

L1 Interference in Students' Translations: A Corpus-Based Analysis of Collocation Errors and Pedagogical Implications

Havid Ardi^{1*}, Ermanto², Novia Juita², Vy Rany³

¹English Department FBS UNP, ²Indonesian Department FBS UNP, ³Super Student Education Cambodian School

*Corresponding Author, email: havid_a@fbs.unp.ac.id

Received: March, 25, 2025

Revised: April 10, 2025

Accepted: May, 29, 2025

Abstract

Collocation is challenging for translators as the words differ across languages. This study examines collocation errors in translations produced by English Department students who took the Indonesian-English Translations subject. The text discussed the Minangkabau tradition written in Indonesian. Employing a corpus-based approach, the research analyzes students' translations to identify recurring collocational mismatches, classify error types, and first-language (L1) interference in English as an L2, which is studied in terms of the insufficient mastery of English phraseology. The translations made by English Department students were compiled from student submissions and compared against reference corpora (e.g., COCA, BNC) as natural collocations. Tools such as AntConc and Kortara were used to quantify deviations and categorize errors into four primary types: (1) verb-noun, (2) adjective-noun mismatches, (3) unnatural noun-noun phrases, (4) adverb-adjective, (5) verb-preposition, and (6) clause base. Findings reveal that 58.94% of errors arise from noun-noun combinations, where students applied Indonesian syntactic or lexical patterns to English, resulting in unnatural collocations. The study highlights the pedagogical need for explicit collocation instruction in translation training, especially for language pairs with significant structural and cultural differences. It advocates for incorporating corpus tools into classrooms to enhance students' awareness of natural collocations and reduce L1 interference.

Keywords: collocation errors, corpus linguistics, L1 interference, pedagogical strategies, translation studies

Introduction

The way certain words go or co-occur together in a language frequently becomes a problem for other speakers of the language to produce natural sentences or expressions (Bui, 2021). Linguistically, this phenomenon is known as collocation. The use of collocations—habitual word combinations that are natural to native speakers—plays a central role in shaping the fluency, clarity, and acceptability of English discourse. In both written and spoken forms, collocations reflect not only linguistic competence but also cultural norms embedded in language use.

The problem occurs when two or more words appear in higher frequency, which differs from the mother tongue or First Language (L1) of the foreign language learners (Demirel & Kazazoğlu, 2015; Laufer & Girsai, 2008). For non-native speakers, writers, and translators, especially those translating from first languages with distinct syntactic and lexical patterns, producing appropriate collocations in English often presents challenges (Kheirzadeh & Marandi, 2014). These issues become particularly visible in academic writing and translation,

where the aim is to convey culturally specific concepts in a linguistically accurate and natural way in the target language. An inappropriate collocation may produce unacceptable translation since the expression is not common in the target language. For example, it is common to write or express “*minum obat*” in Indonesian, but it cannot be translated into *drink medicine in English. It should be translated into “take the medicine”.

Several studies have explored the importance of collocations in writing activities (Frankenberg-Garcia, 2018; Harta, Bay, & Ali, 2021; Laufer & Waldman, 2011), second language acquisition, translation studies (Febriana & Maisarah, 2021), and corpus linguistics (Liu & Tang, 2024). Some researchers focus their studies on the difficulty faced by English language learners and translators in mastering collocations at the university level (El-Siddig, 2023; Harta et al., 2021). Moreover, in translation contexts, the researchers have also shown that the translators tend to underuse or misuse collocational patterns, which is culturally bound texts in the translating of the Quran (Obeidat & Mahadi, 2020) and recipe (Setiarini, 2017). The previous studies inspired this research to explore students’ challenges in translating L1 into L2 since the source texts were culturally embedded concepts and using local collocations that must be translated from L1 (Indonesia) into L2 (English). Therefore, translators tend to use direct translations, which often disrupt natural collocational flow.

Despite these contributions, limited research has examined collocational errors in academic or cultural expository texts produced by non-native speakers of L2, especially when describing local cultural phenomena into L2. Previous researchers did not explore collocation in academic settings, which have culturally embedded explanations into L2 yet. Their works have primarily focused on professional translators, learner corpora in writing, professional translations, or literary texts, and the analysis of texts that straddle academic explanation and cultural interpretation remains underrepresented.

Furthermore, the utilization of corpora in various aspects has also been widely reported by researchers so far. For example, Ermanto, JR, Ardi, Juita, & Jamaluddin (2024) related to linguistic aspects, but this research has not yet examined the collocation aspect. In addition, several corpus-based translation studies have been conducted by various experts (Al-Salman & Haider, 2024; Park et al., 2022), but they also do not expose the translation of collocations from L1 to L2. Mostly, translation studies scholars focus on native speakers of their mother tongue (Mohammed, 2022).

