BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

A Corpus-Based Analysis of Collocation Translation by Arab EFL Learners

Bader Altamimi , ***Abdulaziz Sanosi*** *

Department of English Language and Literature, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia

ABSTRACT

Collocations are essential linguistic structures that enhance communication, making them important for EFL learners and translators. This importance has prompted previous research to analyze their function and significance. Research indicates that understanding collocations does not necessarily correlate with vocabulary acquisition, underscoring the need to teach collocations independently and instruct EFL learners in their use and translation. This need stems from EFL learners' significant difficulties in conceptualizing, using, and translating collocations. To enrich the existing literature and broaden research methodologies, the present study utilized corpus linguistic techniques to assess the accuracy of collocation translations by Arab EFL learners. Data were gathered from a sample of 176 students via a data hub, where they translated 15 English sentences that included 21 collocations. This resulted in a corpus of approximately 22,500 tokens and 4,000 collocations. Accuracy was evaluated based on the frequency of translating collocations into their appropriate Arabic equivalents and an investigation into any statistically significant differences. The results revealed moderate accuracy in collocation translation, likely due to ineffective translation strategies. Additionally, no statistically significant differences were observed in accuracy rates among various types of collocations. These findings have implications for the teaching of collocations and translation. They may also serve as a basis for further research into developing collocation competence among EFL learners and future translators.

Keywords: Collocation Competence; Collocation Instruction; Collocation Translation; Corpus Linguistics; Multiword-

*CORRESPONDING AUTHOR:

Abdulaziz Sanosi, Department of English Language and Literature, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia; Email: a.assanosi@psau.edu.sa

ARTICLE INFO

Received: 4 April 2025 | Revised: 17 April 2025 | Accepted: 20 April 2025 | Published Online: 26 April 2025
DOI: https://doi.org/10.30564/fls.v7i5.9494

CITATION

Altamimi, B., Sanosi, A., 2025. A Corpus-Based Analysis of Collocation Translation by Arab EFL Learners. Forum for Linguistic Studies. 7(5): 327–339. DOI: https://doi.org/10.30564/fls.v7i5.9494

Units; Phraseology

1. Introduction

In English and many other languages, specific lexical structures are made up of multiple words, known as Multi-word Units. These units function as communicative tools, with their meanings derived not from standard semantic conventions but from factors such as their co-occurrence frequency, idiomatic nature, or pragmatic roles. Collocations serve as a good example of this concept. They are defined as vocabulary items that frequently appear alongside particular other items, showing syntactic relationships and a certain level of semantic opacity [1]. Additionally, they should occur at a frequency greater than the chance would predict [2]. It was early proved that "a knowledge of collocations is essential to full communicative mastery of English" [3], as it largely "contributes to vocabulary knowledge and hence to language proficiency" [4]. Accordingly, collocation has become a topic of interest for a considerable body of research.

In the context of second language use and learning, some research studied the effect of collocation competence on translation accuracy [5], and other research examined translation difficulties that face translators and EFL (English as a foreign language) learners when translating collocations [6, 7]. In contrast, other studies investigated the relationship between vocabulary knowledge and collocation [8]. Such studies have led to an overall consensus that collocation is essential for translation accuracy and that EFL learners face considerable difficulties in realising [9, 10], using and translating collocation [11, 12]. Further, it was found that knowledge of collocation is not an inevitable result of vocabulary development and, therefore, needs to be taught in its own right [3]. Yet, further research is required to support or refute these findings by applying different research methods and utilizing newly standardized rubrics.

Starting from this point, the present study aims to explore the translation of collocations by Arab EFL learners at the university level. Utilizing a corpus linguistic approach, the study will first examine the accuracy of these translations in relation to their conformity with Arabic collocations as defined by a well-established reference and reviewed by specialists in relevant fields [13]. Secondly, the study aims to determine whether there are statistically significant differences in accuracy rates when translating three types of collocations. This will involve testing the null hypothesis:

$\mathbf{H_0}$. *There are no significant differences in accuracy rates among the different types of collocations.*

Depending on the results, the study will either accept this hypothesis or reject it in favour of the alternative hypothesis:

$\mathbf{H_1}$. *There are significant differences in accuracy rates among the different types of collocations.*

The findings of this study are considered significant as they provide additional evidence regarding the use and translation of collocations by Arab EFL learners. This research is expected to contribute to the field by employing diverse methodologies and tools. Furthermore, it is anticipated to offer implications for teachers, students, and curriculum designers by emphasizing the importance of teaching collocations and identifying which types or aspects should be prioritized if selection is necessary. The study's hypotheses and expected implications are based on a robust and well-established body of literature on collocations, which will be discussed in the following section.

2. Literature Review

In the previous decades, several research studies have investigated the concept of collocation in English and across languages. This research was intended to either establish or develop theories relating to collocation or to contextualize existing theories and investigate different variables related to collocation use and translation.

2.1. Theoretical Framework

Research on collocation drew on a rich literature in which the concept was developed and contextualized in linguistics, discourse analysis, and corpus linguistics. This research includes both the use and translation of collocation.

2.1.1. Collocations

Collocation is a linguistic phenomenon which can be found in all languages. The study of the co-occurrence of

certain words started in the eighteenth century in the biblical concordance by Alexander Cruden (who listed, for example, the co-occurrences of the word *dry* with *ground*)[14]. In addition, the basic idea of collocation was studied by the linguists of the Prague school, who referred to collocation using the term automation[15]. However, the concept of collocation was first introduced into modern linguistics by the British linguist John Firth[15, 16]. Nevertheless, the term became more popular in later corpus linguistics studies like those by Sinclair[15].

