Toying with Style: Can GenAI Mimic a Literary Translator's Voice?

Beniamin Sopot
SALIS/CTTS
Dublin City University
Ireland
beniamin.sopot2@mail.dcu.ie

Dorothy Kenny
SALIS/CTTS
Dublin City University
Ireland
dorothy.kenny@dcu.ie

Abstract

In this paper we explore the potential of generative AI to produce stylistically-aware literary translations. We use three prompting strategies (zero-shot, one-shot and few-shot) to attempt to elicit translations in the style of Polish literary translator Jan Rybicki from ChatGPT 5.4 and Claude Sonnet 4.6. We test these strategies on a 1,000-word extract from a novel for which we have a previously unpublished translation by Rybicki. Stylometric analyses (based on average sentence length, sentence length distribution, and most frequent words) shows that none of the prompts we tested, regardless of the strategy, brought either LLM significantly closer to Rybicki's style.

1 Introduction

The use of artificial intelligence (AI) systems in the process of literary translation is now a commercial reality, albeit a controversial one (Kenny, 2025). While academic research in the area has tended to focus on technical challenges related to training and evaluating the neural machine translation (NMT) models and—since 2022—large language models (LLMs) used in AI-assisted literary translation, a growing body of work is also concerned with the impact of automation on the style of translated literary texts. Such work includes studies that explore how AI systems can emulate the style of individual human translators. The current research contributes to this body of work by investigating the potential of two LLMs to produce translations in the style of well-known Polish literary translator, Jan Rybicki.

This paper serves as a pilot study for a broader research project that seeks to explore the potential of personalized AI-assisted translation to augment the human translator's performance and experience of translation. One of the research questions asks whether personalized AI translation models can capture a human translator's literary style. As analysts, we learn about Rybicki's style from a large corpus of his previous translations, which he has shared with us for the purposes of this project. This corpus can also be used to supply prompting data and human reference translations for our experiments with stylistically-aware AI-assisted translation, as illustrated in this paper. Ultimately this research hopes to contribute to the currently intense professional and media debate about the use of AI in literary translation and other creative practices. It is conducted in a context in which others active in the creative industries routinely move to protect their (literal) voice and likeness (see, for example, Equity 2026), but where conversational LLMs already claim to be able to produce translations in the (metaphorical) voice of individual human translators (Kenny, in preparation). We proceed from the vantage point of Computer-Assisted Literary Translation or 'CALT', which is "founded on the idea that the possibilities held out by translation technologies…should be explored by both scholars and practitioners of the art" (Rothwell et al., 2024: 6; emphasis in the original); and, in keeping with the tenets of CALT, we are open to the idea that such technologies have the potential to augment rather than displace human translation. Whether LLMs really can emulate an individual translator's style, however, and whether such LLMs might be useful to the literary translator in question – for example by producing a draft translation that consistently uses his/her style – can only be ascertained through targeted research of the kind we conduct here.

2 Previous Research

2.1 NMT, literary translation and translator style

Previous work in AI-assisted literary translation investigates how 'generic' (often free online) machine translation (MT) can be applied to literary text, and on how bespoke 'literary-adapted' MT systems can be created by training or fine-tuning engines using (large quantities of) literary texts. Kenny (2025) provides an overview of the latter research. Literary-adapted NMT systems that aim to emulate the style of an individual human translator are rare, but not unheard of. Hansen and Esperança-Rodier (2022), for example, showed that, notwithstanding the flaws in its output, a bespoke English-French NMT system fine-tuned on translations by Nathalie Serval produced texts bearing some lexical and syntactic features that were consistent with her style. Likewise, Yirmibeşoğlu et al. (2023) fine-tuned a pre-trained generic English-Turkish NMT model using data from a corpus of literary translations by Nihal Yeginobalı (1927-2020) and concluded that their fine-tuned model captured the human translator's stylistic features "much better" than the pre-trained model. Yirmibeşoğlu et al. (ibid.) operationalized 'style' using a set of 29 textual features that are relatively easily quantified. These include word-level features (type-token ratio, number of unique words, mean word length, etc.), sentence-level features (mean sentence length, standard deviation of sentence lengths, etc.), morphological features (mean morphemes per sentence, etc.) and the frequency of certain forms or 'keywords', that is, forms found to be particularly common or uncommon in the translator's work. These features were represented as normalized frequencies in a 29-dimensional vector for each text, allowing texts translated under different conditions to be easily compared. Yirmibeşoğlu et al. found that fine tuning on Yeginobalı's works captured the translator's style "18-40% better in terms of cosine similarity" than the pre-trained model. We share Yirmibeşoğlu et al.'s (2023) interest in quantifiable style and keywords, but are committed to working with living translators who might be able to interact with any models we build or outputs we produce, just as Nathalie Serval does in Hansen and Esperança-Rodier's (2022) experiments, and consistent with Kenny's work with Hans-Christian Oeser (Kenny and Winters, 2020; Winters and Kenny, 2024).

2.2 LLMs, literary translation and style

With the launch of ChatGPT in late 2022 and the subsequent popularization of generative AI, research into AI-assisted literary translation shifted focus to LLMs, with studies usually: comparing LLM outputs to NMT and/or human translation (HT); testing LLMs' ability to translate whole paragraphs or documents as opposed to isolated sentences; and testing various LLM finetuning and prompting strategies. Thai et al. (2022) and Karpinska and Iyyer (2023), both of which had Polish as one of their target languages, are good examples of this early work. The former used LLMs to post-edit NMT output; the latter to translate literary paragraphs directly. The applicability of LLMs to multiple tasks in the translation workflow has been embraced by Wu et al. (2025), who argue that multi-agent collaboration, as exemplified by their TRANSAGENTS framework, is particularly effective for translating longer literary texts. Zhang et al. (2025) problematize findings in this and related research on the basis that the evaluation methods they use are not always adequate for literary texts. Even in a source as thoughtful as Zhang et al. (2025), however, the notion of 'style' remains vague. The term designates both a category of translation error and a sometimes nebulous property of literary texts, referred to by the authors in their experimental prompts. Zhang et al. (2025: 10970), nevertheless, isolate 'style' as one of the features of literary texts to which future automatic evaluation metrics will have to attend.

