AI, HUMANS, AND THE SPACE IN BETWEEN: A CORPUS-BASED EXAMINATION OF STANCE, HEDGING, AND LEXICAL BUNDLES IN ELT WRITING

Öğr. Gör. Abdullah YOLDAŞ

Hakkari Üniversitesi, Yabancı Diller Yüksekokulu

Öğr. Gör. Fırat KESKİN

Hakkari Üniversitesi, Yabancı Diller Yüksekokulu

ABSTRACT

The integration of Large Language Models (LLMs) into academia has formed a "grey zone" of authorship between humans and machines. The purpose of this study is to examine this "space in between" through a comparison of corpus studies of stance, hedging, and lexical bundles in English Language Teaching (ELT) writing. The research is based on a specialized corpus consisting of 90 texts, which have been divided into Human Corpus (HC), AI-Generated Corpus (AIC), and AI-Likely Corpus (ALC) categories, with which the study demonstrates that AI-assisted writing produces a novel 'synthetic' register. The research demonstrates that a core difference between the texts is the presence of authors. In human-authored texts the presence of a self was used to signal a root modality (ability) and a personal recount of a story. In contrast, AI-likely authored texts self-mention authors without real stance and instead used the self-mention for structural purposes and thus had a "hallucination of voice." Moreover, hybrid texts signal an "epistemic shift" toward a vague possibility of improvement and a "listicle effect" which describes a disproportionate use of fixed and rigid enumerative structures with frames of justification. These results point to a concerning homogenization of student discourse, suggesting that assessment frameworks must be revised to prioritize the functional nuances of organic human authorship over the algorithmic structural rigidity of AI-mediated texts.

Keywords: AI writing; AI-likely text, corpus linguistics; hedging; lexical bundles; ELT discourse

INTRODUCTION

The creation of AI tools like ChatGPT and Gemini has completely changed the ways in which academic writing is produced, assessed, and understood in the higher education context. How these AI models assist and even completely write student papers raises concerns about authorship, style uniformity, academic integrity, and how the discipline of writing should be taught in the future. As it becomes more and more difficult to distinguish human writing from machine writing, new problems in detecting, evaluating, and teaching writing come to the fore in the field of English Language Teaching (Stokel-Walkoer, 2022; Liang et al., 2023).

The current research investigates three types of texts in this new context: (1) student-authored essays produced in supervised, non-AI conditions, (2) essays generated directly from LLMs, and (3) student essays that, according to three different detection tools, have characteristics of being largely AI-authored. While the research question is mainly aimed at using AI to better understand how different dimensions of academic writing can be reshaped, this understanding is key to achieving more equitable assessment in multilingual classrooms for weaker non-native English writers (Liang et al., 2023).

This study analyzes three discourse features relevant to writing proficiency: stance expressions, hedges, and lexical bundles, which have been pointed out as crucial elements in establishing authorial presence, epistemic commitment, and fluency and cohesion within a text (Hyland, 1998, 2003, 2005; Biber et al., 1999). Comparing these features in human, AI, and AI-like writing through corpus-based methodologies, this research investigates possible implications of AI in academic writing conventions in ELT contexts. This study aims to explore the implications of AI in higher education, specifically in the area of redesigning assessment frameworks, managing academic integrity, and teaching AI literacy to future English Teachers.

The present study, therefore, addresses the following research questions:

RQ1. To what extent do human-written, AI-generated, and AI-likely texts differ in their use and frequency of stance markers?

RQ2. How do the three sub-corpora differ in their use and distribution of hedging devices?

RQ3. What similarities and differences emerge in the frequency and functional types of 4-word lexical bundles across the three sub-corpora?

RQ4. To what degree can these linguistic features distinguish human writing from AI-generated texts, and where do AI-likely texts position themselves between these two extremes?

RQ5. Do AI-likely student texts exhibit linguistic patterns that position them closer to human writing or to AI-generated writing?