This study aims to address that gap by analyzing a culturally rooted English text that explains the Minangkabau concept of *bagurau*. The text, though informative and well-structured in content, demonstrates several instances of unnatural or awkward collocational use. Such usage not only affects the readability of the text but also impacts the effectiveness of cross-cultural communication. The analysis focuses on how these collocational patterns reflect deeper issues in translating oral cultural concepts into academic English. The study of collocation using corpora in exploring collocation in language studies has been done in various areas, such as English language learning (Çakmak & Ersanlı, 2021), linguistic studies (Khoo & Johnkhan, 2018), translation studies (Bermúdez Bausela, 2016; Mohammed, 2022). Collocation is one of the objects being explored in corpus-based studies. Collocation may be studied in writing and translation studies. Collocation refers to the natural combination of words combination sequence in a language (Kemppanen, 2025, p. 17).

The purpose of this study is to examine collocational problems in the selected text and to categorize the types of errors based on their lexical or grammatical nature. This study investigates collocation errors in translations produced by English department students, focusing on Indonesian-to-English translations of a text discussing the Minangkabau tradition. Through this analysis, the study seeks to contribute to the understanding of how non-native academic writing, especially that which interprets local traditions, can be improved through

more appropriate collocational use. Ultimately, this research highlights the need for increased awareness and instruction on collocation in both academic writing and translation training.

Methods

The researchers used a descriptive analytical method to analyze collocation issues in translation. The data were compiled using English translation materials translated by Indonesian students at the English Department of Universitas Negeri Padang. This study examines collocation errors in translations produced by English department students, with a focus on Indonesian-to-English translations of a text discussing the Minangkabau cultural concept of *bagurau*. The group of students was chosen by using cluster random sampling. Using a corpus-based approach, the researchers analyzed 24 student translations to identify recurring collocational mismatches, classify error types, and the causes of collocation errors. A specialized corpus was compiled from student submissions and compared against reference corpora (e.g., BNC and COCA) and the collocation dictionary (English Collocation Dictionary) to identify natural collocations. Corpus tools, AntConc 3.5.9 and Kortara, were employed to quantify deviations and categorize errors into four primary types: (1) adjective-noun mismatches (e.g., “sharp expressions” vs. “sharp satire”), (2) verb-object discordance (e.g., “play words” vs. “engage in wordplay”), (3) unnatural noun phrases (e.g., “form of performance” vs. “performance art”), and (4) literal translations of L1 idioms (e.g., “good at talking” vs. “eloquent”).

Scholars advocate for an extensive methodological approach that combines various research tools and analytical frameworks to overcome the limitations of narrow methodologies and to address the multifaceted nature of translation research (Sutter & Lefer, 2019). This approach enhances the robustness and generalizability of findings, enabling deeper insights into translation processes, participant roles, and contextual influences (Saldanha & Sharon O'Brien, 2013).

Result and Discussion

Result

In this section, students' translations of the oral tradition in Minangkabau into English (L2) were analysed and displayed. There were 4,163 tokens and 335 word types in English produced by 24 students. The collected texts were analysed by using AntConc. The source language was from an Indonesian abstract. The collocation errors identified in students' translation texts were verified using dictionaries and web corpora, including the English Collocation Dictionary, the British National Corpus (BNC), and the Corpus of Contemporary American English (COCA). These dictionaries and web corpora were used to validate and confirm the obtained collocation forms, determining whether the collocations were correct or in error.

The analysis of the corpus of student translations of a culturally embedded Indonesian text reveals a range of collocational problems, which can be classified based on linguistic categories such as lexical and grammatical collocations. From a lexical standpoint, numerous errors emerge in common patterns like verb + noun, noun + noun, and adjective + noun combinations. For instance, the frequently used phrase “used as a form of performance” demonstrates an unnatural verb + noun collocation, where the verb “used” mismatches with the abstract noun “performance”. This prominent issue found in the translation of this local cultural text into L2 in this corpus is verb-noun collocations, the verb “used” does not naturally collocate with “a form of performance.” It can be seen in the following example that this is the result of literal translation from the Source Text (ST) into Target Text (TT) without considering the natural collocation in the target language.

ST: Dari penelitian tidaklah diketahui kapan persisnya kata *bagurau* (bergurau) digunakan sebagai bentuk pertunjukkan.

TT : *The exact time when the word *bagurau* (joking or bantering) began to be used as a form of performance ... (F2)

The expression "to be used as a form of performance" is **grammatically acceptable**, but from a **collocational** and **stylistic** standpoint, it is **unnatural** in native English. The phrase sounds vague and academic-translated like as a result of Indonesian (L1) interference (digunakan sebagai bentuk pertunjukan). There are 23 instances of the word "used" in the corpus data, 20 of them were combined with "form of performance" (Figure 1). This unnatural collocation is because the students translated the verb "digunakan" from Indonesian into "used" in English (see Figure 2) and combined it with the word "performance" or "form of performance". However, the verb "used" commonly collocates with tools or objects, methods or strategies, or functions with purposes in English. Therefore, the result of the translation sounds unnatural in English text.

