

Lexical Error Patterns in L2 Writing: Evidence from English Majors at a University in Vietnam

Le Thi Giao Chi, *Nguyen Hoang Nhi

The University of Danang – University of Foreign Language Studies, Vietnam

*Corresponding author

DOI: <https://dx.doi.org/10.51244/IJRSI.2026.1305000043>

Received: 24 April 2026; Accepted: 07 May 2026; Published: 26 May 2026

ABSTRACT

Academic writing has always been a challenge for English as a Foreign Language (EFL) learners, particularly at the tertiary level where lexical accuracy and appropriateness are essential. Among the most recurring difficulties, lexical errors emerge as indicators of gaps in vocabulary knowledge and instability within learners' interlanguage systems. Despite growing interest in L2 writing, relatively few studies have examined lexical errors in authentic, school-based contexts. This study investigates patterns of lexical errors in a learner corpus comprising 40 argumentative essays written by English majors at a public university in Vietnam. Adopting a descriptive, corpus-based approach, the analysis draws on a taxonomy adapted from Llach (2005), categorizing errors into semantic, syntactic, orthographic, and pragmatic types, with further distinction between interlingual and intralingual influences. The findings reveal that orthographic errors are the most prevalent, and that intralingual errors significantly outnumber those attributable to L1 transfer. These patterns indicate that learners' difficulties are primarily developmental, reflecting incomplete lexical representations and unstable form-meaning mappings. The results further point to limited consolidation of word forms, morphological patterns, and context-appropriate usage. These findings highlight the need for pedagogical approaches that prioritize lexical stability through systematic vocabulary instruction with emphasis on morphological awareness, orthographic regularities, and contextualized lexical repertoire. Addressing these underlying dimensions of lexical competence may contribute to reducing recurrent errors and enhancing both the accuracy and fluency of academic writing in EFL contexts.

Keywords: *lexical errors, EFL context, taxonomy of lexical errors, academic writing, Vietnamese learners*

BACKGROUND TO THE STUDY

Writing is a foundational mode of human communication that extends beyond the functions of oral discourse, both in its structural complexity and its capacity to preserve knowledge over time. Rather than serving merely as a transcription of speech, writing reshapes cognition and enables the development and transmission of ideas (Manchón & Polio, 2022). In academic settings, writing plays a pivotal role in knowledge construction and dissemination, particularly in global academia, where English is the prevailing medium of scholarly exchange. For EFL learners, developing proficiency in academic writing remains a persistent challenge. Since vocabulary constitutes a core component of language use, lexical errors, ranging from inappropriate word choice, spelling inaccuracies to collocational misuse, can significantly affect the clarity, coherence, and overall quality of written texts (Cordero & Arguedas, 2020).

In the context of Vietnam's increasing integration into the global academic community, English has assumed a central role in scholarly exchange and communication. As noted by Hoang (2020), the number of Vietnamese academics having publications in English has grown significantly, reflecting broader institutional as well as national priorities. In this context, Vietnamese students are increasingly expected to develop a high level of academic writing proficiency to meet these evolving standards. As English is instructed as a foreign language, many students continue to encounter challenges in producing accurate and effective written texts (Le & Trinh, 2025).

At the University of Danang - University of Foreign Language Studies (UD-UFLS), academic writing constitutes a central component of English language programmes. English majors are systematically required to produce a range of essays and research-oriented texts as part of their coursework, reflecting the institutional emphasis on advanced written proficiency. The quality of these written inputs serves not only as an indicator of individual linguistic competence but also a measure of the institution's academic standards. While previous studies conducted within and beyond this context have examined aspects of lexical use like lexical bundles (Le & Ta, 2010) or specific error types like collocation (Tran, 2021; Do & Le, 2023), these studies tend to focus on isolated features or standardized testing contexts. Moreover, many studies attribute lexical errors primarily to L1 interference (Nguyen & Vu, 2024), overlooking the broader range of developmental and intralingual factors that may influence learners' interlanguage systems.

Statement of the Problem and Research Purpose

The present study investigates lexical error patterns in L2 writings amongst English majors at UD-UFLS. The study aims to identify common types of lexical errors and examine the underlying factors contributing to their occurrence. In doing so, it seeks to provide a more comprehensive understanding of lexical development in EFL academic writing and to inform more targeted, effective vocabulary instruction. Accordingly, the study addresses two research questions: (i) What common lexical errors can be identified in L2 writings by English majors at UD-UFLS? and (ii) What underlying factors contribute to the occurrence of these lexical errors?

LITERATURE REVIEW

A substantial body of research in second language acquisition has consistently highlighted the importance of lexical competence in determining writing quality. Early studies primarily focused on lexical richness, measured through vocabulary size, diversity, and sophistication (Laufer & Nation, 1995; Read, 2000). While these dimensions remain important, recent research has shifted attention toward lexical accuracy and the role of lexical errors in shaping the communicative effectiveness of L2 writing (González-Fernández & Schmitt, 2020; Tran, 2021). This shift reflects growing recognition that accurate, context-appropriate vocabulary use is as important as lexical breadth.

In this line of inquiry, corpus-based approaches have advanced, providing empirical insights into learners' actual language use. Studies employing learner corpora have examined various aspects of lexical development, including lexical complication (e.g. Kyle & Crossley, 2014), and lexical diversity as well as formulaic competence (e.g. Paquot & Granger, 2012; Bestgen, 2017; Bestgen & Granger, 2018). These studies demonstrate that higher-quality writing is typically associated with greater lexical variation, more sophisticated vocabulary use, and more accurate deployment of formulaic sequences. However, corpus-based research has been inclined to measure lexical richness and complexity rather than systematically investigating lexical errors and their underlying causes. As a result, the relationship between lexical error patterns and learners' developing interlanguage systems remains insufficiently explored.

