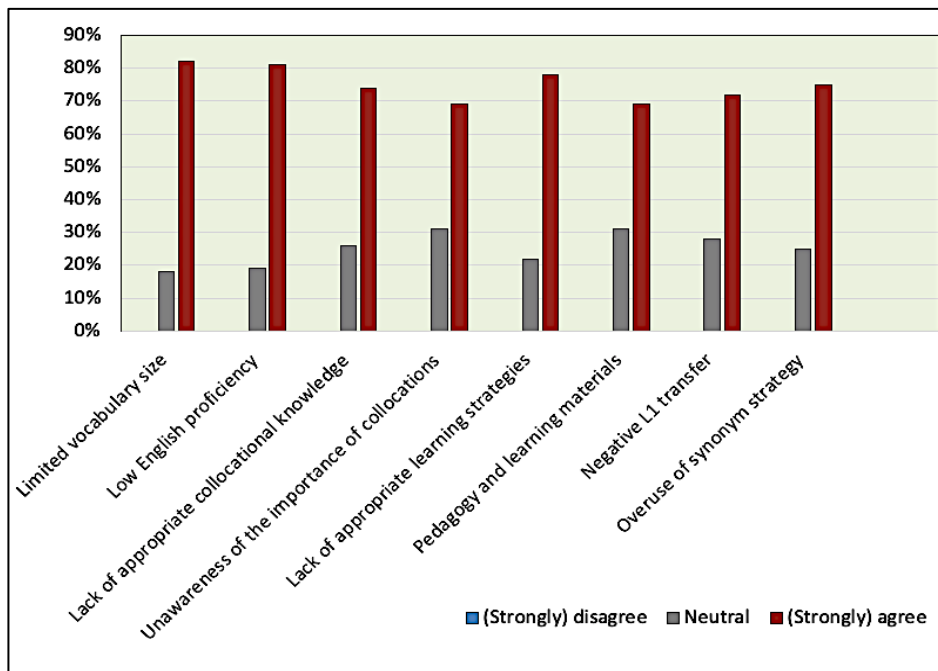


Figure 1. Students' opinions on causes of their difficulties in using English collocations.

As illustrated from Table 3 and Figure 1, the majority of the students, namely more than 70% of the participants in this study admitted that their difficulties in using English collocations resulted from a variety of causes. The most contributing factor was from their limited vocabulary size ($M=4.24$, $SD=.734$), followed by their low English proficiency ($M=4.17$, $SD=.729$). The third cause was due to the students' lack of learning strategies ($M=4.13$, $SD=.645$) with about 78% of the students considering it as one of the primary factors affecting their use of English collocations. In addition, their L1 negative transfer ($M=4.06$, $SD=.790$) was perceived as a major problem with 72% of the students strongly agreeing. Apart from this, roughly three-quarters of the participants thought that other causes influencing their misuse of English collocations included overuse of synonym strategy ($M=4.02$, $SD=.727$), and lack of appropriate collocational knowledge ($M=4.02$, $SD=.598$). Besides, nearly 70% of the students concurred that pedagogy and learning materials ($M=3.98$, $SD=.779$) along with their unawareness of the importance of collocations ($M=3.93$, $SD=.744$) could be decisive factors triggering their difficulties in using English lexical collocations as well.

Moreover, drawing from the results obtained from the open-ended question of the survey, which was further answered by twenty-six students, there were several important factors leading to the students' difficulties in using English lexical collocations. First and foremost, most of them acknowledged that their lack of collocational knowledge associated with their limited size of vocabulary caused their misuse and inappropriate choices of English lexical collocations. For instance, one participant elucidated, "I do not know how to learn and apply collocations effectively in both writing and speaking." Another student stated, "I really lack vocabulary as well as collocational knowledge to totally understand collocation-related issues." Plus, they shared that they had not known a variety of collocations coupled with their uncertainty in the contextual usage of English collocations and their poor memory of various English collocations, which contributed to their lack of confidence in using English collocations appropriately. As one of the participants said, "I could not use collocations effectively due to my limited memory of English collocations." Another idea was added, "I am not really sure how to use English collocations appropriately in different contexts." One more participant proffered the reason for her misuse of English lexical collocations in her academic essay, "I have not known many collocations and could hardly remember collocations for a long

