Malaysian Journal of Social Sciences and Humanities (MJSSH) (e-ISSN : 2504-8562)
2026, Volume 11, Issue 8, e004139
DOI: https://doi.org/10.47405/mjssh.v11i8.4139

RESEARCH ARTICLE

Comparing Linguistic Features of AI-Generated and Students-Produced Essays: A Corpus-Based Comparison

Rafidah Kamarudin*[1], Roslina Abdul Aziz[2], Amalia Qistina Castaneda Abdullah[3], Muhamad Izzat Rahim[4], Nur Athirah Zaidi[5]

[1]Academy of Language Studies, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Kuala Pilah, 72000 Kuala Pilah, Negeri Sembilan, Malaysia.
Email: fid@uitm.edu.my
[2]Academy of Language Studies, Universiti Teknologi MARA (UiTM) Cawangan Pahang, Kampus Jengka, 26400 Bandar Jengka, Pahang, Malaysia.
Email: leenaziz@uitm.edu.my
[3]Academy of Language Studies, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Kuala Pilah, 72000 Kuala Pilah, Negeri Sembilan, Malaysia.
Email: amali608@uitm.edu.my
[4]Academy of Language Studies, Centre of Foundation Studies, Universiti Teknologi MARA (UiTM) Cawangan Selangor, Kampus Dengkil, 43800 Dengkil, Selangor, Malaysia.
Email: izzatrahim@uitm.edu.my
[5]Academy of Language Studies, Universiti Teknologi MARA (UiTM) Cawangan Negeri Sembilan, Kampus Kuala Pilah, 72000 Kuala Pilah, Negeri Sembilan, Malaysia.
Email: Athirahzaidi@uitm.edu.my

CORRESPONDING AUTHOR (*):
Rafidah Kamarudin
(fid@uitm.edu.my)

KEYWORDS:
Artificial Intelligence (AI)
Academic discourse
Corpus-based study
Discourse markers
Comparative analysis

CITATION:
Rafidah, K., Roslina, A. A., Amalia Qistina, C. A., Muhamad Izzat, R., & Nur Athirah, Z. (2026). Comparing Linguistic Features of AI-Generated and Students-Produced Essays: A Corpus-Based Comparison. *Malaysian Journal of Social Sciences and Humanities (MJSSH), 11*(8), e004139. https://doi.org/10.47405/mjssh.v11i8.4139

ABSTRACT

The integration of artificial intelligence (AI) into language teaching and learning has concerned language teachers especially in distinguishing the content produced by AI and by learners. Previous studies commonly emphasised on the grammatical and lexical aspects; however, less attention is given to research specifically examining on discourse markers which are essential for textual cohesion and coherence. Thus, this corpus-based study is conducted to compare the use of discourse markers in academic essays produced by AI and student writers. Two comparable corpora are created: the first corpus consists of 44 essays generated by AI and another corpus comprises of 44 essays written by university students without any support from AI. The LancsBox X was used to analyse the frequency, variety, and functional distribution of discourse markers categories (additive/elaborative, contrastive, causal, and sequential). Results indicate a significant difference in discourse markers use between the two groups. AI-generated texts are found to exhibit over-reliance on elaborative markers, meanwhile, inferential and temporal markers are found more frequently in the student essays. These findings suggest that although the academic essays produced by AI are structurally flawless, it lacks the nuanced, stylistic and argumentative depth found in the students' essay.

1. Introduction

Recently, we have been witnessing the rapid development of technology where it has brought great changes in various sectors including the education scene. This can be seen more clearly especially with the emergence and use of Artificial Intelligence (AI) and digital tools in the context of teaching and learning. Undeniably, the use of AI in teaching and learning brings the light to educators as it not only facilitates during the lesson preparation process, but it also enhances the overall experiences in the classroom. However, they also raise concerns among language teachers particularly in distinguishing academic writing produced by AI and that written by humans. This issue is relevant in academic context as it reflects the authenticity, originality and overall writing competence in student writing, highlighting the importance of distinguishing AI generated texts from those written by the students.

This issue surrounding the unethical and unregulated use of artificial intelligence have prompted researchers to conduct studies that compare the differences between human-written text to AI generated text. Although many researchers have identified differences between AI - generated and human written texts (Herbold et al. 2023, Liu et al., 2023), it is challenging for most teachers to recognise the difference between the fluency of AI-generated written work and the underlying reasoning of human written work (Liu et al., 2023). In addition, most studies have only looked at grammatical accuracy and vocabulary choice (Yildiz et al., 2025, Georgiou, 2024), but have not fully compared the overall organisation of the text. Due to this lack of depth in studies, there is a need to conduct both surface level and deeper analysis of textual features of written discourse, particularly with respect to how discourse markers are used in expressing coherence and cohesion of the text. Moreover, little research has been undertaken to examine how discourse markers are used in AI-generated texts when compared to a sample of written essays from students. As a result of this lack of research on the role of discourse markers in writing produced by an AI system, language teachers have limited means of assessing the features of a written AI text that distinguish it from an authentic student-produced text. Without this knowledge, it is difficult for educators to assess how well a piece of writing is written by students or how to build pedagogical methods that would assist students in developing their written output. Therefore, there is a need to conduct a more comprehensive investigation in the area of discourse markers in AI-generated essays and student-generated essays.

