

Corpus-Assisted Analysis of Lexical and Grammatical Features in ESL Students' Argumentative Essays

Noor Ain Saufi^{1*}, Abdul Azim Mahda², Sheela Faizura Nik Fauzi³, Nur Farah Fadhliah Mahmud⁴,
Aiman Zulaikha Mohd Fadzli⁵, Nur Nadhirah Mijan⁶

¹Akademi Pengajian Bahasa, UiTM Cawangan Sarawak

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.102100182>

Received: 05 September 2026; Accepted: 10 September 2026; Published: 06 October 2026

ABSTRACT

Argumentative writing requires ESL learners to balance clear reasoning with precise vocabulary and grammatical accuracy. This multi-layered task remains a major hurdle in language acquisition. While vocabulary and grammar are widely recognised as twin pillars of effective argumentation, research rarely examines how they interact within a single learner corpus, particularly among Malaysian ESL writers. Addressing this gap, this study uses a corpus-assisted approach to analyse the combined lexical and grammatical profiles of undergraduate argumentative essays from Universiti Teknologi MARA (UiTM). The dataset comprises 24 prompt-controlled essays (7,429 tokens) ranging between 300 and 350 words. Using TagAnt for part-of-speech tagging and AntConc for frequency distribution, collocation, and wordlist analysis, the study mapped key linguistic patterns across the texts. Lexically, students relied heavily on general nouns for core concepts, evaluative adjectives to express stance and comparison, and common verbs to frame actions and opinions. Grammatically, nouns (25.91%) and verbs (18.83%) dominated the text structure, supplemented by conjunctions, adjectives, adverbs, determiners, infinitival to, and personal pronouns to establish clausal connections and maintain coherence. These patterns highlight how vocabulary and syntax work in tandem to construct persuasive arguments, offering a nuanced view of Malaysian ESL writing that extends beyond conventional rubric scoring or error counting. The paper concludes with practical teaching strategies and directions for expanding corpus research in broader learner contexts.

Keywords: Corpus-Assisted Analysis; Lexical Features; Grammatical Features; Argumentative Essays; ESL Learners

INTRODUCTION

Writing is a crucial component of English language learning, serving as a medium through which students communicate their ideas, express their opinions, and present analytical arguments (Kozhakhmet et al., 2022). Writing is a comprehensive process that requires careful steps of planning, drafting, organising and revising. Hence, mastering this skill is essential for academic success, yet it poses a significant challenge for ESL learners who must simultaneously manage content development and linguistic accuracy. Students commonly encounter several challenges in writing, including limited ideas and insufficient content development, inadequate use of connectives to ensure logical organisation, inaccurate word choice and inappropriate use of idiomatic expressions, weak sentence structures, inappropriate use of first-person pronouns that affect formality and objectivity, and insufficient or missing citations in referencing (Pablo & Lasaten, 2018).

In the context of higher education, argumentative writing is particularly important because it requires students not only to produce an essay but also to develop critical thinking skills by constructing logical arguments and defending their positions with sound reasoning and supporting evidence. For many ESL learners, the challenge of producing a coherent argumentative essay while ensuring accurate grammatical structures and appropriate lexical choices often results in language errors and structural weaknesses, which

may negatively affect their academic performance. Moreover, many students perceive argumentative essay writing as a stressful task, particularly those who experience writing anxiety, as the fear of failure or negative evaluation can hinder their ability to begin or complete the writing process (Ulfa & Wahyudi, 2023).

Furthermore, it is important to recognise that both lexical and grammatical competence are equally essential for producing effective argumentative writing. The student's ability to select precise vocabulary and employ accurate grammatical structures would highly impact the clarity, coherence and persuasive power of their arguments. For instance, a study by Hidayati et al. (2025) reveals that grammatical competence serves as a crucial foundation for effective writing, whereas extensive vocabulary knowledge enhances lexical flexibility which enables students to diversify their language use, avoid redundancy, and adapt their word choices to academic contexts. Hence, a systematic examination of these linguistic features is necessary to gain a comprehensive understanding of ESL students' argumentative writing.

Although previous studies have acknowledged the importance of these two components, limited research has comprehensively examined how lexical and grammatical features interact to shape students' argumentative writing. In addition, many previous studies have assessed argumentative essays primarily through writing rubrics or error analysis. However, relatively few have employed a corpus-assisted approach to provide a comprehensive description of the lexical and grammatical features found in a learner corpus of ESL argumentative essays, particularly within the Malaysian context. To address this gap, the present study examines the lexical and grammatical characteristics of Malaysian ESL students' argumentative essays using a corpus-assisted analysis.

Accordingly, the present study seeks to answer the following research questions:

1. What lexical characteristics are evident in ESL students' argumentative essays?
2. What grammatical characteristics are evident in ESL students' argumentative essays?

LITERATURE REVIEW

Interlanguage and Learner Corpus Research (LCR)

Corpus-based research consistently positions lexical bundles as structural bedrock in Second Language (L2) argumentative writing (Etfita et al., 2025; Oktavianti & Prayogi, 2022; Xia & Kessler, 2025). Functioning as prefabricated formulas for discourse organisation and argument construction, these recurrent multi-word sequences enable student writers to project stance, frame empirical evidence, and navigate complex logical progressions. However, effective deployment requires not only a high volume of formulaic language as it demands a calibrated alignment between functional distribution, structural diversity, and register conventions.

Learner argumentative corpora may nominally display the full functional spectrum of multi-word sequences, yet their actual deployment in text remains starkly skewed. The imbalance is primarily compositional. Student writers heavily rely on research-oriented bundles, which describe topic-specific entities and real-world actions, while neglecting the text-oriented markers needed to organise academic prose (Oktavianti & Prayogi, 2022). Without these textual cues, essays risk feeling disjointed. Thus, forcing readers to actively infer the structural connections between evidence and assertion.

Across L2 writing research, raw frequency often masks lexical poverty. As highlighted by Oktavianti and Prayogi (2022) and Etfita et al. (2025), a persistent disparity exists between overall bundle frequency and bundle diversity in novice academic prose. The mechanism is structural. Rather than drawing on a broad, domain-specific phraseology, L2 writers retreat to a limited set of safe, high-utility formulas applied indiscriminately across rhetorical context. Thus, leading to monotonous structural patterns within their essays.

Eftita et al. (2025) traces this heavy reliance on prefabricated chunks back to instructional design, arguing that pedagogical models frequently prioritise rote recall over contextual adaptation. Such framing restricts students' capacity for fluid register control. Lacking practice in register modulation, novice writers remain anchored to speech-like clausal syntax rather than adopting the dense, nominalised phraseology expected in scholarly publication, which is a bottleneck that leaves their prose stalled on an interlanguage plateau.

