The Style of Machines: A Stylometric Study of LLM Generation and Translation

Natalia Resende
Trinity College Dublin
ADAPT Centre
resenden@tcd.ie

Sheila Castilho
Dublin City University
ADAPT Centre
sheila.castilho@dcu.ie

Abstract

This paper investigates whether large language models (LLMs) develop a recognisable stylistic signature and whether such a signature persists across a text generation task and a translation task in different languages. Drawing on stylometric approaches traditionally used in authorship and translator attribution studies, we analyse outputs from ChatGPT-4o in English, Brazilian Portuguese, and Spanish, across original literary composition and translation conditions. The analysis focuses on five feature categories: lexical density, lexical richness, vocabulary sophistication, sentence length, and punctuation style. Our results show a LLM cross-linguistic profile characterised by higher lexical richness and density, rarer vocabulary, and a preference for the em-dash punctuation mark. This pattern appears in both generation and translation, though it is stronger in original text production.

1 Introduction

The notion that individual authors leave a distinctive imprint on their writing is foundational to stylometry and computational literary studies. It has long been argued that each author possesses a set of stylistic idiosyncrasies (i.e., preferences in vocabulary, sentence construction, rhythm, and punctuation) that function as a kind of linguistic fingerprint, consistent enough across texts to be identified and distinguished from those of other authors (Holmes, 1994; Holmes, 1998).

In the translation domain where the translators, who have long been rendered invisible by prevailing norms of equivalence and transparency (Venuti, 1995), have gained more recognition as secondary authors, whose own stylistic features shape the target text (Saldanha, 2014). Baker's (Baker, 2000) proposal for a corpus-based methodology to investigate translator style established that translators exhibit consistent, recoverable patterns of lexical and syntactic choice. These patterns persist across different source texts and source languages, and cannot be attributed entirely to the influence of the original author or the structural constraints of the target language. Corpus stylistic work has confirmed (Mauranen and Kujamäki, 2004; Volansky et al., 2015; Baroni and Bernardini, 2006) and refined these findings, identifying translator fingerprints across a range of linguistic features including reporting verb constructions, the use of emphatic italics, lexical density, and type-token ratio (Saldanha, 2011; Munday, 2008).

In this context, the emergence of large language models (LLMs) as versatile text generators and translators raises a question that the field has only recently begun to address: do these systems develop a writing style of their own?

Early evidence suggests that LLMs might have a "style". Research has shown that ChatGPT tend to systematically overuses lexical items. Juzek and Ward (2025) identified 21 focal words whose usage had spiked anomalously in scientific abstracts following the widespread adoption of ChatGPT, most notoriously the verb *delve* at rates far exceeding their frequency in human writing. They confirmed that these words are disproportionately represented in GPT-generated text relative to human-authored baselines. Kobak and colleagues (2025) found an abrupt increase in a cluster of LLM-

preferred style words when analysing over fourteen million PubMed abstracts between 2010 and 2024. This suggested that at least 10% of 2024 abstracts was processed by an LLM. These findings point not only to occasional word choice differences but to something that could be more systematic: a recognisable and persistent lexical signature that distinguishes LLM output from human texts.

Whether this signature found in LLM-generated texts extends beyond lexical choice to the broader stylistic dimensions studied in authorship and translator attribution research (such as sentence length, lexical richness, lexical density, punctuation preferences, and vocabulary sophistication) remains an open question. Unlike human authors or translators, LLMs are not individuals shaped by biography, mother tongue, or conscious aesthetic choices. They are trained on vast corpora of human-produced texts and refined through reinforcement learning from human feedback; their outputs are thus the product of statistical patterns learned across a broad range of domains, registers, and genres (OpenAI, 2023). Whether this training process produces something that can called a *style* is an open and consequential question.

Building on the established finding that both human authors and human translators leave distinctive stylistic traces in their texts, a key question is whether a comparable stylistic profile emerges for LLMs, and whether such a profile persists across both generative and translation conditions — that is, whether the fingerprint of an LLM is visible not only when the system is composing original text but also when it is operating as a translator. Furthermore, drawing on Baker's (1996) proposal that translated texts tend to exhibit universal features regardless of the source language, we ask whether any such AI stylistic traces persist across the different languages in which the text is generated or translated. Therefore, this paper investigates the following questions:

1. Does ChatGPT-4o exhibit outputs that are stylistically distinguishable from human-authored texts?

2. If such a stylistic tendency exists, is it consistent across generation, translation and language conditions in the texts examined?