LITERATURE REVIEW

Lexical Bundles in Academic and ELT Writing

Lexical bundles-recurrent combinations of multiple words-fundamentally characterize academic writing and demonstrate an author's discourse competence. Lexical bundles are tied to discipline-specific cohesion and fluency, and to the writer's disciplinary identity (Biber et al., 1999). Scholarly literature suggests that systematic and strategic use of lexical bundles is characteristic of proficient academic writers, while inexperienced academic writers and L2 writers depend on a narrower and more restricted lexical bundle (Biber & Barbieri, 2007; Cortes, 2004).

Uçar and Zarfsaz (2022) discuss L2 writing contexts and identify the same kind of tendencies. In the study, it is claimed that Turkish EFL students are less proficient than other writers in terms of the number of bundle types used, heavy reliance on a restricted set of stance bundles, and the underuse of the referential bundles. Other studies based on corpora also demonstrate that the frequency and distribution of lexical bundles depend on the writer's proficiency and discipline and suggest that these bundles can potentially serve as indicators of the writer's proficiency and writing in a particular discipline (Livytska, 2019; Dang, 2020).

In light of these findings, bundles of lexicon stand to appreciate human influence, and AI influences comparatively. While AI-generated prose is structurally coherent and bundles are recognizable, the question remains as to how many of these bundles adhere to the natural academic discourse patterns (Floridi & Chiriatti, 2020; Köbis & Mossink, 2021). For this reason, the current study’s examination of the three corpora may help unearth stylistic fingerprints left by AI.

Hedging, Stance, and Metadiscourse in Academic Writing

One defining feature of academic discourse is hedging and stance taking. As described by Hyland (1998,2005), hedging enables writers to make tentative claims and communicate varying degrees of certainty and epistemic nuance, while stance is concerned with the attitudes, evaluations and presence of the writer. These elements help authors to orient and engage the readership, position the claims, and fulfill the expectations of the field. In second language writing, the use of hedging differs from

the native standard such that L2 writers exhibit either underuse of hedges or a very elementary range of modals, which could lower the interpersonal effectiveness or rhetorical adequacy of the discourse (Hinkel, 2004; Flowerdew, 2002).

Cross-linguistic studies also confirm that the use of hedging differs with respect to the participants' cultural and educational background (Dang, 2020). The difference in the use of hedging also appears in various academic disciplines, indicating that hedging is responsive to contextual and disciplinary requirements (Varttala, 2001).

The growing reliance on AI tools raises the question of how AI handles texts in terms of epistemic certainty, transparency, and the position of the writer. Early studies have suggested that texts generated by LLMs use less rhetorical flexibility or have a tendency to employ formulaic expressions of stance (Floridi & Chiriatti, 2020). Given that effective hedging is a defining characteristic of academic discourse, the comparison of hedging and stance markers in human versus AI-generated and AI-likely texts will help elucidate how AI mediation is transforming academic discourse.

AI, Academic Writing, and Emerging Challenges

The speed of adoption of AI writing tools has highlighted weaknesses in existing detection and assessment systems. AI detection systems have been shown to classify human writing as machine-generated, especially when such writing is produced by nonnative English writers (Liang et al., 2023). Such biases are detrimental to equity in EFL/ESL assessment and call for empirically anchored teaching and learning approaches to writing assessment in the age of AI. Likewise, research assessing AI-generated writing explain the tools' ability to produce writing that is fluid and grammatically correct, but that may also lack sophisticated control of rhetoric, manifesting as overly repetitive or general reasoning (Köbis & Mossink, 2021).

Recently published corpus-based research also investigates the norms of academic writing in the age of AI. For instance, Anderson et al. (2024) analyzed and identified speculated stylistic changes in large collections of academic texts and warned about the standardization of academic writing. However, in ELT contexts, the research discrepancy involving the comparison of AI-generated, AI-assisted, and human-produced student texts is scarce.

This gap will be analyzed through an examination of the use of stance, hedging, and lexical bundles within three different text types. Applying established corpus-linguistic approaches (Baker, 2006; McEnery & Wilson, 2001) and theoretical frameworks from academic discourse (Hyland, 1998, 2005; Biber & Barbieri, 2007; Cortes, 2004), the study aims to provide an evidence-based account of how AI influences the linguistic construction of academic writing in ELT contexts.