time because word combinations seem irrelevant to each other and not to co-occur based on any certain rules.” Furthermore, the students voiced that they usually got confused about phrases having similar meanings and found it too hard to comprehend the appropriate meaning of a certain collocation, which gave rise to their wrong selection of English collocations. For example, one of them elaborated, *“I easily get confused about which words could be combined because they have similar meanings. I do not completely understand the meaning of collocations, which leads to my confusion over how to use them correctly.”* Lastly, the participants also mentioned some other causes of their problems in using English lexical collocations which encompassed their inability in identifying the correct context of use, abusing collocations, misunderstanding collocations as word choice, and their mother tongue’s interference. As one of the participants clarified, *“I often select wrong words to combine because of my language transfer. It is because of my overuse of collocations without being able to determine the context of use for various collocations. I frequently misunderstand collocation as word choice.”* All in all, the participants’ responses to the open-ended question in the survey strongly support the descriptive results achieved from analyzing the questionnaire items in this study. Such alignment confirms that the Vietnamese intermediate EFL students inevitably encountered major challenges in using English lexical collocations due to a multitude of both internal and external influential factors.

Overall, even though there were different elements affecting Vietnamese intermediate EFL students’ use of English lexical collocations in their academic writing, the foremost and primary cause was their limited vocabulary size, closely trailed by their low English proficiency. This finding is in harmony with previous investigations (Duong & Nguyen, 2021; Namvar, 2012), which showed that students’ vocabulary size, coupled with their restricted comprehension of individual lexical items within collocations, and their low language proficiency, served as the principal sources of their issues with English collocations. Besides, the results from this study highlighted that the students’ lack of appropriate collocational knowledge and learning strategies, in conjunction with negative transfer from their native language (L1), resulted in their incorrect employment and selection of English lexical collocations. These research outcomes are consistent with prior investigations (Dokchandra, 2019; Kuo, 2009; Mahmoud, 2005; Namvar, 2012; Thongvitit & Thumawongsa, 2017), which confirmed that deficient collocational knowledge posed a fundamental predicament that led to the L1 negative transfer among learners. This language interference, identified as the predominant cause of students’ collocational errors in prior studies (Ang et al., 2011; Fan, 2009; Farghal & Al-Hamly, 2007; Khoram & Mirsalari, 2020; Phoocharoensil, 2011; Shitu, 2015; Yamashita & Jiang, 2010), has also been echoed in the present research. In essence, these findings coincide with earlier research outcomes that underscored the role of synonymy in fostering students’ collocational errors, compounded by other influential factors, such as an inadequate understanding of target language collocations, a tendency to rely on paraphrasing strategies, unawareness of the significance of collocations, lack of suitable pedagogical approaches and materials, and inability to discern the appropriate usage context for English collocations, which entails the implementation of appropriate interventions to support learners in surmounting these challenges (Boonraksa & Naisena, 2022; Boonyasquan, 2009; Mongkolchai, 2008; Tanihardjo, 2017; Yumanee & Phoocharoensil, 2013).

Suggestions for Improving Students’ Collocational Competence

The descriptive statistics derived from the survey centralizing students’ suggestions for improving collocational competence have been computed and presented in Table 4. Drawing on the research findings and insights from previous studies, several pedagogical practices and suggestions would be discussed in detail. The discussion aims to enhance the effectiveness of teaching and learning English collocations, thereby aiding EFL students in developing their collocational competence for academic writing.

TABLE 4
Students' Suggestions for Improving Collocational Competence

	Mean	Standard Deviation
Develop collocational knowledge through explicit instruction and practical exercises.	4.41	.648
Raise students' greater awareness of the importance of collocations in academic writing in relation to international standards.	4.35	.606
Recommend and instruct students how to use helpful tools to develop their collocational knowledge.	4.39	.630
Provide constructive feedback on how students have used collocations in their writing so that they can adjust and monitor their own acquisition of English collocations.	4.40	.616
Learn and practice the use of collocations regularly based on topics and suitable levels.	4.39	.598