Even though collocation has been thoroughly investigated for more than half a century, the definition of the term is not widely agreed upon and is considered a controversial issue in linguistics[16–19]. Etymologically, *collocation* or *collocate* is derived from the Latin meaning 'place side by side' "which is originally derived from the Latin *col-locare"*[20]. In linguistics, Firth described collocation as the contribution to the discernment of a word's meaning made by "the company it keeps"[21]. Firth states, "Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words" (p. 196). Therefore, one of the meanings of *night* is inferred from its collocation with *dark* and vice versa. Similarly, we can anticipate the limited set of adjectives that frequently accompany *pretty*, such as *girl*, *flower*, *garden*, etc. In addition, Firth is interested in how words are distributed in texts and how some words tend to appear together more frequently than others do. From this point of view on collocation, a word is never isolated from the context in which it appears; instead, it 'predicts' the occurrence of other lexical items.

The seminal work by Firth on the concept of collocation has subsequently been expanded by several linguistics scholars, including Mitchell, Halliday, McIntosh, and Sinclaire[21–25]. McIntosh referred to Firth's term *collocation,* employing the term *range* to denote the degree of compatibility between words[24]. For McIntosh, the range of a word is the possible list of collocates. For example, the range of the word *frozen* may include *water, lake, meal,* etc., but not *paper* since this word contravenes the tolerance of compatibility of the words that collocate with it.

Halliday and Hasan argue that lexical cohesion can be achieved by reiteration (e.g., repetition, synonym, near-synonym, or superordinate) or collocation within the traditional discourse analysis context[26]. According to Halliday and Hasan[26], collocation is "a cover term for the cohesion that results from the co-occurrence of lexical items that are in some way or other typically associated with one another because they tend to occur in similar environments". For example, the word pairs *keyboard*, *mouse,* and *computer* are associated with one another so that they might be found in a text; hence, this text is presumably 'coherent'.

From a corpus linguistics perspective (i.e., frequency-based approach), Halliday and Sinclair proposed definitions of collocation that view collocation as formal co-occurrences[23, 25]. Halliday defined collocation as: "the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur at n removes (a distance of n lexical items) from an item x, the items a, b, c..."[23]. Sinclair expanded upon the concept of co-occurrence by introducing the terms *node*, *span* and *collocates*, where *node* denotes the word under investigation[25], whereas *span* denotes the number of relevant words on each side of the node and *collocates* refer to the lexical items that co-occur with that node within its vicinity. These concepts have been crucial in collocation research since the notion of collocation began to include consecutive co-occurring words and pairs or groups of words that are not necessarily adjacent[25]. These frequency-based definitions proposed by Halliday and Sinclair are adopted in this study.

The contribution of the corpus linguistics approach to collocation studies revealed a strong connection between L2 proficiency and collocation use, though this relationship remains unclear[27]. These studies investigated collocations in corpora from two main angles: (1) the density of collocation within a corpus, which claimed to denote higher proficiency levels of language and (2) the strength of association between the words forming these collocations[1, 8, 28, 29]. As these studies are concerned with L2 learners and users, an important variable to investigate was collocation translation to a congruent or similar collocation in the target language.

2.1.2. Collocation Translation

Within the context of translation studies, which is the focus of this study, collocation is widely examined since it can present various challenges in translation[30]. Like Sinclair[25], Baker views collocation as a "tendency of certain words to co-occur regularly in a given language"[30]. She argues that the tendency for specific lexical items to co-occur is

more likely when these words are related in meaning; therefore, *data* is more likely to collocate with *retrieve*, *mine*, and *process* than with *take*, *pay*, and *play*. However, sometimes the words forming the collocation are not related in meaning; hence, the word *pay* has nothing to do with *visit* regarding meaning in the collocation *pay a visit*. Therefore, she concludes, "patterns of collocation are largely arbitrary and independent of meaning" (p. 53), and this arbitrariness occurs within the English language and other languages such as Arabic. For example, the English verb *deliver* tends to co-occur with several nouns, including *letter*, *lecture, verdict, news*, and *baby*. However, each instance of deliver with each collocate has a different Arabic equivalent (i.e., *yusallimu khitaaban* for *deliver a letter*, *yulqi khutbatan* for *deliver speech*, *yanqilu akhbaaran* for *deliver news*, *yusdiru hukman* for *deliver a verdict*, and *yuwallid imra'tan* for *deliver a baby*)[30].

This arbitrary nature of collocation, added to the fact that they cannot be recognized according to specific syntactic or semantic rules, makes collocation "a problematic area in translation"[31]. Accordingly, EFL learners commit many errors in the translation of collocations because they are generally affected by factors such as negative transfer and using a literal translation[6]. Translators commit such mistakes probably because they "might have a general sense that an equivalent translation can always express a collocation in L1"[32]; however, finding the proper equivalent collocation in the target language is not always possible[31].

Previous research has also investigated other types of problems. For example, Bahumaid identified two main types of collocation translation problems: intralingual and interlingual[7]. Intralingual problems refer to the issues of identifying and conceptualizing collocations per se, while interlingual problems are those related to collocability across languages. The latter type entails awareness of differences across languages at all levels, including differences in vocabulary and grammatical structure. For example, Al-Jarf noted that the typologies of English and Arabic collocation are different[11], yet there are some types of collocation that are very common in both languages. These observations are significant as they highlight the role of cross-linguistic influence and L1 interference in collocation translation accuracy. Ghazala identified three types of collocation as highly frequent in both languages: Adjective + Noun, Noun + Noun, and Verb + Object types[13, 31]. These types are adopted in this study as they were also standardized by previous studies as the most common collocations in the English language[33].