	Type	Rank	Freq	Range
1	used as a form of performance	1	20	20
2	used as a type of performance	2	1	1
3	used in a performance context as	2	1	1
4	used to show the terms of	2	1	1

Figure 1. The combination of "used" in the corpus data

Total Hits: 20 Page Size 100 hits 1 to 20 of 20 hits				
	File	Left Context	Hit	Right Context
1	F1 ...	not certain when the word bagurau (bergurau) was	used as a form	of performance. The term bagurau comes
2	F2 ...	the word bagurau (joking or bantering) began to be	used as a form	of performance is not known from
3	F3 ...	I not find when the word bagurau (joking) was first	used as a form	of performance. One of the characteristics
4	F4 ...	when the word "bagurau" (joking) was persistently	used as a form	of performance. The term "bagurau" emerged
5	F6 EL...	known when exactly the word bagurau (joking) was	used as a form	of performance. The term bagurau emerged
6	F7 ...	own exactly when the word "bagurau" (to joke) was	used as a form	of performance. The term "bagurau" originates
7	F8 ...	own exactly when the word "bagurau" began to be	used as a form	of performance. The term "bagurau" came
8	F9 ...	wn when exactly the word bagurau (bergurau) was	used as a form	of performance. The term bagurau emerged
9	F10 ...	ned exactly when the word "bagurau" (to joke) was	used as a form	of performance. The term "bagurau" originates

Search Query ☒ Words ☐ Case ☐ Regex Results Set All hits Context Size 10 token(s)

used as a form Start ☐ Adv Search

Figure 2. The collocation errors of the verb "used" and noun

Figure 2 indicates that the word "used" is combined with the phrase "a form of performance". There are 20 data have problems in in this collocation. In English, it is more idiomatic expressions would be "recognized as a performance" or "considered as a performance art." This indicates that the student reliance on literal translation strategies, rather than selecting contextually appropriate equivalents in the target language.

This analysis is validated by corpora analysis that available in large corpora, BNC and COCA, that there is no phrase "used as form of performance" except "... used as form ..." (see Figure 3). It is also found that the phrase "used as a performance" or "seen as a performance" are more frequent and natural.

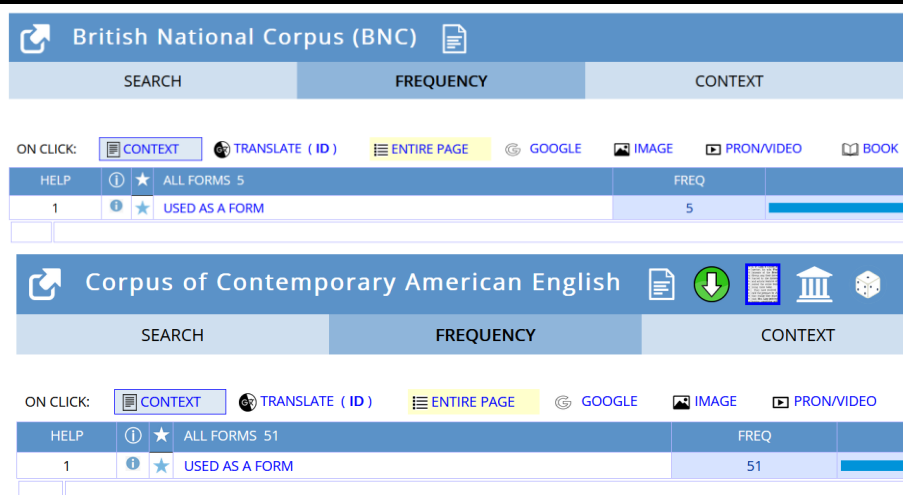


Figure 3. The combination of “used as a form in BNC and COCA”

Based on analysis in BNC and COCA and collocation dictionary, a more idiomatic rendering would be “regarded as a type of performance” or “conceived as a performance.”

Besides, verb + noun collocation problem, the phrase also has another noun + noun collocation problem related to the word “performance”. The phrase “form of performance” is grammatically correct however it is not a strong English collocation. There are 22 total hits of the phrase in the corpus data as can be seen in figure 4.

