HEDGES AND BOOSTERS IN AI AND HUMAN WRITING: A COMPARATIVE ANALYSIS

Elena Shalevska
Faculty of Education – Bitola, University "St. Kliment Ohridski" – Bitola, N. Macedonia
elena.salevska@gmail.com

Abstract: The present qualitative study aims to measure and describe the frequencies and context of use of various hedges and boosters in a self-created corpus of 100 ChatGPT 3.5-generated essays on 100 different topics. The results were then compared to 100 human-written sample essays. The analysis, primarily done via AntConc, shows high frequency use of some hedges (such as "may") and no use of boosters (such as "clearly", "definitely", "certainly" etc.) in AI works, in comparison to human texts which tend to use a wider array of hedges and boosters, more frequently. This discrepancy highlights the distinct linguistic styles between AI-generated and human-produced texts, suggesting potential areas for further development in AI language processing.
Keywords: Artificial Intelligence (AI); Academic Writing; Hedges; Boosters; AI-generated Texts

1. INTRODUCTION

Expressing one's doubt or (un)certainty is essential in academic writing. And such truth-value modifying expressions are generally done via the use of hedges and boosters. Lakoff (1973: 462) conceptualized hedges as linguistic means to either clarify or "fuzzy" the things surrounding them. According to him, hedges can be categorized into two primary groups: (1) expressing uncertainty and (2) lessening the intensity of a statement to enhance its acceptability. Hyland (1998: 350) further defined hedges (such as "possible", "might", "perhaps") as *"a resource for expressing uncertainty, skepticism, and deference in academic contexts"*. According to him, both hedges and boosters are important in academic writing, though boosters (such as "clearly", "obviously", and "of course") can interfere with readers' own interpretations of the text, while also fostering a sense of solidarity. Hedges and boosters highlight the idea that statements convey not only ideas but also the writer's stance toward those ideas and toward the readers, as noted by Halliday (1978). Although hedges are mostly used in the aforementioned ways, they can also be used to convey humility, and respect for colleagues' views (Myers, 1989; Hyland 1996). Thus, their use is multifaceted. Hedges can sometimes even be as part of a wider communicative strategy – they help hide writers/speakers' epistemic attitude (for instance, saying "Your shoes are **a little bit** dirty" instead of saying "Your shoes are dirty") (Markkanen & Schröder, 1997).

As AI becomes increasingly more popular as a tool that produces written content, understanding how the different AI models manage these linguistic features becomes essential. Thus, this study aims to explore the use of hedges and boosters in AI-generated text, comparing it with human-written academic texts to highlight potential disparities and implications for AI language processing.

2. ACADEMIC WRITING AND ARTIFICIAL INTELLIGENCE

With the introduction of advanced AI models like OpenAI's ChatGPT, the world of AI has seen remarkable advancements and thus, a growing interest in analyzing AI-generated content. It seems that ChatGPT can successfully mimic human language patterns. As a result, AI-generated content has become increasingly indistinguishable from human writing.

The model called GPT, according to Dergaa et al. (2023: 616), was first developed and introduced by OpenAI in 2018. The next year, the model was updated and the GPT 2 version became available. Although these initial versions were not as sophisticated, even then, the model was capable of advanced text generation due to its Natural language processing features (NLP). As technology developed, so did the model's capabilities and the ChatGPT 3 version, which was made publicly available in 2022, took the world by storm.

ChatGPT's ability to imitate human writings has opened up new avenues for text-production and analysis, so much so that Khalifa & Albadawy (2024) claim that AI has now significantly revolutionized academic writing and research across various domains. Essays currently remain the most studied form of AI-texts. Fitria (2023) analyzes five ChatGPT-generated essays and finds them to be perfectly acceptable, but Mahama et al. (2023) find that although ChatGPT's essays are adequate, they lack the "human touch" and human creativity that arises from external stimuli that AI cannot grasp. An interesting observation was made by Herbold et al. (2023) who explored AI's ability to write essays and compared them to human essays. Their study found that AI essays are rated higher in terms of quality than essays written by real people. The study, however, also showed that AI essays differ from human ones and are easily detectable. Shalevska & Kostadinovska-Stojchevska (2024) also analyzed AI-generated essays in terms of overall readability and her findings align with Herbold et al. (2023) i.e. she found that AI-generated essays tend to be more complex, with an average Flesch Kincaid Reading Ease score (Kincaid et al., 1945) of about 19.41.

Although some research has been done into the stylistic features of AI-generated essays and how they compare to human texts, no current research focuses on AI's use of hedges and boosters in academic texts – a gap that this particular study aims to address.