Lexical Bundle Profiles and Academic Register

Structural composition differentiates novice L2 prose from mature academic writing just as sharply as bundle frequency of functional range (Guan, 2022; Nguyen, 2023). While expert academic prose relies heavily on dense, noun-headed nominal constructions, learner corpora displays a clear preference for clause-based bundles built on subject-verb frames. Such reliance on clausal architecture marks an earlier developmental phase in academic literacy. Novice writers construct arguments through dynamic verbal patterns, producing a lighter, less compressed syntactic style rather than packing information into compressed noun phrases (Guan, 2022; Nguyen, 2023).

Granular structural analysis makes this pattern concrete as student corpora consistently demonstrate the overuse of specific fragment categories, particularly pronoun or noun phrase plus verb phrase fragments, noun phrase plus of-phrase fragments, and prepositional phrase plus of-phrase fragments (Nguyen, 2023). ESL writers rely on these recurring frame structures as basic grammatical scaffolding during drafting, and the tradeoff is syntactic agility. Due to the overreliance on these familiar templates, learners produce rigid sentence patterns that lack the structural variation of expert prose (Nguyen, 2023).

Register confusion compounds this syntactic rigidity. Guan (2022) observes that ESL undergraduates regularly import conversational bundles, which are formulas typical of informal speech, directly into formal essays. Such intrusion points to an evolving, transitional register awareness. Writers instinctively draw on oral communication strategies to formulate written claims, blurring the boundary between spoken and written registers (Guan, 2022). This modal overlap dilutes the detached, persuasive tone expected in academic writing, highlighting a clear need for explicit instruction in register-bound phraseology.

Language proficiency mediates how these multi-word units are acquired and deployed. In a comparative analysis across EFL, ESL, and native-speaker corpora. Prihantoro and Ishikawa (2025) established a positive correlation between overall bundle mastery and general writing proficiency. Xia and Kessler (2025) observe a parallel trajectory at the move level, where lower-proficiency learners deploy significantly fewer functional bundles within critical argumentative stages. As learners reach intermediate proficiency, their bundle repertoire expands in both volume and functional diversity, allowing them to advance claims and frame supporting data with greater structural sophistication (Xia & Kessler, 2025). Interlanguage development thus manifests as a gradual shift away from beginner patterns toward more specialised phraseological routines.

Grammatical Feature Patterns and Underlying Causes

Grammatical competence provides the structural framework necessary to convey complex argumentative logic clearly and persuasively (Hidayati et al., 2025). L2 writing scholarship remains largely dominated by surface-level error analysis, documenting persistent sentence-level inaccuracies across ELS populations (Escudra, 2024; Setiati et al., 2025). These surface deviations are typically categorised as addition, omission, misordering, and misinformation, which directly compromise syntactic clarity. Yet error distributions are rarely uniform. Empirical evidence indicates that the prevalence of specific error categories shifts markedly across learner cohorts, developmental stages, and institutional settings (Escudra, 2024; Setiati et al., 2025).

Demographic profiles clarify this variation across educational tiers. For instance, at the secondary level, ESL writing exhibits a high concentration of additional errors, followed closely by omission and misinformation (Escudra, 2024). Tertiary datasets display a distinct structural profile. Setiati et al. (2025) found that omission errors predominate, with misordering occurring least frequently. This developmental transition

implies that as academic literacy matures, syntactic vulnerabilities shift. Learners move from over-supplying redundant elements to dropping essential structural markers under the cognitive load of complex sentence production (Escuadra, 2024; Setiati et al., 2025).

Granular structural analyses demonstrate that grammatical inaccuracies cluster within narrow linguistic domains rather than scattering evenly across the grammar. L2 syntactic vulnerabilities concentrate in verb phrase architecture, subject realisation, clausal patterning, preposition selection, and verb complementation (Chávez & Valenzuela Ponce, 2023). These targeted areas show striking persistence over time. Even when completing controlled grammaticality judgment tasks, learners regularly fail to identify or rectify these errors in their own writing (Chávez & Valenzuela Ponce, 2023). Such persistence points to fossilized interlanguage representations rather than fleeting execution slips during the drafting process.

Broader error taxonomies confirm the primacy of syntax in writing evaluation. Evaluating undergraduate argumentative essays, Qadir and Bostanci (2023) identified grammatical errors as the single largest category of textual deviation, which substantially outnumber lexical errors as it appeared least frequently in their dataset. Thus, suggesting that syntax forms the primary barrier instead of lexicon. L2 writers may successfully acquire functional vocabulary, yet framing those lexical items within target-like clause structures remains the principal hurdle in advanced academic writing (Chávez & Valenzuela Ponce, 2023; Qadir & Bostanci, 2023).

Cohesion, Collocation, and Academic Stance in Malaysian ESL Writing

L2 grammatical errors stem from complex cognitive and linguistic drivers rather than simple mechanical slips. Qualitative interviews and error analysis link persistent syntactic flaws directly to first-language interference and overgeneralisation of target-language rules (Qadir & Bostanci, 2023; Setiati, 2025). Yet mechanics capture only part of the struggle. Self-reports from learner writers indicate that grammatical breakdowns coincide with macro-level challenges in structural organisation, lexical selection, and the logical framing of claims (Saputra et al., 2021). Surface errors thus reflect deeper friction between processing constraints and the complex cognitive demands of academic genre construction (Saputra et al., 2021).

The relationship between local mechanics and global structure extends into textual cohesion. High tie density does not guarantee rhetorical coherence. Wandira et al. (2025) found that while student prose frequently overflows with articles, conjunctions, and repeated words, sheer volume correlates poorly with overall writing quality. The distinction lies in execution, where high-scoring essays show stable reference chains, logical connectives, and purposeful lexical reiteration (Wandira et al., 2025). In contrast, weaker essays rely on fragmented reference chains and rigid, repetitive linking devices that fail to unify the underlying argument (Wandira et al., 2025).

Collocational research mirrors this pattern, where phraseological maturity hinges on qualitative precision rather than raw frequency. In a comparative study of ESL essays, Xia et al. (2023) observed that upper-intermediate writers select collocations with higher association strength, superior idiomatic fit, and broader phraseological diversity compared to beginners. The strategic deployment of adverb-verb and verb-preposition combinations in their writing also serves as a primary discriminator between performance tiers, which reflects an advanced grasp of academic register norms (Xia et al., 2023).

In Malaysian higher education, these lexical and structural limitations are directly related to rhetorical stance. Malaysian ESL undergraduates exhibit a narrow stance repertoire, relying almost exclusively on basic modal verbs like *may* or *can* for epistemic hedging (Manokaran, 2026). Boosters and attitude markers are largely absent. Without these rhetorical tools, students frequently advance categorical assertions without necessary modal qualifications (Manokaran, 2026). This limited hedging inventory weakens the academic caution required in persuasive writing, identifying stance expression as an urgent target for classroom instruction (Manokaran, 2026).