This paper is organised as follows. The Methodology (section 2) outlines the corpus design (section 2.1) and data collection procedures (section 2.3). The results are then presented (section 3) followed a discussion of the findings in relation to stylistics and their implications for translation studies (section 4). Finally, we present our concluding remarks and directions for future work (section 5).

Language	Author	Translator	# words
EN	Human	—	1831
		Human	1235
		AI	1164
	AI	—	1467
PT-Br	Human	—	1257
		Human	2035
		AI	1754
	AI	—	2006
ES	Human	—	1036
		Human	1898
		AI	1357
	AI	—	1450

Table 1: Corpus structure: twelve text categories across language, authorship, and translation conditions.

2 Methodology

2.1 Corpus Design

To investigate whether ChatGPT exhibits a distinctive writing style and whether that style persists across authorship and translation roles, we constructed a small but balanced parallel corpus comprising twelve text categories across three languages: English (EN), Brazilian-Portuguese (PT-Br), and Spanish (ES). The corpus covers two conditions, original authorship and translation, and two agent types: human and AI (ChatGPT-4o).

Three human authors, one writing in EN, one in PT-Br, and one in ES, produced original short stories. The same AI system (ChatGPT-4o) also produced one original story in each language. Each human-authored story then served as a source text for two translations: one produced by a human translator and one by ChatGPT-4o. This yields twelve text categories, as summarised in Table 1.

The source text for translations (both human translation and AI translation) into PT-Br and ES was the short story originally produced in EN by the human author. Conversely, the source text for the human and AI translations into EN was the short story written in PT-Br by a human author. This asymmetry reflects a methodological choice grounded in the global dynamics of literary trans-

lation. EN is consistently ranked as the world's dominant source language for translation: according to the UNESCO Index Translationum dataset (UNESCO,), it has historically occupied the first position among source languages, and (CSA Research, 2020) found that English is involved in 19 of the top 20 most translated language pairs worldwide. This dominance is even more pronounced in literary translation, where the flow of texts tends to run from English into other languages far more often than the reverse (Venuti, 1995; Heilbron, 1999). The PT-Br-authored story was chosen as the source for the EN translation because it was the first human-written text produced in a language other than EN as part of the data collection process, making it the natural starting point for exploring translation into the globally dominant language.

2.2 Rationale

Before presenting the results, it is worth making the comparative logic of the study explicit. The two conditions, translation and original authorship, are analysed here as two sources of evidence. In the translation condition, the human and AI texts share the same source text, meaning that source language structure and author style are held constant across both agents. Any systematic difference observed between human and AI translations across three language pairs cannot be attributed to source text properties, and we interpret this as suggestive evidence of a cross-linguistic AI stylistic tendency. The authorship condition is necessarily less controlled: with one author per language, individual stylistic variation cannot be fully disentangled from language-level differences. However, the authorship findings are not used here as independent evidence for an AI signature; rather, if the same patterns detected in the more controlled translation comparison also appear in the authorship condition generated by chatGPT-4o, we interpret this convergence as further support for the claim. The translation condition provides the cleaner test; the authorship condition shows whether the tendency generalises beyond the translation task.

2.3 Data Collection Procedures

Authorship condition. Both human authors and ChatGPT-4o received identical written instructions for text production. The only exception was the prohibition on the use of technological assistance, which applied exclusively to human participants and was not included in the AI prompt.

- a third-person narrator,
- a target audience of adults (18+), a target length of 1,000–1,200 words,
- a fixed narrative brief (a character unearths a mysterious buried object that triggers reflections on family history),
- and an explicit prohibition on the use of any technological assistance including AI tools or writing enhancement software (only humans received this instruction).

Human authors were recruited on the basis of their prior experience in creative writing and their language background (one L1 EN speaker, one L1 PT-Br speaker, and one L1 ES speaker).

Translation condition. Human translators received written guidelines instructing them to translate each source story without recourse to any machine translation or AI-assisted tool (including Google Translate, DeepL, or ChatGPT), while allowing the use of print or online dictionaries. ChatGPT-4o received a minimal prompt: *"Translate this short story into [target language]"*. Human translators were additionally instructed to preserve the names of characters as given in the source text. This constraint was introduced to ensure comparability with ChatGPT-4o's behaviour, which, based on preliminary tests, tends not to alter character names under minimal prompting conditions. No further stylistic constraints were imposed on the AI or on the human translators in the translation condition, in order to reflect the zero-shot setting that characterises typical LLM-based translation practice.