3. BROADER CONTEXT

To provide a foundation for comparison between human-written and AI-generated texts, the following studies on hedges and boosters in academic writing were considered:

Hyland's findings (1998), as some of the earliest in this field, highlight the prevalence of hedging in scientific articles, and emphasize the role hedges play negotiating claims with caution and courtesy. This is a theme that Hyland further explores in his analysis of metadiscourse, including both hedges and boosters, a key element of managing reader-writer interaction (Hyland, 2005). Similarly, Salager-Meyer (1994) identifies the communicative function of hedges in medical discourse, emphasizing their importance in conveying uncertainty and professionalism. Varttala (2001) extends this analysis to disciplinary and audience variation, suggesting that the use of hedges is context-dependent and serves a purpose broader than just semantics. Meanwhile, Crompton (1997) critically addresses the theoretical complexities surrounding hedging, arguing for a more nuanced understanding of its role in academic writing.

Some studies have even suggested that gender can influence the use of hedges and boosters. Thus, in some contexts, it is said that women might use hedges more frequently to express uncertainty or politeness, while men might use boosters to express certainty and assertiveness (Holmes, 1984; Lakoff, 1973; Mulac, Bradac, & Gibbons, 2001). However, it is important to note that these findings are not universal and can vary significantly depending on the context and cultural factors.

4. METHODOLOGY

This qualitative study relies on corpus linguistics analysis, as its primary method. Aarts and Meijs (1984) first introduce the term corpus linguistics, and Leech (1992: 116) further defines it stating that: *"computer corpus linguistics defines not just a newly emerging methodology for studying language, but a new research enterprise, and in fact a new philosophical approach to the subject"* . Whichever way one defines it, corpus linguistics serves as an excellent tool to study both original and pre-existing compiled data i.e. corpus.

A corpus can be defined as a collection of machine-readable authentic texts (including transcripts of spoken data) that is sampled to be representative of a particular natural language or language variety (McEnery et al. 2006: 5). Since no ready-made corpus of AI-generated essays could be accessed online as of February 2024, this study uses an original, self-created corpus of texts generated using ChatGPT-3.5, a readily available, free-to-use, advanced language model developed by OpenAI.

A total of 100 essays were produced for the purpose of this study, each in response to a prompt from an online educational website. The author prompted the model herself and collected the responses i.e. essays. The goal was to create a corpus with texts on a broad spectrum of topics to best portray the model's writing abilities.

Once generated, the essays were compiled into a digital corpus. Each essay was formatted to maintain consistency in presentation, such as cohesive format (.txt or plain text file format). Any identifying markers or prompt instructions were removed to ensure that the analysis focused solely on the essays generated by ChatGPT-3.5. The corpus in its entirety is currently available on Google Drive.

The primary tool for analysis was AntConc, a free software tool used by a number of researchers (Nation & Anthony, 2016; Anthony, 2017; Schmitt et al., 2021 etc.) aimed at corpus analysis. This particular software was chosen primarily for its availability and its comprehensive range of features, including word frequency lists, and concordance views. Descriptive statistics was also used to summarize the qualitative findings.

Two word lists of 5 frequently used hedges and 5 boosters were created by the author:

a) **Hedges:**
Possibly – Indicates that something is one of several possibilities, not stating it as a fact.
Might – Suggests that something could happen or be true, but without commitment.
May – Similar to "Might".
Could – Used to indicate potentiality or possibility, without asserting it will happen.
Seem/s/ed – Implies an appearance or impression of something, without strong assertion.

b) **Boosters:**
Clearly – Indicates that something is obvious or certain, without ambiguity.
Definitely –Used to state something without any doubt or reservation.

Certainly – Expresses a high level of assurance or confidence in a statement.
Claim/s/ed/ing – States or asserts that something is the case.
Obviously – Suggests that something is easily perceived or understood as true or certain.

Then, the concordance tool was used to search for each item of the lists. In addition, the “Word Frequency“ feature was used to see how often hedges and boosters appear across the corpus.

4.1. Limitations

It’s important to acknowledge this study’s limitations, to ensure it can be redone in the future, considering all the ways in which it was limited. Firstly, the study includes a limited, self-made corpus of only 200 essays with a total of 69.320 words. Secondly, the AI corpus consists of essays written exclusively by the free version of the ChatGPT model – ChatGPT 3.5. Additionally, the human-written corpus is about 35.4% larger in word quantity than the AI one. Lastly, ChatGPT’s underlying language-generation strategies to do with the use of hedges and boosters, are not considered. This would have helped determine whether the model’s training or biases encoded in the code architecture influence the use of hedges and boosters in any way.