Total Hits: 23 Page Size 100 hits 1 to 23 of 23 hits				
	File	Left Context	Hit	Right Context
1	F1 ...	he word bagurau (bergurau) was used as a	form	of performance. The term bagurau comes from the cultural
2	F4 ...	agurau” (joking) was persistently used as a	form	of performance. The term “bagurau” emerged from the cultural
3	F6 El...	y the word bagurau (joking) was used as a	form	of performance. The term bagurau emerged from the cultural
4	F7 ...	he word “bagurau” (to joke) was used as a	form	of performance. The term “bagurau” originates from the cultural
5	F8 ...	the word “bagurau” began to be used as a	form	of performance. The term “bagurau” came from from the
6	F9 ...	he word bagurau (bergurau) was used as a	form	of performance. The term bagurau emerged from the cultural
7	F10 ...	he word “bagurau” (to joke) was used as a	form	of performance. The term “bagurau” originates from the cultural
8	F11 ...	he word bagurau (bergurau) was used as a	form	of performance. The term bagurau emerged from the cultural
9	F12 ...	ly the word bagurau (“joke”) was used as a	form	of performance. The term bagurau emerged from the cultural
10	F13 ...	bagurau (bergurau) began to be used as a	form	of performance. The term bagurau originates from the oral
11	F15 ...	the word bagurau (bergurau) is used as a	form	of performance. The term of bagurau appears from the
12	F16 ...	he word of bagurau (joking) was used as a	form	of performance. The term of bagurau appear from the
13	F18 ...	he word bagurau (bergurau) was used as a	form	of performance. The term bagurau emerged from the cultural
14	F19 ...	y the word bagurau (joking) was used as a	form	of performance. The term bagurau emerged from the cultural
15	M3 ...	y the word bagurau (joking) was used as a	form	of performance. The term bagurau is from the Minangkabau

Search Query ☒ Words ☐ Case ☐ Regex Results Set All hits Context Size 10 token(s)

performance Start ☐ Adv Search

Figure 4. The combination of “form of performance”

This phrase is better replaced by other stronger alternative collocations, such as type of performance, style of performance, mode of performance, or just performance.

Another recurrent issue lies in **adjective + noun** collocations, such as “*sharp language*” “*offending words*”, which misrepresent the intended meaning of *sindiran* or *bahasa tajam*. This kind of translation can be found in the following example:

ST: Tradisi bercakap-cakap atau budaya bercerita dalam suasana yang akrab, sindir-sindiran melalui ungkapan-ungkapan bahasa yang tajam merupakan kebiasaan yang sudah umum

TT: The tradition of conversation or storytelling culture in an intimate atmosphere, satire through sharp language expressions is a common and widely known habit in Minangkabau society. (F6)

Total Hits: 15 Page Size 100 hits 1 to 15 of 15 hits

File	Left Context	Hit	Right Context
1 F4 ...	storytelling in a friendly atmosphere, satire through	sharp language	expressions is a common and widely known habit
2 F5 ...	culture in an intimate atmosphere, sarcasm through	sharp language	expressions is a common and widely recognized habit
3 F6 El...	g culture in an intimate atmosphere, satire through	sharp language	expressions is a common and widely known habit
4 F9 ...	culture in an intimate atmosphere, sarcasm through	sharp language	expressions is a common and widely recognized habit
5 F12 ...	g culture in an intimate situation, sarcasm through	sharp language	expressions is a common and widely recognized habit
6 F18 ...	storytelling in a friendly atmosphere, satire through	sharp language	expressions is a common and widely known habit
7 M4 ...	culture in an intimate atmosphere, sarcasm through	sharp language	expressions is a common and widely recognized habit
8 M5 ...	culture in an intimate atmosphere, sarcasm through	sharp language	expressions is a common and widely recognized habit
9 F16 ...	ing stories in friendly atmosphere, sarcasm through	sharp language	expressions are common and widely known habit in
10 F3 ...	in intimate atmosphere, including allusions through	sharp language	expressions, is very common in Minangkabau society.

Figure 5. The combination of “language and sharp”

There are 15 students' translation texts (Figure 5) that make this collocation error. This phrase is a direct translation from Indonesian structures like “*ungkapan bahasa yang tajam*”. In English, these were likely translated based on semantic approximations rather than idiomatic norms. Therefore, appropriate substitutes of the translation would be “*witty remarks*”, “*biting sarcasm*”, or “*sarcastic expressions*.”

Another collocation error is in the combination of adverb and adjective collocation. The example can be seen in Figure 6.

ST: Kebiasaan masyarakat Minangkabau untuk berkumpul bersama sambil bercerita dan bercanda, dengan tema-tema pembicaraan yang saling sindir-menyindir, bahkan juga bisa saling (mencemooh), ...

TT: The habit of the Minangkabau society to gather together while telling stories and jokes, with themes of conversation that are mutually sarcastic, can even be mutually (mock), ... (F12)

The phrase “*mutually sarcastic*” (is also problematic; English does not usually pair “mutually” with “sarcastic.” These collocation error is detected in 8 translation. A more appropriate phrasing might be “*exchanging sarcastic remarks*” or “*engaging in playful teasing*.”