2.4 Feature Extraction

Five categories of stylistic features were extracted computationally from each text:

1. **Lexical density.** Computed as the ratio of content words (nouns, verbs, adjectives, and adverbs) to the total number of tokens in each text. Higher values indicate a greater proportion of information-carrying words and hence denser, more compact phrasing.
2. **Lexical richness (MATTR).** Measured using the Moving-Average Type-Token Ratio

(Covington and McFall, 2010), which computes the mean TTR over a sliding window (50 words in our experiment) of fixed length across the text. Unlike simple TTR, MATTR is robust to differences in text length, making it suitable for comparing documents of varying sizes.

3. **Vocabulary sophistication.** Operationalised as the mean Zipf frequency (van Heuven et al., 2014) of the content words (nouns, verbs, adjectives, and adverbs) in each text, estimated using the `wordfreq` Python library (v3.1.1) (Speer, 2022) for EN, PT-Br, and ES. Zipf frequency is defined as $\log_{10}$(frequency per billion tokens); lower values therefore correspond to rarer, more sophisticated vocabulary.

4. **Mean sentence length.** Calculated as the average number of tokens per sentence across each document.

5. **Punctuation style.** Computed as per-token ratios for a set of typographically salient marks including em-dashes (–) and excluding the hyphens, single quotation marks (excluding apostrophes), and ellipses (. . .). These marks were selected on the basis of their known sensitivity to authorial and translator style (Baker, 2000).

3 Results

We present results for the five feature categories introduced in Section 2.4, organised around the two research questions: (i) whether GPT-4o output differ from human-authored texts, and (ii) whether this stylistic signature (if any) persists across both the original writing, translation and language.

3.1 Lexical Features

Lexical richness (MTTR). Figure 1 shows the results for lexical richness. We note that, ChatGPT originals consistently display higher MTTR values than their human counterparts across all three languages. The gap is most pronounced in EN, where the GPT original (MTTR = 0.808) exceeds the human original (0.742) and both translation texts. In PT-Br and ES the difference is smaller but in the same direction ($\Delta = +3\%$ and $+1\%$, respectively). In the translation condition, GPT translations are marginally richer than human translations in all

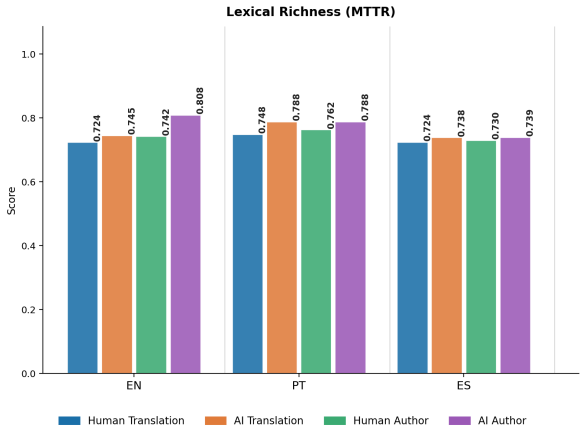

Figure 1: Lexical richness (MTTR) across all conditions and languages.

languages (e.g. EN: 0.745 vs. 0.724; PT: 0.788 vs. 0.748), suggesting that the lexical-richness signature of the system is present, albeit attenuated, even when GPT operates as a translator rather than an author.

Lexical density. Figure 2 shows the results for lexical density. We can see that, GPT originals are denser than human originals in all three languages (EN: 0.549 vs. 0.474; PT: 0.540 vs. 0.485; ES: 0.495 vs. 0.466), indicating a higher proportion of content words relative to total tokens. In the translation condition, the human-GPT difference is smaller and less consistent, though GPT translations show slightly higher density in all languages.

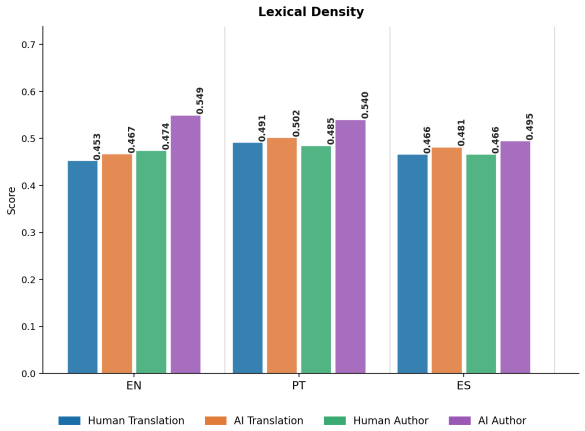

Figure 2: Lexical density across all conditions and languages.