Despite these limitations, this study manages to provide general yet important insights into the use of hedges and boosters in AI writings.

4.2. Conflicts of interest

The author hereby declares no conflict of interest including any financial (direct of indirect), personal or other relationships with other people or organizations within three years from the commencement of any work.

5. RESULTS AND DISCUSSION

The 100 AI-generated essays in the corpus yield a total number of words (also referred to as “running tokens” in AntConc) of 29.442 with 4126 types of distinct words among them. The 100 human-written essays, on the other hand, are longer and yield a total of 39.878 tokens with 3877 distinct words. However, although human-written essays in the corpus are longer they have fewer distinct words compared to AI-generated essays, which might indicate a more diverse, varied vocabulary in AI-generated texts. This conclusion can also be drawn if one considers the type-token ratio (TTR) for each corpus:

<u>AI Essays Type-Token Ratio</u> (TTR1) = (Distinct Words / Total Words) * 100
TTR1 = (4,126 / 29,442) * 100 = **14.00%**
<u>Human Essays TTR2</u> = (3,877 / 39,878) * 100 = **9.71%**
Thus, the difference between TTR1 and TTR2 (in %) can be calculated:
Percentage Difference = |(TTR1 - TTR2) / TTR1| * 100
Percentage Difference = |(14.00 - 9.71) / 14.00| * 100
Total Percentage Difference ≈ |(4.29 / 14.00)| * 100 ≈ **30.64%**

These numbers show that although the human-written essays are about 35% longer in terms of total tokens, they have a significantly lower type-token ratio compared to the AI-generated essays. This implies that human-written essays might sacrifice vocabulary diversity for extended content. In contrast, AI-generated essays appear to prioritize lexical diversity even within a limited word count.

5.1. Hedges

The first category that is considered in terms of stylistics and AI-limitations, is the use of hedges in both corpora, calculated by AntConc:

Table 1: Use of hedges in AI- and human-written texts

Hedge	Total no. of occurrences in AI essays	Total no. of occurrences in human essays
Possibly	0	0
Might	6	28
May	134	57
Could	16	57
Seem/s/ed	1+1+1 = 3	3+5+4 = 12

AI-generated essays show a notable reliance on “may“ as a hedge, with 134 total occurrences (in comparison to the minimal use of other hedges), as in “providing support and resources to struggling parents may be a more effective and compassionate approach.” (AI essay no. 14); or “achieving consensus on contentious issues may be arduous.” (2) and “Couples may be less inclined to stay in unfulfilling or unhappy relationships.” (71)
This could indicate that the AI model has a tendency to default to “may“ when expressing uncertainty or possibility, possibly reflecting its programming or the training materials used.
Human-written essays, on the other hand, display a broader and more balanced use of different hedges. Thus, “might,“ “may,“ and “could“ are used fairly equally.

5.1.1. Frequency of use

The frequency of each hedge per 1,000 tokens in both corpora is as follows:

a) AI-generated essays

- Might: ≈ 0.203 occurrences per 1,000 tokens (as in “Older couples **might** approach marriage with a greater sense of commitment”, (98));
- May: ≈ 4.54 occurrences per 1,000 tokens (as in “Studies suggest that coffee consumption **may** be linked to a reduced risk of certain neurodegenerative issues” (30));
- Could: ≈ 0.542 occurrences per 1,000 tokens (as in “The presence of smartphones in the classroom **could** lead to disruptions” (72));
- Seem/s/ed: ≈ 0.102 occurrences per 1,000 tokens (as in “Some individuals **seem** naturally inclined towards leadership” (90)).

Total frequency of hedges per 1,000 tokens = 0.203 (Might) + 4.54 (May) + 0.542 (Could) + 0.102 (Seem) = **5.387 occurrences per 1,000 tokens**.

b) Human-written essays:

- Might: ≈ 0.702 occurrences per 1,000 tokens (as in “The person who may look ugly **might** be a person with a golden heart” (human-written essay no. 57));
- May: ≈ 1.429 occurrences per 1,000 tokens (as in “Still, there are other types of games that **may** be harmful” (66));
- Could: ≈ 1.429 occurrences per 1,000 tokens (as in “The surplus of money could be used for other purposes” (16));
- Seem/s/ed: ≈ 0.301 occurrences per 1,000 tokens (“...practice well before their match, to improve where they **seem** to lack” (53)).

Total frequency of hedges per 1,000 tokens = 0.702 (Might) + 1.429 (May) + 1.429 (Could) + 0.301 (Seem) = **3.861 occurrences per 1,000 tokens**.