2.3 Stylometry in translation analysis

Of the hundreds of research articles that now examine the interface of literary translation and NMT and/or LLMs, many mention style in passing, but few are primarily concerned with style. Those that do focus on style tend to make use of the quantitative methods developed in stylometry, the application of which to literary translation was pioneered by Jan Rybicki (Rybicki, 2006, 2012; Rybicki and Heydel, 2013), the translator whose work we study in this paper.

Examples include Lee (2022), who conducted a stylometric study of unigrams, bigrams and tri-grams in NMT translations (by Google Translate, Bing Translation and Papago) and human

translations of modern Korean novels. But while Lee (ibid.) differentiated between NMT systems, 'human translation' remained an undifferentiated mass. Rybicki (2025), meanwhile, uses stylometric techniques to study English translations of classic French literature done by Google Translate and DeepL, on the one hand, and humans, on the other. While distributions of the most frequent words (MFW) in the texts under comparison were not found to distinguish between machine-generated and human translations, Rybicki finds that sentence length distributions can be used to discriminate between the two modes, for "more traditional" texts at least (ibid. 271). Kong and Macken (2025), for their part, initially considered 447 features in their study of three sets of translations into Chinese of *Peter Pan* (HT, NMT and LLM-based respectively). These features include lexical (e.g. TTR, lexical density, PoS tag distribution), syntactic (average sentence length, dependency structures), readability (measures designed to assess a text's difficulty and comprehensibility for the target Chinese-speaking audience; measures of concreteness), and *n*-gram features (word-level or PoS uni-, bi- and tri-grams), to which they add features considered important in the translation of creative texts (e.g. related to repetition, rhythm, and translatability, as well as a 'miscellaneous' category). They find that their approach can distinguish relatively easily between human translations and automatic translations.Yao et al. (2025), finally, use stylometric techniques to compare human and LLM-based translations of Chinese online literature, finding that "GPT-4 translations closely mirror human translations in lexical, syntactic and content features" (ibid. 1)

While the above stylometric studies are rich in ideas on how to operationalize style, and how to conduct experimental research, providing guidance, for example, on prompting, none of them is interested in investigating whether the styles of an individual, named, human translator can be replicated by either NMT or LLMs. While we look to them for procedural guidance, we break new ground by applying stylometric techniques in the study of how LLMs can be used to mimic human translator styles. It should be noted at this point, however, that individual human translator style has been the subject of much research in translation studies since at least the early 2000s. Space restrictions prevent us from elaborating here. Instead, the interested reader is referred to (Kenny and Winters, 2024).

3 Research Design

This pilot study aims at establishing appropriate prompting strategies to produce AI-assisted, style-aware AI translations. Much recent attention has been dedicated to prompt engineering in LLM-based MT, with Tang et al. (2025) arguing that models are sensitive to prompts and that providing appropriate examples and instructions is essential for obtaining quality outputs. Our analysis thus evaluates the effectiveness of zero-, one- and few-shot prompting strategies in the creation of style-aware literary translation outputs. We structure prompts following the template provided (for few-shot prompts) by Tang et al. (ibid.). Our study further attempts to *quantify* – using stylometric techniques via the stylo package (Eder et al., 2016) – the extent to which we can personalize two LLMs using selected prompts. As outlined by Herrmann et al. (2015) and illustrated above, hundreds of features are available for the quantitative characterization of texts. In this study, and drawing on the studies reviewed earlier, we rely on average sentence length, sentence length distribution, lexical diversity based on standardized type-token ratios, and most frequent words. We also provisionally determine some features of Rybicki's style based on the passages selected for use in our experiment and compare them to those of the two LLMs in use.

3.1 Selected LLMs and prompting strategies

This study compares two LLM systems – ChatGPT 5.4 and Claude Sonnet 4.6 – to assess whether either or both can be prompted to output translations in Jan Rybicki's style. ChatGPTwas selected due to its popularity and large market share, with web traffic analyses suggesting it is the most frequently visited LLM chatbot (Similarweb, 2026). Claude was chosen as it is described by Lokalise as the best for creative translation in testing rounds involving English to Polish translation (Wolff, 2026).

ChatGPT was accessed via Arena (arena.ai), a platform created to evaluate AI models (Chiang et al., 2024). Arena offers free access to the best ChatGPT model. To minimize potential bias resulting from previous interactions, a new Chrome Incognito session was opened to input a new prompt in a new Arena account. After selecting 'Direct' chat, 'gpt-5.4-mini-high' was chosen. Interactions with Claude were performed directly

on the claude.ai website using a standard free account with cross-chat memory disabled to reduce risk of bias from previous interactions. The best available model for generating translations with Claude was Sonnet 4.6 Adaptive. As this model was also the most advanced one available via Arena, we opted for direct interaction on the claude.ai website.

The prompts used in our interactions with the LLMs are listed in section 3.4, in our explanations of the various texts used in our stylometric analysis. As indicated above, the one-shot and few-shot prompts were designed based on the few-shot prompt in Tang et al. (2025).

3.2 Test data

As our source text (ST), we chose a 1,000-word passage (44 sentences) from Ceridwen Dovey's debut novel Blood Kin, which had been previously translated by Rybicki but never published. This created a unique research opportunity where there is a high certainty that the human translator's output had never been used to train any LLM systems. LLM training data is often obscured, with suspicions arising that non-open-access data has been used to train these models (Nature, 2026); therefore, any published translation may have already been used as training data, limiting the validity of evaluation of the LLM's actual capabilities (can a model translate the text well because it's good at translating or because it has access to an already existing translation of said text?). The ST was accessed through a publicly available ebook.