	File	Left Context	Hit	Right Context
1	F5 ...	ies and jokes, with themes of conversation that are	mutually sarcastic,	can even be mutually (mock), in a dialogical
2	F6 El...	ies and jokes, with themes of conversation that are	mutually sarcastic,	can even be mutually (mock), in a dialogical
3	F9 ...	ies and jokes, with themes of conversation that are	mutually sarcastic,	can even be mutually (mock), in a dialogical
4	F11 ...	ies and jokes, with themes of conversation that are	mutually sarcastic,	can even be mutually. (scoffing) in a dialogic
5	F12 ...	ies and jokes, with themes of conversation that are	mutually sarcastic,	can even be mutually (mock), in a dialogical
6	M4 ...	ies and jokes, with themes of conversation that are	mutually sarcastic,	can even be mutually (mock), in a dialogical
7	M5 ...	ies and jokes, with themes of conversation that are	mutually sarcastic,	can even be mutually PHQFLPHHK (mocking), in a
8	F4 ...	ies and joking, with topics of conversation that are	mutually sarcastic,	and can even be mutually (mocking), in a

Figure 6. The combination of “mutually and sarcastic”

There are some other collocational problems that come from literal translation and unfamiliarity with natural English word combinations. Similarly, the errors can be detected by comparing the result of N-Gram or Cluster in AntConc with collocation in BNC and COCA and Collocation dictionary. Another collocation errors were analysed as follows: “*namely the oral cultural tradition*,” the word “namely” is too formal and awkward in this position, and “oral tradition” is the standard collocation—not “oral cultural tradition.”

Moreover, five learners resorted to circumlocutions like “*people who are really good at talking*”, which reflect an overgeneralization strategy due to a lack of collocational resources. It is another example of L1 interference—that is a literal translation of “*dikenal sebagai masyarakat yang suka dan pintar bicara*”. The translation is grammatically correct but stylistically awkward. It illustrates an informal lexical collocation, and semantically understandable—however it is stylistically inappropriate for academic writing and more idiomatic options can be translated into “*eloquent speakers*” or “*a talkative and articulate community*.” These errors are often rooted in negative transfer from the source language, where learners rely on direct equivalence rather than contextually and collocationally appropriate alternatives.

The translation also demonstrates noun-noun and adjective-noun mismatches, as seen in “dialogical atmosphere.” Although grammatically correct, the term is overly academic and not typically used in general English to describe casual verbal interaction. A more natural phrase would be “conversational setting” or “informal atmosphere.” Likewise, the phrase “satire through sharp expressions” illustrates an unusual collocation of “satire” with “sharp expressions,” which may reflect a calque from the Indonesian expression “*ungkapan tajam*”. In English, “biting satire” or “sharp wit” would better convey the intended meaning while aligning with natural collocational patterns.

Additionally, verb-preposition collocational errors appear in constructions such as “comes from the cultural tradition.” While understandable, native usage would favor verbs like “originates in” or “derives from.” These subtle deviations, while not strictly ungrammatical, hinder the fluency and idiomaticity of the translation, which are critical factors in achieving equivalence in translation studies.

Overall, the translation reflects an interlingual interference wherein source-language patterns influence target-language production, particularly in collocational behavior. These findings highlight the importance of collocational competence in translation and support prior research that emphasizes the need for corpus-informed resources and training in collocation use (Baker, 1992; Nesselhauf, 2005). Addressing such collocational issues is essential for improving the naturalness and acceptability of translated texts in academic and literary domains.

The result reveals 151 instances of lexical collocation errors in students' argumentative essays. The 151 errors are spread out into six types of collocation errors;. The results are provided in table 1.

Table 1. The types of collocation errors

Types of collocation errors	Number	Percentage
Verb + Noun / Pronoun (PP),	20	13.25%
Adjective + Noun,	25	16.56%
Adverb + Adjective,	8	5.30%
Noun + Noun,	89	58.94%
Verb + Prep	4	2.65%
Clause	5	3.31%
Number	151	100.00%

The most dominant type of collocation error was found in Noun + Noun combinations, comprising 89 instances, or 58.94% of the total. This indicates a strong tendency among translators to mispair nouns in ways that do not align with idiomatic or collocational English usage. Errors such as “*oral culture tradition*” instead of “*oral tradition*”, “*Minangkabau society*” instead of “*Minangkabau people*” and “*performance concept*” instead of “*concept of performance*” are typical examples, revealing a direct translation pattern from Indonesian to English without proper collocational awareness.

The second most frequent error type was Adjective + Noun collocations, with a total for 25 cases (16.56%). These errors often involved awkward or uncommon modifiers such as “*sharp language expressions*” instead of the more idiomatic “*sarcastic remarks*” or “*biting comments*.” Such issues suggest a lexical mismatch due to either over-literal translation or limited awareness of natural adjective-noun pairings in English.

Verb + Noun / Pronoun errors were also significant, with 20 occurrences (13.25%). A recurrent example includes “*used as a form of performance*”, which, although grammatically correct, is semantically awkward and unnatural. More suitable alternatives would be “*regarded as a type of performance*” or “*categorized as a performance*.” This pattern suggests a partial understanding of verb complementation in collocational contexts.