From a pedagogical perspective, early work by Ferris (1995) established that targeted feedback on specific error types leads to more effective improvement in writing accuracy than generalized correction. Though influential as it may sound, such research primarily addresses the outcomes of error correction rather than the cognitive and developmental mechanisms. More recent studies attempt to bridge this gap by examining lexical inaccuracies as manifestations of learners' evolving interlanguage. For example, Cordero and Arguedas (2020) found that although learners can expand their lexical repertoire, achieving precision in word choice remains a big challenge, particularly in terms of semantic misuse. Similarly, Quinn (2020) demonstrated that even advanced learners struggle with collocations and phraseological patterns, often relying on one-to-one lexical substitutions without sufficient attention to contextual constraints. These findings are compatible with corpus-based evidence suggesting that learners often lack depth of lexical knowledge, particularly in relation to collocational behavior and usage patterns (Kyle & Crossley, 2014; Granger & Paquot, 2015).

From a theoretical standpoint, such patterns can be understood through the lens of interlanguage theory (Selinker, 2009) and models of lexical acquisition (Jiang, 2000). For instance, Nation (2022) emphasize that lexical competence entails not only breadth but also depth, including knowledge of collocations, register, and context-based usage. Similarly, González-Fernández and Schmitt (2020) demonstrates that learners frequently establish

basic form-meaning connections while struggling to integrate morphological and syntactic information, leading to unstable lexical representations. From a dynamic systems perspective, lexical development is characterized by variability and non-linear progression, further explaining the persistence of such errors (Fogal & Verspoor, 2020).

In the Vietnamese EFL context, existing research has largely examined specific sub-types of lexical errors, such as collocational inaccuracies or performance in standardized writing tasks (Do & Le, 2023; Tran, 2021). Moreover, these studies often attribute lexical errors predominantly to L1 interference (Nguyen & Vu, 2024), potentially overlooking the role of intralingual and developmental factors. Put it differently, there remains an absence of localized, corpus-based research that provides a comprehensive account of lexical error patterns in authentic academic writing produced by English majors. Addressing these gaps, the present study adopts a corpus-based approach to analyse lexical error patterns in argumentative essays written by English majors at UD-UFLS. By integrating a comprehensive error taxonomy with an interlanguage-informed perspective, the study aims to provide a more nuanced understanding of lexical development, thus informing pedagogical practices that enhance lexical stability in EFL academic writing.

THEORETICAL BACKGROUND

Lexical Competence in L2 Writing

Lexical competence is widely recognised as a central component of second language acquisition (Read, 2000; González-Fernández & Schmitt, 2020). In academic writing, it is best understood as a multidimensional construct comprising both *breadth* (vocabulary size) and *depth* (quality of lexical knowledge), including semantic precision, collocational control, morphological awareness, and syntactic behavior (Wesche & Paribakht, 1996). This perspective positions vocabulary not as static knowledge, but as a dynamic resource deployed in meaning construction.

In L2 writing research, lexical competence is operationalised through measures such as lexical diversity, sophistication, and originality, which have been shown to correlate with writing quality (Laufer & Nation, 1995). These measures, however, primarily capture the distributional aspects of lexical use and do not fully account for the appropriateness of lexical selection in context.

This limitation foregrounds lexical accuracy as a core dimension of lexical competence. As argued by Read (2000), effective writing depends not only on the lexical range but also on accurate and contextually appropriate deployment. His framework highlights four features of successful lexical performance: (i) a substantial presence of content words, (ii) appropriate use of less frequent vocabulary, (iii) lexical variety with minimal redundancy, and (iv) a low incidence of lexical errors. While the first three reflect productive capacity, the fourth draws attention to deviation that may disrupt meaning making. Accordingly, lexical errors constitute a critical interface between lexical knowledge and language use, providing a principled entry point for examining learners' lexical competence in writing.

Lexical Errors

Lexical errors are widely regarded as particularly consequential forms of linguistic deviance due to their direct impact on meaning. Unlike grammatical inaccuracies, which may not always impede comprehension, inappropriate lexical choices can distort propositional content and reduce communicative effectiveness (Llach, 2005; Boers, 2021).

Such errors may extend beyond surface-level issues (e.g., spelling) to include incorrect word forms, semantic imprecision, as well as improper word choice or collocation. These deviations typically reflect incomplete or unstable lexical knowledge, especially in L2 contexts (González-Fernández & Schmitt, 2020). As such, lexical errors should not be treated merely as isolated inaccuracies, but as indicators of lexical development.

To capture this complexity, a systematic and qualitative approach to error analysis is necessary. The present study adopts a taxonomy-based framework proposed by Llach (2005) to examine lexical errors in relation to their linguistic properties and underlying processes.

Taxonomy of Lexical Errors

Llach (2005) proposes a comprehensive taxonomy that classifies errors into four categories - semantic, syntactic, orthographic, and pragmatic, while distinguishing between interlingual (L1-based) and intralingual (developmental) sources. This framework is adopted in the present study due to its capacity to relate observable errors to inherent psycholinguistic processes.

The taxonomy conceptualises lexical competence as comprising four corresponding dimensions (i.e., semantic, syntactic, orthographic, and pragmatic), thereby providing a principled basis for analysing how lexical knowledge is represented and deployed. This multidimensional approach enables a systematic examination of both the form and source of lexical errors.

Interlingual and Intralingual Errors

This study is further grounded in the interlanguage framework (Guo, 2022), viewing learner language as a developing system shaped by both cross-linguistic influence and internal restructuring processes. Within this perspective, lexical errors are interpreted as evidence of evolving linguistic competence rather than mere communicative failure.