Habtoor & Al-Swaidan found that familiarity with English collocations, in their own right, is positively correlated with the ability to translate such terms into Arabic rather than from familiarity with vocabulary in general[9]. Therefore, raising students' collocation competence is envisaged as a critical factor in improving EFL trainees' translation accuracy[5]. Such competence includes an awareness of the congruences and incongruences between collocation in both source and target languages and should be on different language levels, including grammar and vocabulary[11]. For this reason, the translation course strategies should incorporate approaches that can adequately accommodate these congruencies and incongruencies.

2.1.3. EFL Learners' Translation Strategies

Researchers have identified various strategies translators employ when converting texts from a source to a target language[9]. Several frameworks have been proposed to classify these strategies, the most notable being those outlined by Newmark and Baker[30, 33]. Newmark introduced eight strategies to address translators' challenges when dealing with different types of texts[33]. These strategies range from straightforward approaches, such as word-for-word and literal translation, to techniques that prioritize conveying the overall meaning of the text, such as adaptation, free translation, and idiomatic translation.

Similarly, Baker identified eight strategies to handle non-equivalence between source and target languages[30]. These include using more general or neutral terms, paraphrasing, omitting information, or providing illustrations. While methods like adaptation and idiomatic translation are often effective for rendering formulaic language, using other strategies, such as those suggested by Baker, may compromise the integrity of collocations, leading to linguistically accurate but non-formulaic expressions in the target language. As a result, educators and curriculum developers are encouraged to prioritize strategies that preserve the cultural and idiomatic essence of collocations. The task of successfully translating idiomatic expressions, including collocation, involves not just identifying the idioms but also how to translate them and choosing appropriate strategies for different types of text[34].

2.2. Previous Studies

The use of formulaic expressions by Arab EFL Learners has been investigated frequently in recent years. Most of these studies found that Arab EFL learners find using and applying such expressions quite challenging. For example, Sanosi found that the use of lexical bundles by Arab EFL learners differs significantly from native speakers and that even advanced users show no development in such use [35]. Additionally, Jamshed et al. found that learning idiomatic expressions is challenging for Arab EFL learners [36], and they traced that to several causes, including lack of familiarity with the concepts of idioms, variation in literal interpretation, and lack of Arabic parallel. Regarding collocation, several studies have shown similar results regarding problems in their conceptualization, use and interpretation. In this regard, El-Dakhs found that Arab learners' collocational competence is notably unsatisfactory [4], although it is gradually improving with increased language exposure. She also found that learners were more confident in their use of Verb + Noun collocations than Adjective + Noun collocations.

Collocation translation by Arab EFL learners was also a topic of interest for a considerable body of research. In this regard, Zienel'aabdin & Ahmed conducted a study that incorporated 20 undergraduate students and utilized a translation test of 14 items to measure collocation translation [37]. The results revealed that students' translation was poor, resulting in unnatural structures mainly because the participants relied entirely on literal and word-for-word translation strategies. Similarly, Al Nakhala administered a multi-question test including both subjective and objective translation questions to measure the translation of collocation into Arabic [12], finding that students faced considerable difficulties, mainly caused by word-for-word translation and negative transfer from their L1. Although the study identifies L1 transfer as the leading cause of translation errors, it considers this transfer primarily from a linguistic perspective, focusing on structural or syntactic differences such as preposition addition, omission, or word order. However, this approach is insufficient, as mistranslation of collocations often involves not only surface linguistic differences but also deeper cultural and pragmatic factors. In a more detailed study, Al-Jarf used the corpus linguistics approach to identify the types and sources of collocation translation errors and the strategies learners use to translate verb + preposition English collocations into Arabic [11]. Her results showed that the participants mistranslated certain prepositions in word + preposition collocations. The learners substituted prepositions mistakenly, added incorrect ones and omitted other prepositions. Further, other interlingual and intralingual errors appeared.

From a different perspective, Habtoor & Al-Swaidan conducted a study to investigate EFL learners' familiarity with collocations and their effect on their translation of collocations [9]. Using a collocation test administered to 40 students, the researchers found that learners' familiarity with collocations positively correlated with their ability to translate them into Arabic. These results are confirmed by Metwally and Hamad [10], who conducted a literature review that included several studies and found a strong correlation between collocation competence and quality of translation.

Overall, most studies agreed that Arab EFL learners' collocation competence is unsatisfactory, resulting in low levels of collocation translation as presented in their translations. Though these studies were relatively sufficient and comprehensive, they are general, and few of them focused on the typology of collocations analyzed or the strategies utilized. Also, the standardization of Arabic collocation is unclear. Accordingly, there is a paucity of sources that reference common and agreed-upon sets and taxonomies of Arabic collocations. These subtle shortcomings are meant to be addressed by the current study using standardized Arabic collocations referenced by specialists in translation and applied linguistics. The data was then analyzed using appropriate statistical tests to generate robust results. These procedures are described in the following section.

3. Methods

3.1. Research Design

The present study uses the corpus linguistic method to gather and analyze its data. This method was deemed suitable for the study as it incorporates analyzing the frequency and range distribution of specific linguistic categories in a collection of texts [38]. Corpora are considered optimal for providing robust evidence on language use since they contain authentic data [39]. Using this approach is also supported by the fact that Systematic Functional Grammar, of which collocation is a fundamental construct, always calls for analysis based on real language use [40]. Accordingly,

the researchers compiled a learner corpus to investigate the collocation translation output and evaluate it on a rubric of expected translation according to a reference dictionary[13].