3.3 Bilingual examples used in LLM prompts

The bilingual examples used in our prompts were selected based on the above-mentioned corpus of Jan Rybicki's translations. Within the corpus, we identified three translations of contemporary works that used diverse narrative styles (espionage realism, metafiction and speculative fiction). Selecting contemporary translations reduced the risk of diachronic stylistic variation in Rybicki's work. The three translations were based on source texts (STs) by three different authors to minimise ST-author stylistic influence. All three ST examples were accessed using publicly available ebooks. The excerpts from each ST were selected from the first few pages of the work, in such a way as to ensure the structural integrity of the excerpts, and to allow continuous sections of around 400 words (see Table 1). The translation of the selected excerpt was then located in Rybicki's corpus.

source text	author	excerpt length
Agent Running in the Field	John le Carré	359 words
The Gum Thief	Douglas Coupland	407 words
Tanglewreck	Jeanette Winterson	454 words

Table 1. ST-data used with Rybicki's translation in prompts

3.4 Translated texts

We saved the translated texts on which our stylometric analysis is based under the following filenames (with the prompt that elicited each translation (if applicable) given in the parentheses):

- Rybicki (Jan Rybicki's human translation)
- GPT_control (ChatGPT's output prompted as follows: "Translate the below text from English to Polish")
- GPT_style.prompt (ChatGPT's output prompted as follows: "Translate the below text to Polish in Jan Rybicki's translation style")
- GPT_1shot (ChatGPT's output prompted as follows: "I will give you examples of text fragments, where the first one is in English and the second one is Jan Rybicki's translation of the first fragment to Polish. These sentences will be displayed below. … Translate the below text to Polish in Jan Rybick's translation style." with one example of a passage from Le Carré's 'Agent Running in the Field' and Rybicki's translation of that passage into Polish).
- GPT_3shot (ChatGPT's output prompted as follows: "I will give you examples of text fragments, where the first one is in English and the second one is Jan Rybicki's translation of the first fragment to Polish. These sentences will be displayed below. … Translate the below text to Polish in Jan Rybicki's translation style." with three examples of passages from: Le

Carré's 'Agent Running in the Field', Coupland's 'The Gum Thief' and Winterson's 'Tanglewreck', and Rybicki's translations of these passages into Polish).

- Claude_control (Claude's output prompted as follows: "Translate the below text from English to Polish")
- Claude_style.prompt (Claude's output prompted as follows: "Translate the below text to Polish in Jan Rybicki's translation style")
- Claude_1shot (Claude's output prompted as follows: "I will give you examples of text fragments, where the first one is in English and the second one is Jan Rybicki's translation of the first fragment to Polish. These sentences will be displayed below. … Translate the below text to Polish in Jan Rybicki's translation style." with one example of a passage from Le Carré's 'Agent Running in the Field' and Rybicki's translation of that passage into Polish).
- Claude_3shot (Claude's output prompted as follows: "I will give you examples of text fragments, where the first one is in English and the second one is Jan Rybicki's translation of the first fragment to Polish. These sentences will be displayed below. … Translate the below text to Polish in Jan Rybicki's translation style." with three examples of passages from: Le Carré's 'Agent Running in the Field', Coupland's 'The Gum Thief' and Winterson's 'Tanglewreck', and Rybicki's translations of these passages into Polish).

3.5 Stylometric analysis

Eder et al.'s (2016) R Studio package 'stylo' was used to quantitatively analyse and compare AI outputs with Rybicki's translation. A 100 MFW cluster tree was created to assess whether model outputs are stylistically similar to Rybicki's translation based on their distributions of most frequently-used words. Further inquiry into word frequency resulted in a bootstrap consensus tree (BCT) to validate the findings. To investigate lexical diversity, the type-token ratio metric was applied. Lastly, sentence length was approached from two angles: sentence length distribution and average sentence length analyses were conducted to discover Rybicki's stylistic tendencies. All processing was based on default settings in stylo.

4 Results

Both 100 MFW (Fig. 1 in the Appendices) and BCT (Fig. 2) analyses clearly confirm that Rybicki's translation is set completely apart from all LLM-based machine translations; in the 'corpus cluster analysis' his text joins the tree at a distance of over 2.0, which is the maximum value possible and shows how far the text is from the MT outputs. The Delta distance matrix is further evidence of Rybick's translation's outlier position with distances between 1.46–1.65. Within the group of MT outputs, both GPT and Claude show a symmetrical pattern: three translations from each system cluster within their own model family. The BCT across multiple word frequency ranges (100–500 MFW) further verifies the findings (Fig. 2). While the MFW100 figure presents some grouping patterns among LLM outputs, the BCT shows weak cluster stability under resampling. This suggests that examined prompts did not produce consistent stylometric profiles, although this may be amplified by output sizes below 1,000 words. The BCT further supports the argument that prompt conditioning does not reliably bring LLM outputs closer to the human translator's style. In short, based on Burrow's Delta method, none of the prompts, regardless of the strategy, brought any LLM output significantly closer to Rybicki's style.

To further inspect the word frequency, an analysis of most frequent words was run. Among the results, two words were of particular interest due to their relatively similar use among models and striking distance from Rybicki's text, namely, the words 'jego' and 'się'. Both words are pronouns, 'jego' is a third-person masculine singular pronoun, whereas 'się' is a common reflexive particle, indicative of the Polish language. These stylistic features distinguish Rybicki's texts from the AI outputs. Possessive pronouns are often naturally omitted in Polish, but AI models tend to translate them literally, suggesting English interference. Moreover, Rybicki translated chapter titles 'His portraitist', 'His chef' and 'His barber' omitting the pronouns ('Portrecista', 'Kucharz' and 'Fryzjer'), exercising his translator's agency by making an active decision – which none of the models replicated in its output. In Rybicki's translation 'jego' has a relative frequency of 0.83%, which is significantly lower than in the

LLM translations, which range from 1.53% to 2.67%, with two of the ChatGPT outputs using the word three times more frequently than Rybicki (Fig. 3).

A reverse phenomenon can be observed in Fig. 4 presenting the frequency of the word 'się'. As 'się' is indicative of the Polish language, less frequent usage may suggest less natural text. Rybicki's significantly higher usage of this word (2.97%) points to a more idiomatic use of structures such as reflexive verbs or impersonal constructions, characteristic of Polish prose. Typography and punctuation can be indicative of one's personal style. Some LLM models use certain punctuation marks so much that it becomes an indicative feature of their outputs (Przystalski et al., 2026). Certain stylistic features of LLM outputs are particularly apparent. For example, recent research notes that LLMs are characterised by an increased use of em dashes compared to traditional human writing, potentially resulting from specific training data, tokenization and reinforcement learning processes (Hawkinson, 2025). Indeed, all genAI translations in this corpus contained dashes (see Fig. 5), whereas Rybicki's text does not use them at all, favouring conjunctions or splitting sentences into shorter ones. Interestingly, ChatGPT outputs use 4 to 5 dashes within the whole text, but Claude's versions range from 6 to 16 dashes, with the few-shot prompt resulting in the lowest value of 6 (whereas ChatGPT's few-shot prompt resulted in one more dash than the rest of the outputs).