L1-oriented (interlingual) errors arise from cross-linguistic influence, whereby learners project L1 semantic boundaries, collocational patterns, and usage constraints onto the target language. This often results in inappropriate lexical selection and non-target-like usage due to assumed equivalence between L1 and L2 forms (Duskova, 1969, cited in Llach, 2005).

By contrast, L2-oriented (intralingual) lexical errors stem from the internal complexity of the target language and the learner's developing interlanguage system. These include overgeneralisation, simplification, and confusion between formally or semantically similar items, reflecting incomplete lexical differentiation (Laufer, 1990, cited in Llach, 2005).

Semantic Errors

Semantic errors involve a mismatch between the meaning of a selected lexical item and the actual meaning required by the context. These errors arise when learners grasp the general concept but fail to delineate the precise semantic boundaries of the target word (Zimmermann, 1986). These errors may result from interlingual transfer or intralingual confusion between semantically related items.

In L2 writing, such errors commonly manifest as inappropriate word choice, where a word form is linguistically valid but contextually irrelevant. These semantic inaccuracies are considered as disruptive type of lexical deviance (Cordero & Arguedas, 2020). As Boers (2021) notes, these semantic boundaries are resistant to rule-based instruction, as they require learners to restructure conceptual mappings shaped by their native language and to develop sensitivity to conventionalised patterns of usage in the target language.

Syntactic Errors

Syntactic errors pertain to the grammatical behaviour of a lexical item within sentence structures. These errors occur when semantically appropriate words are used in grammatically inappropriate manners, including incorrect word class assignment or violations of collocational and structural constraints (Lennon, 2009). That being said, syntactic errors reflect a breakdown not in meaning selection, but in the integration of lexical items into well-formed patterns.

L1-oriented syntactic errors result from the transfer of L1 grammatical patterns and word order into L2 writing, while L2-oriented errors reflect difficulties in mastering the syntactic constraints. Evidence suggests that even learners with substantial vocabulary knowledge may struggle with collocational accuracy, indicating that syntactic control remains a persistent challenge in L2 writing (Do & Le, 2023).

Orthographic Errors

Orthographic errors refer to inaccuracies in representing L2 spelling system and morphological conventions. These errors are often phonologically induced, arising from incorrect sound-spelling mapping rather than mispronunciation.

L1-oriented orthographic errors reflect cross-linguistic influence, particularly in the assumption of orthographic correspondence between languages. This is especially relevant for Vietnamese learners, given the typological contrast between Vietnamese as an isolating language and English as a morphologically complex system marked by derivational and inflectional topologies (Lieber, 2021).

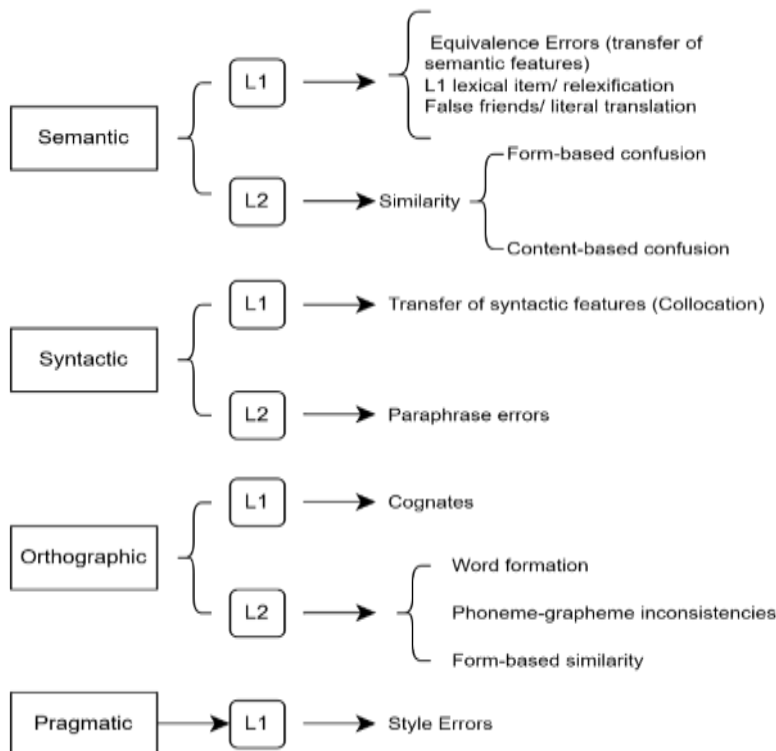
L2-oriented orthographic errors, by contrast, stem from the irregularity of English orthography, including phoneme-grapheme inconsistency and morphological complexity, either in morphological marking (e.g., prefixes & suffixes) or in formal similarity of lexical items.

Pragmatic Errors

Pragmatic errors involve inappropriate lexical use according to context, register, and discourse (Llach, 2005). Even when semantically and syntactically correct, lexical choices may be pragmatically ineffective if they fail to meet with the stylistic expectations of academic writing.

These deviations are often associated with pragmatic transfer, where learners apply L1 sociolinguistic norms, rhetorical preferences, and interpersonal conventions to the target language. This frequently results in the use of informal or affectively marked expressions in academic contexts. For Hinkel (2020), EFL learners tend to select lexical items that adequately convey propositional meaning but carry informal, conversational, or even affectively marked connotations (e.g., *boss* instead of *employer*, or *a huge deal* instead of *a significant issue*). While such errors may not obscure meaning directly, they reduce appropriateness and weaken rhetorical nuance. *Figure 1* shows the lexical error taxonomy adapted from Llach (2005).