3.2. Participants

College students from Prince Sattam bin Abdulaziz University, Saudi Arabia, participated in the study. They were at their final levels 6, 7, and 8 at the time of intervention. All students had completed at least two translation courses providing the main requirements for translation from English to Arabic and vice versa; however, no special training on collocation translation or translation of any formulaic structures was provided to them. Overall, the student levels range from upper intermediate to advanced, given that they reach these levels in the English language and literature program. The intact class sampling method was used where all the sections underwent the study in a structured test atmosphere.

3.3. Data Collection

The data for this study was collected by the Collocation Translation Data Hub (CTD Hub), which was developed for research purposes and distributed to the participants through the Blackboard learning management system and QR code. The students were asked to translate the sentence without using any electronic or classic translation tool and were given 45 minutes to complete the task.

3.3.1. The CTD Hub

The CTD Hub is a tool designed to collect various versions of EFL students' translations of 15 sentences containing different English collocations. The tool was designed by the researchers and underwent different stages of quality control to guarantee maximum validity and reliability. After that, the final version of the tool was launched through Google Forms to collect students' responses. The tool incorporated 15 sentences and 21 collocations distributed, as shown in **Table 1** below.

A copy of the tool is presented in **Appendix A**.

3.3.2. Validity

To guarantee maximum validity, the researchers built the sentences that comprise the tool by referring first to McCarthy and O'Dell and (Oxford Online Collocation Dictionary, n.d) to ensure that the sentences contain proper and common collocations that the respondents can spot[41]. Further, the researchers refer to the work of Ghazala to ensure that these English collocations have equivalent or close collocations in Arabic accessible to learners at their current levels[13]. Initially, the tool incorporated 30 sentences, each containing at least one collocation. The tool was then provided to four referees majoring in translation and applied linguistics to test its validity. The referees provided suggestions to enhance the tool, which was then updated accordingly and resulted in 15 sentences.

Table 1. Details of Collocation in the CTD Hub.

Collocation Type	Count
Adjective + Noun*	7
Noun + Noun	8
Verb & Noun	9
Total	**24**

* *Note*. In Arabic, the adjective follows the noun it modifies and matches it in gender, definiteness, number and case; collocations of this type in Arabic are in Noun + Adjective structure.

3.3.3. Reliability

A pilot test was conducted where the tool was administered to 22 students of different levels. The responses were analyzed to develop the tool by deleting the confusing collocations and restructuring the sentences to avoid interference between collocations. Then, the cases where the translation of collocation may be because of a literal translation were determined. In other words, the approved collocations are those that stem from a higher knowledge of collocation, not those that may result from literal or hasty translation. For example, the researchers excluded the collocation *hard times, school year,* and *health problems* as they do not necessarily indicate collocation knowledge while retaining collocations such as *busy day, traffic jam,* and *heavy rain* as their Arabic equivalent collocations are different from the literal translation which is a fundamental aspect of collocation.

3.4. Corpus

The CTD Hub was linked to Google Forms and administered in real-time classes as part of different translation courses. The responses were then downloaded in spreadsheet format. The 188 responses were filtered to account for the irrelevant answers and texts that include English characters and remove any other invalid entries. Ultimately, 174 rows in the spreadsheet were approved. The 174 student transla-

tions were extracted using Python and stored in separate text files. The files were stored in a folder to form the corpus of the study, which was titled 'English to Arabic Collocation Translation' (EACOLT). EACOLT details are provided in **Table 2** below, and the corpus raw files, database files, and extracting code are available online.

Table 2. EACOLT Corpus Details.

Metric	Value
Files	174
Tokens	22,481
Types	2,032
Type/token ratio	0.09
Average token per file	129

The low type/token ratio indicates that only 9% of the words of the corpus are unique, which was anticipated given that all the respondents may translate the sentences similarly, including using collocations.

3.5. Data Analysis

The researchers used AntConc software to analyze the corpus[42]. The N-Gram tool was first used to detect common collocations in the corpus that meet the following criteria: (1) Appear with a frequency cut-off of ten occurrences. (2) Appear at a range percent of 17% (14 files), and (3) Be relevant to the expected collocation translation list. Ultimately, 21 collocations relevant to the English collocations in the CTD Hub and the expected translated list were selected and analyzed. They represent the findings of the study, which will be presented and discussed below. The accuracy rate (AR) is calculated using the formula: $\mathrm{AR} = \frac{\mathrm{FOC}}{\mathrm{TR}} \times \mathbf{100}$. The formula involves dividing the Frequency of Occurrence of the collocation (FOC) by the Total number of Responses (TR). The result of this division is then multiplied by 100 to express the accuracy rate as a percentage. Further, an Analysis of Variance (ANOVA) test was conducted to examine the differences in accuracy rates among the three types of collocation: Adjective + Noun, Noun + Noun, and Noun + Verb. This process was meant to check if these differences are statistically significant to accept or reject the study's hypothesis accordingly. The alpha level (α) was set to 0.05, which is common in humanities and social sciences research.