METHODOLOGY

Data Collection

The corpus used for this study consisted of 24 files which were compiled from the argumentative essays produced by undergraduate students from Universiti Teknologi MARA (UiTM). In total, the corpus contained 7429 tokens, representing the overall number of running words in the collected texts. All participants wrote on the same topic and followed the same target length of 300 to 350 words. Students were tasked with writing an argumentative essay based on a prompt that required them to evaluate whether online news or printed newspapers serve as the better source of information. Keeping the topic, genre, and word count constant across all essays was intentional, which ensured that variations in student performance were not influenced by differing levels of background knowledge or essay length. This uniformity allowed for a fairer, more direct comparison across the sample.

Prior to analysis, all essays were transcribed into Word format (.docx) to create a clean and standardised digital corpus. Additionally, the essays were transcribed verbatim to ensure that the original texts were preserved. No corrections, modifications, or alterations were made to the students' writing, including grammatical errors, spelling mistakes, punctuation, and lexical choices. This approach ensured that the corpus accurately represented the students' authentic language use.

Data Analysis

This study utilised two tools for data analysis namely TagAnt and AntConc. TagAnt is a tagging tool that can process TextFile (.txt), Word (.docx), and PDF files to generate linguistic annotations for each word, including the word form, part-of-speech (POS), POS tag, and Lemma (Anthony, 2022). Meanwhile, AntConc is a multi-platform corpus analysis toolkit that provides various functions, including concordance analysis, word and keyword frequency generation, cluster and lexical bundle analysis, and word distribution visualisation (Anthony, 2004).

In this study, TagAnt was primarily employed for part-of-speech (POS) tagging, which involves assigning each word a grammatical category based on its meaning and its relationship with the surrounding words within a phrase, sentence, or paragraph. Meanwhile, AntConc was used as the corpus analysis tool to generate wordlists, conduct collocation analysis, and examine concordance lines. These functions enabled the study to identify frequently occurring lexical items, explore patterns of word co-occurrence, and examine the linguistic contexts in which the selected words were used throughout the corpus. In addition, the GraphColl tool in #LancsBox X was also utilised to examine the distribution of collocates across the corpus. This tool was used to visualise the relationships between the node words and its collocates which allowed the distribution of collocational patterns to be identified more clearly.

For RQ1, a wordlist and collocation analysis were conducted to identify the frequently occurring lexical items and examine the words that commonly co-occur with the selected target words. For RQ2, an analysis of the POS-tag frequency was employed to determine the distribution of grammatical categories and the words associated with the selected POS tags.

Corpus Preparation and Analysis Settings

Before conducting the corpus analysis, a stopword list was created to remove high-frequency function words that were not relevant to the analysis. For example, words such as the, a, and an were included in the stopword list and excluded from the frequency analysis. The stopword list was applied to exclude these words from the word frequency results and allow greater focus on the content words in the corpus. Moreover, as this study retained the students' original writing including all errors, a manual check was conducted to ensure the accuracy and reliability of the tagged data.

Furthermore, relevant settings were used in AntConc which included the parameters for word frequency and collocation analysis to ensure consistency throughout the analysis. This study primarily utilised the default

settings in AntConc for both Wordlist and Collocation Analysis. For instance, the minimum frequency and range for Wordlist were set to 1 and the results were sorted by frequency. Meanwhile, the Collocate tool employed the likelihood measure of “Log-Likelihood (4-term)” with the threshold set to $p < 0.05$ (3.84 with Bonferroni) and the Window Span was set to 5L-5R.

FINDINGS AND DISCUSSION

Lexical Characteristics of ESL Students' Argumentative Essays

Wordlist of Frequently Used Lexical Items

Table 1 shows the top 100 most frequently used lexical items identified in the ESL students' argumentative essays.

Table 1 Wordlist of Frequently Used Lexical Items in ESL students' argumentative essays

Type	Freq	Type	Freq	Type	Freq
news	308	do	22	nowadays	14
online	228	easy	22	secondly	14
information	116	faster	22	updates	14
printed	109	platform	22	believe	13
people	103	save	22	day	13
more	87	conclusion	21	era	13
newspaper	82	use	21	firstly	13
because	78	while	21	give	13
better	64	however	20	go	13
newspapers	64	know	20	modern	13
read	51	media	20	most	13
they	47	need	20	much	13
time	44	today	20	only	13
readers	43	just	19	smartphones	13
source	42	money	19	internet	12
websites	39	our	19	now	12
many	38	will	19	other	12
some	36	about	18	provide	12
access	35	anytime	18	unlike	12
not	35	easier	18	convenient	11
get	29	buy	17	current	11
still	29	them	17	environment	11
digital	28	there	17	helps	11
their	28	through	17	home	11
from	27	all	16	new	11

than	27	fast	16	which	11
platforms	26	lot	16	around	10
using	26	way	16	find	10
world	26	free	15	like	10
anywhere	25	help	15	make	10
example	25	phone	15	reduce	10
paper	25	prefer	15	right	10
reading	25	social	15		
your	23	important	14		

Based on the generated wordlist above, the words can be categorised into several broad lexical groups. The first group relates with general nouns which mainly refer to people, objects, media, technology, places or general concepts. This includes words such as people, news, newspaper, readers, websites, media, world, home and a few more. This suggests that the students' argumentative essays mainly focus on everyday topics and issues related to news, media, technology and society. Moreover, the high frequencies of news, online, information and printed indicate that the essays strongly focus on digital and traditional sources of information. Another significant group that appears from the list is evaluative adjectives which refer to words that are used to express opinion, judgement or comparison. For example, better, easy, easier, faster, important and convenient. This suggests that students frequently use these adjectives to express their opinions, make comparisons and evaluate the advantages and disadvantages of an issue when writing their argumentative essays.

In addition, a number of common verbs also frequently appear in the list which are used to describe actions, processes and personal positions. For instance, use, get, make, give, and do are verbs that can be applied to various topics. Meanwhile, the use of believe and prefer is particularly relevant to argumentative writing because these verbs allow students to express their positions, personal opinions, and preferences on the issues.

Collocation Analysis of Frequently Used Lexical Items

Collocation refers to a combination of words that frequently occur together, with one central word serving as the node, while the words that commonly occur before or after it are referred to as its collocates (Sanosi, 2025). In this section, eight high-frequency lexical items were selected as nodes for further collocation analysis, comprising two items from each of the identified lexical groups. The selected words were news and people for general nouns, better and easy for adjectives, and using and reading for verbs. The analysis was conducted to examine the words that frequently co-occur with these selected nodes and to identify recurring lexical patterns in the students' argumentative essays.