Figure 1. Taxonomy of Lexical Errors (Adapted from Llach (2005))



RESEARCH METHODOLOGY

This study adopts a descriptive, corpus-based research design that integrates qualitative error analysis with quantitative frequency tabulation. Rather than relying on holistic writing scores, the study employs qualitative

procedures to identify and categorize lexical deviations, including their underlying psycholinguistic processes (L1 and L2-oriented errors). Quantitative analysis is then applied to determine the frequency and distribution of these error types. This combined approach enables the identification of dominant errors patterns while providing both descriptive and explanatory insights (Creswell, 2009).

The research was conducted at the University of Danang - University of Foreign Language Studies (UD-UFLS). The learner corpus comprises 40 argumentative essays randomly selected from the Writing Skill C1.2 course within the Faculty of English. The data were drawn from handwritten final-term examinations completed by undergraduate English majors at B2-C1 proficiency levels according to the Common European Framework of Reference for Languages (CEFR). This controlled dataset ensures that the identified lexical errors reflect learner proficiency rather than variation in genre or task conditions. As noted by Flowerdew (2004), small, specialized corpora are particularly effective for analyzing specific genres and pedagogical contexts. What is more, focusing on a single cohort and task type (i.e., academic essays) allows the researcher to yield more precise identification of lexical error patterns. *Table 1* details the corpus size and word count of the UD-UFLS learner corpus.

Table 1. Summary of Corpus Size and Word Count for UD-UFLS English Majors' Essays

Group	Number of Essays	Min	Max	Mean	SD
UD-UFLS	40	212	416	299.20	50.72

The analytical framework combines the interlanguage perspective (Guo, 2022) with the adapted version of Llach's (2005) taxonomy. This enables a two-stage analysis in which errors are first classified according to their linguistic properties - semantic, syntactic, orthographic, and pragmatic - and then interpreted in terms of their psycholinguistic origins. Given the focus on error formation, manual annotation was employed in place of automated corpus tools. Each essay was systematically examined, and all lexical deviations were coded according to the four listed categories, followed by classification as either interlingual (L1-oriented) or intralingual (L2-oriented). This dual-layered coding procedure allows for systematic analysis of both the form and source of lexical errors.

To ensure data integrity, a strict transcription protocol was established. All lexical, morphological, and syntactic features were transcribed verbatim to preserve the authenticity of learner production. Words that had been crossed out or revised during the examination were deliberately excluded from the final corpus, ensuring that the analysis reflects only the learners' final lexical choices.

RESEARCH FINDINGS AND DISCUSSION

The following section presents the study findings in conjunction with their discussion, beginning with a quantitative overview of the identified lexical deviations. This is complemented by a qualitative analysis of some examples illustrating typical lexical error patterns. These cases highlight recurrent lexical difficulties and provide insight into the challenges encountered by students in L2 writing.

Figure 2. Distribution of Lexical Error Types in Percentages

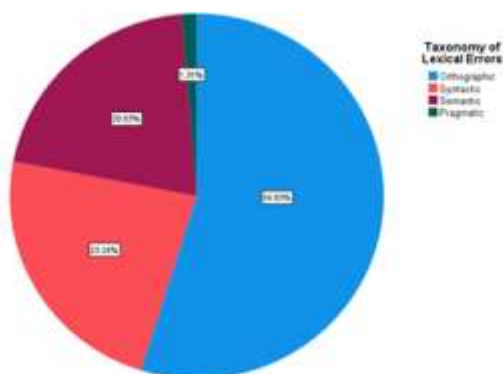


Table 2. Classification of Lexical Errors by Category and Linguistic Origin

Error Type	Subtype	Frequency (=n)	Percentage (%)
Orthographic	L2 Word formation	161	54.83
	L2 Phoneme-grapheme inconsistencies	45	
	L2 Similary (Form-based confusion)	4	
Syntactic	L1 Transfer	43	23.24
	L2 Paraphrase	46	
Semantic	L1 Equivalence	61	20.63
	L2 Similarity (Form-based confusion)	10	
	L2 Similarity (Content-based confusion)	8	
Pragmatic	L1 Style	5	1.31
Total		383	100

Figure 2 and **Table 1** present the distribution of the 383 lexical errors identified across 40 sampled essays drawn from the broader corpus of L2 writing produced by English majors at UD-UFLS. The findings reveal a pronounced concentration of errors in the orthographic category, which accounts for over half of the identified errors (almost 55%). Syntactic and semantic errors follow at 23.24% and 20.63% respectively, while pragmatic errors constitute only a marginal proportion (1.3%). This distribution indicates that difficulties related to word form and spelling remain the most salient challenge in students’ academic writing.

Orthographic Errors

Orthographic errors constitute the largest proportion of the lexical deviations in the dataset, accounting for 54.83% (210 instances) of the total errors identified. Notably, all orthographic errors were classified as L2-oriented (intra-lingual), indicating that these inaccuracies stem from learners’ developing knowledge of the English lexical system rather than direct L1 transfer. These errors can be further categorized into three main subtypes: *word formation*, *phoneme-grapheme inconsistencies*, and *form-based confusion errors*.

L2 Word Formation

Errors in L2 word formation represent the majority of orthographic errors, with 161 occurrences. These errors primarily involve incorrect derivation, inflection, and the misuse of morphological forms.

(1) Artificial Intelligence (AI) are gradually replace traditional document due to quick response for student's demand on homework, paper or even report.

Example (1) illustrates two typical word formation errors. First, the learner produces the base verb *replace** instead of the required progressive form (*are gradually replacing*), suggesting a simplification strategy to reduce processing demands. Second, the learner generates the non-standard noun *response** by overgeneralizing the derivational suffix *-ence*. This reflects an attempt to extend morphological rules analogically, albeit inaccurately.