4. Findings

The present study aims to show the level of accuracy and correctness of collocation translation from English to Arabic and whether EFL learners retain common Arabic collocations when translating equivalent or similar English ones. The researchers used EACOLT, which is a corpus of 22,500 words by an Arab EFL translator compiled by the researchers for this research. The corpus incorporated 174 entries, 2,610 sentences, and 4,176 collocations. After analyzing the frequency and the range percent of the common two-word N-grams in the corpus, several N-grams were revealed. However, since the N-Gram feature captures all co-occurrences of specific words, it may include combinations that do not meet the semantic and linguistic pragmatic criteria necessary to qualify as collocations; the researchers conducted manual filtering to elect those collocations that were relevant to the data as presented in the CTD Hub and the expected translation of the Arabic collocation list. This process returned 21 N-grams of Arabic collocations relevant to the CTD Hub's embedded collocations. The summary of these collocations is presented in **Table 3** below.

Table 3. A Summary of Common Relevant N-Gram in the EACOLT Corpus.

Type	Count	Accuracy Rate
Adjective + Noun	7	52.79%
Noun + Noun	7	53.04%
Verb + Noun	7	30.54%
Total	**21**	**45.46%**

Table 3 reveals that both Adjective + Noun and Noun + Noun collocation types were translated with a medium accuracy rate. In contrast, collocations of the Verb + Noun type were translated with a lower accuracy rate. In other words, the students were mostly successful in realizing Noun + Noun and Adjective + Noun collocations and translating them using Arabic collations as expected by the researchers, while their translation to most Verb + Noun translations was unsuccessful in terms of their compatibility with the expected Arabic translation. To display the detailed findings of the accuracy of the translation, **Table 4** presents the results of the first type.

The findings reveal that (*bright future*) collocation had the highest AR, around 78%. Contrastively, the collocation

Table 4. Frequency and Accuracy Analysis of Adjective + Noun Collocations.

Collocation	Expected Translation	Frequency	Accuracy Rate
bright future	مستقبل مشرق	135	77.59%
junk food	وجبات سريعة	117	67.24%
mutual trust	ثقة متبادلة	110	63.22%
knockout blow	ضربة قاضية	110	63.22%
heavy rain	أمطار غزيرة	105	60.34%
supreme happiness	سعادة غامرة	39	22.41%
busy day	يوم حافل	27	15.52%
	Total	**643**	**52.79%**

(*busy day*) seems more complicated to translate to its suggested equivalent, with an AR of only 16%. This distribution reveals a broad variation among ARs. Overall, the average accuracy rate of translation for the Adjective + Noun type is medium (around 53%). This result also matches the Noun + Noun translation results, as shown in **Table 5** below.

Table 5. Frequency and Accuracy Analysis of Noun + Noun Collocations.

Collocation	Expected Translation	Frequency	Accuracy Rate
university degree	شهادة جامعية	108	62.07%
traffic jam	ازدحام مروري	89	51.15%
herd of sheep	قطيع من الأغنام	100	57.47%
rush hours	ساعات الذروة	95	54.60%
training course	دورة تدريبية	81	46.55%
weather forecasts	توقعات الطقس	70	40.23%
placement test	اختبار تحديد المستوى	103	59.20%
	Total	**646**	**53.04%**

The results indicate similar results to Adjective + Noun translation findings with a close mean average of around 53%. However, the Noun + Noun collocations demonstrate a tighter range of ARs, with the highest being 62% (*university degree*) and the lowest at 40% (*weather forecasts*). The last type of collocation, Verb + Noun collocations, demonstrates different results, as shown in **Table 6**.

Although one collocation (*catch attention*) has a relatively high AR (65%), which is higher than the top AR of the highest Noun + Noun collocation, the overall findings show that the Verb + Noun collocations have the lowest average AR (31%) compared to the Adjective + Noun and Noun + Noun collocations. Further, the results reveal a wider range of ARs, with the highest being 65% (*catch attention*) and the lowest at 6% (*break the news*).

To find the significance of the differences between the accuracy rates of the three types of collocations, an ANOVA test was conducted to examine the differences in accuracy rates among three types: Adjective + Noun, Noun + Noun, and Noun-Verb. The results of which are presented in **Table 7** below.

Although the mean accuracy of rates is superficially different, the results indicated no statistically significant differences between the groups, $F(2, 18) = 3.056$, $p = 0.072$. Since the p-value of 0.072 is greater than the conventional alpha level of 0.05, we fail to reject the null hypothesis that *there are no significant differences in accuracy rates among the different types.* In other words, there is not enough evidence to conclude that the accuracy rates differ significantly among the types, and the challenges that Arab EFL learners face in translating English collocations to equivalent and common Arabic collocations are the same and are not affected by collocation type.

5. Discussion

The current study aimed to investigate the accuracy and equivalence of collocation translations among Arab EFL

Table 6. Frequency and Accuracy Analysis of Verb + Noun Collocations.

Collocation	Expected Translation	Frequency	Accuracy Rate
catch attention	يجذب الانتباه	113	64.94%
keep a promise	يفي بالوعد	99	56.90%
make decision	يتخذ قرار	63	36.21%
took a time	استغرق وقت	36	20.69%
pay attention	يلقي الانتباه	25	14.37%
stay home	يبقى بالبيت	25	14.37%
break the news	نقل الأخبار	11	6.32%
	Total	**372**	**30.54%**

Table 7. ANOVA Results for AR by Type.

Source	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2335.940	2	1167.970	3.056	0.072
Within Groups	6878.673	18	382.148		
Total	9214.613	20			

learners compared to the expected Arabic collocations. Data analysis using AR and ANOVA tests revealed that these learners encountered moderate challenges when translating Adjective + Noun and Noun + Noun collocations and faced greater difficulties with Verb + Noun collocations. Although the average accuracy rates for translating collocations varied, there were no statistically significant differences among these averages, indicating that overall, the learners experienced challenges in translating collocations.