The survey results clearly indicated that almost all the students agreed with all the proposed recommendations. To be more precise, the students were in strong agreement with the idea of explicitly instructing English collocations and directly integrating them into classroom activities because this would enable learners to better understand, acquire, and use collocations appropriately in their compositions, which was considered as one of the most promising and practical applications ($M=4.41$, $SD=.648$). Such a suggestion is indeed in accordance with earlier explorations (Choi, 2019; Namvar, 2012; Phoocharoensil, 2011; Siengsanoh, 2021), which recommended that EFL instructors should design and develop collocation lessons based on the genuine difficulties confronting EFL learners. On top of that, the teachers should highlight the transfer-based problems resulting from non-congruence in the target language and the learners' mother tongue, with a particular focus on the common deviations occurring in their interlanguage (Fan, 2009; Farghal & Al-Hamly, 2007; Nesselhauf, 2003). Hence, one potentially beneficial classroom activity for students is to present typical topic-based collocations relevant to each lesson's content. This can be done initially in the target language, followed by asking the students to identify equivalents in their native language, which can be accomplished through collocation-based completion tasks. These collocation-oriented translation exercises aim to highlight any similarities and differences in collocational usage between the target language and the students' first language. Through such comparisons, learners are expected to increase their awareness of the transfer-related issues, leading to heightened sensitivity towards English collocational usage. This process corresponds with the noticing hypothesis, posited by Schmidt (1994), suggesting that learners acquire language through their conscious identification of target language knowledge. In addition, there should be a greater emphasis on the lexical approach for teaching collocations (Farrokh, 2012; Yumanee & Phoocharoensil, 2013), which means that learners should be explicitly instructed word combinations or chunks instead of individual words. It is especially important for learners to be exposed to authentic texts and materials in order to construct their own collocational knowledge (Boonraksa & Naisena, 2022; Duong & Nguyen, 2021; Namvar, 2012).

Most significantly, the students from this study strongly proposed that instructors should offer regular constructive feedback on their use of collocations in writing so that they could adjust and monitor their own mastery of English collocations ($M=4.40$, $SD=.616$). Another practical suggestion from the students was that instructors should recommend and guide them on using helpful tools to develop their collocational knowledge ($M=4.39$, $SD=.630$). With the advancement of technology, the employment of corpus-based tools and concordancers, namely the COCA and the BNC, have been highly recommended for teaching academic vocabulary, particularly collocations, to EFL and ESL learners (Ang et al., 2011; Daskalovska, 2015; Duong & Nguyen, 2021; Fan, 2009; Farrokh, 2012) apart from conventional printed materials and dictionaries (e.g., BBI Dictionary of English Word Combinations, Collocations Dictionary for Students of English, Longman Dictionary of Contemporary English). These computer-based tools could provide students with easy access to electronic databases, allowing them to not only determine the frequency of a specific collocation in authentic texts, but also comprehend how it is used in a given context. Accordingly, teachers should properly instruct their students on how to use common corpora and provide them with adequate opportunities to be exposed to corpus analysis practices (Duong & Nguyen, 2021; Gao et al., 2019; Yamashita & Jiang, 2010). This exposure can be facilitated by extracting sentences containing target collocations and then editing them to match their learners' proficiency levels. Another particularly effective

classroom activity is to engage students in identifying collocational patterns or understanding how certain words collocate with others in authentic texts through using specific corpora such as the BNC or the COCA. Such a practical exercise is presumed to assist students in comprehending and subsequently using English lexical collocations appropriately in their academic writing.

Moreover, the students from this investigation wholeheartedly agreed that learners should learn and practice using English collocations on a certain topic of their appropriate level on a daily basis ($M=4.39$, $SD=.598$). This recommendation aligns with other prior studies (Duong & Nguyen, 2021; Kuo, 2009; Tanihardjo, 2017), suggesting that EFL students could enhance their collocational knowledge through their regular exposure to the target language's collocations by reading, listening, and taking notes of useful collocations in various authentic contexts. Last but not least, almost all the students were of the same opinion that it is crucial to raise learners' awareness about the importance of using English collocations in their academic writing by explicitly emphasizing its core value in relation to the international standards for assessing academic essays ($M=4.35$, $SD=.606$). Additionally, Farrokh (2012), Dokchandra (2019), and Trinant and Yodkamlue (2019) especially stressed that teachers play a substantial role in helping the learners notice the usage of English collocations in authentic materials, which requires both adequate time of exposure and sufficient input.