Less frequent but still notable were errors in Adverb + Adjective combinations, with 8 instances (5.30%). Phrases like “*mutually sarcastic*” and “*really good at talking*” reflect non-native patterns that deviate from idiomatic usage, such as “*exchanging sarcastic remarks*” or “*eloquent speakers*.” These errors may stem from learners applying Indonesian adverbial patterns directly to English.

Discussion

The analysis of students' translations of a culturally embedded Indonesian text reveals a range of collocational problems, which can be classified based on linguistic categories such as lexical and grammatical collocations. From a lexical standpoint, numerous errors emerge in common patterns like noun + noun, adjective + noun, and verb + noun combinations. In the students' English renderings, these combinations frequently appear unnatural or unidiomatic. The problem in pairing the words illustrates a problematic collocation combination among the students. It indicates that students still face difficulties in producing natural collocations in L2. The collocation produced in L2 fails to reflect natural English usage. For instance, learners often apply direct translation strategies from Indonesian, assuming structural or lexical equivalence, which leads to awkward or semantically imprecise phrases in English. This finding supported previous research conducted by Bartan that L1 plays an important role in the collocation errors in the target language (Bartan, 2019).

The study underscores the pedagogical need for explicit collocation instruction in translation training, particularly for language pairs with significant structural and cultural

disparities. It advocates for integrating corpus tools into classrooms to raise students' awareness of natural collocations and mitigate L1 interference. Practical recommendations include designing collocation-focused exercises, contrastive analysis tasks, and discipline-specific phrase banks to enhance translation accuracy and fluency. In this case, according to Abdullah et al. (2021) In teaching collocation, both lexis and grammar should be learned and taught simultaneously.

Overall, the collocation errors found in the student's translation can be caused by several keys: interlanguage interference, literal translation strategies, limited exposure to authentic collocational patterns, and mismatch in register whenever the expression is translated literally. These findings align with previous studies on L2 collocational acquisition, which emphasize the challenge of mastering formulaic sequences and idiomaticity in translation. From a pedagogical perspective, these issues highlight the importance of integrating corpus-informed approaches and explicit collocation instruction in translator training programs. Some cultural expressions in the ST demand not only lexical equivalence but also pragmatic awareness and sensitivity to discourse conventions in the target language. Thus, building learners' collocational competence is essential for achieving both linguistic accuracy and cultural fidelity in translation practice (Feng, 2020, p. 105). It is because knowledge of collocations plays a significant role in language users' communicative competence, especially in using language correctly in the correct place.

These findings align with previous research on second-language collocational competence. For instance, Frankenberg-Garcia (2018) emphasizes lexical competence, which is indicated by the ability to arrange words correctly in a text. Similarly, Ağçam & Özkan (2015) found that L2 learners frequently underuse or misuse adjectives due to their unfamiliarity with combinations. One of the primary difficulties in translation is not the absence of vocabulary but the mismatch of conceptual frameworks across languages and cultures (Corpas, 2015). Such mismatches demonstrate that collocational errors are not merely linguistic but also cultural and conceptual.

In this case, the collocation difficulties arise because the collocation patterns in L1 differ from those in L2. As a result, students are often influenced by the collocational pattern of their first language. In relation to this phenomenon, Gablasova, Brezina, & McEnery (2017) propose a method to measure and interpret collocation use in both L1 and L2 production. This information will be important for the students in writing and translating into the language they learned. Moreover, based on the current research findings and previous studies, these challenges should be addressed by integrating a corpus-informed approach and explicit instruction in collocation into translation subjects (Mcenery & Xiao, 2011).

Comparatively, studies such as Abdullah et al. (Abdullah et al., 2021) stress specifically Verb-Noun collocations in academic writing should not only include the semantic elements, but also the syntactical elements of the verbs. Moreover, unlike more mechanical translation errors, these collocational failures risk misrepresenting the communicative intention and social function of the source expression, thus affecting translation fidelity.

These results suggest several implications for both translation studies and language pedagogy. First, collocational competence should be treated as a core skill in translator training (Herteg & Popescu, 2013), especially for learners working with culturally loaded texts. Second, corpus-informed approaches and contrastive analysis can help learners become more sensitive to native-like usage and the collocational behaviour of culturally specific terms. Lastly, the teaching of translation should go beyond syntax and lexis, incorporating pragmatic, cultural, and genre-based instruction to equip students with the tools needed for accurate and natural translation (McEnery, Brezina, Gablasova, & Banerjee, 2019).

These findings align with previous studies on L2 collocational acquisition, which emphasize the challenge of mastering formulaic sequences and idiomaticity in translation. From

a pedagogical perspective, these issues highlight the importance of integrating corpus-informed approaches and explicit collocation instruction in translator training programs. Thus, building learners' collocational competence is essential for achieving both linguistic accuracy and cultural fidelity in translation practice (Frankenberg-Garcia, 2018; Harta et al., 2021). Pedagogically, this reminds instruction on English collocations, idiomatic expressions, and natural phraseology to internalize typical English word combinations rather than relying on direct equivalents from their L1. It is in line with research conducted in developing intercultural communication through corpus-based pedagogic learning activities (Herțeg & Popescu, 2013).