To analyse the lexical diversity of the outputs and compare them with Rybicki's translation, the TTR (Fig. 6) measure was applied. The results indicate similarity in terms of lexical variety, with all models' outputs displaying similar diversity to Rybicki's version (values between 0.627 and 0.672, with Rybicki's text having a value of 0.63). Although all texts have similar values, it's interesting to see Claude's scores are consistently higher than GPT's ones, suggesting Claude's higher vocabulary richness and potential wordiness.

On a sentence level, Rybicki's translation is distant from all AI outputs across various metrics. On average, as presented in Fig. 7, his sentence length is significantly shorter than LLM ones, with a mean sentence length of 15.38 words for his translation and between 17.87 (one-shot prompted Claude's output) and 19.68 (few-shot prompted ChatGPT's output) for MT versions. This pattern can be further observed in sentence length distribution (see Fig. 8), where the short sentence peak is the highest for Rybicki's text and a steeper decline toward longer sentences indicates a differentiating density of concise sentences across his translation and a smaller number of longer sentences overall. As for absolute values, Rybicki's translation has 26 short sentences of 10 words or less and only 8 long sentences of 25 words or more. To contrast it with the AI outputs, all models produced between 19 and 22 short sentences and between 14 and 16 long sentences. This clearly indicates Rybicki's stylistic preference for shorter sentences in comparison with the LLMs' more unified sentence length distribution. Lastly, the total number of sentences per translation yet again indicates Rybicki's stylistic distance from all AI-generated outputs. His text consists of 58 sentences, whereas the AI versions have between 47 and 50 sentences, which is consistent with Rybicki's tendency to use shorter sentences (for example, through splitting longer sentences into shorter ones) leading to the higher total number. (Recall that the ST contains 44 sentences.)

5 Discussion

Having inspected the common quantitative metrics with stylo, the results suggest that no prompting strategy resulted in stylistic convergence. Apart from the token-type ratio, where all outputs have similar lexical diversity (with Claude showing slightly higher and ChatGPT equal or slightly lower richness compared with Rybicki's translation), all analysed metrics isolate Rybicki's style from LLM outputs. As suggested through Burrow's Delta and BCT metrics, the human translator remains an outlier, with his text being structurally different from eight AI-generated machine translations. AI models, on the other hand, are more stylistically stable internally with both ChatGPT and Claude outputs creating their own clusters in terms of lexical diversity and most frequently used words. The observed stylistic differences are driven more by the system's internal behaviour than prompt variation, as the generated outputs exhibit a relatively stable style across four prompting strategies.

We investigated various prompts: two zero-shot prompts where one asked LLMs to generate an English to Polish translation and the other asked them to do so in Jan Rybicki's translation style; one one-shot prompt and one few-shot

prompt with 3 examples of Rybicki's past translations. None of the prompting strategies led to an output that would be stylometrically related to the human translation, maintaining the stylistic space between the HT and MT. These findings suggest that style-related conditioning through prompting does not result in measurable improvement in output distribution, with the prompting instructions failing to override the underlying text-generation processes of the studied models. These findings are consistent with existing research on LLM style in creative texts, such as O'Sullivan (2025), which concludes that human creative writing is distinctive from LLM outputs, which tend to cluster together tightly by model and exhibit less stylistic diversity.

To complement this quantitative analysis, the next step would be to inspect the selected translations qualitatively using non-statistical metrics like storytelling or the use of rhetorical devices, such as irony, to determine whether they are in any way replicated in the MT versions.

6 Conclusions

This pilot study aimed at investigating whether two LLMs, ChatGPT and Claude, are capable of producing stylistically-aware machine translations imitating a specific literary translator's voice – in this case that of Jan Rybicki. The idea was not to reproduce exactly the translator's stylometric profile using LLMs, but rather to use stylometry to test whether prompting LLMs in given ways gets them to produce output that gets closer to the human translator's stylometric profile. If such prompts were effective in getting the LLM in question to approximate our translator's style, then this would suggest an avenue that could be pursued in our ongoing efforts to personalise CALT systems for use by literary translators.The results clearly show that none of the tested methods resulted in an improvement of stylometric distance from Rybicki's text, consistently indicating that human translation remains distinctive from AI translation. This suggests that prompting may not be an effective strategy to improve stylistic awareness of LLM-based translation. The study contributes to ongoing research on AI-assisted literary translation and demonstrates, through the lens of stylometric measures, the existing limitations of zero-, one- and few-shot prompting techniques in AI-generated translation.

References

Chiang, Wei-Li, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. 2024. Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. https://doi.org/10.48550/arXiv.2403.04132

Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. Stylometry with R: A Package for Computational Text Analysis, *The R Journal*, 8: 107–21.

Equity. 2026. Don't Sign On Set: 8 Steps to protect yourself from AI exploitation. https://www.equity.org.uk/advice-and-support/know-your-rights/ai-toolkit/dont-sign-on-set-8-steps-to-protect-yourself-from-ai-exploitation (Accessed: 21 April 2026).

Hansen, Damien, and Emmanuelle Esperança-Rodier. 2022. Human-Adapted MT for Literary Texts: Reality or Fantasy? In Sheila Castilho, Rocío Caro Quintana, Maria Stasimioti, and Vilelmini Sosoni (eds) *NeTTT 2022*, 178–190, Rhodes, Greece. ⟨hal-04038025⟩

Hawkinson, Eric. 2025. The Em Dash as a Site of Contest Between AI Determinism and Human Agency. *Together Research*, vol. 2, no. 1. https://doi.org/10.62883/FILJ2955

Herrmann, J. Berenike, Karina van Dalen-Oskam, and Christof Schöch. 2015. Revisiting Style, a Key Concept in Literary Studies. *Journal of Literary Theory* 9 (1): 25–52.