In addition to the above-mentioned suggestions, to improve the quality of using English collocations in academic writing, the students in this current research put forth several additional recommendations that were elaborated upon through their responses to the open-ended survey question. Among the twenty-six comments collected from the students, nine suggested that learners should engage with English collocations everyday by communicating with peers and learning from teachers. For instance, one of them proposed, *“Get exposed to collocations on a regular basis. Learn how to use it by making sentences often. Improve it daily with the teacher’s help.”* Another participant urged, *“Try to include collocation in daily communication. Currently, I think every student should apply it to their writing and practice speaking every day to get used to using collocation.”* Besides, the rest of the participants underscored the importance of developing collocational knowledge through self-study. They recommended using various reliable and trusted sources like newspapers and movies. These participants advised, *“Read newspapers or read books in English. Read many books on collocation. Find documents from reliable sources.”* Lastly, all of them advocated that teachers should incorporate collocation lectures into classrooms and design practical activities to expand students' collocational knowledge. Their suggestions included, *“Teachers should devise numerous activities aimed at expanding students' grasp of collocations. Regularly mention collocations in their lectures.”*

In a nutshell, it is apparent that although Vietnamese intermediate EFL students frequently use English lexical collocations in their argumentative essays, they still encounter various challenges in their usage. To be exact, the students' difficulties in utilizing English collocations are rooted in a range of key factors involving their limited vocabulary, low proficiency in English, lack of appropriate collocational knowledge, and negative transfer. In the context of EFL education in Asia, these challenges may extend to cultural and linguistic differences, differing educational systems, and resources across countries (Khoram & Mirsalari, 2020; Mongkolchai, 2008). Drawing upon the identified problems and the underlying causes, it is crucial for Vietnamese EFL instructors, as well as EFL educators across Asia to adapt appropriate pedagogical practices and strategies to suit the specific needs and contexts of the students, taking into account factors such as learners' cultural backgrounds, language proficiency levels, and access to resources. Furthermore, fostering a supportive learning environment that encourages active participation, critical thinking, and continuous practice is essential for the successful acquisition of collocational knowledge and other aspects of English language proficiency. Consequently, addressing these challenges and implementing effective pedagogical approaches can help to empower EFL students to become more confident and proficient in using English collocations in academic writing.

Conclusion

To recapitulate briefly, this descriptive corpus-based study sheds light on the intricate patterns of English lexical collocations utilized in academic compositions among Vietnamese intermediate EFL learners. Through meticulous analysis and exploration, the research has offered evidence-based insights not only into the most frequently used collocations, namely adjective + noun collocations, but also into the specific challenges faced by Vietnamese intermediate EFL students in using English lexical collocations in their academic essays, specifically noun + noun, adjective + noun, noun + verb, and verb + adverb collocations. These findings are indeed in line with previous studies (Begagić, 2014; Kuo, 2009; Namvar, 2012; Phoocharoensil, 2011; Siengsanoh, 2021; Umair et al., 2023), which also underscores the need to raise the students' awareness of the challenging collocational patterns through targeted pedagogical interventions. Besides, providing sufficient input and appropriate practices with the incorporation of corpus-based tools and concordances such as the BNC and the COCA is crucial to strengthen the learners' collocational competencies (Ang et al., 2011; Daskalovska, 2015; Duong & Nguyen, 2021; Fan, 2009; Farrokh, 2012; Gao et al., 2019; Yamashita & Jiang, 2010).

Although this study has been carefully conducted, a few limitations still exist. The fundamental constraint is its restricted scope. Given that this research primarily concentrated on examining students' utilization of English lexical collocations in their academic essays, other aspects and classifications of English collocations, notably grammatical collocations, have not been explicitly taken into account. These concerns obviously exceed the boundaries of the current investigation. Thus, there is more room for future research to delve into exploring how EFL students employ both grammatical and lexical collocations in their academic writing as well as speaking in order to gain a comprehensive understanding of L2 learners' use of collocations in the target language. Another restraint relates to the exclusive focus on argumentative or persuasive essays in this study, which has led to overlooking other types of academic written prose, such as cause-effect essays, compare-contrast essays, or problem-solution essays, to name a few. It is argued that since the predominant use of a specific category of collocations might hinge on the genre and task requirements, examining the application of collocations across different types of academic essays would yield deeper insights into this domain. Lastly, though both qualitative and quantitative methods of data collection and analysis were employed in this investigation to address the research questions, it is believed that to further understand the students' challenges and the underlying causes of their difficulties in using English collocations, additional data gathered from in-depth interviews with the prospective participants should be integrated. In addition, it is important to qualitatively analyze students' actual usage of collocations in their written discourse to further corroborate and triangulate the research findings in this field of study.