As it can be seen, the total frequency of hedge use in AI-generated essays (5.387 occurrences per 1,000 tokens) is higher than in human-written essays (3.861 occurrences per 1,000 tokens). In addition, AI-generated essays show a significantly higher frequency of the hedge “may,“ with approximately 4.54 occurrences per 1,000 tokens, than human-written essays with 1.429 occurrences per 1,000 tokens. Yet, AI-generated essays show lower frequencies of “might“ and “could,“ – with approximately 0.203 and 0.542 occurrences per 1,000 tokens, respectively, while human-written essays show a more comparable frequency for both hedges, with around 0.702 occurrences per 1,000 tokens. It is interesting to note that both AI-generated and human-written essays include the different inflected forms of “seem“ as a hedge, albeit with relatively low frequencies i.e. approximately 0.102 occurrences per 1,000 tokens for AI-generated essays and 0.301 occurrences per 1,000 tokens for human-written essays.

5.2. Boosters

The use of the key boosters in both corpora, as calculated by AntConc is as follows:

Table 2: Use of boosters in AI- and human-written texts

Booster	Total no. of occurrences in AI essays	Total no. of occurrences in human essays
Clearly	0	3
Definitely	0	12
Certainly	0	9
Claim/s/ing/ed	0*	3+0+0+1 = 4

Obviously	0	4

*Both “claim” and “claims” are used in the AI-texts corpus, but in their noun form, and not as a verb booster.

AI-generated essays show almost no use of boosters, with “clearly,“ “definitely,“ “certainly,“ and “obviously“ not being used at all, and “claim“ appearing only twice in the corpus. This suggests that the AI may be less adept at or less inclined to use, so called, strong language. Again, this could be due to its programming, as the model tends to avoid strong assertions without having access to clear evidence in its training data.
Human-written essays, on the other hand, use boosters more frequently. The use of adverbs like “definitely,“ “certainly,“ and “obviously“ indicates a human tendency to assert opinions or facts with confidence. Such use contributes to a stronger, more assertive tone in human writing.

5.2.1. Frequency of use

The frequency of each booster in proportion to the total number of words or tokens in the human-written essays corpus is as follows:

- Clearly: ≈ 0.075 occurrences per 1,000 tokens (as in “Thus, we can **clearly** see why some have claimed that…” (Human-written essay no. 7);
- Definitely: ≈ 0.301 occurrences per 1,000 tokens (as in “In the globalization era, language is **definitely** not enough to communicate…” (84));
- Certainly: ≈ 0.226 occurrences per 1,000 tokens (as in “If time travel was possibly, I would certainly like to go back to…” (95));
- Claim/s/ing/ed: ≈ 0.100 occurrences per 1,000 tokens (as in “Some people **claim** that universities should give the same amount of money…” (9));
- Obviously: ≈ 0.100 occurrences per 1,000 tokens (as in “In this case, the decision is **obviously** influential…” (67)).

As indicated by the data above, AI-generated essays include no boosters whatsoever. This suggests that the AI may be less inclined to use strong language or make assertive statements. In contrast, human-written essays display more frequent usage of boosters. In the human-written-essays corpus, the adverbs like “definitely,“ “certainly,“ and “obviously“ are used more frequently, to assert opinions or facts with confidence. This contributes to a stronger, more assertive tone in human writing. The use of “claim“ in all of its inflected forms also suggests a tendency to assert arguments or positions in a clear and direct manner. The total frequency of booster use in human-written essays is **0.802 occurrences per 1,000 tokens**.

5.3. Implications

AI-generated text, represented by ChatGPT-3.5 in this case, tends to be more cautious, possibly overly reliant on certain hedges like “may,“ and not inclined towards the use of other expressions of certainty or emphasis. In contrast, human writing exhibits a more balanced use of linguistic devices to convey both certainty and doubt.
These differences may contribute to the way AI-texts are perceived in terms of credibility, persuasiveness, or reliability. The findings could also reflect the impact of the training data on the linguistics output as the model tends to adhere to more conservative language patterns. Thus, we need to further highlight the importance of diverse and comprehensive training data for AI language models.

6. CONCLUSION

The study has revealed that AI, represented by ChatGPT-3.5, shows a conservative approach when generating texts, mainly depending on hedges like “may”, while completely avoiding boosters. This pattern suggests a cautious, perhaps too restrained, style in AI-generated texts. In contrast, human-written essays include a wider array of hedges and unlike AI models, they do include boosters such as “clearly” and “definitely”. The results, as such, emphasize the need for improved training of the AI models so they can have an even better grasp of academic language's intricacies. Still, further research may need to identify the hedging and boosting patterns across different AI models, genres of text, or even languages.