1.1. Research Objectives

The research objectives of the present study are as shown: -

i. to compare the frequency of occurrence of discourse markers in both student – produced and AI – generated essay corpora
ii. to examine the range of discourse markers between the two corpora
iii. to evaluate the contextual appropriateness of the discourse markers as used in both corpora

2. Literature Review

2.1. Theoretical Foundations: Corpus Linguistics in Writing Research

The analysis of extensive text sample can be done systematically by utilising corpus linguistics framework (Biber, 1993; Biber, Conrad, & Reppen, 1998). With the help of computational tools such as *L2 Syntactic Complexity Analyser* and *Coh-Metrix*, language

researchers can conduct in depth examination of text by looking at their features like syntactic complexity, lexical diversity and lexical density. Such text features are the focus of many studies in academic and learner writing, especially in second language and foreign language contexts (ESL & EFL) where differences show cognitive and proficiency differences (Lu, 2010; Crossley et al., 2015). From the perspective of ESL and EFL research, corpus analyses have been extensively used to examine argumentative structure, metadiscourse markers, and cohesive devices in student essays (Mat Zali et al., 2020). Such studies position linguistic features within broader discourse functions relevant to academic writing (Hyland, 2005).

2.2. Linguistic Profiles of Student-Produced Essays

Studies conducted in Malaysian tertiary contexts identify persistent features of ESL student writing relevant to comparative analysis with AI-generated texts. For instance, research on Universiti Teknologi MARA (UiTM) undergraduate essays demonstrates that discourse organisation and rhetorical clarity significantly influence assessment outcomes, with rhetorical organisation being particularly salient for evaluators (Kamaludin, 2014).

Corpus based investigations based on digital text collections about undergraduate writing in Malaysia show the high value of linguistic signals and organisational cues when separating stronger essays from weaker essays. In the detailed examination conducted by Mohamed et al. (2021), results indicated that weaker essays frequently employ too many specific text-guiding elements if compared to stronger essays. This pattern exists because differences appear in how writers direct the focus of people reading and how writers build their reasoning.

It has been observed that Malaysian ESL learners experience many difficulties regarding the variety of vocabulary and the correctness of language rules. These difficulties are encountered by Malaysian ESL learners often, which shows in mistakes during the use of action words and a small understanding of how words sit together (Abdullah, 2015). While reviewing the argumentative writing of students, Malaysian researchers Ting et al. (2011) see that Malaysian university learners use basic linking words and basic sentence patterns most of the time. Complex arrangements of words, such as sentences about things that might happen, appear very rarely in the argumentative writing of students. The fact that students depend on these simple tools makes smaller the growth of their reasoning and stops the deep level of thinking from appearing.

Collectively, these findings establish a baseline for student-produced essays in Malaysian ESL contexts: moderate lexical diversity, frequent use of basic cohesive devices, variable syntactic complexity, and metadiscourse patterns indicative of developmental writing stages.

2.3. AI-Generated Essays: Linguistic Characteristics and Comparative Findings

Extensive scientific examinations done recently about the divergence between AI-generated and human-authored essays provide useful knowledge for researchers using corpora. One major study was produced by Herbold et al. (2023), which compared AI-generated and human-authored essays to find a special "stylistic signature" for AI. The researchers claim that AI uses a high variety of words and uses nouns often, but lacked the small signs of a human voice. Because these signs are missing, AI-generated essays are defined differently from the original human-authored essays.

Liu et al. (2023) showed similar results when analysing AI-generated and human-authored essays using the ArguGPT, a research framework and a specialised dataset designed to detect and analyse argumentative essays generated by an AI corpus. The researchers observed that even though generative models are good at making sentences with complex grammar, the human-authored essays remain richer in words and have more variety in word choice. Despite these differences, teaching professionals may face difficulties in distinguishing between the smoothness of AI-generated and human-authored essays and the deep thoughts found in human-authored essays. If the teaching professionals do not have special training, identifying the difference remains a hard task because the AI-generated and human-authored essays may appear similar on the surface level.

Further research, including studies featured in *Frontiers in Education*, has begun profiling these differences by integrating metrics for syntactic complexity and readability (Amirjalili et al., 2024; Fredrick & Craven, 2025). The consensus appears to be that AI outputs achieve impressive structural coherence and surface-level "polish," yet they frequently lack the conceptual depth and rhetorical engagement seen in L2 student work. The consensus appears to be that AI outputs achieve impressive structural coherence and surface-level "polish," yet they frequently lack the conceptual depth and rhetorical engagement seen in L2 student work. Students typically utilise a broader range of modals and personalised argumentation markers to establish a clear stance (Jiang & Hyland, 2024).

Furthermore, applied linguists use social science corpora to analyse academic writing; they look closely at structural and lexical choices, revealing that AI-generated academic texts often fall back on formulaic lexical patterns and an over-reliance on "prestige" vocabulary, yet they fail to utilise subordination as effectively as human scholars' writers. Ultimately, these linguistic fingerprints (i.e. spanning lexical, syntactic, and discourse-level features) provide a roadmap for distinguishing machine-generated content from human scholarship, even as the models themselves grow more sophisticated (Mizumoto & Eguchi, 2023).

Table 1: Comparative Findings between AI-Generated Academic Text and Human Expert Scholar Text

Metric Type	AI-Generated Academic Text	Human Expert Scholar Text
Lexical Bundles	High frequency of generic, fixed transition templates (*"In conclusion, it is important to..."*).	Lower frequency of fixed templates; high variability in transition choices.
Syntactic Complexity	High ratio of clauses per sentence (**Clausal Subordination**).	High ratio of complex nominals and modifiers per noun phrase (**Phrasal Density**).
Vocabulary Distribution	Homogeneous; over-relying on a specific subset	Heterogeneous; vocabulary shifts drastically depending on the specific field of study.

Metric Type	AI-Generated Academic Text	Human Expert Scholar Text
	of "prestigious" words across texts.	