This finding aligns with previous research on Arab EFL learners' translation of collocations. For instance, Abdul Rida and Al-Riyahi reported that Iraqi EFL learners frequently made lexical collocation errors [6], particularly with Verb + Noun collocations, followed by Adjective + Noun and Noun + Noun collocations. However, it contrasts with the findings on the use of collocation by Arab EFL learners, such as those found by El-Dakhs [4], who observed that Arab EFL learners exhibited greater confidence in using Verb + Noun collocations compared to Adjective + Noun collocations. This discrepancy may stem from differences in how learners approach usage versus translation, as the latter involves additional cognitive tasks and applying specific translation strategies.

While it is apparent that considerable challenges face Arab EFL learners in providing proper translation across all types of collocations, most participants successfully translated a few collocations. These collocations were found among different types, such as *bright future, university degree,* and *catch attention.* This suggests that these particular collocations are well-established and frequently used in both languages, making it easier for translators to convey the intended meaning accurately. This conclusion is based on the fact that these specific collocations have no distinct structural or semantic features from the other collocations, making their wide use the only possible interpretation for their comparatively easy realization by students. Considering the context of these collocations, it might also be that contextual clues in the sentences aid their translation. For instance, in the sentence "*Many students hope to have a bright future after they get their university degree*," which was used to elicit the Arabic collocations *mustqbalun mushriq* and *shahādtun jāmi'īyah*, there are contextual hints that would help translate them into equivalent Arabic collocations. This result supports the claims found by Wu et al. that congruent collocations [32], which are highly frequent in both languages, are more accessible to L2 language learners.

Inaccurate translations of collocations often result from the reliance on literal translation strategies, leading to outcomes that appear unnatural or nonsensical. This issue is evident across all types of collocations. For instance, the collocation *busy day* has frequently been translated into Arabic as *yawman mashghūlan* (preoccupied day), *yawmun mumtly'un* (packed day), or *yawmun muzdaḥmun* (congested day). In this case, misinterpreting the context may affect the translation of the collocation. The sentence "*Today was a busy day. It was filled with meetings and phone calls*," intended to

elicit the standard Arabic collocation *yawmun ḥāfil*, contains the word "filled," which can lead to incorrect choices like *mumtly'un* or *mu'aba'un (packed or congested)*. Adopting contextual clues is a common strategy among translators, as noted by Al Jarf[11]. However, this approach often leads to errors, since translators may misinterpret the context or translate words literally, which compromises the accuracy of collocation translation.

On the other hand, *bright colours* have been translated to *alwān lāmi'ah (*shiny colours*)* and *alwān muḍī'ah (*luminous colours*)*. These literal translations fail to capture the intended meaning, as the adjectives used convey nuances that differ from the original expression. The problem is even more pronounced with Verb + Noun collocations. For example, the phrase *take time* is rarely translated accurately into its expected equivalent in Arabic. Instead, it is often rendered with phrases like *itkhadh waqtan* (adopted time), *taṭalaba waqtan* (required time), or *iḥtāja waqtan* (needed time), which are semantically unrelated and produce awkward results. These observations that literal and word-for-word translation is a common strategy that violates collocation and which, in this study, made student translators produce inaccurate and unnatural collocation are compatible with the findings of previous studies [12, 37].

Additional challenges in translating collocations arise from misusing specific translation strategies, such as paraphrasing, omission, and illustration, as outlined by Baker [30]. For instance, the collocation *weather forecast* has been translated into Arabic as *al'arsād al-jawwī* (meteorology), *akhbār al-ṭaqs* (weather news), and *al-nashrah al-jawyyah* (weather report). These translations reflect the translation-by-illustration strategy but fail to capture the precise meaning of the original collocation. Similarly, the collocation *pay attention* is often translated using the translation-by-omission strategy, resulting in phrases like *intabih* or *intabihū* (be careful), where a single word with an embedded pronoun is used to replace the two-word collocation.

These findings align with those of Elnoty [34], who highlighted paraphrasing as a commonly used strategy among Arab translators when dealing with idiomatic expressions and that Arab EFL students frequently employ various strategies from Baker's (2011) taxonomy. While these strategies can be effective in general translation tasks, their application to idiomatic structures may compromise the naturalness of the translated text, especially when an equivalent collocation already exists in the target language. This highlights the impact of cross-linguistic influence on collocation translation and its role in conveying the intended meaning. Therefore, instructors should emphasize linguistic, cultural, and pragmatic differences to help students recognise and address potential L1 interference that could hinder accurate translation. Previous studies, such as Al Nakhala [12], have shown that negative L1 transfer often contributes to inaccurate translation. However, this influence extends beyond linguistic or structural factors to include cultural aspects as well. Errors in collocation translation occur across all types, regardless of whether their structures align with English. For instance, while English Adjective + Noun collocations typically follow the (Adjective + Noun) order, e.g. heavy rain, the equivalent structure in Arabic is (Noun + Adjective), e.g *amṭārun ghazyratun*. In contrast, verb-noun collocations like "*take time*" and its Arabic equivalent "*istaghraqa waqtan*" share the same structure in both languages. Despite this structural similarity, more errors were observed in translating verb-noun collocations. This suggests that cultural factors, rather than purely linguistic interference, play a significant role in collocation translation accuracy.

The results implied that to improve EFL translation accuracy, the concept of collocation should be highlighted clearly to students, with extensive examples from the source and target language. Students should also be taught that different translation strategies should be applied, but for idiomatic expression, starting with the translation-by-cultural-substitution strategy is more effective and helps generate more fluent and natural texts in the target language. However, the fact that some of the investigated collocations were translated with high ARs demonstrated the need for raising students' awareness of common collocations in both languages, as this knowledge can be the base for teaching with these strategies.