Karpinska, Marzena and Mohit Iyyer. 2023. Large language models effectively leverage document-level context for literary translation, but critical errors persist. https://doi.org/10.48550/arXiv.2304.03245

Kenny, Dorothy. In preparation. Do it in Style! Human and automatic translation revisited.

Kenny, Dorothy. 2025. Literary machine translation: from taboo to controversy. In Stefan Baumgarten and Michael Tieber (eds) *The Routledge Handbook of Translation Technology and Society*. 381–395, Routledge, New York. 10.4324/9781003271314-33

Kenny, Dorothy, and Marion Winters. 2024. Customisation, Personalisation and Style in Literary Machine Translation. In *Translation, Interpreting and Technological Change: Innovations in Research, Practice and Training*. Ursula Böser, Sharon Deane-Cox, and Marion Winters (eds). 59–79. Bloomsbury, London. https://doi.org/10.5040/9781350212978

Lee, Changsoo. 2021. How do machine translators measure up to human literary translators in stylometric tests? *Digital Scholarship in the Humanities*

37(3): 813–829. https://doi.org/10.1093/llc/fqab091

O'Sullivan, James. 2025. Stylometric comparisons of human versus AI-generated creative writing. Humanities and Social Sciences Communications, 12(1), 1708. https://doi.org/10.1057/s41599-025-05986-3

Przystalski, K., J. K Argasiński, I. Grabska-Gradzińska, and J. K Ochab. 2026. Stylometry recognizes human and LLM-generated texts in short samples. *Expert Systems with Applications*, 296, 129001. https://doi.org/10.1016/j.eswa.2025.129001

Rybicki, Jan. 2006. Burrowing into Translation: Character Idiolects in Henryk Sienkiewicz's Trilogy and Its Two English Translations. *Literary and Linguistic Computing* 21 (1):91–103.

Rybicki, Jan. 2012. The Great Mystery of the (Almost) Invisible Translator: Stylometry in Translation. In *Quantitative Methods in Corpus-Based Translation Studies*, edited by Michael Oakes and Meng Ji, 231–248. John Benjamins.

Rybicki, Jan. 2025. Can Machine Translation of Literary Texts Fool Stylometry? *Digital Scholarship in the Humanities* 40 (1): 268–276.

Rybicki, Jan, and Magda Heydel. 2013. The stylistics and stylometry of collaborative translation: Woolf's *Night and Day* in Polish. *Literary and Linguistic Computing*, 28(4):708–717. https://doi.org/10.1093/llc/fqt027

Similarweb. 2026. Top AI Chatbots and Tools Websites Ranking. Available at: https://www.similarweb.com/top-websites/ai-chatbots-and-tools/ (Accessed: 21 April 2026).

Tang, Minghua, Chen Bian, Liming Yang, and Xueling Zhong. 2025. Key-concept thinking prompting for improved reasoning in large language models. *Neurocomputing*, 656:130986, https://doi.org/10.1016/j.neucom.2025.130986.

Thai, Katherine, Marzena Karpinska, Kalpesh Krishna, William Ray, Moira. Inghilleri, John Wieting and Mohit Iyyer. 2022. Exploring Document-Level Literary Machine Translation with Parallel Paragraphs from World Literature, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 9882-9902, Association for Computational Linguistics, https://aclanthology.org/2022.emnlp-main.672.pdf

Winters, Marion and Dorothy Kenny. 2024. Mark my keywords: a translator-specific exploration of style in literary machine translation. In Andrew Rothwell, Andy Way and Roy Youdale (eds) *Computer-Assisted Literary Translation*. 69—87, Routledge, New York. 10.4324/9781003357391-5

Wolff, Rachel. 2026. What is the best LLM for translation? A comparison of top AI translation models. Lokalise Blog. Available at: https://lokalise.com/blog/what-is-the-best-llm-for-translation/ (Accessed: 21 April 2026).

Wu, Minghao, Jiahao Xu, Yulin Yuan, Gholamreza Haffari, Longyue Wan, Weihua Luo, Kaifu Zhang. 2025. (Perhaps) Beyond Human Translation: Harnessing Multi-Agent Collaboration for Translating Ultra-Long Literary Texts. *Transactions of the Association for Computational Linguistics*, vol. 13, pp. 901–922. MIT Press, Cambridge, MA. https://doi.org/10.1162/tacl.a.25

Yao, Xiaofang, Yong-Bin Kang and Anthony McCosker. 2025. Missing the human touch? A computational stylometric analysis of GPT-4 translations of online Chinese literature. *Translation Spaces*. https://doi-org.dcu.idm.oclc.org/10.1075/ts.24043.yao

Yirmibeşoğlu, Zeynep, Olgun Dursun, Harun Dallı, Mehmet Şahin, Ena Hodzik, Sabri Gürses, and Tunga Güngör. 2023. Incorporating human translator style into English-Turkish literary machine translation, arXiv preprint arXiv:2307.11457