Vocabulary sophistication. In figure 3, we note that the lower Zipf frequency values indicate rarer, more sophisticated vocabulary. GPT originals use markedly rarer words than human originals in EN

(mean Zipf: 3.20 vs. 3.56) and PT-Br (3.51 vs. 3.70), while the difference in ES is negligible (5.65 vs. 5.78). In the translation condition the pattern reverses slightly: human translations tend to use words of similar or slightly lower frequency than GPT translations, suggesting that the lexical sophistication gap is a feature of GPT's generative mode rather than its translation mode.

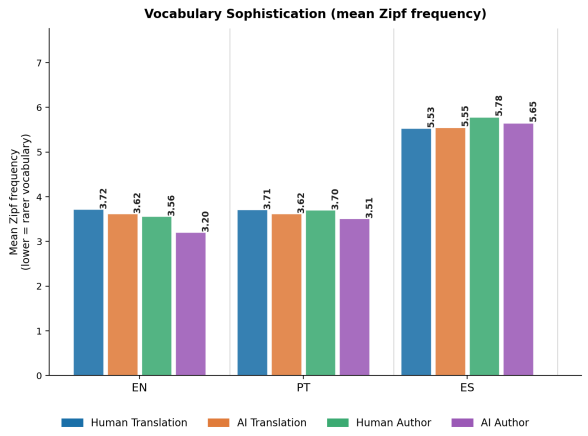

Figure 3: Vocabulary sophistication (mean Zipf frequency) across all conditions and languages. Lower values indicate rarer, more sophisticated vocabulary.

3.2 Sentence Length

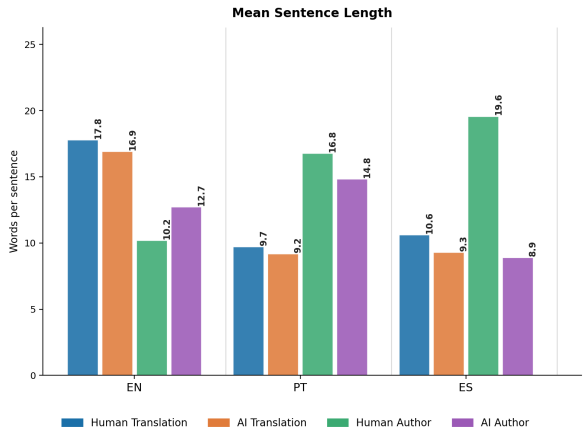

Figure 4: Mean sentence length (words per sentence) across all conditions and languages.

Mean sentence length (Figure 4) shows the largest absolute differences in the dataset. In the original writing condition, GPT originals are systematically shorter-sentenced than human originals in two of three languages: ES and PT-Br. The ES contrast is particularly striking: the human original has a mean sentence length nearly 2.2 times that of the GPT original (Human: 19.6 vs. GPT: 8.9 words). In EN, however, the GPT original is slightly longer-sentenced than the human original (GPT: 12.7 vs.; Human: 10.2 words). In the translation condition, the pattern diverges markedly between languages. EN human translations are longer-sentenced than GPT translations (Human Tr.: 17.8; GPT Tr.: 16.9 words). A possible explanation is that the PT-Br source text already featured longer sentences on average (16.8 words), which may have influenced both translations independently of any agent-specific tendency. In PT-Br and ES, by contrast, both human and GPT translations produce shorter sentences than the human originals written in those languages, which may also reflect the influence of the shorter-sentenced EN source text on both translations rather than any intrinsic property of the target language or the translating agent.

3.3 Punctuation Style

Figure 5 shows the per-1,000-token rates for em-dashes, commas, and periods across all conditions.