METHODOLOGY

Research Design

This study employs a comparative corpus-based research design to investigate the linguistic variations in stance, hedging, and lexical bundles across three distinct corpora: human-authored, AI-generated, and AI-likely (hybrid) texts. The research is exploratory in nature, and it aims to characterize the "space in between" fully human and fully machine-generated academic writing within an ELT context.

To ensure the comparability of the datasets and control for topic-based lexical variation, all texts in the corpus focus on the same subject matter: "Reading Skills and Strategies in EFL Learning." By keeping the topic the same, the study isolates authorship as the primary variable that could affect linguistic choices.

Corpus Compilation

The study utilizes a specialized small corpus comprising 90 texts in total which were divided equally into three sub-corpora (N=30 for each). The texts are short academic paragraphs (approx. 150-200 words each), typical of 2nd-year undergraduate writing tasks. Table 1 details the composition of the corpus.

Table 1. *Composition of the Study Corpus*

Sub-Corpus	N (Texts)	Source	Production Condition	Selection Criteria
Human Corpus (HC)	30	2nd-Year ELT Students	In-class, pen-and-paper, supervised.	No technology access; verified human authorship.
AI-Generated Corpus (AIC)	30	ChatGPT (GPT-4) & Gemini	Systematically prompted generation.	Diverse "student personas" simulated via prompt engineering.
AI-Likely Student Corpus (ALC)	30	Student Assignments (LMS)	Unsupervised, digital submission.	Topic alignment + AI Detection Score (80-100%).

The Human Corpus (HC)

The Human Corpus serves as the baseline for "authentic" learner language. Data was collected from 30 second-year ELT students during a "Reading Skills" course. Students were asked to write a single paragraph that reflects their preferred reading strategies. To ensure ecological validity and zero AI interference, the task was completed in a controlled classroom environment using pen and paper, without access to mobile devices or computers.

The AI-Generated Corpus (AIC)

The AI-Generated Corpus consists of 30 short academic paragraphs produced by ChatGPT (GPT-4) and Gemini 3 Pro as a response to the same writing prompt used for the human participants: *"Describe one pre-reading strategy that you would like to use in your reading classes and explain the reasons."* To maintain full comparability with the Human Corpus, no stylistic constraints, personas, or proficiency level instructions were given to the model. Each response was required to be approximately 150–180 words and written in a formal academic tone appropriate for undergraduate ELT coursework. The model was prompted several times to generate 30 unique paragraphs, which ensures natural lexical and syntactic variation without artificial manipulation. All texts were saved in plain-text format and incorporated into the corpus exactly as produced by the model, without any post-editing or human intervention.

The AI-Likely Student Corpus (ALC)

The AI-Likely Student Corpus represents the "space in between"—texts that are submitted by students but exhibit strong linguistic elements of AI. These texts were selected from online

assignments that were submitted via Google Classroom for the same "Reading Skills" course. The selection process involved a systematic two-step filtering:

1. **Topic Consistency:** Only assignments addressing the specific topic of "Reading Strategies" were considered to maintain lexical comparability with the HC and AIC.

2. **Detection Threshold:** Texts were screened using popular AI detection tools. Only submissions flagged with an 80% to 100% probability of being AI-generated were included. These texts are treated as "hybrid" or "AI-mediated," acknowledging the possibility of human editing or heavy reliance on GenAI tools.

Data Processing and Analysis

Prior to analysis, all texts were digitized (for the HC), cleaned of metadata (student names, dates, etc.), and converted into plain text (.txt) format.

The analysis focuses on three specific linguistic features:

1. **Stance and Hedging:** Using Hyland's (2005) interactional metadiscourse framework, the texts are annotated for markers of hedging (e.g., *might, perhaps*) and boosters (e.g., *definitely, in fact*).

2. **Lexical Bundles:** Following Biber et al.'s (1999) multidimensional model, the corpus is analyzed for 4-word lexical bundles (e.g., *on the other hand, it is important to*).

Quantitative analysis involves frequency counts and normalization (per 1,000 words) to allow for comparison across three corpora. Qualitative analysis is then applied to interpret the functional use of identified markers, specifically looking for patterns of "artificial perfection" versus "human nuance".