The influence of the mother tongue (Indonesian) leads to applying L1 grammatical and lexical rules inappropriately to English (L2), causing errors in morphology, syntax, and semantics. Language instructors should raise learners' awareness of cross-linguistic differences and train them to recognize and overcome interference through contrastive analysis and error correction strategies, including incorrect sentence structures and improper collocation used. The material about collocation should be integrated with writing, grammar, and translation practice, emphasizing sentence construction and morphological rules to reduce errors and improve accuracy (Harta et al., 2021). Many errors arise from misunderstanding the context or pragmatic meaning of the source text, leading to inappropriate lexical choices or misinterpretation.

Conclusion

This study has explored the challenges of collocational difficulties in Indonesian-English translation. The analysis was conducted on students' translations of the culturally specific term *bagurau* as the primary data. The analysis revealed that many of the translation outputs contained unnatural or awkward collocations, particularly within common lexical patterns such as verb + noun, noun + noun, and adjective + noun combinations. These errors were frequently the result of literal translation strategies, first-language interference, and a limited awareness of native-like collocational norms in English. The findings emphasize that achieving idiomatic and fluent translations requires more than grammatical accuracy. It involves a deep understanding of collocational restrictions, cultural connotations, and pragmatic appropriateness in the target language. The inaccurate renderings of this term into formal or performative English expressions not only distort its meaning but also reflect the translators' reliance on structurally equivalent, but contextually inappropriate, lexical items.

The implications of this research are twofold. First, it calls for greater emphasis on collocation instruction within translator experience and knowledge, particularly for learners working with culture-specific source texts. Second, it supports the integration of corpus-based tools, such as BNC and COCA, and contrastive linguistic analysis to expose students to authentic usage patterns and prevent reliance on intuition or literal transfer. Language educators and trainers should promote the learning of collocations not just as fixed expressions but as context-bound realizations of meaning that vary across genres and cultures. Finally, this study contributes to the broader field of translation and applied linguistics by highlighting the intersection between lexical competence and intercultural communication. Future research may extend this work by exploring collocational patterns across different genres (e.g., literature, political texts, news discourse), or by assessing how corpus-informed learning impacts student performance in translating culturally embedded terms. Understanding and addressing collocational problems is essential not only for producing more natural translations but also for fostering translators' deeper engagement with both languages involved.

References

- Abdullah, S., Abdul, R., & Kamaruddin, R. (2021). Lexical verbs in verb-noun collocations: Empirical evidence from a Malay ESL Learner Corpus. *3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies*, 27(4), 144–156. <https://doi.org/10.17576/3L-2021-2704-11>
- Ağçam, R., & Özkan, M. (2015). A Corpus-based Study on Evaluation Adjectives in Academic English. *Procedia - Social and Behavioral Sciences*, 199, 3–11. <https://doi.org/10.1016/j.sbspro.2015.07.480>
- Al-Salman, S., & Haider, A. S. (2024). Assessing the accuracy of MT and AI tools in translating humanities or social sciences Arabic research titles into English: Evidence from Google Translate, Gemini, and ChatGPT. *International Journal of Data and Network Science*, 8(4), 2483–2498. <https://doi.org/10.5267/J.IJDNS.2024.5.009>
- Bartan, Ö. Ş. (2019). Lexical collocation errors in literary translation. *Dil Dergisi*, 170(1), 71–86. Retrieved from <https://www.ceeol.com/search/article-detail?id=777747>
- Bermúdez Bausela, M. (2016). The importance of corpora in translation studies: A practical case. In A. Pareja-Lora, C. Calle-Martínez, & P. Rodríguez-Arancón (Eds.), *New perspectives on teaching and working with languages in the digital era* (pp. 363–374). Dublin: Research-publishing.net. <https://doi.org/10.14705/rpnet.2016.tislid2014.448>
- Bui, L. T. (2021). The role of collocations in the English teaching and learning. *International Journal of TESOL & Education*, 1(2), 99–109. <https://doi.org/10.11250/ijte.01.02.006>
- Çakmak, Z., & Ersanlı, C. Y. (2021). *A corpus-based study on the authenticity of dialogues in the B1-B2 levels ELT coursebooks used in Turkey*. 13(3), 3212–3230.
- Corpas, G. (2015). Translating English verbal collocations into Spanish. *Lingvisticae Investigationes*, 38(2), 229–262. <https://doi.org/10.1075/li.38.2.03cor>
- Demirel, E. T., & Kazazoğlu, S. (2015). The comparison of collocation use by Turkish and Asian learners of English: The case of TCSE corpus and icnale corpus. *Procedia - Social and Behavioral Sciences*, 174, 2278–2284. <https://doi.org/10.1016/j.sbspro.2015.01.887>
- El-Siddig, M. (2023). Difficulties in translating and employing word collocation for undergraduate students at Al-Baha University. *Journal of Umm Al-Qura University for Language Sciences and Literature*, (32), 211–218. <https://doi.org/10.54940/ll60338357>
- Ermanto, JR, V. T. D., Ardi, H., Juita, N., & Jamaluddin, N. (2024). Quantitative lexical comparison of novels of Indonesian writers before independence and the post-reformation period: study using a corpus linguistic approach. *Cogent Arts and Humanities*, 11(1). <https://doi.org/10.1080/23311983.2024.2412404>
- Febriana, I., & Maisarah, I. (2021). An analysis of students' collocation translation in EFL classroom. *International Journal of Language Pedagogy*, 1(1), 46–53. <https://doi.org/10.24036/ijolp.v1i1.10>
- Feng, H. (2020). *Form, meaning and function in collocation*. Oxon: Routledge. <https://doi.org/10.4324/9780429318368-1>
-