Em-dash. The em-dash is the single most discriminating feature in the dataset. Human translators and human authors use it rarely or not at all: every human text scores 0.0 em-dashes per 1,000 tokens in the translation condition across all three languages, and the human originals score 3.2 (EN), 0.0 (PT-Br), and 0.0 (ES). GPT, by contrast, uses em-dashes extensively: GPT translations score 10.1 (EN), 27.3 (PT), and 25.3 (ES) per 1,000 tokens, and GPT originals score 12.5 (EN), 9.5 (PT-Br), and 2.9 (ES). Within this corpus, this constitutes a near-categorical stylistic distinction between human and GPT-4o output across both conditions and all three languages. We acknowledge, however, that em-dash frequency varies considerably among human authors and that GPT-4o's tendency to favour it reflects a default behaviour under zero-shot prompting conditions, one that could be suppressed through explicit stylistic instructions. Nevertheless, the consistency of this pattern across three languages and both authorship and translation conditions, without any interference in the prompting, suggests that em-dash overuse is a feature of GPT-4o's default stylistic output, even if it does not constitute a universal or immutable distinction between human and AI writing.

Comma and period. Comma usage is high in the EN human translation (94.3/1k) relative to both

the GPT EN translation (70.1/1k) and the EN human original (44.6/1k). GPT EN translation uses commas at an intermediate rate. In PT-Br and ES, the pattern is reversed: human translations use fewer commas than GPT translations, and the originals use more commas than the translations. A possible explanation is that the human translation into EN was influenced by the amount of comma introduced in the source text in PT-Br but ChatGPT not so much. Similarly, the low frequency of comma introduced in the EN source influenced the human translation into ES and PT-Br to be reduced compared to their original counterparts in ES and PT-Br in which the authors tend to introduce much more comma. Period frequency broadly mirrors the inverse of comma usage, with GPT translations in PT-Br and ES using more periods than their human counterparts (PT: 108.2 vs. 86.7/1k; ES: 105.1 vs. 95.8/1k), consistent with GPT's shorter mean sentence lengths in those languages.

3.4 Summary: GPT vs. Human Relative Differences

Figure 6 presents a heatmap of the percentage difference between GPT and human values for each feature across the six conditions (three languages $\times$ two tasks). The pattern that emerges is consistent: GPT shows a robust positive divergence from human output on em-dash usage in the translation task ($+\infty$ % over a human baseline of zero in PT and ES; $+24\%$ in EN original), and positive divergence on MTTR and lexical density across both tasks. Vocabulary sophistication divergence is most pronounced in the original writing task (EN: -10%, indicating rarer words in GPT), while sentence length divergence is largest in the ES original condition (-55%, indicating shorter GPT sentences). Crucially, the direction of divergence on em-dash, lexical richness, and density is consistent across both tasks, suggesting that GPT stylistic signature is present in both the authorship and translation conditions at the lexical level and punctuation level.

4 Discussion

4.1 A Recognisable GPT Stylistic Signature

The results suggest that, in the texts examined, GPT-4o exhibits a stylistic tendency that sets it apart from the human authors and translators across linguistic dimensions. This signature is not reducible to a single feature but is instead distributed across lexical and punctuation domains. Three components are particularly robust.

First, and most strikingly, the em-dash functions as a near-categorical marker of GPT authorship. Its complete absence from all human translation texts and from human original writing in PT-Br and ES, contrasted with its consistent and substantial use by GPT in both tasks and all three languages, suggests that it is not a translational borrowing from any particular source language or a genre convention but a idiosyncratic preference of the default model under zero shot prompting. This finding extends earlier work on GPT lexical preferences (Juzek and Ward, 2025; Kobak et al., 2025) from the vocabulary domain to punctuation: just as GPT overuses certain words, it overuses certain typographic conventions.

Second, GPT originals are consistently richer in lexical terms than human originals. Higher MTTR and lexical density values, alongside rarer average word frequencies, paint a picture of GPT as an over-lexicalised author that deploys a broader vocabulary and a higher proportion of content words than human literary authors writing under equivalent conditions. This aligns with the observation that LLM outputs tend toward a particular kind of formal, informationally dense register (OpenAI, 2023), though the mechanism in a literary prompting context deserves further investigation.

Third, the GPT stylistic lexical tendencies are present in both the original writing and translation conditions, though it is more pronounced in original authorship. This is consistent with the hypothesis that when GPT operates as a translator, the source text exerts a partial normalising constraint on its output, pulling certain surface features closer to the source author's style and this is visible when we analyse the influence of the sentence length of the source texts influence the length of the sentences in the target text. Even so, the em-dash signature is not suppressed by the translation task, suggesting that some stylistic preferences are robust enough to persist despite source text influence.