The findings of these studies should be generalized, considering some potential limitations. First, the relatively small data retrieved from single-institution participants might not provide comprehensive insights into the actual research practice of Arab students. This potential limitation may affect the generalizability of the results and can be addressed by applying the study with the same tool to a larger sample from different Arab countries, as cultural differences in the Arab

World may affect the research output. Future research should also investigate professional translators' practice regarding collocation translation. Using the same research tool among professional translators can reveal comparable results that highlight the effect of professional practice on improving collocation translation and thus approve the impact of teaching practice on translation and whether it can be adopted as a potential reason for low levels of ARs of collocation translation. Other relevant research can also compare the current research output to data from different institutions to support or refute the findings.

6. Conclusions

This study aimed to investigate the accuracy of translation of collocation by Arab EFL learners. The accuracy is assessed in terms of the match between the provided English collocation and the equivalent Arabic ones. The findings revealed that students face challenges in translating most English collocations accurately and compatibly. Different strategies, such as illustration and omission, applied by the students caused that inaccuracy. The presence of some collocations with high ARs demonstrates that although collocation translation is generally challenging, the problem is not so significant when it comes to translating common expressions. These findings underscore the importance of incorporating dynamic strategies in EFL classrooms to enhance awareness of collocation, revealing their cultural aspects and highlighting common collocations in the target and source languages. Teaching strategies should also focus on applying suitable strategies for each type of translation, as translating idiomatic language requires more cultural awareness and sensitivity. The findings can be generalized as long as potential limitations such as data variability are considered. To overcome these limitations, future research could include participants from a greater variety of learning settings, and professional translators could be engaged to understand better the reality of translation of collocation by Arabic EFL learners.

Author Contributions

Conceptualization, A.S.; methodology, A.S.; software, A.S.; validation, B.A.; formal analysis, A.S.; investigation, B.A.; resources, B.A.; data curation, B.A.; writing—original draft preparation, B.A.; writing—review and editing, B.A. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding

This study is supported via funding from Prince Sattam bin Abdulaziz University project number [PSAU/2025/R/1446].

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Prince Sattam bin Abdulaziz University (PSAU/2025/R/1446).

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

Data is available upon request.

Acknowledgments

The authors thank the deanship of scientific research and higher studies for funding the research project. They would also thank participants at the CTD hub for their effective participation.

Conflicts of Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

The CTD Hub

PLEASE TRANSLATE THE FOLLOWING SENTENCES INTO ARABIC

1. It is important to make a decision before the deadline.

……

2. He took a long time to heal from his broken heart.

……

3. The bright colours helped to catch many customers' attention.

……

4. Please pay attention during the training course to understand the method.

……

5. It was difficult to break the news to my friend.

……

6. The weather forecasts have warned residents to stay home because of the heavy rain and thunderstorms.

……

7. Today was a busy day. It was filled with meetings and phone calls.

……

8. Many students hope to have a bright future after they get their university degree.

……

9. My close friend always supports me during tough times.

……

10. Eating too much junk food can lead to health problems and make you feel tired.

……

11. Many boxing matches ended with a knockout blow.

……

12. We got stuck in a traffic jam during rush hours.

……

13. The student felt supreme happiness upon receiving an A+ on the placement test.

……

14. It is necessary to keep a promise, as it builds mutual trust.

……

15. The farmer guided the herd of sheep across the green pasture.

……

References

[1] Granger, S., Bestgen, Y., 2014. The use of collocations by intermediate vs. advanced nonnative writers: A bigram-based study. International Review of Applied Linguistics in Language Teaching, 52(3), 229–252.

[2] Nattinger, J.R., DeCarrico, J.S., 1992. Lexical Phrases and Language Teaching. Oxford University Press: Oxford, UK.

[3] Bahns, J., Eldaw, M., 1993. Should we teach students collocations? System, 21(1), 101–114.

[4] El-Dakhs, D., 2015. The lexical collocational competence of Arab undergraduate EFL learners. International Journal of English Linguistics, 5(5), 60–75. DOI: https://doi.org/10.5539/ijel.v5n5p60

[5] Ansari, S., Ghaffarof, S., 2013. The effect of collocational competence on translation accuracy of translation trainees. Journal of Advanced English Language Teaching, 1(3), 76–84.

[6] Abdul Rida, N., Al-Riyahi, A., 2011. Lexical collocational errors in the writings of Iraqi EFL learners. Journal of College of Arts University of Basrah, 58, 24–51.

[7] Bahumaid, S., 2006. Collocation in English–Arabic Translation. Babel, 52(2), 133–152. DOI: https://doi.org/10.1075/babel.52.2.03bah

[8] Men, H., 2015. Vocabulary Increase and Collocation Learning: A Corpus-Based Cross-Sectional Study [Doctoral dissertation]. University of Birmingham: Birmingham, UK.

[9] Habtoor, S., Al-Swaidan, R., 2019. Familiarity with collocations in EFL context and strategies utilized in translating them into Arabic. Journal of Language Teaching and Research, 10(4), 724–737.

[10] Metwally, A.A., Hamad, M.M., 2023. The impact of collocation competence on translation quality: Literature review. Arab World English Journal for Translation and Literary Studies, 7(1), 175–186. DOI: https://doi.org/10.24093/awejtls/vol7no1.13

[11] Al-Jarf, R., 2022. Undergraduate student-translators' difficulties in translating English word + preposition collocations to Arabic. International Journal of Linguistics Studies, 2(2), 60–72.