Data Analysis Tools and Procedures

The quantitative and qualitative analyses of the corpus were completed using AntConc Version 4.2 (Anthony, 2022). AntConc was chosen based on its ability to produce concordances, word lists, and N-grams, all of which are necessary in the analysis of lexical bundles and metadiscourse markers.

To extract lexical bundles, AntConc's N-Gram function was employed. Following the parameters defined by Biber et al. (1999), the search settings were defined in order to locate bundles of 4 words. To avoid the analysis of random patterns of usage and to ensure the bundles exemplified a stable pattern, a frequency threshold was used, where only bundles occurring 3 or more times in the sub corpus were included in the analysis. The bundles were subsequently assigned to Biber's structural and functional (referential, discourse organization, and stance) classifications.

A computer-assisted manual analysis was used to identify markers of stance and hedging.

- Identification: The query list used in AntConc was an established list of interactional metadiscourse markers proposed by Hyland (2005).

- Contextual Verification: The usage of specific linguistic markers is polysemous, as the term "may" denotes both permission and possibility. The "Concordance" tool was employed to analyze each occurrence in context. Only markers functioning as verbs of hedge (tentative) or booster (certain) were coded.

Raw frequency counts of the specified features were adjusted to a standardized rate of occurrence per 1,000 words to enable valid cross-comparison of the three sub-corpora, which are of different sizes. To test what frequency of difference was statistically significant between the Human, AI-Generated, and AI-Likely corpora, Log-Likelihood (LL) tests were performed using the calculator developed by Rayson (2016), as LL is widely considered more reliable for small corpora than standard Chi-square tests.

FINDINGS

Within the three sub-corpora analyzed for the corpus study, the Human Corpus (HC) contains students' genuine essays regarding pre-reading strategies; the AI-Generated Corpus (AIC) contains pre-reading strategy essays generated by large language models; the AI-Likely Corpus (ALC) contains students' essays that were flagged for possible AI generation.

The total size of the analyzed corpus was approximately 9,500 words. To ensure comparability, all frequency counts reported below are normalized per 1,000 words (ptw).

Stance Markers: The "I" vs. The "It" (Addressing RQ1)

RQ1 asked what the differences were in the corpora regarding the use of stance markers. Analyzing **self-mentions** and **boosters** demonstrated differences among the sub-corpora that showed substantial differences between AI and human writing, with the ALC being the most complicated and somewhat positioned in between.

Table 2. *Normalized Frequency of Key Stance Markers (per 1,000 words)*

Feature Category	Marker	Human Corpus (HC)	AI-Generated (AIC)	AI-Likely (ALC)	Log-Likelihood (HC vs AIC)
Self-Mention	*I / My / Me*	**48.2**	0.0	12.4	+184.5 (Sig. $p<0.001$)
Attitudinal	*In my opinion*	**4.5**	0.0	2.8	+14.2 (Sig. $p<0.001$)
Boosters	*Clearly / Obvious*	1.2	2.5	**5.1**	+9.8 (Sig. $p<0.01$)
Causality	*Because*	11.8	8.4	**15.2**	+6.1 (Sig. $p<0.05$)

Human Corpus (HC): There is a revealed authorial presence to the corpus. The students framed their arguments using self-mentions saying, "I would like to use," or "In my opinion." This remains a subjective and narrative stance.

AI-Generated Corpus (AIC): AIC, however, is characterized by what we refer to as "detached authority." There was no use of first person pronouns (I, me, my), but rather, the AI expressed stance through a more objective assertion, using impersonal sentence openers like, "This strategy is effective because," and "This builds confidence."

AI-Likely Corpus (ALC): The ALC showed what we refer to as a "hybrid" stance. This corpus showed some self-mentions, like, "I think," but in far lower frequencies than the HC (12.4 vs. 48.2). However, the ALC overused boosters, like "clearly" and "obvious," more than the other two groups, which suggests an attempt at mimicking academic certainty, which resulted in a somewhat repetitive, overly assertive tone.

Hedging Devices: Nuance vs. Certainty (Addressing RQ2)

RQ2 examined the pattern concerning the distribution of hedging devices. In order to differentiate the strategies in the three sub-corpora, Table 4 depicts the normalized frequencies for the top hedging

devices. There is a clear split in the data; while the human students hedge to communicate personal ambiguities or abilities, AI, on the other hand, hedges to propose a situation theoretically.