Despite some inevitable restrictions, the results obtained from this research were particularly beneficial for both teachers and learners in the EFL context. As for the students, the research outcomes would help them concretely realize the types of commonly used collocations, and the obstacles they frequently face when using English lexical collocations. More importantly, based on the research findings coupled with the pedagogical implications, students could identify suitable ways to learn and practice their English collocations, enabling them to use the target collocations more effectively in their academic essays. What is more, these results would greatly assist EFL instructors in not only understanding the reality of English lexical collocation usage in academic compositions by intermediate EFL learners in general and by Vietnamese EFL students in particular, but also identifying the root causes of their problems. By incorporating tailored instructional strategies and providing sufficient opportunities for practice and feedback to tackle these challenges, language instructors can help students bolster their collocational competence. This, in turn, will contribute to the effectiveness of academic writing programs and the advancement of pedagogical practices in EFL writing instruction, as well as promote cross-cultural understanding in English language education.

Acknowledgements

I would like to express my sincere gratitude to the reviewers for their insightful and constructive feedback throughout the revision process, which greatly assisted in finalizing the manuscript. I also deeply appreciate the editor's thorough guidance and support. My heartfelt thanks are further extended to everyone who offered comments, encouragement, and insights to this study. Without these invaluable contributions, this paper would not have been possible.

The Author

Hong Thi Tuyet Vo is a TESOL lecturer at the School of Foreign Languages, Can Tho University, Vietnam. Her main research interests include applied linguistics, corpus linguistics, sociolinguistics, second language writing, second language acquisition, EAP and ESP, and foreign language teaching.

Department of General English and ESP
School of Foreign Languages
Can Tho University
Can Tho, Vietnam
Tel: +84 765309879
Email: vtthong@ctu.edu.vn

References

- Ang, L. H., Rahim, H. A., Tan, K. H., & Salehuddin, K. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis. *3L: The Southeast Asian Journal of English Language Studies*, 17(Special Issues), 31–44. <https://tinyurl.com/mrwwt8nd>
- Begagić, M. (2014). English language students' productive and receptive knowledge of collocations. *Explorations in English Language and Linguistics*, 2(1), 46–67. <https://doi.org/10.1515/exell-2016-0003>
- Benson, M., Benson, E., & Ilson, R. (1997). *The BBI combinatory dictionary of English* (2nd ed.). John Benjamins.
- Benson, M., Benson, E., & Ilson, R. (2010). *The BBI combinatory dictionary of English* (3rd ed.). John Benjamins.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5–35. <https://doi.org/10.5054/tq.2011.244483>
- Boonraksa, T., & Naisena, S. (2022). A study on English collocation errors of Thai EFL students. *English Language Teaching*, 15(1), 164–177. <https://doi.org/10.5539/elt.v15n1p164>
- Boonyasquan, S. (2009). An analysis of collocational violations in translation. *MANUTSAT PARITAT: Journal of Humanities*, 27(2), 79–91. <https://ejournals.swu.ac.th/>
- Choi, W. (2019). A corpus-based study on “Delexical verb+ noun” collocations made by Korean learners of English. *The Journal of Asia TEFL*, 16(1), 279–293. <http://dx.doi.org/10.18823/asiatefl.2019.16.1.18.279>
- Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415–443. <https://doi.org/10.17239/jowr-2020.11.03.01>
- Crosthwaite, P., Hutchison, S., Wijayatilake, C., & De Souza, N. (2017). *Mindset for IELTS: Student book 1*. Cambridge University Press.
- Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130–144. <https://doi.org/10.1080/09588221.2013.803982>

- Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, 25(4), 447–464. <https://doi.org/10.1093/lc/fqq018>
- Demir, C. (2017). Lexical collocations in English: A comparative study of native and non-native scholars of English. *Journal of Language and Linguistic Studies*, 13(1), 75–87. <https://search.informit.org/doi/pdf/10.3316/informit.059189707722764>
- Dokchandra, D. (2019). Thai EFL learners' collocational competence and their perceptions of collocational difficulty. *Theory and Practice in Language Studies*, 9(7), 776–784. <http://dx.doi.org/10.17507/tpls.0907.04>
- Duong, D. T. H., & Nguyen, N. D. T. (2021). Using collocations to enhance academic writing: A survey study at Van Lang University. In W. Strielkowski, J. M. Black, S. A. Butterfield, C. Chang, J. Cheng, F. P. Dumanig, R. Al-Mabuk, R. Al-Mabuk, M. Urban, & S. Webb (Eds.), *Proceedings of the 17th International Conference of the Asia Association of Computer-Assisted Language Learning (AsiaCALL 2021)* (pp. 275–287). Atlantis Press. <https://doi.org/10.2991/assehr.k.210226.035>
- Fan, M. (2009). An exploratory study of collocational use by ESL students—A task-based approach. *System*, 37(1), 110–123. <https://doi.org/10.1016/j.system.2008.06.004>
- Farghal, M., & Al-Hamly, M. (2007). Lexical collocations in EFL writing. *The Journal of Asia TEFL*, 4(2), 69–94. <https://shorturl.at/tSEgT>
- Farrokh, P. (2012). Raising awareness of collocation in ESL/EFL classrooms. *Journal of Studies in Education*, 2(3), 55–74. <https://doi.org/10.5296/jse.v2i3.1615>
- Firth, J. (1957). Modes of meaning. In F. Palmer (Ed.), *Papers in linguistics* (pp. 190–215). Oxford University Press.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155–179. <https://doi.org/10.1111/lang.12225>
- Gao, B., Zhu, L., & Luo, Z. A. (2019). Corpus-based study of collocation production of Chinese EFL learners. *International Journal of Education*, 11(4), 66–86. <https://doi.org/10.5296/ije.v11i4.15895>
- González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *International Journal of Applied Linguistics*, 166(1), 94–126. <https://doi.org/10.1075/itl.166.1.03fer>
- Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, 52(3), 229–252. <https://doi.org/10.1515/iral-2014-0011>
- Gries, S. T. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics*, 1(1), 1–27. <https://doi.org/10.1075/arcl.1.02gri>
- Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In D. M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47–69). Language Teaching Publications.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
- Hosseinpour, R. M., Behzadi, F., & Damavandi, A. A. (2023). Lexical bundles in EAP research articles of L1 and L2 English authors: A contrastive analysis. *The Journal of Asia TEFL*, 20(2), 215–243. <http://dx.doi.org/10.18823/asiatefl.2023.20.2.1.215>
- Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Max Niemeyer Verlag.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- Ishikawa, S. I. (2024). Lexicogrammar of the L2 English essays written by Asian college students: A corpus-based study. *The Journal of Asia TEFL*, 21(1), 119–138. <http://dx.doi.org/10.18823/asiatefl.2024.21.1.7.119>