4.2 Implications for Translation Studies

From a translation studies perspective, these findings reinforce the view that the translator, whether human or artificial, is not a stylistically neutral medium but an agent whose own preferences leave traceable marks on the target text (Baker, 2000;

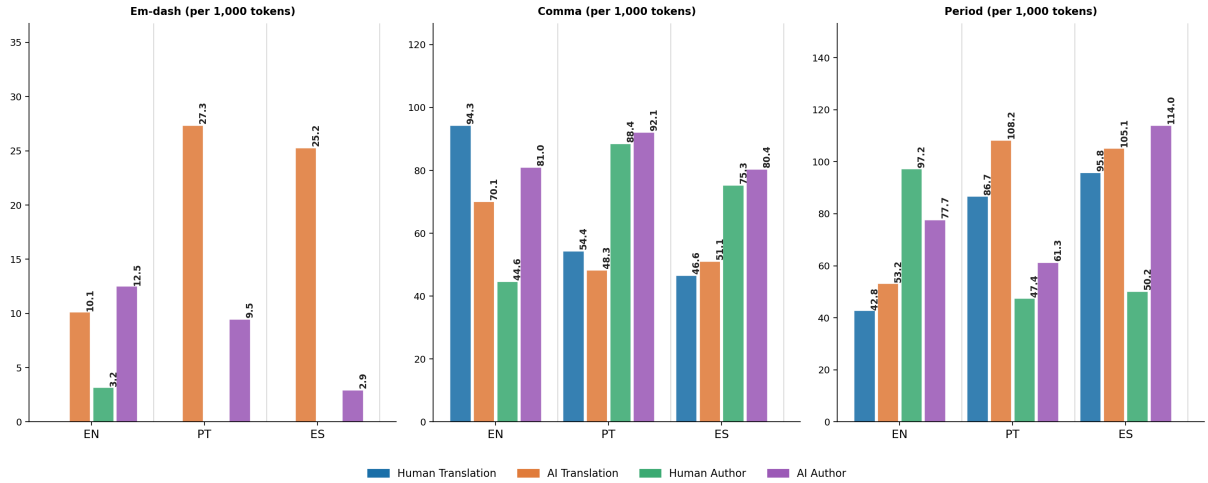

Figure 5: Punctuation signatures: em-dash, comma, and period rates (per 1,000 tokens) across all conditions and languages. Note the near-zero em-dash usage in all human-authored texts versus the consistently elevated rates in GPT outputs.

Baker, 1996; Baker, 1993). The finding that GPT's punctuation preferences persist even in the translation condition implies that researchers and practitioners evaluating MT output should be alert not only to semantic accuracy and fluency but to the stylistic imprint that the system imposes on target texts.

The sentence length results also raise questions about source language interference in AI-mediated translation. The pattern observed in the EN translations, where both human and GPT translations exhibit longer sentences than their respective target language norms, is consistent with the classical phenomenon of interference (Toury, 1995), whereby source text syntactic patterns leave traceable marks on the target text. GPT-4o proves no less susceptible to this effect than human translators, suggesting that source text influence operates similarly across both human and AI-mediated translation. The PT-Br and ES translation conditions point in the same direction: in both languages, translations produce shorter sentences than the human and AI originals written in those languages, possibly as a result of interference from the shorter-sentenced EN source text. This cross-linguistic consistency suggests that source language interference in sentence length is not an isolated phenomenon but a broader tendency in translation, regardless of the target language or the agent involved.

4.3 Limitations

Several limitations should be noted. The corpus is small, twelve texts, each of approximately 1,000-2,000 words, and results cannot be taken as statistically representative. The human participants, though carefully selected, each represent a single authorial voice, meaning that observed human-GPT differences could partly reflect individual idiosyncrasy rather than a categorical human-AI contrast. The analysis is based on surface-level stylometric features and does not address deeper syntactic or narrative dimensions of style. Finally, the study covers a single AI system (ChatGPT-4o) at a single point in time; the stylistic properties of LLMs are known to shift across model versions and fine-tuning iterations. Future work should address these limitations through larger, more diverse corpora, multiple AI systems, and the inclusion of inferential statistics. Additionally, all AI outputs were generated using a minimal zero-shot prompt, which may not reflect the stylistic behaviour of ChatGPT-4o under the more directive prompting strategies commonly used in practice. Users can readily instruct LLMs to suppress specific stylistic features, such as em-dash usage, or to preserve source text punctuation conventions; however, whether more explicit,i.e., detailed prompting would also attenuate the lexical patterns observed here remains an open question for future research. Finally, no replication runs were conducted for the AI-generated texts; it cannot be excluded that different runs would yield different stylistic profiles. We also acknowledge that the absence of qualitative analysis reflects a deliberate methodological choice consistent with the corpus stylistic tradition within which this study is situated (Stamatatos, 2009; Burrows, 2002; Baker, 1996), in