Appendices

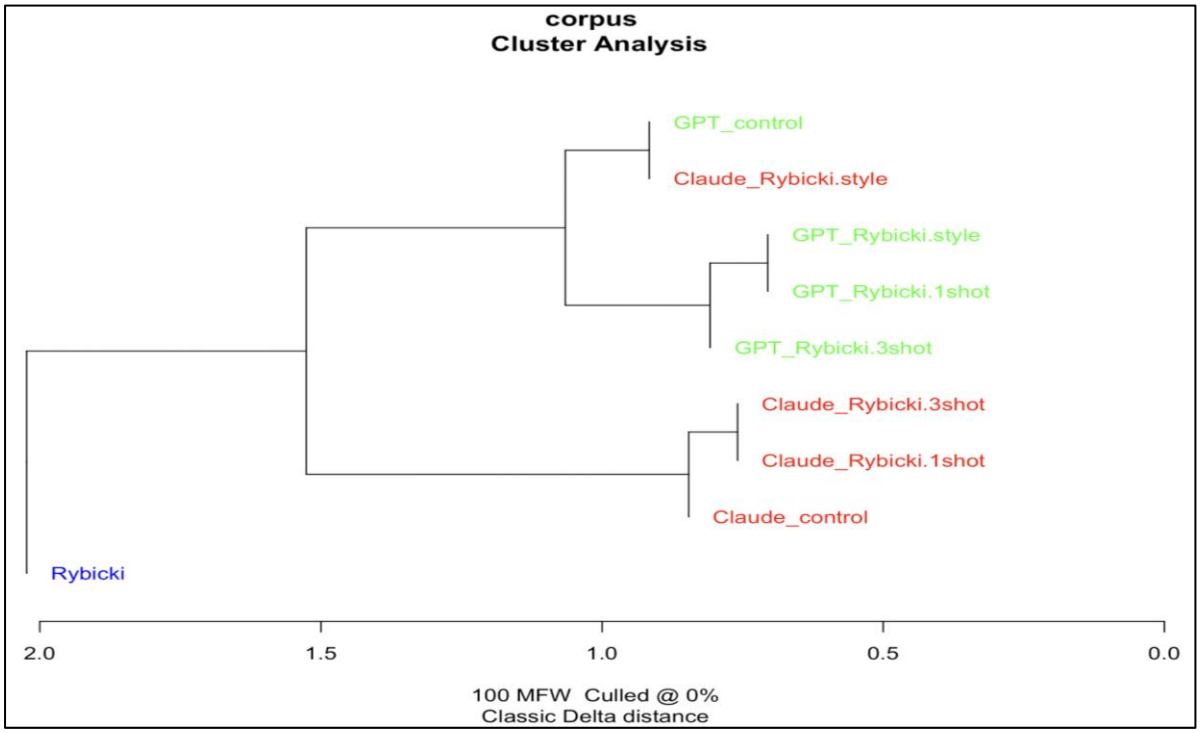

Fig. 1: 100 MFW Cluster Analysis

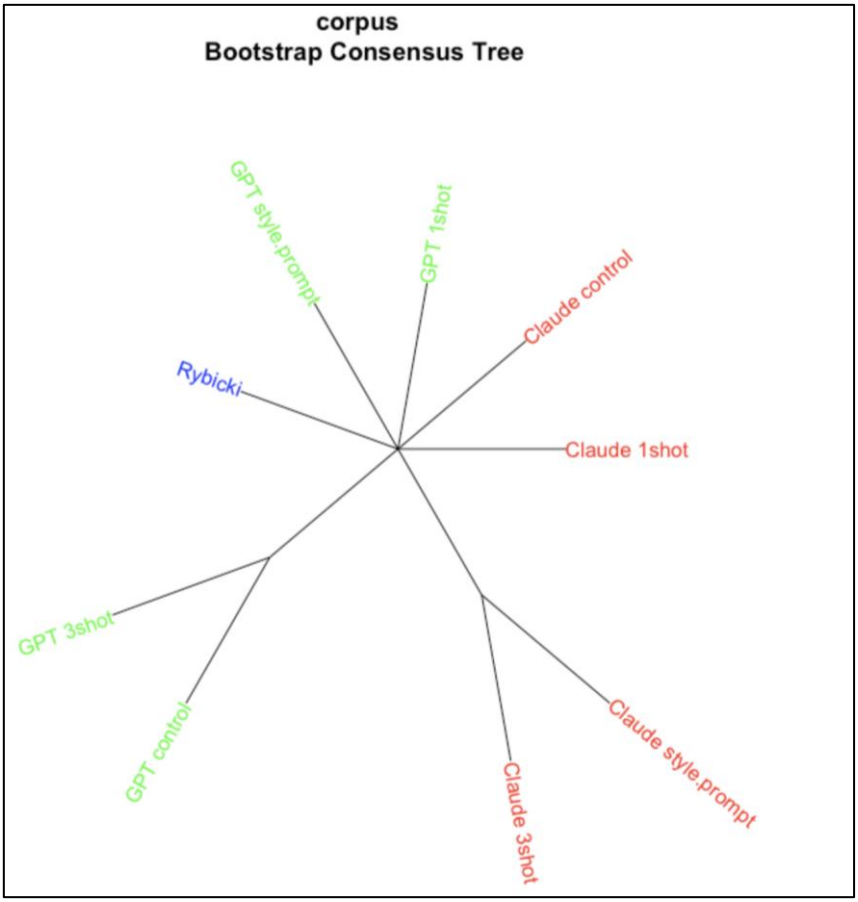

Fig. 2: Bootstrap Consensus Tree (BCT)

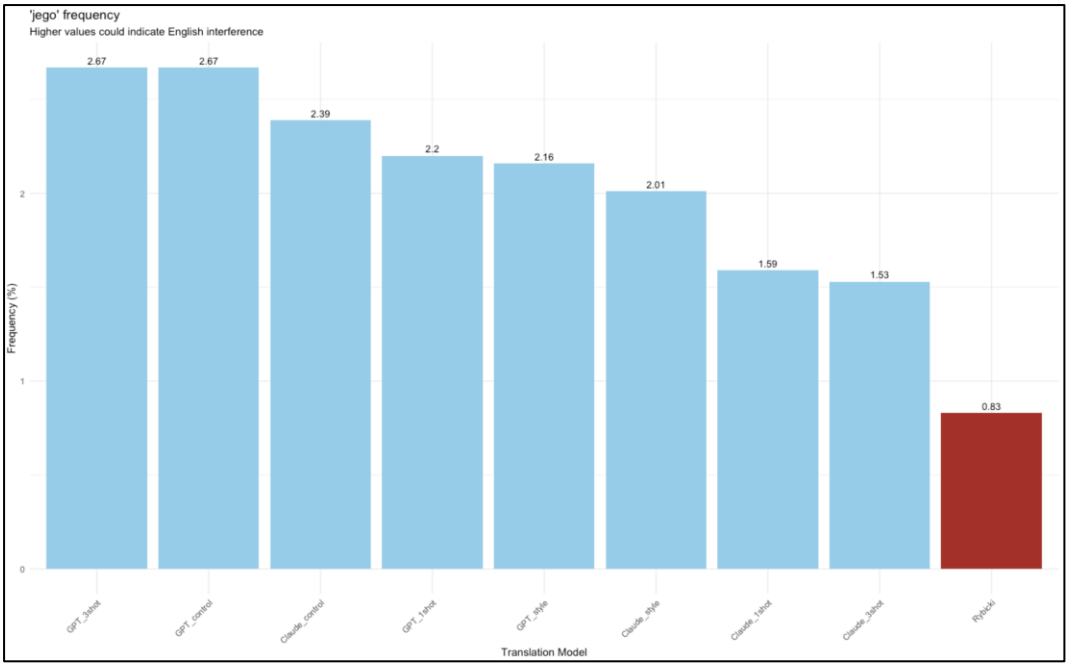

Fig. 3: Frequency of the pronoun ‘jego’