(2) Lastly, not every student can have access to proper equipments required for a usable learning experience.

Example (2) shows a failure to recognize the uncountable nature of *equipment*, resulting in an inappropriate plural form. These patterns suggest that learners are actively engaging in rule-based morphological processing, but their understanding of derivational and inflectional constraints remains unstable. Such errors represent developmental hypotheses about the target language rather than simple performance mistakes (Zhang, 2022).

The high frequency of these errors may also be partially explained by typological differences: Vietnamese, as an analytic language, lacks the inflectional and derivational morphology characteristic of English (Lieber, 2021). Accordingly, students may acquire entirely new morphological mechanisms, leading to overgeneralization or simplification during the developmental process, these are strategies of L2 learning to manage cognitive load by EFL learners (Guo, 2022).

L2 Phoneme-Grapheme Inconsistencies

A total of 45 errors were identified in this category. These errors arise when learners rely on phonological representations to infer spelling, resulting in mismatches between sound and conventional orthographic forms. This demonstrates how internal phonological representations can directly influence orthographic inaccuracies.

(3) *Addition, online learning can also save time and **conviniient** for people who work and study at the same time.*

In Example (3), *conviniient* is used instead of *convenient*. This reflects a reliance on phonological approximation rather than stored orthographic knowledge, particularly in contexts where vowel reduction obscures clear sound-spelling correspondences.

(4) *In my standpoint, I wholeheartedly align with the **later** statement.*

For Example (4), the learner confuses *later** and *latter*, two lexemes with similar phonological forms but distinct meanings and orthographic representations. This substitution affects textual cohesion, as the intended referential meaning is compromised. These errors highlight the challenges posed by the non-transparent phoneme-grapheme mapping in English. From a lexical acquisition perspective (Jiang, 2000), they indicate incomplete integration between phonological and orthographic representations in the mental lexicon. Given that Vietnamese exhibits more consistent sound-symbol correspondences due to its mono-syllabic nature (Lieber, 2021), learners may rely on phonological strategies that are less effective in English, leading to systematic orthographic inaccuracies.

L2 Similarity (Form-based Confusion)

Errors of form-based similarity occur when learners confuse lexemes that share orthographic or phonological resemblance, resulting in incorrect lexical selection. Consider the following instance.

(5) *people should be more careful when using social networking media these days to avoid unwanted **incidences** like cheating or loss of privacy.*

In Example (5), *incidences** is used instead of *incidents*, reflecting confusion between two visually and phonologically similar forms.

(6) *...especially if the account is of important personnels or big **bussinesses**.*

Example (6) showcases orthographic confusion involving consonant doubling (*bussinesses**), which is likely influenced by analogical reasoning based on familiar word patterns. Unlike word formation errors, these are *substitution* ones, showing that learners retrieve an incorrect lexical item despite having access to a semantically appropriate target. Such patterns suggest that learners' lexical breadth is expanding, as they attempt to use more advanced vocabulary, but their lexical depth remains underdeveloped, particularly in distinguishing closely related forms (González-Fernández & Schmitt, 2020).

Syntactic Errors

Syntactic errors constitute the second most frequent category, accounting for 23.24% (89 instances) of the entire dataset. Unlike orthographic errors, this category reflects a relatively balanced distribution between *L1-oriented* (43 cases) and *L2-oriented* (46 cases) sources. This suggests that learners' interlanguage development at the

syntactic level is shaped by both cross-linguistic influence and intralingual restructuring. Two principal subtypes were thus identified: *L1 transfer* and *L2 paraphrase errors*.

L1 Transfer

L1-oriented syntactic errors represent a substantial proportion of structural deviations, with 43 instances, accounting for nearly half of all syntactic deviations. These errors primarily arise from the transfer of Vietnamese grammatical patterns into English.

(7) *For example, I can apply teach assist for **Center English**, and help children...*

Example (7) illustrates a typical case of word order transfer. The phrase *Center English** reflects the Vietnamese structure *trung tâm tiếng Anh*, where the head noun precedes its modifier. In contrast, English requires a modifier–head structure (*English center*). This error demonstrates that learners are not producing random deviations but are systematically applying L1 syntactic rules within their L2 production.

(8) *For example, in poor area, some **families even difficult** to build a stable life.*

Example (8) reflects copula omission, a well-documented feature of Vietnamese syntax, where adjectives can function as predicates without a linking verb. The absence of the copula (*are*) indicates that learners are transferring this structural feature into English, resulting in non-target-like constructions. From an interlanguage perspective, these errors represent *systematic transfer strategies*, where learners rely on familiar L1 structures to compensate for incomplete L2 knowledge (Guo, 2022). This suggests that lexical development alone is insufficient, that is learners must acquire syntactic environments in which lexical items operate. In this sense, it is important that targeted instruction explicitly shows contrasts between the Vietnamese and English grammatical patterns, which is to support further interlanguage development.

L2 Paraphrase

L2 paraphrase errors, with 46 instances, originate from learners' internal processing of English rather than L1 influence. Unlike L1 transfer, these errors occur when learners attempt construct meaning using available lexical resources but fail to conform to the collocational and grammatical constraints of English.

(9) *In addition, online learning **affects on** students' performance.*

For Example (9), the addition of the preposition *on* reflects confusion between the verb *affect* and the noun *effect*, the latter of which commonly co-occurs with *on*. This indicates partial knowledge of lexical relationships but incomplete control over their syntactic realization.

(10) *However, this development **poses certain threats on** online users, ...*

In contrast to the overuse of prepositions as seen in the previous examples, (10) demonstrates incorrect preposition selection where *threats on** is used rather than the standard collocation *threats to*. This likely results from the assumption that semantically related expressions such as *impact on* or *influence on* are analogous.