Table 3. *Distribution of Key Hedging Devices (per 1,000 words)*

Functional Category	Marker	Human Corpus (HC)	AI-Generated (AIC)	AI-Likely (ALC)	Log-Likelihood (HC vs AIC)
Informal Uncertainty	*Maybe / Kind of*	6.4	0.0	1.2	+22.4 (Sig. p<0.001)
General Potentiality	*Can / Could*	14.5	18.2	24.6	-4.1 (Sig. p<0.05)
Probability	*May / Might*	1.1	5.4	3.8	-12.6 (Sig. p<0.001)
Frequency Hedges	*Often / Usually*	2.3	7.8	5.2	-9.4 (Sig. p<0.01)

The analysis reveals three distinct patterns in how these markers are deployed:

1. **Informal Hedges (HC Dominance):** "Conversational hedges" is a term that is especially applicable to the Human Corpus. Human students utilized the hedge markers like *maybe* and *kinda* (e.g., *"maybe connecting the lessons"* [1], *"kinda addicted"* [2]). These hedges do not occur in the AI texts and their frequency is very low in the AI-Likely corpus, which additionally is proof that the mediation of AI is purging the students' discourse of the inherent "fuzziness" of human language.

2. **Precise Probability (AIC Dominance):** The AI-Generated corpus prefers more formal markers of probability such as *may* and *might* (e.g., *"students who might lack background"*). This indicates the LLM's training in which they are expected to produce discourse that is academically careful and precise.

3. **The "Can" Anomaly (ALC Dominance):** The AI-Likely corpus exhibits an unusually high frequency of the modal *can* (24.6 ptw). However, the quantitative similarity masks a functional difference.

Qualitative Analysis of Modal Function: Beyond frequency counts, analysis of concordances shows a stark functional difference in the deployment of the modal can between the two corpora. In the **Human Corpus**, can is predominantly used to convey ability or permission (**Root Modality)**. For instance, students write "I can choose my shy students" or "You can express yourselves better". In contrast, can in the **AI-Likely Corpus (and AIC)** is used predominantly to convey theoretical possibility **(Epistemic Modality)** and is often situated within repetitive formulaic phrases. These include: "It can be boring," "This method can be more based on memorization," and "It can be helpful.". This functional shift—from the human agent's ability to the abstract entity's possibility—serves as a strong linguistic fingerprint of AI mediation in student writing.

Lexical Bundles: The Structural Fingerprint (Addressing RQ3)

RQ3 examined 4-word lexical bundles. This feature provided the strongest evidence for distinguishing the corpora.

Table 4. *Top 5 Functional Lexical Bundles by Sub-Corpus*

Rank	Human Corpus (HC) (Personal & Action-Oriented)	AI-Generated (AIC) (Abstract & Causative)	AI-Likely (ALC) (Formulaic & Structural)
1	*I would like to*	*is effective because it*	*one of the most*
2	*students to understand the*	*allows students to*	*advantages and disadvantages of*
3	*pre-reading activity in*	*background knowledge to*	*grammar translation method is*
4	*because it helps the*	*the main idea of*	*listening and speaking skills*
5	*in my reading class*	*play a vital role*	*in the target language*

HC Bundles: Human bundles were heavily **verb-based** and **personal**, focusing on the writer's intended actions (*I would like to*) or specific classroom context (*in my reading class*).

AIC Bundles: AI bundles were **abstract** and **causative**. The bundle *is effective because* appeared repeatedly and functioned as a "justification frame." The AI commonly used *noun phrase + preposition* bundles (*the main idea of*, *a visual representation of*).

ALC Bundles: The ALC was dominated by **list-making** and **definition** bundles. Phrases like *one of the most* and *advantages and disadvantages* were ubiquitous. This reflects a "template-like" quality where the writer (or tool) fills in a structural skeleton.

Distinguishing Features and the "Space In Between" (Addressing RQ4 & RQ5)

RQ4 and RQ5 sought to position the AI-Likely texts relative to the other two groups.