Although comparative studies on AI-generated versus student essays are increasingly represented in Scopus and WoS indexes, few have examined lower-proficiency cohorts, such as second-semester diploma students in UiTM ESL settings. Most existing research has focused on university undergraduates or test-based corpora, such as TOEFL or GRE prompts in ArguGPT, resulting in a gap in understanding how AI-generated texts compare to those produced by developing academic writers (ESL/EFL) early in their tertiary education.

This gap is significant because linguistic patterns, including lexical choice, clause complexity, and cohesion, vary according to proficiency level. Malaysian ESL writing research, including metadiscourse analyses and lexical error studies, Mat Zali et al. (2022), identifies systematic developmental trends in early academic writing that may interact differently with AI-generated text features than in higher-proficiency university essays Ramachandran et al. (2024) For example, diploma students' essays may demonstrate lower lexical density and fewer complex subordination structures than both AI-generated texts and more advanced student corpora, resulting in distinct contrastive profiles that can be quantified using corpus metrics. Combining cohesive device tracking with traditional structural and lexical complexity indices provides a more robust framework, making it pragmatically essential for profiling early-stage academic writers in this specific context.

2.4. Critical Gaps and Rationale for the Present Study

The present study aims to compare student writing with AI-generated text. Even though studies on ESL writing and AI-generated texts are abundant, they are commonly treated as individual texts and not compared to each other. In addition, comprehensive comparative corpus analyses from the perspective of Malaysian ESL diploma writing remain scarce.

Previous studies explicitly examine how low- to mid-proficiency ESL writers differ from AI-generated texts, despite the fact that proficiency mediates the presence of argumentative coherence. Malaysian academic writing research offers insights into structural and rhetorical challenges but seldom situates these within AI-assisted contexts that are crucial for informed pedagogical responses.

By addressing these gaps, this study aims to contribute to the literature on ESL writing and the utilisation of AI in language education. Furthermore, the findings of the present study can provide evidence-based insights that can be the starting point for further investigation on ESL writing and AI-generated texts.

2.5. Summary

The literature to date reveals that corpus linguistics provides robust tools to profile linguistic features crucial for comparative text analyses. Within ESL settings, student-produced essays consistently exhibit characteristic patterns in cohesion, lexical choice,

and developmental discourse signals. Conversely, AI-generated texts differ systematically from human writing in structural diversity, syntactic complexity, and overall stylistic signatures. Despite these insights, there remains a notable gap in comparing AI and student essays within Malaysian ESL diploma contexts, particularly through the use of integrated, multidimensional measures. Identifying these contrasts through corpus analysis not only addresses theoretical questions regarding language generation and proficiency but also directly informs pedagogical practices in ESL writing.

3. **Research Methods**

3.1. Data and Corpus Compilation

Two comparable corpora were compiled to facilitate a systematic comparison between AI-generated and student-produced essays. The AI corpus comprised essays generated using a large language model, while the comparison corpus consisted of essays written by students under non-AI conditions. Purposive sampling technique was employed for data collection, whereby all the students involved were enrolled to a proficiency English language course taught by the researchers. All texts were controlled for topic, genre (expository), and length (250–300 words) to ensure comparability.

For the AI-generated corpus, 50 students were instructed to use ChatGPT to produce an expository essay based on a standardised prompt. Students were instructed not to edit or revise the generated output; the essays were copied verbatim into a standard submission template and submitted via an online folder provided by the lecturers.

For the student-produced corpus, the same students wrote an expository essay on the same topic in a supervised classroom setting. The use of AI tools or external assistance was prohibited. Students completed the task by hand within one hour and submitted the scripts directly to the lecturers for digitisation.

The controlled production conditions ensured comparability, enabling a reliable examination of linguistic and rhetorical features across AI-generated and human-authored essays. The final count of the AI-generated essay is 44 and 47 for student-produced essays. Others were excluded from the datasets as they did not meet the criteria set by the research. Table 2 below summarises the composition of the corpora.

Table 2: Composition of AI-generated and student-produced essays corpora

	No of Essays	**No of words**
AI-Generated Essays Corpus (AI_C)	44	8292
Student-Produced Essays Corpus (SP_C)	47	9763

3.2. Identification of Discourse Markers

Discourse markers were identified based on an established functional taxonomy adapted from Fraser (2009) and Hyland (2005). These frameworks were selected for their widely recognised categorisation of discourse markers according to their pragmatic and rhetorical functions in academic discourse. In this study, discourse markers were classified into four main functional categories:

- Elaborative markers (e.g. *and, moreover, in addition*)
- Contrastive markers (e.g. *however, but, on the other hand*)
- Inferential markers (e.g. *because, therefore, as a result*)
- Temporal markers (e.g. *firstly, finally, another*)

Following this, an automated corpus analysis tools, in particular the Words tool in LanscBox X (Brezina & Platt, 2025) was used to extract discourse markers from both corpora. To ensure accuracy, concordance lines were manually checked to exclude non-discourse uses (e.g. and used within noun phrases such as "hide and seek").

3.3. Analytical Procedures

The next step of analysis is to normalise the frequency counts. In order to facilitate cross-corpora comparison, the frequency counts of the retrieved discourse markers were normalised per 1,000 words. Following this, the frequency of total discourse markers and the frequency of DM per functional category (EM, CM, IM, and TM) in both corpora were computed and compared to analyse the DM diversity. Measures of variety (type–token ratio for discourse markers) were calculated to examine DM range. To evaluate contextual appropriateness and functional effectiveness of DM, concordance analysis using the KWIC tool was then administered on selected samples extracted from the two corpora.