[12] Al Nakhala, A., 2018. The problems of translation of collocations from English to Arabic and vice versa of English students at QOU. Journal of University of Palestine for Research and Studies, 8(1), 1–32.

[13] Ghazala, H., 2007. Dar El-ilm's Dictionary of Collocations. Dar Al-ilm: Beirut, Lebanon.

[14] Kennedy, G., 1998. An Introduction to Corpus Linguistics. Longman: London, UK.

[15] McEnery, T., Wilson, A., 2001. Corpus Linguistics: An Introduction. Edinburgh University Press: Edinburgh, UK.

[16] Xiao, R., Li, W., 2015. Collocation. In: Biber, D., Reppen, R. (eds.). The Cambridge Handbook of English Corpus Linguistics. Cambridge University Press: Cambridge, UK. 106–124.

[17] Bartsch, S., 2004. Structural and Functional Properties of Collocations in English: A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-occurrence. Narr: Tübingen, Germany.

[18] Evert, S., 2008. Baker's most frequently used strategies for translating idioms in Maḥfūẓ's works: Al-Thulāthīyah, Al-Liṣṣ wa-al-kilāb, and Mīrāmār. Journal of Scientific Research in Arts, 7, 17–46.

[19] Seretan, V., 2011. Syntax-Based Collocation Extraction. Springer: Berlin, Germany.

[20] Hoad, T.F., 1996. The Concise Oxford Dictionary of English Etymology. Oxford University Press: Oxford, UK.

[21] Firth, J., 1957. A synopsis of linguistic theory, 1930–1955. Studies in Linguistic Analysis, 10–32.

[22] Mitchell, T., 1975. Principles of Firthian Linguistics. Longman: London, UK.

[23] Bazell, C.E., 1966. In Memory of J.R. Firth. In: Bazell, C., Catford, J., Halliday, M., et al. (eds.). Longman: London, UK.

[24] McIntosh, A., 1961. Patterns and ranges. Language, 37(3), 325–337.

[25] Sinclair, J., 1966. Beginning the study of lexis. In: Bazell, C., Catford, J., Halliday, M., et al. (eds.). In Memory of J.R. Firth. Longman: London, UK. pp. 410–430.

[26] Halliday, M., Hasan, R., 1976. Cohesion in English. Longman: London, UK.

[27] Gablasova, D., Brezina, V., 2024. Adjective + noun collocations in L2 spoken English: How robust is the role of proficiency? International Journal of Learner Corpus Research, 11(1), 79–113.

[28] Brezina, V., Fox, L., 2021. Adjective + noun collocations in L2 and L1 speech: Evidence. In: Granger, S. (ed.). Perspectives on the L2 Phrasicon: The View from Learner Corpora. Multilingual Matters: Bristol, UK. pp. 152–177.

[29] Eguchi, M., Kyle, K., 2023. L2 collocation profiles and their relationship with vocabulary proficiency: A learner corpus approach. Journal of Second Language Writing, 60, 100975.

[30] Baker, M., 2011. In Other Words: A Coursebook on Translation, 2nd ed. Routledge: London, UK.

[31] Brashi, A.S., 2005. Arabic Collocations: Implications for Translation [Doctoral dissertation]. University of Western Sydney: Sydney, Australia.

[32] Wu, Y.P., Abdhalim, H., Hamzah, M., 2024. Assessing the roles of L1 and meaning knowledge in processing and production of English collocations among Chinese EFL learners. 3L: Language, Linguistics, Literature, 30(1), 175–190.

[33] Newmark, P., 1988. A Textbook of Translation. Prentice Hall: London, UK.

[34] Elnoty, N., 2020. Baker's most frequently used strategies for translating idioms in Maḥfūẓ's works: Al-Thulāthīyah, Al-Liṣṣ wa-al-kilāb, and Mīrāmār. Journal of Scientific Research in Arts, 7, 17–46.

[35] Sanosi, A.B., 2018. The use and development of lexical bundles in Arab EFL writing: A corpus-driven study. Journal of Language and Education, 8(2), 108–123.

[36] Jamshed, M., Almashy, A., Saleem, M., et al., 2024. Assessing Saudi EFL learners' use and understanding of English idioms: Perceptions, problems, and patterns of learning. Journal of Language Teaching and Research, 15(3), 1001–1009.

[37] Zienel'aabdin, S., Ahmed, G., 2017. The problem of translating English collocations into Arabic language. International Journal of English Language, Literature and Humanities, 5(3), 200–209.

[38] Tognini-Bonelli, E., 2001. Corpus Linguistics at Work. John Benjamins: Amsterdam, Netherlands.

[39] Meyer, C., 2023. English Corpus Linguistics: An Introduction. Cambridge University Press: Cambridge, UK.

[40] Yang, B., Li, W., 2020. Systemic Functional Linguistics: Interaction, methodology, and applicability. In: Yang, B., Li, W. (eds.). Corpus-Based Approaches to Grammar, Media, and Health Discourse: Systemic Functional and Other Perspectives. Springer: Berlin, Germany. pp. 1–14.

[41] McCarthy, M., O'Dell, F., 2017. English Collocations in Use, 2nd ed. Cambridge University Press: Cambridge, UK.

[42] Anthony, L., 2024. AntConc (Version 4.2.0) [Computer software]. Waseda University: Tokyo, Japan. Available from: https://www.laurenceanthony.net/software/antconc/