Table 2 Collocation Analysis of General Nouns

Node	Collocates	Freq	LL
News	online	218	136.239
	news	34	98.740
	better	52	21.611
People	some	22	33.323
	access	25	14.536

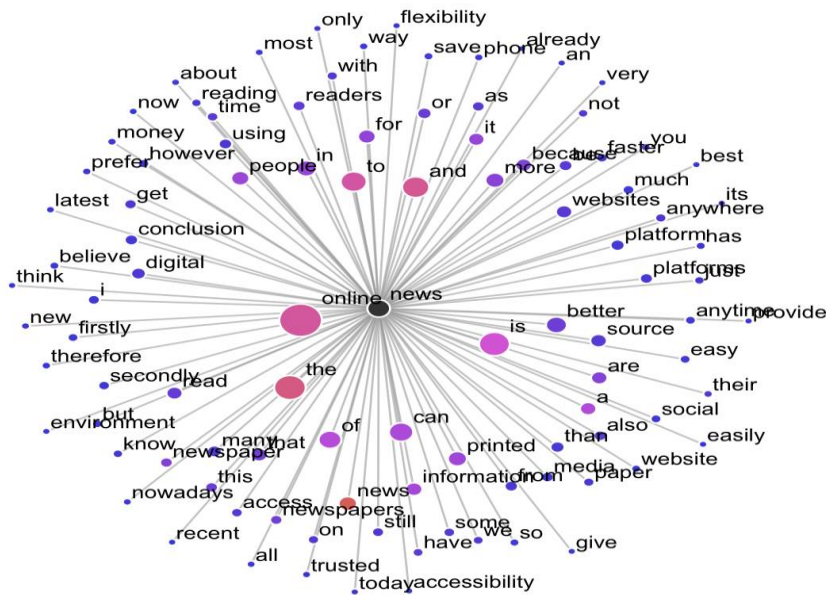


Figure 1 Collocation distribution for news

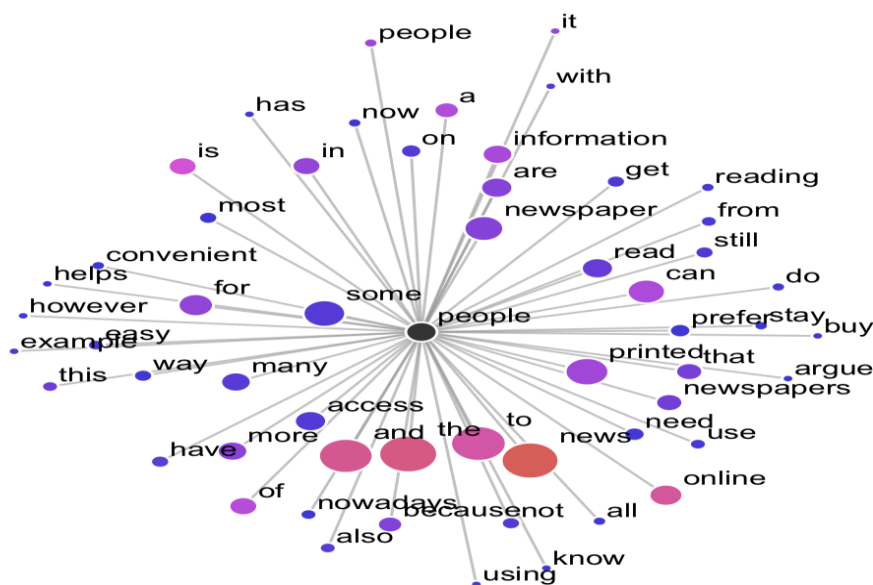


Figure 2 Collocation distribution for people

Table 2 above presents the collocates for the nouns *news* and *people* as found in the students' argumentative essays. The word *online* is the strongest collocate for *news* with 218 occurrences and LL of 136.239. This strong association suggests that students frequently discussed news in relation to online sources or platforms. Besides, the word *better* also frequently co-occurs with *news* which indicates that students use evaluative words particularly when comparing or evaluating different sources of news. Meanwhile, the word *people* highly collocates with *some* (F:22, LL:33.323) which might suggest that students usually use the phrase *some people* when presenting a particular group or viewpoint in their writing. *Access* is another word that co-occurs with *people* which is potentially associated with the concept of access to information, technology or other resources. Furthermore, an examination of collocation distribution across the corpus further shows that the words *online* and *some* are widely distributed throughout the data. Specifically, *online* functions as a primary modifier when defining news mediums, frequently pairing with nodes like *news* (218 occurrences) and *better* (50 occurrences) to establish contrast against physical print. On the other hand, *some* serves a distinct stance-building function, co-occurring predominantly with *people* (22 occurrences) to

introduce specific reader perspectives or frame counterarguments. The concordance lines for these collocates include:

“...online **news** is a better news source of information..”

“...online **news** is much better than printed...”

“...**people** can read news anytime...”

“...some **people** still prefer printed news...”

Table 3 Collocation Analysis of Adjectives

Node	Collocates	Freq	LL
Better	source	29	73.221
	online	50	35.458
	information	33	34.752
	because	25	30.579
	than	13	24.280
	news	52	21.021
	think	6	15.891
	conclusion	9	15.016
	printed	23	14.814
Easy	access	11	33.516
	activity	2	13.487