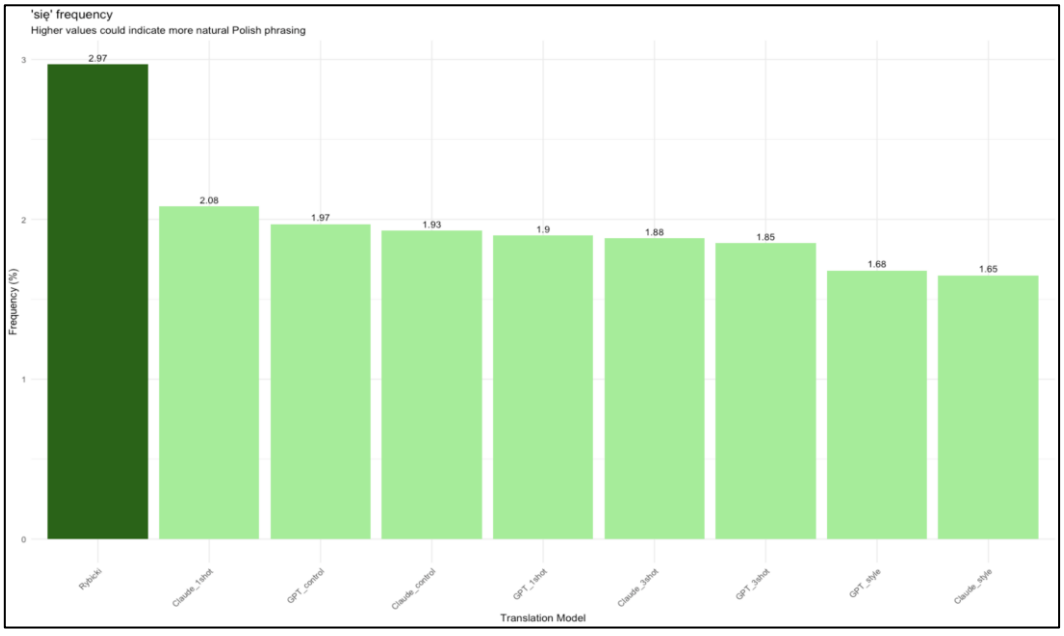

Fig. 4: Frequency of the particle ‘się’