- Khoram, A., & Mirsalari, S. A. (2020). An investigation of Iranian EFL learners' knowledge of English lexical collocations across academic levels. *The Journal of Asia TEFL*, 17(1), 87–104. <http://dx.doi.org/10.18823/asiatefl.2020.17.1.6.87>
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., ... & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7–36. <https://link.springer.com/article/10.1007/s40607-014-0009-9>
- Kim, E. S., & Lee, E. J. (2021). A corpus-based analysis of lexical bundles between English L1 and L2 writers in medical journal abstracts: Focusing on structures and functions. *The Journal of Asia TEFL*, 18(1), 142–160. <http://dx.doi.org/10.18823/asiatefl.2021.18.1.9.142>
- Kuo, C. L. (2009). An analysis of the use of collocation by intermediate EFL college students in Taiwan. *ARECLS*, 6, 141–155. https://research.ncl.ac.uk/media/sites/researchwebsites/arecls/kuo_vol6.pdf
- Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5), 721–753. <https://doi.org/10.1093/applin/amy012>
- Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice*. Language Teaching Publications.
- Lewis, M. (2000). *Teaching collocation: Further developments in the lexical approach*. Language Teaching Publications.
- Mahmoud, A. (2005). Collocation errors made by Arab learners of English. *Asian EFL Journal*, 5(2), 117–126. <https://www.asian-efl-journal.com/monthly-editions-new/collocation-errors-made-by-arab-learners-of-english/index.htm>
- Mayor, M. (2009). *Longman dictionary of contemporary English* (5th ed.). Pearson.
- McCarthy, M., & O'Dell, F. (2005). *English collocations in use*. Cambridge University Press.
- McCarthy, M., & O'Dell, F. (2008). *English collocations in use: How words work together for fluent and natural English: Self-study and classroom use*. Cambridge University Press.
- McCarthy, M., & O'Dell, F. (2017). *English collocations in use* (2nd ed.). Cambridge University Press.
- McIntosh, C., Francis, B., & Poole, R. (2009). *Oxford collocations dictionary for students of English* (2nd ed.). Oxford University Press.
- Men, H. (2018). *Vocabulary increase and collocation learning: A corpus-based cross-sectional study of Chinese learners of English*. Springer.
- Mongkolchai, A. (2008). *A study of university students' ability in using English collocations* [Master's thesis, Srinakharinwirot University]. [http://thesis.swu.ac.th/swuthesis/Eng\(M.A.\)/Angkana_M.pdf](http://thesis.swu.ac.th/swuthesis/Eng(M.A.)/Angkana_M.pdf)
- Namvar, F. (2012). The relationship between language proficiency and use of collocation by Iranian EFL students. *3L, Language, Linguistics, Literature*, 18(3), 41–52. <https://core.ac.uk/download/pdf/11494429.pdf>
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223–242. <https://doi.org/10.1093/applin/24.2.223>
- Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130–149. <https://doi.org/10.1017/S0267190512000098>
- Phoocharoensil, S. (2011). Collocational errors in EFL learners' interlanguage. *Journal of Education and Practice*, 2(3), 103–120. <https://tinyurl.com/9wfpfn9k4>
- Phoocharoensil, S. (2014). Exploring learners' developing L2 collocational competence. *Theory and Practice in Language Studies*, 4(12), 2533–2540. <https://doi.org/10.4304/tpls.4.12.2533-2540>
- Saito, K. (2020). Multi-or single-word units? The role of collocation use in comprehensible and contextually appropriate second language speech. *Language Learning*, 70(2), 548–588. <https://doi.org/10.1111/lang.12387>
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11–26. <https://tinyurl.com/3kjw83ra>

- Shitu, F. M. (2015). Collocation errors in English as second language (ESL) writing. *International Journal of Cognitive and Language Sciences*, 9(9), 3270–3277. <https://doi.org/10.5281/zenodo.1110888>
- Siengsanoh, B. (2021). Lexical collocational use by Thai EFL learners in writing. *LEARN Journal: Language Education and Acquisition Research Network*, 14(2), 171–193. <https://so04.tci-thaijo.org/index.php/LEARN/index>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Stubbs, M. (2007). An example of frequent English phraseology: distributions, structures and functions. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on* (pp. 89–105). Brill. https://doi.org/10.1163/9789401204347_007
- Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, 47(2), 420–430. <https://www.jstor.org/stable/43267799>
- Tanihardjo, J. (2017). Indonesian EFL students' production of English lexical collocation in writing. *Journal of English Language and Culture*, 7(2), 64–72. <http://dx.doi.org/10.30813/jelc.v7i2.1025>
- Thongvitit, S., & Thumawongsa, N. (2017). A corpus-based study of English collocations found in the abstracts of research articles written by Thai EFL writers. *International Journal of Social Science and Humanity*, 7(12), 751–755. <https://bit.ly/44RPaLC>
- Trinant, K., & Yodkamlue, B. (2019). Lexical collocations in a sample corpus of nursing research articles (SCNRA). *Suranaree Journal of Social Science*, 13(1), 45–72. <https://doi.org/10.55766/OBXC3400>
- Trng, V. T., & Thao, T. T. (2021). A study on the use of English collocation in writing by students at Thai Nguyen University. *International Journal of Social Science and Human Research*, 4(5), 1044–1049. <https://doi.org/10.47191/ijsshr/v4-i5-19>
- Umair, H. M., Imran, M., & Sarwat, S. (2023). An analysis of the use of collocations in English essay writing at undergraduate level. *Pakistan Languages and Humanities Review*, 7(1), 43–54. [https://doi.org/10.47205/plhr.2023\(7-1\)05](https://doi.org/10.47205/plhr.2023(7-1)05)
- Wei, Y. (1999, March 9–13). *Teaching collocations for productive vocabulary development* [Conference session]. Annual Meeting of the Teachers of English to Speakers of Other Languages, New York, NY, United States. <https://files.eric.ed.gov/fulltext/ED457690.pdf>
- Wood, D. (2002). Formulaic language acquisition and production: Implications for teaching. *TESL Canada Journal*, 20(1), 01–15. <https://doi.org/10.18806/tesl.v20i1.935>
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, 44(4), 647–668. <https://doi.org/10.5054/tq.2010.235998>
- Yuliawati, S., Ekawati, D., & Mawarrani, R. E. (2021). Investigating lexical bundles in the corpora of English and Indonesian research articles with the sketch engine. *Jurnal Sosioteknologi*, 20(2), 188–200. <https://tinyurl.com/6ywj2br>
- Yumanee, C., & Phoocharoensil, S. (2013). Analysis of collocational errors of Thai EFL students. *LEARN Journal: Language Education and Acquisition Research Network*, 6(1), 88–98. <https://so04.tci-thaijo.org/index.php/LEARN/article/view/102723>