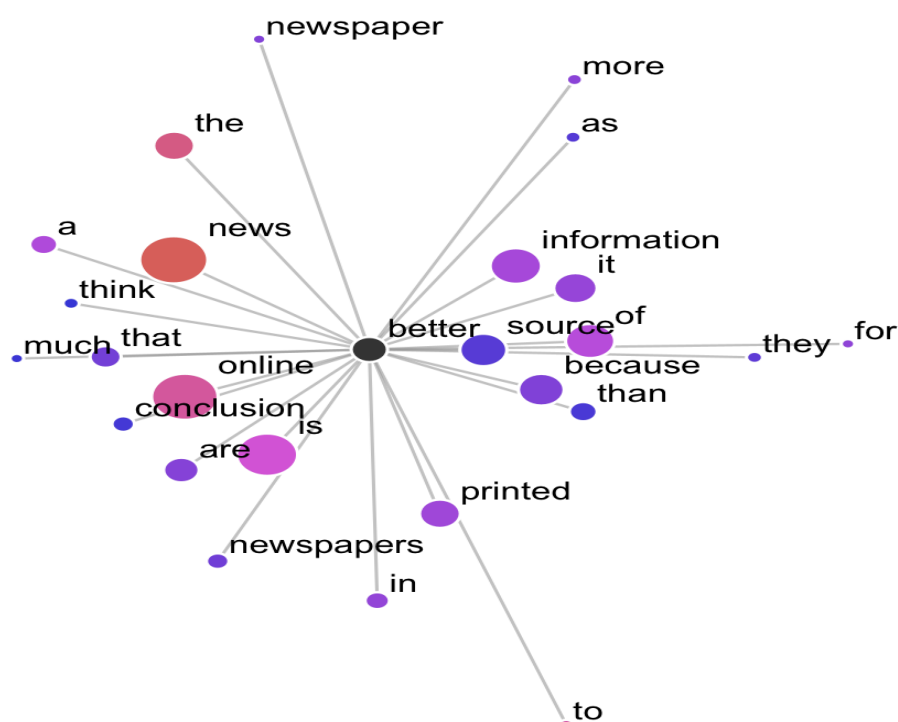


Figure 3 Collocation distribution for better

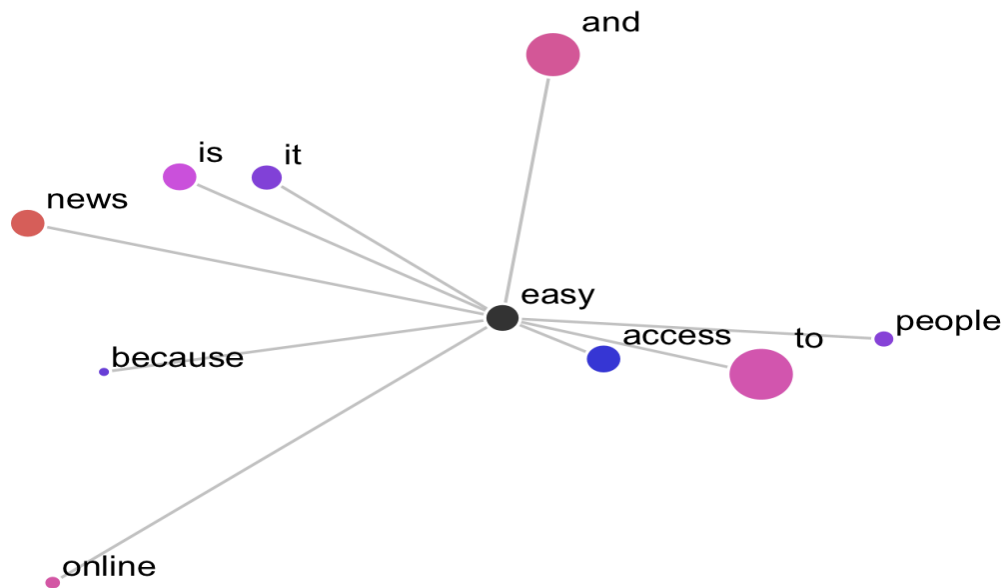


Figure 4 Collocation distribution for easy

The table above highlights the collocates for the adjectives better and easy. The term source highly collocates with better, with 29 occurrences and LL of 73.221. This strong association suggests that students often evaluate or compare different information sources in their writings. Meanwhile, the frequent occurrence of online and information suggest that students usually use comparative language when discussing online information and its sources. The collocate further indicates the use of comparative forms such as better than, while the collocate because shows that students often provide reasonings to support their arguments. On the other hand, easy highly collocates with access (F:11, LL:33.516) which indicates that students frequently discuss the ease of accessing information or resources. Additionally, the analysis of collocation distribution for better and easy reveals that they are consistently deployed across the corpus to frame comparative arguments and emphasise the convenience of digital media. Specifically, better functions as a primary evaluative marker when students weigh the merits of online news against traditional print, while easy reinforces the accessibility of digital platforms. The concordance lines for these collocates include:

“...online news is the **better** source of information...”

“...online news is **better** because they are...”

“...**easy** to access and quickly give...”

“...digital news is so **easy** and we can do it anytime...”

Table 4 Collocation Analysis of Verbs

Node	Collocates	Freq	LL
Using	benefits	4	17.840
	computer	3	14.721
	tablets	4	13.263
Reading	actually	3	22.107
	improve	4	18.144
	skills	4	18.144

	enjoy	3	14.951
	newspaper	11	14.517
	included	2	12.889

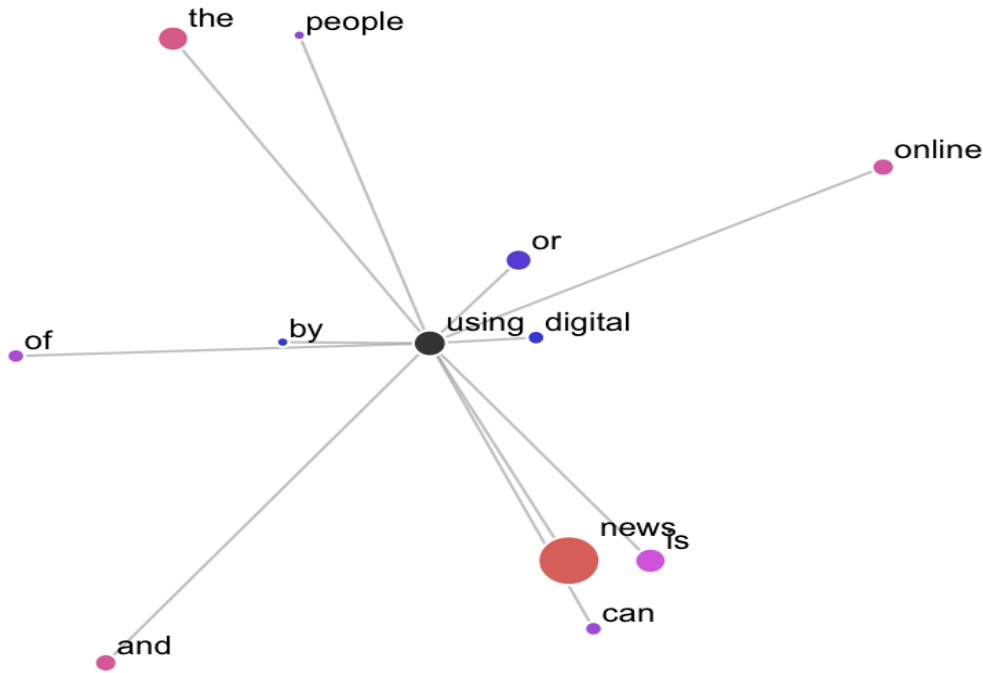


Figure 5 Collocation distribution for using

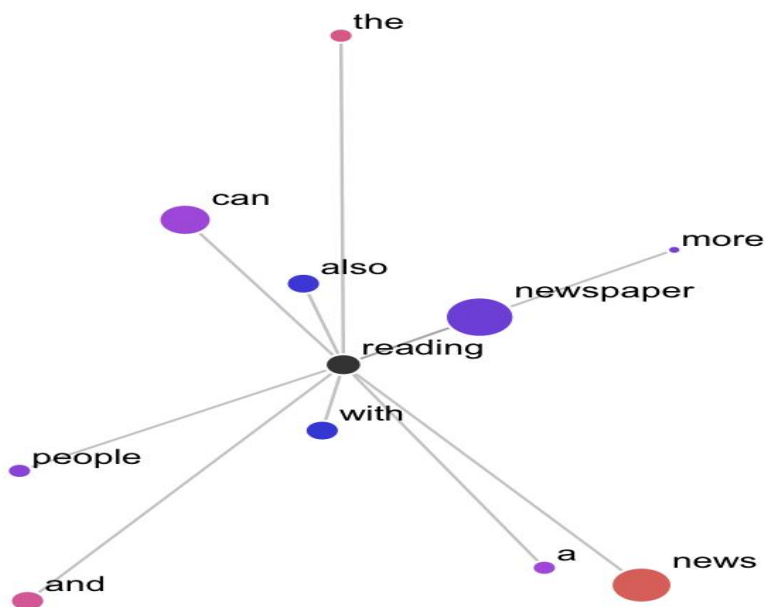


Figure 6 Collocation distribution for reading

The table above displays the collocates for the verbs using and reading. The word benefits is identified to be highly co-occur with the node using, with 4 frequencies and LL of 17.840. This suggests that students often discuss the benefits associated with certain types of sources or technology. Next, the collocated computer and tablets further indicate that using is often associated with digital devices. This pattern shows that technology is a crucial context in students' writing where they discuss the use and advantages of different resources. Meanwhile, the node reading strongly collocates with LL of 22.107 even though it only occurred

three times in the corpus. Moreover, the collocates improve and skills show that there is a strong association between reading and the development of skills. In addition, enjoy is also one of the collocates which suggest that reading is associated with a form of enjoyment while newspaper indicates the common context for reading-related discussions.