(11) *Regarding large-scale businesses that establish online platforms and social media, the main responsibility **lying with them** is to invest more in the security system and...*

In Example (11), the learner selects *lying with* instead of the more appropriate *resting with*. While semantically related, the chosen verb does not conform to the conventional collocation in formal academic usage.

Semantic Errors

Semantic errors account for 20.63% (or 79 instances) of the total, with a clear predominance of L1-oriented errors (61 cases) over L2-oriented errors (18 cases). This distribution suggests that lexical meaning selection

remains strongly influenced by learners' first language, particularly in contexts requiring precise and contextually relevant word choice.

L1 Equivalence

Semantic L1 equivalence errors represent the most frequent subtype within the semantic category, with more than two-thirds of the total semantic errors. These errors occur when learners rely on direct translation from Vietnamese, resulting in semantically imprecise or non-standard expressions in English.

(12) ...*I think classroom teaching will **have more relationship** than online learning...*

As can be seen in Example (12), the phrase *have more relationship** reflects a direct mapping from Vietnamese (*có nhiều mối quan hệ*), where general-purpose verbs such as *có* (have) are widely used. However, English language requires more specific collocations, such as *build* or *foster relationships*, particularly in academic contexts.

(13) *You should not click to any **strange links** because it can lead to the loss of personal accounts.*

Example (13) illustrates a case of semantic overextension where *strange* is used as a direct translation of '*lạ*' in Vietnamese, which can mean both *strange* and *suspicious*. While semantically related, *strange* lacks the contextual specificity of *suspicious*, which is more appropriate in cybersecurity discourse. These errors indicate that learners rely on broad semantic mappings derived from their L1 rather than selecting contextually precise lexical items in English. From an interlanguage perspective, this reflects a stage where lexical items are mediated through L1 conceptual categories (Jiang, 2000), resulting in semantic approximation rather than precision. Such patterns suggest that improving lexical accuracy requires not only expanding vocabulary but also developing sensitivity to register, collocation, and disciplinary usage.

L2 Similarity (Form-based Confusion)

In contrast to errors based on meaning, L2 similarity (form-based) errors occur when learners select a similar morphological form but inappropriate grammatical or semantic function.

(14) ...*as online learning can help with **distant learning** and class management while the traditional classroom...*

As clearly seen in (14), the student uses the adjective *distant* (describing a far-off place) instead of the noun-modifying form *distance* (as in *distance learning*), indicating confusion between related word forms within the same lexical family.

(15) *Many people, especially internet users, have concern about safety when accessing any websites or applications on **electric** devices.*

Example (15) demonstrates confusion between *electric* and *electronic*, two lexemes with overlapping meanings but distinct usage domains. Errors like this reflect partial control over the morphological and lexical distinctions, suggesting that learners recognize word families but have not fully internalized their functional differences. Such patterns align with Jiang's (2000) notion of incomplete lexical integration, where form and meaning are not yet fully aligned.

L2 Similarity (Content-based Confusion)

L2 similarity (content-based) errors arise from confusion between semantically related words within the English lexicon. Consider the following examples.

(16) *People can lose contact with their friends, relatives, colleagues and **counterparts**.*

(17) *That development creates chances for **modern crime** to hack and steal money from banking apps on phones.*

In (16), *counterparts* is used inappropriately to refer to general social or professional contacts whereas the term typically denotes individuals in equivalent positions across organizations. Similarly, (17) illustrates confusion between an abstract noun (*crime*) and its agentive form (*criminals*), resulting in a mismatch between intended meaning and lexical choice.

Unlike L1-based errors, these patterns indicate that learners are operating within the English lexical system, attempting to differentiate between semantically related items. While L1 equivalence errors show students' adherence to L1 system, from a developmental perspective, such errors can be interpreted as a positive indicator of lexical expansion when learners attempt to move beyond basic vocabulary and engage with more nuanced lexical distinctions. The persistence of these errors highlights the need for greater emphasis on semantic precision and contextual appropriateness in vocabulary instruction.

Pragmatic Errors

Finally, pragmatic errors were the least frequent, representing only 1.30% (5 instances) across 40 essays. All these identified cases were classified as L1-oriented, specifically reflecting pragmatic transfer at the level of style and register. These errors arise when learners draw on discourse conventions from Vietnamese - informal, expressive, or emphatic modes of communication – and apply them inappropriately within formal English academic writing.

(18) *By judging their friends just from the click of a button, many **straight up determine** whether they want to interact with that person or not without even directly interacting with them in real life.*

In Sample (18), the expression *straight up determine* introduces a colloquial tone that is misaligned with the conventions of academic discourse. While the sentence is grammatically acceptable, the use of *straight up* reflects an informal intensifier more typical of spoken interaction. This suggests a transfer of pragmatic norms associated with immediacy or emphasis in Vietnamese (e.g., *ngay lập tức* or *thẳng thừng*), resulting in a stylistic mismatch rather than a semantic or grammatical error.

(19) *This leads to many cases that are hacked and scammed over a year and even leads to **serious situations like death**.*

Example (19) illustrates a different type of pragmatic deviation, characterized by overly direct and emotionally loaded expression. In academic writing, sensitive topics are typically conveyed through more neutral and technical formulations (e.g., *fatalities* or *loss of life*). The use of *death* in this context, while not incorrect in meaning, lacks the degree of formality and rhetorical distancing expected in scholarly discourse.