4. Results

4.1. Discourse Marker Use in AI-Generated Essays

Tables 3 and 4 below summarise the raw frequency, range and normalised frequency (over 1000) of DM in AI-generated essays. As shown in the Table 3 AI-generated essays recorded an overall dominance of elaborative markers (EM) (514/ 76%), with a very heavy reliance on *and* (432 tokens). Meanwhile, temporal markers (TM) recorded the least usage with only 33 tokens (4.9%). Inferential markers (IM) both recorded moderate usage with 62 and 60 occurrences respectively.

There also appears to be an imbalance distribution of DM within each category, in the EM category only *and* appears to be overused, while other connectors especially the former alternatives (*furthermore, additionally, moreover*) were underused. A similar trend is observed in contrastive markers CM, IM and TM, whereby each category relies on a very restricted variety of DM (CM- *while*; IM-*in conclusion, therefore*; TM-*another*).

In terms of range of occurrences, *and* recorded the highest percentage of about 98 percent, followed by *while* with about 89 percent, *also* with 75 percent, *in conclusio*n and *therefore* sharing 59 percent each and *in addition* with 55 percent. These figures indicate a wider range of usage of *similar* basic connectors across the AI-generated texts in achieving text clarity and coherence. Overall, the findings suggest that although AI-generated essays demonstrate consistent cohesion, they rely on a restricted and repetitive repertoire of discourse markers, reflecting limited functional and stylistic variation. This corroborates with previous research findings that AI tools (e.g., ChatGPT, Grammarly) reliably enhance surface features such as syntax and cohesion, however, the vocabulary breadth and lexical diversity in AI-supported writing are still limited (Conde et al., 2024; Goyibova et al., 2025; Shi et al., 2025).

Table 3: Distribution of DM in AI-generated essays according to category

Category	Discourse Marker	Frequency	Range (%)	Normalised Freq
Elaborative Markers	And	432	97.73	52.1
	Also	46	75	5.5
	in addition	24	54.55	2.9
	Furthermore	6	13.64	0.7
	Similarly	3	6.82	0.4
	Additionally	1	2.27	0.1
	Moreover	1	2.27	0.1
	Besides	1	2.27	0.1
	Total	514		62.0
Contrastive Markers	While	44	88.64	5.3
	Although	7	15.91	0.8
	Instead	5	11.36	0.6
	However	3	6.82	0.4
	But	2	4.55	0.2
	Despite	1	2.27	0.1
	Total	62		7.5
Inferential Markers	in conclusion	27	59.09	3.3
	Therefore	26	59.09	3.1
	Because	5	11.36	0.6
	as a result	2	4.55	0.2
	Total	60		7.2
Temporal Markers	another.	22	50	2.7
	one (of the)	5	11.36	0.6
	finally,	3	6.82	0.4
	Lastly	3	6.82	0.4
	Total	33		4.0

Table 4: Overall distribution of DM in AI-generated essays

Category	Raw Frequency	Percentage	Normalised Freq
Elaborative Markers	514	76.8	62.0
Contrastive Markers	62	9.3	7.5
Inferential Markers	60	9.0	7.2
Temporal Markers	33	4.9	4.0
Total	669	100	80.7

4.2. Discourse Marker Use in Student-Produced Essays

Tables 5 and 6 presents the distribution of DM in student-produced essays. Similar to the trend in the AI-generated essay, EM appears to be the most dominant, recording 367 tokens (62%), A substantial proportion of these tokens is contributed by the connector *and (287 tokens)*, indicating a heavy reliance on a single additive marker to achieve cohesion. Ranking second is IM with 100 tokens (16.8%), while CM and TM are moderately represented with 60 (10%) and 69 tokens (11%) respectively.

There also appears to be high-frequency dependence on a very restricted pool of DM from each category for instance, *and* overwhelmingly dominates the EM category (285 tokens), *but* is the primary marker for contrast (CM: 26 tokens), *in conclusion* and *so* account for much of the inferential usage (IM: 35 and 27 tokens respectively), and sequencing or temporal marker (TM) is largely achieved through *firstly* (30 tokens) and *secondly* (22 tokens). This trend is similar to that recorded in AI-generated essays, whereby cohesion is achieved by a restricted and basic DM repertoire. This suggests that cohesion in both datasets is achieved through a limited and relatively basic set of DM.

As for the range of usage, several DM appear to be used comparatively more frequently across the texts, among them include *and* (98%), *in conclusion* (74%), *firstly* (64%), *also* (60%), while others appear to be underused, indicating limited functional variety. Overall, the findings point to a narrow discourse marker repertoire, reflecting constrained lexical and functional diversity in structuring ideas and signaling relationships in students writing.

Table 5: Distribution of DM in student-produced essays according to category

Category	Discourse Marker	Frequency	Range (%)	Normalised Freq
Elaborative Markers	And	285	97.87	29.19
	Also	51	59.57	5.22
	Furthermore	10	21.28	1.02
	addition (In addition)	9	17.02	0.92
	Besides	5	10.64	0.51
	Another	5	10.64	0.51
	Moreover	2	4.26	0.20
	Total	367		37.59
Contrastive Markers	But	26	40.43	2.66
	While	13	25.53	1.33
	However	10	17.02	1.02
	Instead	7	14.89	0.72
	Still	2	4.26	0.20
	Despite	2	4.26	0.20
	Total	60		6.15
Inferential	conclusion (In conclusion)	35	74.47	3.58

Markers				
	So	27	34.04	2.77
	Therefore	14	25.53	1.43
	Then	6	10.64	0.61
	Hence	6	10.64	0.61
	Thus	5	8.51	0.51
	Conclude	5	8.51	0.51
	nutshell (In a nutshell)	1	21.3	0.10
	Consequently	1	2.13	0.10
		100		10.24
Temporal Markers	Firstly	30	63.83	3.07
	Secondly	22	46.81	2.25
	First	8	17.02	0.82
	Next	6	12.77	0.61
	Lastly	1	2.13	0.10
	Finally	1	2.13	0.10
	Second	1	2.13	0.10
	Total	69		7.07

Table 6: Overall distribution of DM in student-produced essays

Category	Frequency	Percentage	Normalised Freq
Elaborative Markers	367	61.6	37.6
Contrastive Markers	60	10.1	6.1
Inferential Markers	100	16.8	10.2
Temporal Markers	69	11.6	7.1
Total	596	100.0	61.0

4.3. Comparison of DM Usage Across Corpora

4.3.1 Comparison of Types and Tokens

In discerning the convergence and divergence in the usage of DM in AI-generated and student-written essays, type-token analysis was administered. Table 7 below summarises the findings.