However, an examination on the distribution of the collocates shows that digital is more widely distributed across the corpus. This indicates that although benefits demonstrates stronger collocational association with using, but digital occurs more consistently throughout the corpus. Similarly, collocation analysis identifies actually as the strongest collocate of reading, which indicates a strong lexical association between the two words. But, the dispersion analysis shows that newspaper is widely distributed across the corpus. This suggests that while actually shows a stronger statistical association with reading, newspaper occurs more consistently throughout the texts. The concordance lines for these collocates include:

“...major advantage of **using** online news platforms...”

“...access articles **using** their phone...”

“...makes **reading** news more fun...”

“...improving **reading** skills and critical thinking...”

Grammatical Characteristics of ESL Students' Argumentative Essays

Frequency of Grammatical Features

The table below presents the frequency and percentage of the frequently occurring parts of speech (POS) in the students' argumentative essays.

Table 5 Frequency and Percentage of POS Categories

Broader Category	POS	POS Tags Combined	Frequency	Percentage
Nouns		NN + NNS	1925	25.91%
Verbs		VB + VBP + VBZ+ VBN + MD	1399	18.83%
Conjunctions		IN + CC	1248	16.80%
Adjectives		JJ + JJR	888	11.95%
Adverbs		RB + RBR	638	8.59%
Determiners		DT	534	7.19%
Infinitive Marker		TO	381	5.13%
Personal Pronouns		PRP	341	4.59%

The POS analysis reveals that Nouns and Verbs are among the most frequently occurring grammatical categories in the students' argumentative writing. The high frequency of nouns suggests that the students frequently refer to specific topics, concepts, people, and issues, while the frequent use of verbs reflects their need to express actions, processes, and relationships when developing their arguments. This is consistent with a view by Antón-Méndez (2020) where sentence planning depends on an understanding of the action being described and the verb used to express it, highlighting the important role of verbs in structuring and conveying meaning in argumentative writing. Besides, the high frequency of Conjunctions indicates that students effectively use these grammatical elements to link words, phrases, clauses, and sentences, which promotes greater coherence and a smoother flow in their writing (Unubi, 2016). Furthermore, the use of

Adjectives is important as they allow students to describe and evaluate issues while Adverbs can indicate how students modify their claims, actions or statements. On the other hand, the use of Determiners makes their claims more specific, clear, and grammatically precise while Infinitive “To” allows students to explain why something is done or what an action is intended to achieve. Lastly, the use of Personal Pronouns allows students to refer to themselves, their readers, or other individuals and groups without repeatedly using nouns. In the subsequent sections, the high frequency POS categories such as Noun, Verb and Conjunction will be analysed and discussed in more detail.

Analysis of Noun Forms

As indicated in the previous section, nouns were found to be the most frequently occurring POS category, with a total frequency of 1925, accounting for approximately 25.91% of the corpus. This category comprises NN (singular or mass noun) and NNS (plural noun). The table below presents the frequency, percentage and examples of the noun forms identified.

Table 6 Frequency and Percentage of Noun Forms

Noun form	Frequency	Percentage	Examples
NN	1319	68.52%	news, newspaper, coverage, paper, era, information, story
NNS	606	31.48%	websites, smartphones, media, clips, pictures, videos, standards, skills
Total	1925	100%	

For NN form, the frequently occurring nouns include news, newspaper, coverage, paper, era, information and story. The frequent occurrence of these nouns is closely related to the essay topic, which focuses on physical newspapers and online news. Nouns such as news, newspaper, coverage and paper directly refer to students' discussion on traditional and online news, while information and story refer to the content presented in the sources. For NNS form, the frequently occurring nouns include websites, smartphones, media, clips, pictures, videos, standards and skills. Similar to the NN form, several of these nouns are associated with the essay topic, particularly the discussion of online news and digital media. Nouns such as websites, smartphones, media, clips, pictures, and videos reflect the students' discussion of digital platforms and multimedia content used to access and present news. Meanwhile, standards and skills may reflect discussions of the quality of news and the abilities required to access or evaluate information. The collocates of some of the nouns and their interpretations are presented in the table below.

Analysis of Verb Forms

Verbs were found to be one of the most frequently occurring POS categories, with a total frequency of 1399, accounting for approximately 18.83% of the corpus. This category comprises VB (verb, base form), VBP (verb, non-third-person singular present), VBZ (verb, third-person singular present), VBN (verb, past participle) and MD (modal auxiliary verb). The table below presents the frequency, percentage and examples of the verb forms identified.

Table 7 Frequency and Percentage of Verb Forms

Verb form	Frequency	Percentage	Examples
VB	456	32.59%	leave, review, understand, report, watch, download
VBP	318	22.73%	people need, we have, you have, people tend, readers receive
VBZ	268	19.16%	news is, this saves, newspaper has, information helps, everyone prefers

VCN	188	13.44%	printed, used, checked, limited, restricted, transformed
MD	169	12.08%	can, would, might, must, will, should
Total	1399	100%	

For VC, the frequently used verbs include leave, review, understand, report, watch and download. These verbs indicate that students often use base-form verbs to describe various actions related to accessing, evaluating, and consuming news. For example, review and understanding reflect the process of evaluating and understanding news content, while watch and download are associated with accessing news through digital platforms. On the other hand, VBP form includes verb phrases such as people need, we have, you have, people tend and readers receive. These patterns show that students use present tense verbs to describe general actions, behaviours, and relationships related to news consumption. Moreover, the frequent adoption of plural subjects such as people and readers reflect that students often make general statements about how individuals access and receive news.

Next, VBZ form commonly consists of phrases such as news is, this saves, newspaper has, information helps and everyone prefers. These phrases demonstrate that students often use third-person singular present tense verbs when describing the characteristics, functions, and effects of newspapers and news information. For VBN form, the commonly used verbs include printed, used, checked, limited, restricted and transformed. These verbs were used to display some passive actions related to how news is produced, accessed, evaluated, and transformed, particularly when comparing traditional newspapers with online news. Moreover, the use of passive forms would help to shift the focus from the subjects performing the actions to the actions or processes themselves. Lastly, students also use some MD form in their writings which include can, would, might, must, will and should. These modal verbs are particularly important in argumentative writing because they allow students to express possibility, ability, necessity, prediction, and recommendations. For instance, can may be used to describe the benefits of online news, while must and should can be used to express strong recommendations or obligations regarding the production and consumption of news.