DISCUSSION

From the presentation of the findings regarding different types of lexical errors, it is interesting to note that all orthographic errors were L2-oriented (intra-lingual), which stem from learners' developing knowledge of the English lexical system rather than direct L1 transfer. This exclusive classification of orthographic errors as L2-oriented is particularly noteworthy, suggesting that UD-UFLS students have largely progressed beyond reliance on L1 phonological transfer and are instead encountering challenges associated with the internal complexities of the target language. In line with Zhang's (2022) model of lexical development, these errors may reflect incomplete integration of form, meaning, and morphological information within the mental lexicon.

The high frequency of L2 word formation errors (161 instances) further supports this interpretation. Learners frequently attempt to generate more sophisticated vocabulary by applying derivational affixes (e.g., *-ence*, *-tion*) in non-target-like manners, resulting in inaccurate lexical representation. Such patterns indicate that while learners are actively expanding their productive vocabulary, their knowledge of morphological rules remains unstable. This aligns with Fernández and Schmitt's (2020) distinction between breadth and depth of vocabulary

knowledge, suggesting that learners possess sufficient lexical breadth to attempt more advanced forms, yet lack the depth required for accurate and context-relevant usage.

Alongside orthographic errors, syntactic and semantic errors reveal a complex interplay between L1 transfer and L2 developmental processes. Given that Vietnamese is an analytic language devoid of morphological inflection (Lieber, 2021), the typological distance between the learners' L1 and English strongly contributes to these challenges. While syntactic errors demonstrate the continued influence of Vietnamese structural patterns (e.g., *Center English* instead of *English Center*), semantic errors highlight the role of conceptual transfer and lexical approximation. At the same time, L2-oriented errors in both categories reflect learners' increasing engagement with the English system, albeit with incomplete mastery of its grammatical and collocational constraints. These patterns of transfer errors suggest that learners possess sufficient lexical resources to express complex ideas but lack mastery of collocational restrictions and syntactic compatibility. The findings echo the lexical perspective by González-Fernández and Schmitt (2020) and Nation (2022), stating that such errors reflect limitations in depth of vocabulary knowledge, particularly in understanding how words function in combination in context. They also add empirical insights into the interlanguage framework (Guo, 2022), treating L2 paraphrase errors as developmental approximations. The results align with Quinn (2020) suggesting that pedagogical interventions should move beyond vocabulary expansion to focus on collocational competence, syntactic integration, and lexical precision, thereby supporting the stabilization of learners' interlanguage and enhancing the overall quality of academic writing.

The low frequency of *pragmatic errors* (only 5 instances, making up only 1.3 %) suggests that students at the B2 to C1 level generally possess a functional awareness of the appropriate style for essay writing, even if the formal accuracy of their word choices may be inconsistent. Though meagre in number, these error patterns indicate that pragmatic competence extends beyond grammatical accuracy and lexical choice to include control over register, tone, and discourse conventions. From an interlanguage perspective (Guo, 2022), such errors reflect a stage in which learners have developed sufficient linguistic resources but have not yet fully internalized the sociolinguistic norms governing language use in specific contexts.

CONCLUSIONS AND IMPLICATIONS

This study provides a corpus-based analysis of lexical errors in the academic writings of English majors at UD-UFLS, drawing on 40 argumentative essays. By combining quantitative frequency analysis with qualitative classification based on Llach (2005), the research offers a systematic account of how lexical errors are distributed and what they reveal about learners' developing language systems.

The findings indicate that lexical difficulties at the B2-C1 level are predominantly intralingual and developmental rather than driven by direct L1 transfer. While orthographic errors remain the most frequent, followed by syntactic and semantic deviations, pragmatic errors are comparatively rare. These patterns suggest that learners have largely moved beyond reliance on L1-based strategies and are actively engaging with the structural and lexical complexities of English. However, their interlanguage remains unstable at the level of form-meaning integration, morphological control, and collocational accuracy, leading to recurring deviations in academic writing.

From a theoretical perspective, the study contributes to interlanguage research by demonstrating that lexical errors are not merely surface-level inaccuracies but reflect systematic developmental processes. The predominance of intralingual errors supports the view that learners are constructing and refining hypotheses about the target language, particularly in relation to morphological patterns, lexical selection, and syntactic compatibility. By linking error types to underlying issues in lexical representation and acquisition, the study provides a multi-faceted understanding of how vocabulary knowledge develops in advanced EFL contexts.

Empirically, the study contributes a localized learner corpus perspective from Vietnamese higher education, an area that remains underrepresented in corpus-based lexical research. Unlike studies that focus narrowly on specific error types, this research offers a comprehensive, multi-dimensional analysis across orthographic, syntactic, semantic, and pragmatic domains, thereby capturing the complexity of lexical performance in authentic EFL academic writing.

Pedagogically, the findings suggest that vocabulary instruction at advanced levels should move beyond the expansion of lexical breadth to prioritize lexical stability and depth. In particular, greater emphasis is needed on morphological awareness and word formation, collocational competence and syntactic constraints, as well as context-relevant word choice and academic register. Such an approach requires a shift from translation-based or decontextualized vocabulary teaching toward integrated, usage-based instruction that supports the consolidation of form-meaning relationships and reduces error recurrence in writing.

Despite these contributions, the study is subject to several limitations. The relatively small sample size (40 essays) limits the generalizability of the findings, and the classification of lexical errors inevitably involves interpretive judgment, particularly where lexical and grammatical boundaries overlap. Future research should therefore aim to expand the dataset across genres and proficiency levels to enhance representativeness. Longitudinal designs would be valuable in tracing the development and potential fossilization of lexical errors over time. In addition, combining corpus analysis with qualitative methods, such as stimulated recall, could provide deeper insight into the cognitive processes underlying learners' lexical choices.