The "Robot" Structure with a "Human" Mask: The ALC texts show a distinct structural pattern that aligns closer to the AI-Generated corpus than the Human corpus. The texts in the ALC frequently follow a rigid *Introduction > Pros > Cons > Conclusion* format. This macro-structure is highly similar to the standard output of earlier LLM iterations (like GPT-3.5) when asked to "discuss" a topic.

Linguistic Positioning (RQ5): Log-Likelihood comparisons place the ALC in a hybrid position:

1. Closer to AI: In terms of **Lexical Diversity** and **Sentence Length**, the ALC resembles the AIC. The sentences are grammatically polished and utilize complex subordination (*Although... however...*) more successfully than the HC.

2. Closer to Human: In terms of **Repetitiveness**, the ALC mimics lower-proficiency human writing. It tends to reuse the same nouns (*method*, *students*, *grammar*) a lot within a single paragraph, which lacks the cohesive variation found in the high-proficiency AIC.

The "Hallucination" of Personal Voice: A critical finding is the "Formulaic Self-Mention" in the ALC. While the Human corpus uses "I" to narrate personal teaching desires ("I am a little too concerned about..."), the ALC uses "I" primarily to mark transitions in the argument ("I think it is an effective method", "Finally, I mentioned..."). This suggests that the ALC texts may be the result of students prompting AI to "write an essay giving my opinion," which results in a synthetically personalized text.

Summary of Findings

The comparative study agrees that AI mediation causes certain types of text distortion that distinguish the "hybrid" (AI-Likely) texts from both authentic student writing and pure AI generation:

1. The "Hallucination" of Voice: Genuine human text is structurally self-referential (I, my) in its expression, and is, therefore, able to construct an authorial voice. AI-Likely text, as an example of structural self-reference, uses these to flag internal text transitions (e.g., *In my opinion, Finaly, I mentioned*) and, in doing so, creates an illusion of authorial voice.

2. Epistemic Shift (Ability vs. Possibility): A critical functional divergence was observed in hedging. Human writers use modals like *can* to express personal ability (Root Modality), whereas AI-Likely texts mostly use them to denote theoretical possibility (Epistemic Modality), often within repetitive formulaic structures.

3. The "Listicle" Effect: AI-Likely texts are noted to use more than average of enumerative lexical bundles, advanced articulative resources, and rigid macro structures (e.g., *First, Second, One of the advantages, Pros/Cons*), offering greater instances of simplified, binary, rather than cohesive, argumentative discourse.

4. Rise of the "Justification Frame": Unlike human students who often state their preference, AI-mediated outputs rely on explicit causal markers (This is effective because…) to justify every claim. This reflects the "Explainable AI" training set rather than Natural Argumentative structures that learners use.

DISCUSSION

This study sought to understand the linguistic divide that exists between human authorship and machine-generated texts by examining the linguistic features of stance, hedging, and lexical bundles. The results indicate that the hybrid texts likely generated by AI do not simply fall within a continuum of lower-level novice L2 texts and higher-level texts generated by sophisticated AI. Rather, they convey a unique and perhaps perverse construct of the academically accepted writing genre characterized by "synthetic personalization" and "structural rigidity". These outcomes help to shed light on the concept of the authorial voice and its evaluation in the context of GenAI.

The "Hallucination" of Voice and the Loss of Narrative Identity

One of the most noticeable differences observed was the functional deployment of stance markers, specifically self-mentions (I, my). While the Human Corpus (HC) utilized these markers to construct a narrative identity and express personal desire (e.g., *"I would like to"*), the AI-Likely Corpus (ALC) used them almost exclusively as introductory and organizational phrases (e.g., *"In my opinion," "Finally, I mentioned"*).

This finding contradicts Hyland's (2005) argument regarding interactional metadiscourse. Hyland claims that stance markers indicate instances of authorial presence and evaluation. However, in the ALC, the use of "I" does not indicate an authentic authorial self, but rather a simulation of one. This points to the warning Floridi and Chiriatti (2020) gave regarding LLM texts displaying "detached authority." In the context of texts likely written by AI, we notice the "hallucination of voice" (mimicking an authentic engagement) without any of the narrative intent one would expect of genuine, student-authored text. For L2 writers, as noted by Hinkel (2004), who frequently find it difficult to articulate an authoritative voice, AI tools seem to provide a shortcut: they present the voice of an academic stance without the personal engagement that true discourse competence requires.