There exists a slight divergence in the overall usage of DM in AI-generated essays (699 tokens) than in student-written essays (596 tokens), suggesting perhaps more repetitions of fixed frameworks in AI – generated texts as suggested by Shi (2024).

Table 7: Comparison of types and tokens of DM across corpora

	AI-Generated Corpus		Student-Produced Corpus	
Category	**Types**	**Tokens**	**Types**	**Tokens**
Elaborative Markers	8	514	7	367
Contrastive Markers	6	62	6	60
Inferential Markers	4	60	8	100
Temporal Markers	4	33	5	69
Total	22	669	26	596

As presented in Table 7 and Figure 1, both datasets appear to be dominated by EM, with significantly heavy dependencies on and and also. This could stem from alignment with the expository genre, which calls for more additive markers in topic expansion and idea development. AI-generated essays, however, recorded a higher frequency of EM than student-written essays (514 tokens vs student's 367 tokens), but with almost identical types of DM. These findings suggest a comparatively stronger tendency for repetition of high-frequency DM by AI.

As illustrated by Figure 1, student essays show a higher preference for IM (100 tokens vs AI's 60 tokens) such as in conclusion, so and therefore signalling perhaps an over-reliance of explicit logical connectors; a strategy stemming from the need to demonstrate logic. Student essays also show a higher token of TM specifically ordinal sequencers firstly and secondly (69 tokens vs AI's 33 tokens). This indicates students' gravitation towards overt signposting in achieving text coherence, whereas AI shows preference for a general temporal marker another in introducing new ideas.

Both datasets used almost identical tokens of CM with both recording about 60 tokens each. Despite the overall similar frequency, student writers relied more heavily on basic contrasting connector specifically but, a clear divergence of the preference of while in the AI corpus. This seems to suggest contrasting stylistic and discourse patterns in marking differences and oppositions.

Figure 1: Comparison of total DM tokens by Category

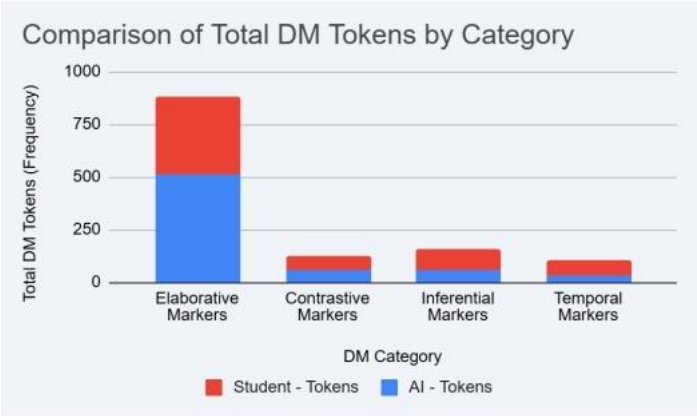
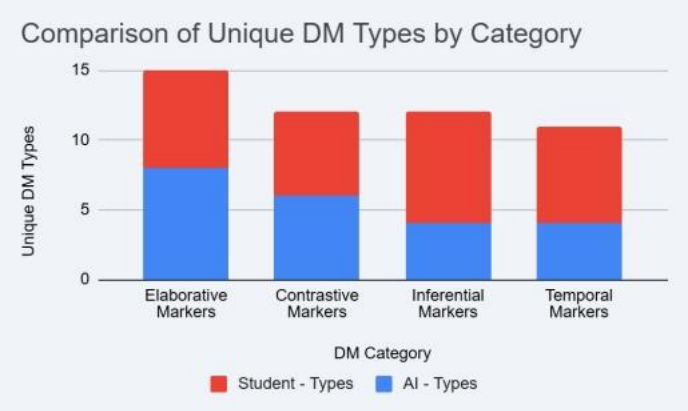


Figure 2 illustrates the comparison of DM types across the datasets. AI essays used a slightly higher count of EM category with 8 types compared to 7 types in the student-written essays. Meanwhile, student essays exceeded the number of types in the IM category by 1-fold (Student-8 types; AI-4 types), indicating a broader vocabulary of

reasoning words which could result from the pressure for establishing visible reasoning since grading criteria might be based on logical connections. Student essays also recorded slightly higher types for the TM category (5 types vs AI's 4 types). Meanwhile, the CM category sees an equal number of types across corpora. As mentioned earlier, even though, they share the same number of types, they diverge in the selection of DM.

Figure 2: Comparison of DM types by category



To determine whether the difference in the distribution of tokens between the AI-generated essays and the student-written essays is statistically significant, a Chi-squared test of independence (X2) was performed. This test evaluates whether the frequency distribution of DM categories (Elaborative, Contrastive, Inferential, and Temporal) differs significantly between the two corpora.

The test was conducted on the contingency table of token frequencies across the four categories:

Table 8: Summary of Statistical Findings

Metric	Value	Interpretation
Chi-Square Statistic (X^2)	43.20	High value suggests strong divergence.
p-value	< 0.00000001	Extremely significant (much lower than 0.05).
Degrees of Freedom (*df*)	3	Based on the 4 categories and 2 groups.
Cramer's V (Effect Size)	0.185	Represents a small-to-moderate effect.