REFERENCES

1. Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, 69, 65-78. <https://doi.org/10.1016/j.system.2017.08.004>
2. Bestgen, Y. & S. Granger (2018). Tracking L2 writers' phraseological development using collgrams: Evidence from a longitudinal EFL corpus. In Hoffmann, S., Sand, A., Arndt-Lappe, S. & Dillmann, L.M. (eds.) *Corpora and Lexis*. Brill, 277-301. <https://brill.com/abstract/title/36203>
3. Boers, F. (2021). *Evaluating second language vocabulary and grammar instruction: A synthesis of the research on teaching words, phrases, and patterns*. First Ed. Routledge. <https://doi.org/10.4324/9781003005605>
4. Cordero, D., & Arguedas, V. N. (2020). Lexical errors in EFL students' written production. *Letras*, 175–198. <https://doi.org/10.15359/rl.2-68.7>
5. Creswell, J. W. (2009). Editorial: Mapping the field of mixed methods research. *Journal of Mixed Methods Research*, 3(2), 95-108. <https://doi.org/10.1177/1558689808330883>
6. Do, N. H. M., & Le, Q. T. (2023). Lexical collocation errors in essay writing: A study into Vietnamese EFL students and their perceptions. *International Journal of Language Instruction*, 2(2), 101-118. <https://doi.org/10.54855/ijli.23226>
7. Ferris, D. (1995). Can advanced ESL students become effective self-editors? *CATESOL Journal*, 8(1), 41-62
8. Flowerdew, L. (2004). The argument for using English for Specific Purposes (ESP) corpora in ESP vocabulary research. *International Journal of Corpus Linguistics*, 9(1), 11-33.
9. Fogal, G. G., & Verspoor, M. H. (Eds.). (2020). *Complex dynamic systems theory and L2 writing development*. John Benjamins Publishing Company. <https://doi.org/10.1075/llt.54>
10. González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics*, 41(4), 481–505. <https://doi.org/10.1093/applin/amy057>
11. Granger, S., & Paquot, M. (2015). Electronic corpora and learner corpus research. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 145-159). Wiley.
12. Guo, Q. (2022). Interlanguage and its implications to second language teaching and learning. *Pacific International Journal*, 5(4), 08–14. <https://doi.org/10.55014/pij.v5i4.223>
13. Hinkel, E. (2020). *Teaching Academic L2 Writing: Practical Techniques in Vocabulary and Grammar* (2nd ed.). Routledge. <https://doi.org/10.4324/9780429437946>
14. Hoang, V. V. (2020). The Roles And Status Of English In Present-Day Vietnam: A Socio-Cultural Analysis. *VNU Journal of Foreign Studies*, 36(1). doi:10.25073/2525-2445/vnufs.4495
15. Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, 21(1), 47-77. <https://doi.org/10.1093/applin/21.1.47>
16. Kyle, K., & Crossley, S. A. (2014). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757-786. <https://doi.org/10.1002/tesq.194>
17. Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322.

18. Le, T. A., & Trinh, N. T. (2025). A student-centered framework for understanding EFL thesis writing difficulties in Vietnam. *International Journal of Learning, Teaching and Educational Research*, 24(11), 1-20. <https://www.ijlter.org/index.php/ijlter/article/view/14690>
19. Le, T. G. C., & Ta, T. T. A. (2010). *A study on lexical bundles in English and Vietnamese scientific writing*. The University of Danang - The University of Foreign Language Studies. Unpublished Master Thesis. <https://data.ufl.udn.vn/handle/UFL/1576>
20. Lennon, P. (2009). Error and the very advanced learner. *International Review of Applied Linguistics in Language Teaching*, 29(1), 31-44.
21. Lieber, R. (2021). *Introducing morphology*. Third Edition. Cambridge University Press.
22. Llach, M. D. P. A. (2005). The relationship of lexical error and their types to the quality of ESL compositions: An empirical study. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, (3), 45-57.
23. Manchón, R. M., & Polio, C. (Eds.). (2022). *The Routledge handbook of second language acquisition and writing*. Routledge. <https://doi.org/10.4324/9780429199691>
24. Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.
25. Nguyen, A. N., & Vu, T. V. (2024). Typical linguistic errors committed by tertiary students in legal written outputs. *East European Journal of Psycholinguistics*, 11(2), 120-141. <https://doi.org/10.29038/eejpl.2024.11.2.ngu>
26. Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130-149.
27. Quinn, C. C. (2020). *L2 writers referencing corpora to address accuracy: A qualitative analysis of learners' lexicogrammatical error corrections* [Doctoral dissertation, University of Birmingham]. UBIRA ETheses. <https://etheses.bham.ac.uk/id/eprint/11084/1/Quinn2020PhD.pdf>
28. Read, J. (2000). *Assessing vocabulary*. Oxford University Press.
29. Selinker, L. (2009). INTERLANGUAGE. *International Review of Applied Linguistics in Language Teaching*, 10(1-4), 209-232. <https://doi.org/10.1515/iral.1972.10.1-4.209>
30. Tran, T. T. (2021). *An analysis of lexical collocation errors in academic IELTS writing task 2 by Vietnamese students in an English center* [Master's thesis, Dalat University]. Dalat University Institutional Repository.
31. Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review*, 53(1), 13-40. <https://doi.org/10.3138/cmlr.53.1.13>
32. Zhang, X. (2022). The relationship between lexical use and L2 writing quality: A case of two genres. *International Journal of Applied Linguistics*, 32(3), 371-396. DOI:10.1111/ijal.12420
33. Zimmermann, R. (1986). Classification and distribution of lexical errors in the written work of German learners of English. *Papers and Studies in Contrastive Linguistics*, 21, 31-40.