-
- Frankenberg-Garcia, A. (2018). Investigating the collocations available to EAP writers. *Journal of English for Academic Purposes*, 35, 93–104. <https://doi.org/10.1016/J.JEAP.2018.07.003>
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67, 155–179. <https://doi.org/10.1111/LANG.12225>
- Harta, I. G. W., Bay, I. W., & Ali, S. W. (2021). An analysis of lexical collocation errors in students' writing. *TRANS-KATA: Journal of Language, Literature, Culture and Education*, 2(1), 15–25. <https://doi.org/10.54923/transkata.v2i1.18>
- Herțeg, C., & Popescu, T. (2013). Developing students intercultural and collocational competence through analysing the business press. A corpus-based pedagogic experiment among Romanian MA students. *Procedia-Social and Behavioral Sciences*, 83, 378–382. <https://doi.org/10.1016/j.sbspro.2013.06.074>
- Kemppanen, H. (2025). Empirical and quantitative approaches to corpus translation studies. In D. Li & J. Corbett (Eds.), *The Routledge Handbook of Corpus Translation Studies* (pp. 11–29). New York: Routledge. <https://doi.org/10.4324/9781003184454>
- Kheirzadeh, S., & Marandi, S. (2014). Concordancing as a tool in learning collocations: The case of Iranian EFL learners. *Procedia-Social and Behavioral Sciences*, 98, 940–949. <https://doi.org/10.1016/j.sbspro.2014.03.503>
- Khoo, C., & Johnkhan, S. (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, 44(4), 491–511. <https://doi.org/10.1177/0165551517703514>
- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694–716. <https://doi.org/10.1093/applin/amn018>
- Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, 61(2), 647–672. <https://doi.org/10.1111/J.1467-9922.2010.00621.X>
- Liu, D., & Tang, X. (2024). Comparative linguistic analysis with Firthian collocations: Cases of synonym differentiation and proficiency assessment. *Lingua*, 306, 103755. <https://doi.org/10.1016/J.LINGUA.2024.103755>
- McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus linguistics, learner corpora, and SLA: Employing technology to analyze language use. *Annual Review of Linguistics*, 39, 74–92. <https://doi.org/10.1017/S0267190519000096>
- Mcenery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In *Handbook of Research in Second Language Teaching and Learning* (pp. 364–380). Taylor & Francis. <https://doi.org/10.4324/9780203836507.ch22>
- Mohammed, T. A. S. (2022). The use of corpora in translation into the second language: A project-based approach. *Frontiers in Education*, 7, 849056. <https://doi.org/10.3389/educ.2022.849056>
- Obeidat, A. M., & Mahadi, T. S. (2020). The English translation of idiomatic collocations in
-

-
- the noble Quran: Problem and solutions. *Issues in Language Studies*, 9(2), 78–93. <https://doi.org/10.33736/ils.2246.2020>
- Park, C., Shim, M., Eo, S., Lee, S., Seo, J., Moon, H., & Lim, H. (2022). Empirical Analysis of Parallel Corpora and In-Depth Analysis Using LIWC. *Applied Sciences* 2022, Vol. 12, Page 5545, 12(11), 5545. <https://doi.org/10.3390/APP12115545>
- Saldanha, G., & Sharon O'Brien. (2013). *Research Methodologies in Translation Studies*. New York: Routledge.
- Setiarini, N. L. P. (2017). The translation of English collocation into Indonesian in a bilingual recipe: verbal node and collocate. *UNNES 6th ELTTLT International Conference Proceedings*, (October), 375–380. Semarang. Retrieved from <http://eltlt.proceedings.id/index.php/eltlt/article/download/151/136>
- Sutter, G. De, & Lefer, M.-A. (2019). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, 1–23. <https://doi.org/10.1080/0907676X.2019.1611891>