Analysis of Conjunctions

Conjunction was found to be the third most frequently used POS category, with a total frequency of 1248, accounting for approximately 16.80% of the corpus. This category comprises IN (conjunction, subordinating or preposition) and CC (conjunction, coordinating). The table below presents the frequency, percentage and examples of the conjunctions identified.

Table 8 Frequency and Percentage of Conjunctions

Noun form	Frequency	Percentage	Examples
IN	944	75.64%	around, over, because, of, in, about, until
CC	304	24.36%	and, but, or, so
Total	1248	100%	

For IN form, the frequently occurring terms include around, over, because, of, in, about and until. The use of prepositions such as around, over, of, in, about and until help students to connect and express relationships between their ideas in order to form a more persuasive argument. Meanwhile, the conjunction because is mainly used to provide reasonings and justifications for their claims, which helps students to better explain the causes or rationale behind their arguments. Meanwhile, the frequently occurring coordinating conjunctions for CC form include and, but, or and so. These conjunctions allow students to connect words, phrases, and clauses when developing their arguments. The frequent use of and indicates that students often add or combine ideas, while but allows them to express contrast, or help them to present alternatives and so is used to express result, consequence and effect.

DISCUSSION

The present study set out to examine the lexical and grammatical characteristics evident in Malaysian ESL students' argumentative essays. With respect to RQ1, the findings indicate that the students' lexical choices were characterised primarily by frequent use of topic-related nouns, evaluative adjectives, common verbs, and recurring collocational patterns. The wordlist showed particularly high frequencies for news (308), online (228), information (116), printed (109), people (103), newspaper (82), and better (64), indicating that students consistently drew on vocabulary associated with news information sources, digital platforms, and comparisons between printed and online news. The frequent occurrence of better, easy, easier, faster, important, and convenient further suggests that evaluation and comparison were central to the students' construction of their arguments. This finding is consistent with the view that lexical resources play an important role in argument construction and stance expression in L2 writing (Etfita et al., 2025; Oktavianti & Prayogi, 2022; Xia & Kessler, 2025). In particular, Nguyen (2023) found that lexical bundles in undergraduate argumentative writing frequently served stance and discourse-organising functions, which reflects the communicative demands of argumentation. Thus, the present findings suggest that the students selected lexical resources that enabled them to identify the subject of discussion, express evaluations, and position themselves in relation to the issues.

A further finding that merits attention is the strong collocational relationship between the lexical items identified in the corpus. The strongest association was observed between news and online ($F = 50$; $LL = 35.458$), and information ($F = 33$; $LL = 34.752$). Similarly, easy was strongly associated with access, whereas using occurred with benefits, computer, and tablets, and reading was associated with improve, skills, enjoy, and newspaper. These patterns suggest that the students' lexical choices were not simply frequent individual words, but were embedded in recurring associations that supported comparison, evaluation, access, and discussion of benefits. This interpretation is supported by Xia et al. (2023), who demonstrate that collocational frequency, association strength, and diversity provide information about L2 argumentative writing. At the same time, the findings may indicate a relatively concentrated lexical repertoire, as several highly frequent items were general-purpose words such as better, easy, use, get, and make. This observation is compatible with the literature reviewed in the study, which notes that L2 writers may rely on familiar, high-utility lexical sequences rather than a wider range of phraseological resources (Etfita et al., 2025; Oktavianti & Prayogi, 2022). Nevertheless, this should be interpreted cautiously because the present study examined lexical frequency and collocation rather than directly measuring lexical diversity.

With respect to RQ2, the findings indicate that nouns, verbs, and conjunctions constituted the most prominent grammatical categories in the students' argumentative essays. Nouns accounted for 25.91% of the corpus, followed by verbs at 18.83% and conjunctions at 16.80%, while adjectives, adverbs, determiners, infinitive to, and personal pronouns occurred at lower frequencies. The predominance of nouns reflects the students' frequent reference to the central entities and concepts associated with the topic, including news, newspaper, coverage, information, websites, smartphones, and media. Meanwhile, verbs such as review, understand, report, watch, and download enabled students to describe actions and processes associated with accessing and evaluating news. This finding is broadly consistent with Antón-Méndez (2022), who highlights the important role of verbs in sentence production and expression of actions. The prominence of nouns and verbs also reflects the functional demands of argumentative writing, where students need to establish the subject of discussion and then describe actions, relationships, effects, or positions associated with the subject.

The relatively high frequency of conjunctions provides further insight into how the students organised their arguments. Conjunctions accounted for 16.80% of the corpus, with because occurring as an important resource for providing reasons and justification, while and, but or, and so enabled students to add ideas, establish contrasts, introduce alternatives, and express consequences. This finding suggests that students had access to grammatical resources for establishing logical relationships between propositions which is consistent with the role of conjunctions in creating connections across ideas (Unubi, 2016). Not only that, the presence of modal verbs such as can, would, might, must, will, and should further indicate that students

were able to express possibility, ability, prediction, necessity, and recommendation. This is particularly relevant to argumentative writing because claims often require different degrees of certainty or force. Nevertheless, the findings necessarily demonstrate effective coherence, since cohesion depends on the purposeful use of logical connectives and other cohesive resources (Wandira et al., 2025).

In light of the RQ1 and RQ2 findings, it can be concluded that lexical and grammatical resources worked in combination to construct the students' argumentative positions. The lexical analysis showed that students used words such as better, online, information, access, and news to establish topics, make comparisons, and express evaluations, while grammatical analysis showed that nouns, verbs, conjunctions, and modal verbs provided the structural resources through which these meanings were expressed and connected. For example, the recurring association between better and source demonstrates lexical evaluation, while the frequent use of because provides a grammatical mechanism for supporting an evaluation with a reason. This interaction supports the broader argument in the Literature Review that lexical selection, grammatical organisation, and the logical framing of claims are closely related in L2 argumentative writing (Saputra et al., 2021). It also corresponds with research that shows phraseological and grammatical patterns contribute to the development of argumentative discourse (Xia et al., 2023; Nguyen, 2023). Therefore, the findings suggest that students' argumentative writing is better understood by examining the interaction between lexical and grammatical resources rather than treating vocabulary and grammar as entirely separate dimensions.

Overall, the study answers RQ1 by demonstrating that students relied on topic-related, evaluative, and collocational lexical resources, and answers RQ2 by demonstrating the prominence of nouns, verbs, and conjunctions, supplemented by other grammatical categories. Taken together these patterns indicate that the students possessed a functional linguistic repertoire for constructing argumentative positions, while also suggesting opportunities to develop greater lexical variation, collocational precision, and more strategic use of grammatical resources.