Epistemic Flattening: From Ability to Theoretical Possibility

The exploration of hedging devices, most notably the modal *can*, revealed an important functional shift. In instances of *can* usage, human students manifested root modality (i.e., ability/agency), while in the case of AI, in addition to AI-likely text, the usage of *can* was overwhelmingly situated within instances of theoretical potential (i.e., epistemic possibility).

This contrast points to the fact that the mediation of AI might be shifting the cognitive parameters of the students' writing. In moving from "what I can do" (agency) to "what can be done" (possibility), AI-likely texts demonstrate greater formal objectivity. However, this comes at the cost of rhetorical nuance. As the literature notes, L2 writers tend to underuse hedges, or to use them from a limited range (Flowerdew, 2002; Hinkel, 2004). The ALC data suggest that while the assistance of AI improves the frequency of modal usage (moving closer to native norms), it also impoverishes the variability of the usage, leading to formulaic, repetitive patterns. This explains the "categorical rigidity" reported in the findings, where the educatşonal elements were simplified into a binary of "pros and cons" rather than the polyphonic exploration of educational themes with the epistemic clarity described by Hyland (1998).

The "Listicle Effect" and the Homogenization of Discourse

The results of the lexical bundle analysis showed the greatest degree of change in structure. The predominance of enumerative and definitional bundles in the ALC, such as "one of the most" and "advantage and disadvantage," starkly contrasts the actionising bundles in HC and the causative bundles in AIC.

This corroborates the suspicion of Anderson et al. (2024) about the homogenization of scholarly discourse. The texts likely produced by AI embody a "Listicle Effect," as the essay structure becomes a quite rigid and formulaic structure of *Introduction* > *Pros* > *Cons* > *Conclusion*. While lexical bundles, as argued by Biber and Barbieri (2007), do signal a disciplinary affiliation, the bundles in the ALC seem to indicate no discipline, other than the "AI accent" in itself. Such standardization is especially worrisome in ELT contexts. While it may indeed assist students in generating texts that are apparently well structured (Köbis & Mossink, 2021), it seems to circumvent the desired developmental path of learning to articulate discourse that organically integrates a coherent structure. The students producing AI-likely texts are not exhibiting the "causative" sophistication of the pure AI model (GPT-4), but a rudimentary, template-driven imitative version of it.

Pedagogical Implications and Assessment

This unique "AI-likely" registry is testing current assessment approaches. As noted by Liang et al. (2023), non-native speakers are already at a disadvantage due to the biases surrounding detection. Our findings suggest these biases are likely to be greater owing to L2 students utilizing AI to attain structures that emphasize grammatical accuracy but lack a personal voice.

If students are exchanging their informal uncertainty for a robotic precision, then educators have to reconsider what is really being valued in the assessment. Current criteria that reward high levels of organizational structure and minimal grammatical errors might be unintentionally privileging the ALC's synthetic voice. In order to retain the space for genuine human learning, ELT pedagogy may have to shift to valuing the imperfections in the Human Corpus, the narrative risk-taking, the agency's true modality, the organic and messier structures of writing, which AI is currently preventing from being expressed.

CONCLUSION

This investigation's primary objective is to define the linguistic 'space in between' through the comparison of three different text types: human-generated, AI-produced, and AI-possible student-authored texts. The outcome substantiates that the incorporation of GenAI in L2 writing does not only produce 'better' or 'worse' writing, but also fosters the emergence of a hybrid register that balances the surface characteristics of academic writing while purging the primary functions associated with human writers' inner voice.

As the comparison suggested, AI-mediated student writing displays 'synthetic authority.' While human student authors employ personal voice through the use of narrative (I want) and root modality (I can), AI-possible texts much rather employ formal voice through structural signposting (In my opinion) and use a more formal, abstract epistemic modality (It can be). Likewise, the 'Listicle Effect' within the texts significantly adds to the prevalence of rigid, structured enumerations, suggesting that AI tools shape students' capacity to organize their writing through algorithms rather than through natural, organic means of creating argument.