BIBLIOGRAPHY

Aarts, J. & Willem M. (1984). Corpus linguistics: Recent developments in the use of computer corpora in English language research. Amsterdam: Rodopi

Anthony, L. (2017). Corpus Linguistics and Vocabulary: A Commentary on Four Studies. Vocabulary Learning and Instruction, 6 (2), 79–87. DOI: 10.7820/vli.v06.2.Anthony

Crompton, P. (1997). Hedging in academic writing: Some theoretical problems. English for Specific Purposes, 16(4), 271-287.

Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From Human Writing to Artificial Intelligence Generated Text: Examining the Prospects and Potential Threats of ChatGPT in Academic Writing. Biology of Sport, 40(2), 615-622. DOI: https://doi.org/10.5114/biolsport.2023.125623

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. ELT Forum: Journal of English Language Teaching, 12(1), 44-58. DOI: https://doi.org/10.15294/elt.v12i1.64069

Halliday, M. A. K. (1978). Language as a Social Semiotic: The Sociological Interpretation of Language and Meaning. London: Edward Arnold

Herbold, S., Hautli-Janisz, A., Heuer, U. et al. (2023) A large-scale comparison of human-written versus ChatGPT-generated essays. Sci Rep 13, 18617. DOI: https://doi.org/10.1038/s41598-023-45644-9

Holmes, J. (1984). Modifying illocutionary force. Journal of Pragmatics, 8(3), 345-365.

Hyland, K. (1996). Talking to the academy: Forms of hedging in science research articles. Written Communication 13(2). pp. 251 281

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. Texts, 18(3),pp. 349-382. Retrieved from: hyland-boosting-hedging-and-the-negotiation-of-academic-knowledge-1998.pdf (wordpress.com) on 25.2.24

Hyland, K. (1998). Hedging in Scientific Research Articles. John Benjamins Publishing Company.

Hyland, K. (2005). Metadiscourse: Exploring Interaction in Writing. Continuum.

Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. Computer Methods and Programs in Biomedicine Update, 5(100145). DOI: https://doi.org/10.1016/j.cmpbup.2024.100145

Lakoff, G. (1973). Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. Journal of Philosophical Logic, 2(4), 458–508. Retrieved from: http://www.jstor.org/stable/30226076 on 4.1.24

Lakoff, R. (1973). Language and woman's place. Language in Society, 2(1), 45-80.

Leech, G. (1992). Corpora and theories of linguistic performance. In J.Svartvik (ed.). 105–122.

Mahama, I., Baidoo-Anu, D.,et al. (2023). ChatGPT in academic writing: A threat to human creativity and academic integrity?1 An exploratory study. Indonesian Journal of Innovation and Applied Sciences (IJIAS), 3(3), 228-239. DOI: https://doi.org/10.47540/ijias.v3i3.10053

Markkanen, R., & Schröder, H. (1997). Hedging: A challenge for pragmatics and discourse analysis. In Markkanen & H. Schröder (eds.), Hedgingand discourse: Approaches to the analysis of a Pragmatic phenomenon in academic texts (pp. 3-18). Berlin: Walter de Gruyter.

Mulac, A., Bradac, J. J., & Gibbons, P. (2001). Empirical support for the gender-as-culture hypothesis: An intercultural analysis of male/female language differences. Human Communication Research, 27(1), 121-152.

Myers, G. (1989). The pragmatics of politeness in scientific articles. Applied Linguistics 10, pp. 1-35.

Nation, P. & Anthony, L. (2016). Measuring vocabulary size. In E. Hinkel (Ed.), Handbook of Research in Second Language Teaching and Learning, Volume III (pp. 355-368). New York: Routledge.

Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. English for Specific Purposes, 13(2), 149-170.

Schmitt, N., Dunn, K., O'Sullivan, B., Anthony, L. , and Kremmel, B. (2021). Introducing knowledge-based vocabulary lists (KVL). TESOL Journal, e622. DOI: https://doi.org/10.1002/tesj.622.

Shalevska, E. & Kostadinovska-Stojchevska, B. (2024). ChatGPT Generated Essays: An Analysis. Annual Conference of PhD Students, University "St. Kliment Ohridski" - Bitola. Retrieved from: https://uklo.edu.mk/doktorski-proekti-2024-zimska-sesija/ on 12.5.24

Varttala, T. (2001). Hedging in Scientifically Oriented Discourse: Exploring Variation According to Discipline and Intended Audience. University of Tampere. Retrieved from: https://trepo.tuni.fi/bitstream/handle/10024/67148/951-44-5195-%093.pdf?sequence=1 on 26.2.24