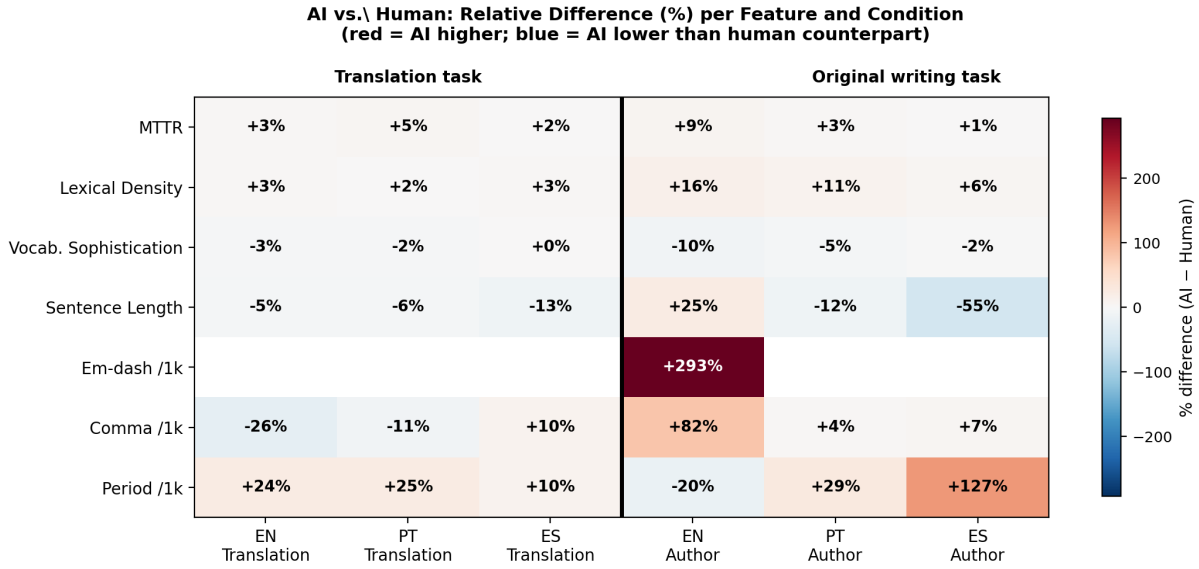

Figure 6: Percentage difference between GPT and human values for each stylometric feature, across translation and original writing conditions in three languages. The vertical black line separates the two task types. Red indicates GPT-higher values; blue indicates GPT-lower values.

which quantitative, corpus-based methods are considered the most appropriate means of detecting stylistic patterns that operate below the threshold of conscious perception and would not be recoverable through qualitative close reading alone.

5 Conclusion

This paper has investigated whether ChatGPT-4o presents a recognisable stylistic signature when producing literary text in original and translation conditions, across EN, PT-Br, and ES. The analysis of five stylometric feature categories (lexical density, lexical richness, vocabulary sophistication, sentence length, and punctuation style) reveals tendencies in the texts examined whereby GPT outputs show higher lexical richness, greater lexical density, rarer vocabulary, and a marked preference for the em-dash. This profile is present in both original authorship and translation conditions, though it varies across languages. In addition, results revel that these stylistic tendencies tend to be more pronounced when GPT functions as an author than when it operates as a translator, suggesting that source text influence partially constrains the model's stylistic autonomy.

In computational stylometry, they provide preliminary evidence that ChatGPT-4o may develop stylistic tendencies detectable by the same methods used to identify human authorial style, warranting investigation at larger scale. In translation studies, they suggest that the AI translator may function as an active stylistic agent whose tendencies are not fully neutralised by the translation task. And in AI language research, they extend the emerging literature on LLMs lexical markers into the domain of punctuation and literary genre, suggesting that the stylistic tendencies of ChatGP'T-4o may be richer and more systematic than isolated lexical observations have so far revealed.

Future work will expand the corpus to include multiple human authors and translators per language, additional AI systems, and more fine-grained syntactic features, while also developing inferential methods appropriate for the small-corpus regime characteristic of literary stylometric research. Furthermore, replication across different story selections and literary traditions will be necessary to assess the generalisability of the findings reported here.