As presented in Table 8, a Chi-square test of independence was conducted to determine whether the distribution of DM categories differs significantly between the AI-generated essays and the student-written essays. Findings of the statistical analysis revealed that there is a statistically significant association between essay type and DM category, χ^2 (3) = 43.20, $p < 0.001$, indicating that the distribution of EM, IM, CM, and TM differed significantly across the two corpora. The extremely small p-value suggests that the variation in DM usage is not due to random chance. Furthermore, the obtained Cramer's V value of 0.185 indicates a small-to-moderate effect size, suggesting that although the magnitude of the association is not large, it is meaningful and reflects distinct DM usage patterns between AI-generated and student-written texts.

Therefore, based on findings shown in Table 8, it demonstrates that AI-generated essays and student-written essays employ DM in systematically different ways. AI-generated essays rely predominantly on EM, reflecting a tendency to extend ideas through additive connectors, resulting in repetitive discourse structures. On the other hand, student-

written essays exhibit greater use of IM and TM, indicating stronger logical sequencing, causal reasoning, and chronological organisation of ideas. Although AI-generated texts are generally cohesive, their reliance on EM suggests a preference for expanding information rather than constructing nuanced argumentative relationships.

These findings support previous studies which argue that, despite producing linguistically fluent and coherent text, generative AI has limitations in replicating the deeper reasoning processes demonstrated in human writing (Amirjalili et al., 2024; Jen & Salam, 2024; Lin, 2023; Zaheer et al., 2025). In particular, the lower use of IM and TM in AI-generated essays suggests that AI is less effective in signalling logical progression, causal relationships, and argumentative development. These discourse functions require critical thinking and purposeful organisation of ideas, which remain more evident in student-produced writing. Therefore, while AI can effectively produce coherent and grammatically accurate text, it cannot fully substitute for domain-specific critical thinking and original argumentation, particularly in the utilisation of IM and TM to indicate logical connections and reasoning.

4.4 Concordance analysis to evaluate contextual appropriateness and functional effectiveness.

Samples of concordance lines presented in Table 9, reveal that the DM used in the AI-generated essays are generally grammatically accurate and contextually appropriate. Connectors that indicate cause-and-effect relationships (thus), contrast (while), addition (furthermore), sequencing (another), and closure (in conclusion) are placed at logical and suitable locations. For example, in conclusion appropriately indicates the final evaluative viewpoint, while therefore successfully introduces a consequence or proposal after a discussion of negative impacts. Similarly, while functions correctly as a concessive marker, it presents balanced arguments by first highlighting the advantages of online gaming before discussing its drawbacks. This shows that AI achieve structural consistency and adhere to the conventions of expository essay writing by frequently using these markers. In addition, although AI can introduce points smoothly by using DMs like furthermore and another, they do not necessarily strengthen arguments, develop reasoning or create subtle transitions. As a result, cohesion is not achieved by using varied syntactic or lexical strategies but mainly through obvious and specific markers.

Table 9: Extracts from AI_C

AI Code	Sample of Concordance Line	Discourse Marker	DM Category
AI-GEN 20	Excessive gaming reduces focus… **Therefore**, it is important to maintain a balanced lifestyle…	Therefore	IM
AI-GEN 39	**While** online games provide entertainment and social interaction, excessive gaming can lead…	While	CM
AI-GEN 26	…Addicted gamers may experience stress… **Furthermore**, excessive gaming can harm…	Furthermore	EM
AI-GEN 35	**Another** serious consequence is the deterioration of social relationships.	Another	TM
AI-GEN 10	**While** playing games can be fun and relaxing, excessive involvement…	While	CM

AI-GEN 11	**While** online games can offer enjoyment… excessive involvement may lead to addiction.	While	CM
AI-GEN 11	**In conclusion**, online gaming should be enjoyed in moderation.	In conclusion	IM
AI-GEN 11	**Therefore**, it is important to manage gaming time wisely…	Therefore	IM

Further analysis of the concordance lines also reveals that although the use of DMs in the AI corpus is structurally and contextually appropriate, the use of repetition and typical language patterns considerably limit their functional effectiveness. AI's reliance on predetermined template or framework is suggested by the very similar sentence pattern found in several samples in the AI-generated essays corpus, such as "While online games provide...". Thus, the development of discourse can be predicted (i.e. problem elaboration followed by a cautious recommendation), as these identical concessive openers are repeated frequently. This finding is consistent with previous study which reported recurring language patterns in AI-supported texts in scientific and medical writing, implying AI's overreliance on generic language by suggesting common templates or typical expressions (Amirjalili et al., 2024; Zaheer et al., 2025; Erliani et al., 2025; Jiang, 2025).

Overall, although DMs used in the AI-generated texts are functionally accurate and coherent, their repetitive and conventionalised deployment reflects limited stylistic variation and a reliance on standard argumentative scaffolding.