These discrepancies present a critical problem concerning ELT assessment. If we continue to overlook the chaotic nature of authentic student language while compensating for the precise structure of AI-mediated texts, we run the risk of legitimizing a "hollowed-out" form of literacy. Thus, this study proposes that the focus of developing writing pedagogy must shift from teaching solely structural templates to the importance of teaching critical AI literacy, as well as the 'imperfections' of voice, risk-taking, and narrative agency that comprise authentic authorship.

LIMITATIONS

While this study demonstrates both depth and breadth across multiple dimensions in this subject, there are specific limitations that need to be clarified in this study.

First, the research is predicated on what is, in essence, a small, specialized corpus of nearly 9,500 words (with the number of sampled texts being 90). The volume of this corpus is adequate for examining stance and hedging subsystems, but it sparsely populates the countless other L2 writing contexts, rendering the resulting data of limited applicability. That is, we cannot ascertain if the "listicle effect" and "epistemic shift" gaps are more than just a result of this limited sample corpus in other L2 writing contexts.

Additionally, the corpus was limited to one single specific discourse under the head "Reading Skills and Strategies." It is acknowledged that having a single topic of study is methodologically important to control for the author variable; however, it must be noted that participants` bundles did indeed exert a positive influence on the particular results. Lexical bundles, such as the one that featured in the domain of this study, "pre-reading activity in," were identifiable. The results, therefore, are not as generalizable to other fields of study and essay/syllabus topics as the one incorporated in the study.

Finally, given that the classification of the AI-Likely Student Corpus (ALC) relies on commercial AI software detection tools, we know some false positives will exist, and though we used a high threshold (80-100% likelihood) for results to be reassuring, we know the literature documents inconsistencies for reliability on these detection algorithms and a lack of a defined threshold. Because of that, we can

never know the true total of the hybrid texts of human vs AI contribution. We can only know the fragments sufficiently aligned to AI detection software.

REFERENCES

Anderson, B. R., Shah, J. H., & Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. In *C&C '24: Proceedings of the 16th Conference on Creativity and Cognition* (pp. 413–425). Association for Computing Machinery. https://doi.org/10.1145/3635636.3656204

Anthony, L. (2022). *AntConc* (Version 4.2) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antconc/

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Biber, D., & Barbieri, F. (2007). Lexical bundles in academic speech and writing. *English for Specific Purposes, 26*(3), 263–281. https://doi.org/10.1016/j.esp.2006.08.001

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397–423.

Dang, P. (2020). Vietnamese majors of English: The use of lexical bundles in argumentative essays. *INELTAL Conference Proceedings*.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines, 30*, 681–694. https://doi.org/10.1007/s11023-020-09548-1

Flowerdew, J. (2002). *Academic discourse*. Longman.

Hinkel, E. (2004). *Teaching academic ESL writing: Practical techniques in vocabulary and grammar*. Lawrence Erlbaum Associates.

Hyland, K. (1998). *Hedging in scientific research articles*. John Benjamins.

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

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

Köbis, N. and Mossink, L. (2021). Artificial intelligence versus maya angelou: experimental evidence that people cannot differentiate ai-generated from human-written poetry. Computers in Human Behavior, 114, 106553. https://doi.org/10.1016/j.chb.2020.106553

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Livytska, I. (2019). Lexical bundles in academic writing of non-native English speakers. *Journal of Language and Education, 5*(1), 69–79.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh University Press.

Rayson, P. (2016). *Log-likelihood calculator* [Computer software]. UCREL. http://ucrel.lancs.ac.uk/llwizard.html

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays—should academics worry? *Nature, 613*(7943), 196–197. https://doi.org/10.1038/d41586-023-00056-7

Uçar, S., & Zarfsaz, E. (2022). A corpus-based study: The employment of lexical bundles in Turkish students' academic writing. *Journal of Language and Linguistic Studies, 18*(1), 654–672.

Varttala, T. (2001). *Hedging in scientifically oriented discourse: Exploring variation according to discipline and intended audience*. University of Tampere.