CONCLUSION

This study examines the lexical and grammatical characteristics of Malaysian ESL students' argumentative essays using a corpus-assisted approach. The analysis was based on 24 argumentative essays written by undergraduate students from Universiti Teknologi MARA (UiTM), comprising a total of 7,429 tokens, with each essay written within the required length of 300 to 350 words. The lexical analysis showed that the students frequently relied on familiar vocabulary to express opinions, comparisons and evaluations, while their grammatical choice helped organise ideas, establish relationships between clauses and maintain coherence across the essays. Taken together, these patterns show that lexical and grammatical resources work closely in shaping how students construct and sustain an argumentative position.

The study therefore contributes to research on Malaysian ESL writing by providing a corpus-based description of the linguistic resources students employ when constructing argumentative texts, extending beyond approaches that focus mainly on rubric scores or language errors. However, the findings are drawn from a relatively small corpus of 24 essays produced within one institutional context, which limits the extent to which the patterns can be generalised to a wider population of ESL learners. Future research could examine larger and more diverse learner corpus and compare students from different proficiency levels, institutions or writing tasks to determine whether similar patterns occur across context. Ultimately, this study highlights the importance of examining lexical and grammatical features together to obtain a more comprehensive understanding of how ESL students develop, connect and communicate their arguments in writing.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to Universiti Teknologi MARA (UiTM) Cawangan Sarawak for providing the institutional support and resources necessary to conduct this study. The authors

also extend their appreciation to the undergraduate students who contributed their argumentative essays, which formed the corpus for this research.

REFERENCES

1. Anthony, L. (2004). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. IWLeL 2004: An Interactive Workshop on Language e-Learning.
2. Anthony, L. (2022). TagAnt (Windows, MacOS, Linux). Center for English Language Education in Science and Engineering.
3. Antón-Méndez, I. (2020). The role of verbs in sentence production. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2020.00189>
4. Chávez, M., & Valenzuela Ponce, K. (2023). Grammatical errors in pre-service English teachers' argumentative essays and their views regarding error correction. *Foro Educacional*, (40), 65–99. <https://doi.org/10.29344/07180772.40.3314>
5. Escuadra, D. (2024). Grammar errors in argumentative writing of junior high school ESL students: An analysis. *Journal of Higher Education Research Disciplines*, 9(1), 49–72. <https://doi.org/10.65023/jherd.v9i1.236>
6. Etfita, F., Wahyuni, S., Ahmad, A., Sudusinghe, W. S., & Gamage, C. K. W. (2025). Revealing lexical bundles of non-native English essay at a private Islamic university: A corpus-based evaluation. *English Learning Innovation*, 6(2), 450–468. <https://doi.org/10.22219/englie.v6i2.40821>
7. Guan, J. (2022). A contrastive study on lexical bundles in argumentative writing by L1-Chinese and L1-English undergraduates. *English Language Teaching*, 15(12), 43–52. <https://doi.org/10.5539/elt.v15n12p43>
8. Hidayati, N., Qibtiyah, L., & Hasanah, U. (2025). The impact of grammar, vocabulary, and reading interest on descriptive writing skills in Islamic higher education. *Jurnal Pendidikan Dan Sastra Inggris*, 5(1). <https://doi.org/10.55606/jupensi.v5i1.5009>
9. Kozhakhmet, E., Tulepova, S., & Stepanov, I. (2022). The importance of writing skills for EFL learners: Common errors and ways of improving. *Педагогика И Методы Обучения*, 61(4). <https://doi.org/10.47344/sdu20bulletin.v61i4.859>
10. Manokaran, J. (2026). Hedging in ESL academic essays: A corpus-based analysis of Malaysian undergraduate writing. *International Journal of Research and Innovation in Social Science*, 10(3), 1264–1276. <https://doi.org/10.47772/IJRISS.2026.100300087>
11. Nguyen, N. X. (2023). Structures and functions of lexical bundles in student and expert timed argumentative writing: A corpus-driven study. *Pertanika Journal of Social Sciences & Humanities*, 31(3), 1015–1039. <https://doi.org/10.47836/pjssh.31.3.06>
12. Oktavianti, I. N., & Prayogi, I. (2022). Discourse functions of lexical bundles in Indonesian EFL learners' argumentative essays: A corpus study. *Studies in English Language and Education*, 9(2), 761–783.
13. Pablo, J. C. I., & Lasaten, R. C. S. (2018). Writing difficulties and quality of academic essays of senior high school students. *Asia Pacific Journal of Multidisciplinary Research*, 6(4).
14. Prihantoro, & Ishikawa, S. (2025). A corpus-based comparison of EFL, ESL, and ENS lexical bundles in the ICNALE corpus. *Indonesian Journal of Applied Linguistics*, 15(2), 398–419. <https://doi.org/10.17509/xq43nf49>
15. Qadir, S. M., & Bostanci, H. B. (2023). An analysis of English as a foreign language undergraduate students' written errors. *Forum for Linguistic Studies*, 5(3), 1881. <https://doi.org/10.59400/fls.v5i3.1881>
16. Sanosi, A. B. (2025). The present study answers the research question: A Corpus-based analysis of collocate directionality in academic English writing. *ScienceDirect*. <https://doi.org/10.1016/j.heliyon.2025.e42088>
17. Saputra, A. B. B., Jumariati, & Febriyanti, E. R. (2021). EFL students' problems in writing argumentative essays. *Proceedings of the 2nd International Conference on Education, Language, Literature, and Arts (ICELLA 2021)*, 587, 8–12. <https://doi.org/10.2991/assehr.k.211021.002>

18. Setiati, A. T., Siregar, E. B. A., Ridwan, M., & Dinamika, S. G. (2025). Investigating grammatical errors in EFL university students' argumentative essays. *International Journal of Social Science and Human Research*, 8(5), 2839–2845. <https://doi.org/10.47191/ijsshr/v8-i5-24>
19. Ulfa, S. M., & Wahyudi, Moh. A. (2023). Students' difficulties in writing an argumentative essay: A qualitative study of EFL learners in a writing class. *English Teaching Journal: A Journal of English Literature Language and Education*, 11(1). <https://doi.org/10.25273/etj.v11i1.15814>
20. Unubi, A. S. (2016). Conjunctions in English: Meaning, types and uses. *International Journal of Social Science and Humanities Research*, 4(3).
21. Wandira, S., Suherman, L. O. A., & Hadiyansah, W. S. D. N. (2025). Grammatical and lexical cohesion in Indonesian EFL argumentative essays. *EDUTREND: Journal of Emerging Issues and Trends in Education*, 3(1), 1–12. <https://doi.org/10.59110/edutrend.830>
22. Xia, D., Chen, Y., & Pae, H. (2023). Lexical and grammatical collocations in beginning and intermediate L2 argumentative essays: A bigram study. *International Review of Applied Linguistics in Language Teaching*, 61(4), 1421–1453. <https://doi.org/10.1515/iral-2021-0188>
23. Xia, D., & Kessler, M. (2025). Making arguments at lower proficiencies: Move–bundle connection in Korean university students' English argumentative writing. *International Journal of Applied Linguistics*, 35(3), 1219–1241. <https://doi.org/10.1111/ijal.12693>