Table 10: Extracts from SP_C

AI Code	Sample of Concordance Line	Discourse Marker	DM Category
STU_ES 43	**To conclude**, online gaming can be harmful if done excessively as it may threaten…	To conclude	IM
STU_ES 1	**Secondly**, is people will always feel unconfidence because some of people playing…	Secondly	TM
STU_ES 1	**In conclusion**, being addicted to video game is a dangerous disease to our life…	In conclusion	IM
STU_ES 8	**In conclusion**, we have to know what is the right time when we want to playing…	In conclusion	IM
STU_ES 1	**Besides that**, online gaming addiction can also cause to unhealthy decision…	Besides that	EM
STU_ES 26	**Secondly**, forget to eat your meals also the danger of online gaming addiction.	Secondly	TM
STU_ES 26	…**Therefore**, teenagers should manage their time wisely so that they can enjoy…	Therefore	IM
STU_ES 38	**As a result**, they will make an unhealthy decision such as meeting them in real life…	As a result	IM
STU_ES 38	…**Therefore**, it is good for us to prevent this problem from now.	Therefore	IM
STU_ES 25	…**But**, there are two dangers of being addicted to online gaming.	But	CM
STU_ES 25	**Firstly**, addiction to online gaming can cause lack of sleep…	Firstly	TM
STU_ES 3	Online game is fun game, most people make it as a hobby **but** it become serious…	But	CM

STU_ES 5	**Firstly**, online gaming addiction can be effect on your academic performance…	Firstly	TM
STU_ES 5	**Secondly**, addiction to the online gaming also make people have difficulty…	Secondly	TM
STU_ES 18	**Firstly**, when you playing for a long time, it could affect your physical health…	Firstly	TM

Table 10 above presents the concordance sample of student-written essays extracted from the corpus. The concordance lines demonstrate that DMs are generally used by the students in contextually appropriate ways to signal sequence, contrast, elaboration, and conclusion. TMs such as firstly and secondly are used by the students to indicate order of points or ideas, particularly in discussing the effects of online gaming on academics, health and social interactions. To summarise points or draw logical consequences, IMs such as to conclude, in conclusion, as a result, and therefore are appropriately used by the students, which indicate their understanding of IMs function in expository writing. CMs such as but are used by the students to highlight opposing viewpoints or to emphasise the advantages of gaming before discussing its negative effects. Similarly, to achieve textual cohesion, the students used EMs like besides that to expand ideas or provide more examples. These patterns indicate that students understand the communicative function of DMs and employ them appropriately to lead readers through their writing.

However, further analysis of the concordance lines extracted from the student writers' corpus shows some inconsistencies in the functional effectiveness of DMs used, which may be due to the varying proficiency levels of the students. For instance, there are cases where students used TMs such as secondly and firstly, in unclear or grammatically incorrect sentences (e.g., "Secondly, is people will always feel unconfidence…"), which affects coherence. Overall, analysis of the samples extracted from the student writers' corpus indicate that the students demonstrate an ability to use DMs in their academic writings, specifically expository writing, but grammatical errors and vague sentences often undermine the effectiveness of these markers.

5. Conclusion

The study has identified the difference in the usage of discourse markers between AI-generated essays and student-produced essays. Although both types of text utilise discourse markers, there are distinct linguistic features that set them apart. AI-generated essays are characterised by the over reliance on elaborative markers. The connector "and" was found to be the dominant marker found in AI-generated essays. Furthermore, the texts exhibit repetition of discourse markers that makes them appear flawless and formulaic. For student-produced essays, they are found to utilise inferential and temporal markers for overt logical signposting. This finding demonstrates the students' ability to utilise discourse markers in their writing. However, grammatical inaccuracies sometimes hinder the effectiveness of the markers within the essays.

Pedagogically, the findings of this study carry significant importance as instructors and learners need to be aware of how the two texts are constructed. Recognising the difference between both texts, instructors can evaluate the authenticity of a text produced by students. Moreover, the findings suggest that instructors should focus on developing students writing skills by helping students to minimise grammatical inaccuracies. This could help to enhance the effectiveness of discourse markers used in students' writing. Other than that, writing pedagogy must prioritise teaching students how to smoothly and

accurately integrate a wider range of cohesive devices into more complex sentence structures, moving them away from predictable text progression.

Considerably, more research is required to have better understanding of this issue. The findings of this study could be used as a starting point for further investigation on differences between students' writing and AI-generated text. Future research could explore the differences in the usage of discourse markers of different academic texts such as expository essay or argumentative essay. Additionally, future research could also include other types of linguistic features in writing such as word choice or sentence structure. These studies could shed more light to this issue, thus, providing better understanding on the differences between AI language generation and human-produced language.

Ethics Approval and Consent to Participate

Ethical considerations were observed throughout the study. The students provided informed consent for their essays to be used as research data. The essays were originally produced as part of regular classroom activities and were analysed only for research purposes after obtaining students' permission. As the data were derived from routine classroom practice and did not involve any intervention beyond normal teaching activities, formal ethical approval was not required.

Acknowledgement

The authors would like to express their sincere gratitude to all participants for their willingness to participate in the study and made this research possible.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflict of Interest

The authors declare there are no conflicts of interest.

References

Abdullah, S. (2015). A data-driven contrastive study on Malay ESL learners' use of lexical verbs and verb-noun collocations in argumentative writing (Doctoral dissertation, Universiti Teknologi MARA, Shah Alam, Malaysia). UiTM Institutional Repository. https://ir.uitm.edu.my/id/eprint/19589/

Adnan, A. H. M. (2023). Corpus-based analysis of lexical errors in paragraph writing among low-proficiency Malaysian ESL diploma students. *Journal of Creative Practices in Language Learning and Teaching, 11(*2), 45–61. https://doi.org/10.24191/jcpilt.v11i2.23412

Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English

literature. *Frontiers in Education*, *9*, Article 1347421. https://doi.org/10.3389/feduc.2024.1347421

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, *8*(4), 243–257. https://doi.org/10.1093/llc/8.4.243

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Brezina, V., & Platt, W. (2025). *#LancsBox X* [Computer software]. Lancaster University. http://lancsbox.lancs.ac.uk