(Received August 25, 2023; Revised February 28, 2024; Accepted June 10, 2024)

Appendix

Typical Examples of Collocations Extracted from the Analysis on Sketch Engine

Details	Left context	Key words in contexts	Right context	Details
ADJECTIVE + NOUN				
doc#0	They need to use computer with	*affordable time	due to produce television,	doc#0
doc#0	</s><s> Firstly, lacking of	*controllable mind	about using these electronic	doc#0
doc#0	of our lives. </s><s> Therefore,	*multifarious people	tend to stay at home and put	doc#0
doc#0	and televisions having a more	*profound presence	in our daily life. </s><s> There	doc#0
ADVERB + ADJECTIVE				
doc#0	health more and more so bad, so	*easy sick	, increase weight, impact	doc#0
doc#0	at the beginning, but then, they	*gradually familiar	with the exercises, they might	
doc#0	is a serious problem it will be	*more worse	if people still not realize this	doc#0
doc#0	about why this fear is	*quite unfound	to some extent. </s><s> Firstly,	doc#0
ADVERB + VERB				
doc#0	who are self-disciplined can	*easily balance	the true that they spend on using	doc#0
doc#0	problem.</s><s>If we cannot	*equally separate	our time that spent on these	doc#0
doc#0	nowadays many adolescents are	*increasingly diagnose	to the myopic because the have	doc#0
doc#0	In short, using electronic	*relevantly make	people balance their life and	doc#0
NOUN + VERB				
doc#0	your own lives. </s><s> The	*idea show	that a number of people of	doc#0
doc#0	convenient. </s><s> But some	*opinion said	that people nowadays tend to do	doc#0
doc#0	a part of our life and many	*people founded	it helpful, while others said	doc#0
doc#0	. </s><s> But numerous of	*survey say	people who do not do activities	doc#0
NOUN + NOUN				
doc#0	/s><s> Secondly, if they have a	*desktop lifestyles	for a long time, they will be	doc#0
doc#0	<s> In contract, there is not less	*drawback aspects	from this.</s><s> People get	doc#0
doc#0	Nowadays, people lived in a	*modernization world	, many electronic inventions	doc#0
doc#0	inventions also have some	*trouble effect	to the people's life. </s><s>	doc#0
VERB + ADVERB				
doc#0	in outdoor activities which	*affect badly	on their health due to computers	doc#0
doc#0	television, computer which	*impact effectively	on person about health problem.	doc#0
doc#0	had serious healths problems	*increase strongly	. </s><s> Some problems they	doc#0
doc#0	<s> Children's eyes will not	*seeing cleanly	and they have to wear a glasses	doc#0
VERB + NOUN				
doc#0	reason, the lazy of people that	*brings failures	for them. </s><s> A good	doc#0
doc#0	sports, hang out with friends	*do cooks	, so on, when they has free time.	doc#0
doc#0	and televisions would	*make decrease	health of many people. </s><s>	doc#0
doc#0	many cause that young people	*rise health	for games or they hurt other	doc#0