6 Acknowledgements

We would like to express our sincere gratitude to the authors Adriana Martins, Lisa Ronan, Ana Olivares and to the translators Thais Gianmarco, Catarina Freitas, and the students of the Masters in Translation Technologies at DCU who contributed to the translation from English into Spanish. All participants generously volunteered their time and creativity in writing and translating the short stories that form the basis of this study. Their collaboration and enthusiasm made this research possible. The authors benefit from being member of the

ADAPT SFI Research Centre at Dublin City University, funded by the Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2.

Data and Materials Availability

The files and materials used in the experiments are available in the following repository.

References

[Baker1993] Baker, Mona. 1993. Corpus linguistics and translation studies: Implications and applications. In Baker, Mona, Gill Francis, and Elena Tognini-Bonelli, editors, *Text and Technology: In Honour of John Sinclair*, pages 233–250. John Benjamins, Amsterdam and Philadelphia.

[Baker1996] Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In *Terminology, LSP and Translation*, page 175. John Benjamins.

[Baker2000] Baker, Mona. 2000. Towards a methodology for investigating the style of a literary translator. *Target: International Journal of Translation Studies*, 12(2):241–266.

[Baroni and Bernardini2006] Baroni, Marco and Silvia Bernardini. 2006. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, 21(3):259–274.

[Burrows2002] Burrows, John F. 2002. 'delta': a measure of stylistic difference and a guide to likely authorship. *Lit. Linguistic Comput.*, 17:267–287.

[Covington and McFall2010] Covington, Michael A. and Joe D. McFall. 2010. Cutting the gordian knot: The moving-average type–token ratio (mattr). *Journal of Quantitative Linguistics*, 17(2):94–100.

[CSA Research2020] CSA Research. 2020. The state of the linguist supply chain: Translators and interpreters in 2020. Technical report, CSA Research.

[Heilbron1999] Heilbron, Johan. 1999. Towards a sociology of translation: Book translations as a cultural world-system. *European Journal of Social Theory*, 2(4):429–444.

[Holmes1994] Holmes, David I. 1994. Authorship attribution. *Computers and the Humanities*, 28(2):87–106.

[Holmes1998] Holmes, David I. 1998. The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing*, 13(3):111–117.

[Juzek and Ward2025] Juzek, Tom S and Zina B. Ward. 2025. Why does ChatGPT "delve" so much? exploring the sources of lexical overrepresentation in large language models. In Rambow, Owen, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert, editors, *Proceedings of the 31st International Conference on Computational Linguistics*, pages 6397–6411, Abu Dhabi, UAE, January. Association for Computational Linguistics.

[Kobak et al.2025] Kobak, Dmitry, Rita González-Márquez, Emőke Ágnes Horvát, and Jan Lause. 2025. Delving into llm-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11(27):eadt3813.

[Mauranen and Kujamäki2004] Mauranen, Anna and Pekka Kujamäki. 2004. *Translation universals: Do they exist?*, volume 48. John Benjamins Publishing.

[Munday2008] Munday, Jeremy. 2008. *Style and Ideology in Translation: Latin American Writing in English*. Routledge Studies in Linguistics. Routledge, New York and London.

[OpenAI2023] OpenAI, R. 2023. Gpt-4 technical report. arxiv 2303.08774. *View in Article*, 2(5):1.

[Saldanha2011] Saldanha, Gabriela. 2011. Translator style: Methodological considerations. *The Translator*, 17(1):25–50.

[Saldanha2014] Saldanha, Gabriela, 2014. *Style in, and of, Translation*, chapter 7, pages 95–106. John Wiley Sons, Ltd.

[Speer2022] Speer, Robyn. 2022. rspeer/wordfreq: v3.0, September.

[Stamatatos2009] Stamatatos, Efstathios. 2009. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, 60(3):538–556.

[Toury1995] Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. John Benjamins, Amsterdam and Philadelphia.

[UNESCO] UNESCO. Index translationum: World bibliography of translation. `https://data.unesco.org/explore/dataset/tran001/information/`. Accessed: 05/2026.

[van Heuven et al.2014] van Heuven, Walter J. B., Pawel Mandera, Emmanuel Keuleers, and Marc Brysbaert. 2014. SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67(6):1176–1190.

[Venuti1995] Venuti, Lawrence. 1995. *The Translator's Invisibility: A History of Translation*. Routledge, London and New York.

[Volansky et al.2015] Volansky, Vered, Noam Ordan, and Shuly Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities*, 30(1):98–118.