Conde, J., Reviriego, P., Salvachúa, J., Martínez, G., Hernández, J. A., & Lombardi, F. (2024). Understanding the impact of artificial intelligence in academic writing: Metadata to the rescue. *Computer*, *57*(1), 105–109. https://doi.org/10.1109/mc.2023.3327330

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*(5), 570–590. https://doi.org/10.1093/applin/amt056

Erliani, Nys. N. P., Eryansyah, & Amrullah. (2025). The Impact of AI Tools on EFL Students' Self-Efficacy in Academic Writing: A Qualitative Study. *Edukasi: Jurnal Pendidikan Dan Pengajaran*, *12*(01), 6–17. https://doi.org/10.19109/s75mbm13

Fredrick, D. R., & Craven, L. (2025). Lexical diversity, syntactic complexity, and readability: A corpus-based analysis of ChatGPT and L2 student essays. *Frontiers in Education*, *10*, Article 1616935. https://doi.org/10.3389/feduc.2025.1616935

Georgiou, G. P. (2024). Differentiating between human-written and AI-generated texts using linguistic features automatically extracted from an online computational tool. *arXiv*. https://doi.org/10.48550/arxiv.2407.03646

Goyibova, N., Muslimov, N., Kannazarova, Z., Kadirova, N., Alautdinova, K., & Ismatullaeva, I. (2025). Exploring the impact of artificial intelligence on academic writing: A bibliometric analysis of trends, advancements, and ethical challenges. *Forum for Linguistic Studies*, *7*(6), 342–360. https://doi.org/10.30564/fls.v7i6.9054

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, *13*(1), Article 18617. https://doi.org/10.1038/s41598-023-45644-9

Herbold, S., Hautli-Janisz, A., Heuer, U., Krell, M. T., Lorbach, J., Dey, M., Zhang, B., Schwinn, A., Trautsch, L., Lütjen, B., & Heidinger, C. (2023). A large-scale comparison of human-written and AI-generated essays. *Scientific Data*, *10*(1), 1–15. https://doi.org/10.1038/s41597-023-02301-w

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.

Jen, S. L., & Salam, A. R. (2024). Using artificial intelligence for essay writing. *Arab World English Journal*, (Special Issue on ChatGPT), 90–99. https://doi.org/10.24093/awej/ChatGPT.5

Jiang, Y. (2024). Interaction and dialogue: Integration and application of artificial intelligence in blended mode writing feedback. *The Internet and Higher Education*, *64*, 100975–100975. https://doi.org/10.1016/j.iheduc.2024.100975

Jiang, F. K., & Hyland, K. (2024). Framing student writing: Metadiscourse in human and AI-generated essays. *Assessing Writing, 62*, Article 100863. https://doi.org/10.1016/j.asw.2024.100863

Kamaludin, P. N. H. B. (2014). *Factors affecting the assessment of ESL students' writing: A case study of UiTM English language lecturers* (Unpublished master's dissertation). Universiti Teknologi MARA.

Lin, Z. (2023, October 19). *Techniques for supercharging academic writing with generative AI*. PsyArXiv. https://doi.org/10.31234/osf.io/9yhwz

Liu, Y., Han, T., Sun, S., Liang, C., & Passonneau, R. J. (2023). ArguGPT: Exploiting LLMs for generating argumentative essays and analyzing their stylistic signatures. *arXiv*. https://doi.org/10.48550/arXiv.2304.14072

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

Mat Zali, M., Mohd Razlan, R., Raja Baniamin, R. M., & Setia, R. (2022). Interactional metadiscourse analysis of ESL learners' essays. *Environment-Behaviour Proceedings Journal*, *7*(SI 9), 55–60. https://doi.org/10.21834/ebpj.v7isi9.4248

Mizumoto, A., & Eguchi, M. (2023). Exploring the linguistic fingerprints of AI-generated academic texts: A corpus-driven comparison with human scholarship. *Journal of English for Academic Purposes, 65*, Article 101287. https://doi.org/10.1016/j.jeap.2023.101287

Mohamed, A. F., Ab Rashid, R., Lateh, N. H. M., & Kurniawan, Y. (2021). The use of metadiscourse in good Malaysian undergraduate persuasive essays. INSANIAH: Online Journal of Language, Communication, and Humanities, 4(1), 1–12.

Ramachandran, S., Jalaluddin, I., & Sharatol Ahmad Shah, S. (2024). Investigating lexical collocational errors in the argumentative essays of Malaysian tertiary ESL learners. *3L: Language, Linguistics, Literature, 30(*1), 112–128. https://doi.org/10.17576/3L-2024-3001-08

Shi, J., Liu, W., & Hu, K. (2025). Exploring how AI literacy and self-regulated learning relate to student writing performance and well-being in generative AI-supported higher education. *Behavioral Sciences*, *15*(5), Article 705. https://doi.org/10.3390/bs15050705

Ting, S. H., Raslie, H., & Jee, L. J. (2011). Case study on the persuasiveness of argument texts written by proficient and less proficient Malaysian undergraduates. *Malaysian Journal of Learning and Instruction*, *8*, 71–92. https://doi.org/10.32890/mjli.8.2011.7627

Yildiz Durak, H., Eğin, F., & Onan, A. (2025). A comparison of human-written versus AI-generated text in discussions at educational settings: Investigating features for ChatGPT, Gemini and BingAI. *European Journal of Education*, *60*, Article e70014. https://doi.org/10.1111/ejed.70014

Zaheer, S., Ma, C., Zhu, Y., & Vasinda, S. (2025). GenAI in academic writing—empowering learners or redefining traditional pedagogical practices?: A systematic review from 2019-2023. *International Journal of Artificial Intelligence (AI) in Teaching and Learning (IJAITL)*, *1*(1), 1–34. https://doi.org/10.4018/IJAITL.373582