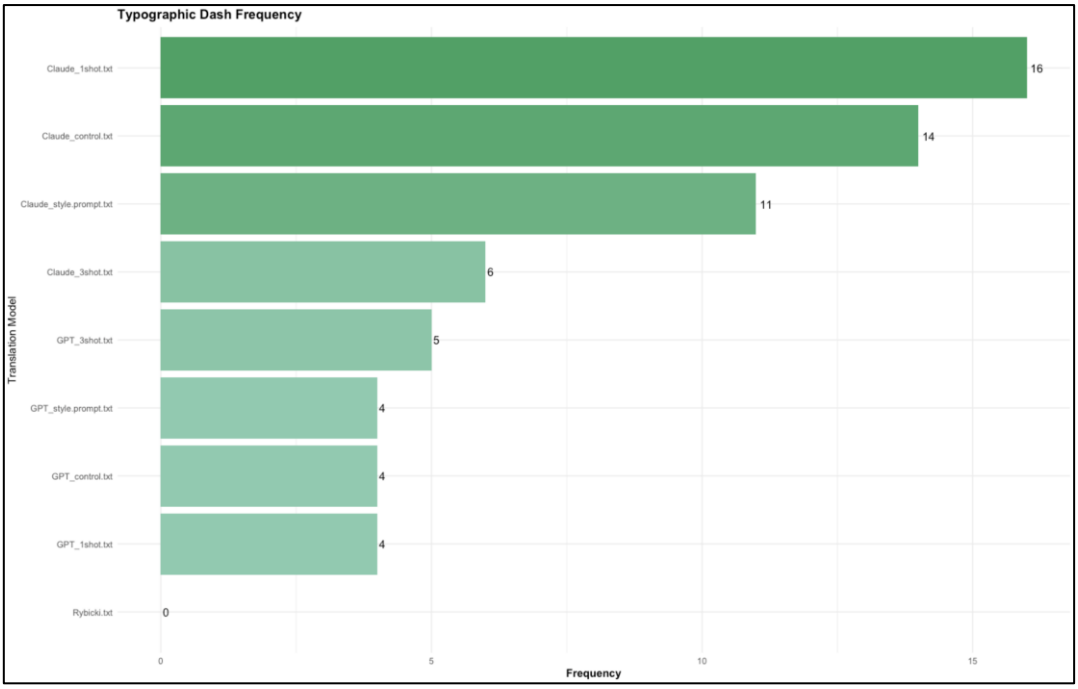

Fig. 5: Frequency of dashes

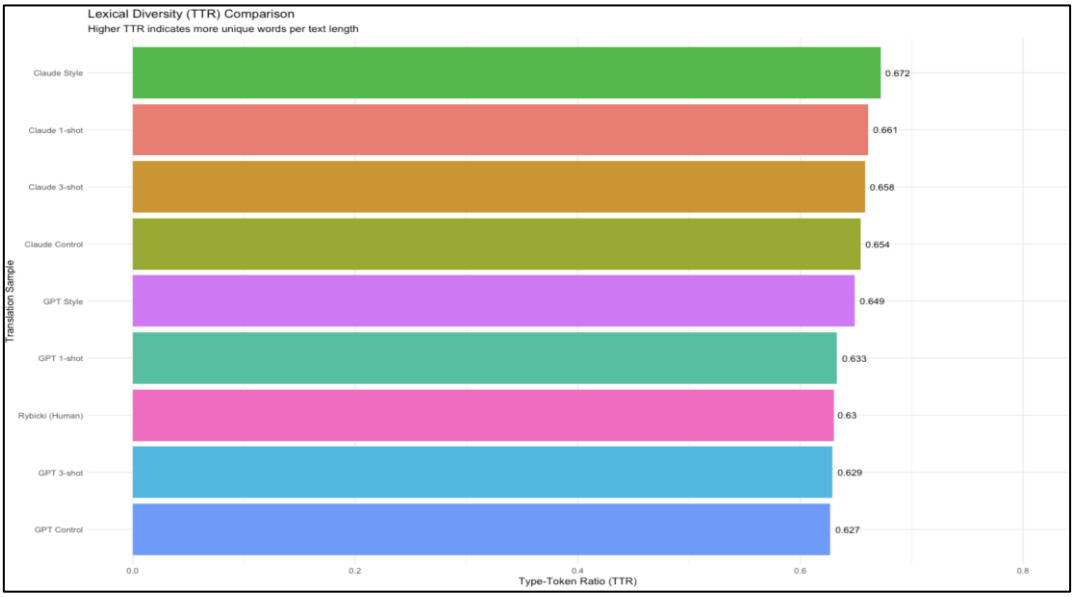

Fig. 6: Model Token-Type Ratio Comparison

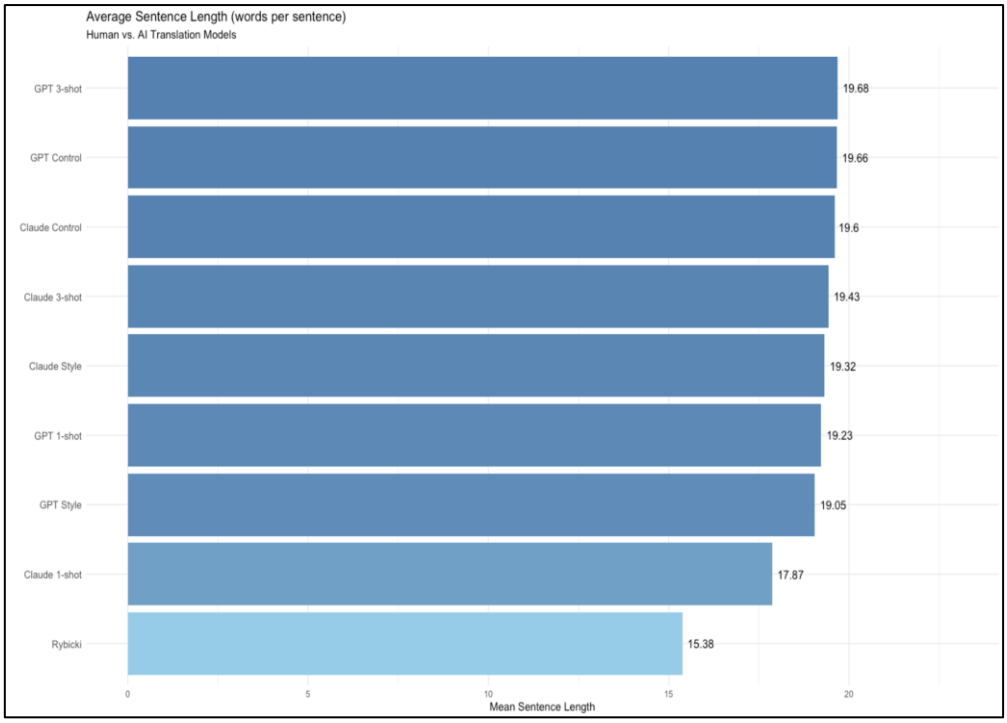

Fig. 7: Average sentence length

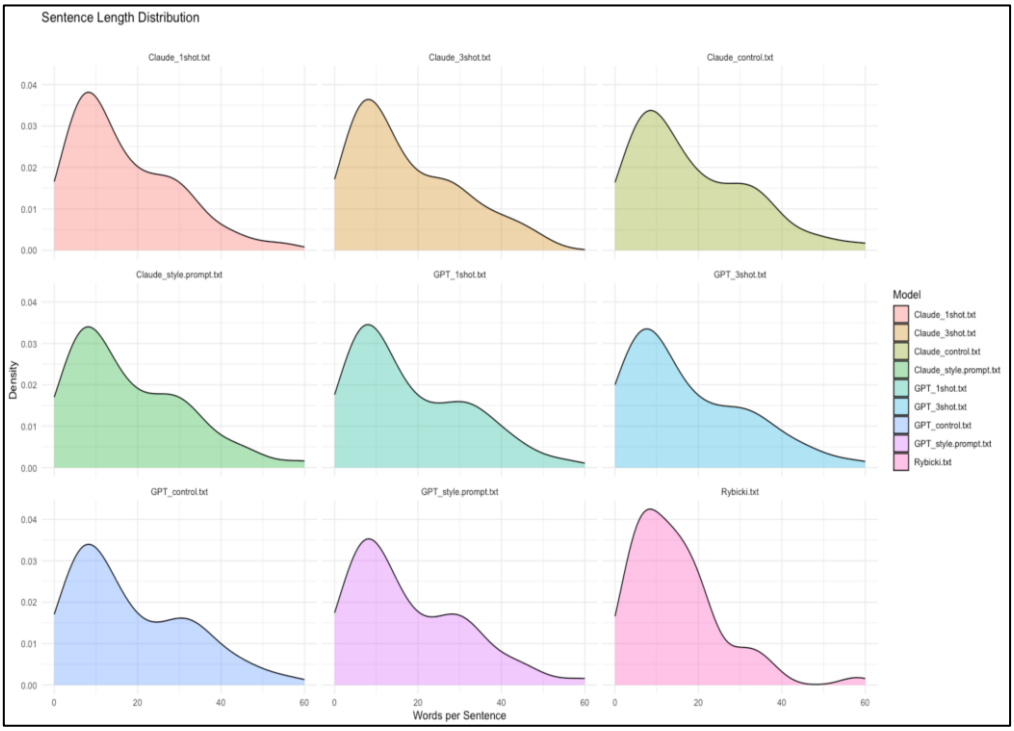

Fig. 8: Sentence length distribution