



OPEN ACCESS

EDITED BY

Lucia Migliorelli,
Marche Polytechnic University, Italy

REVIEWED BY

Paolo Sernani,
University of Macerata, Italy
Cristian Santini,
University of Macerata, Italy

*CORRESPONDENCE

Katherine Elkins
✉ elkinsk@kenyon.edu

RECEIVED 04 June 2024

ACCEPTED 24 July 2024

PUBLISHED 13 August 2024

CITATION

Elkins K (2024) In search of a translator: using AI to evaluate what's lost in translation. *Front. Comput. Sci.* 6:1444021. doi: 10.3389/fcomp.2024.1444021

COPYRIGHT

© 2024 Elkins. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

In search of a translator: using AI to evaluate what's lost in translation

Katherine Elkins*

KDHLab, Integrated Program in Humane Studies, Kenyon College, Gambier, OH, United States

Machine translation metrics often fall short in capturing the challenges of literary translation in which translators play a creative role. Large Language Models (LLMs) like GPT4o and Mistral offer new approaches to assessing how well a translation mirrors the reading experience from one language to another. Our case study focuses on the first volume of Marcel Proust's "A la recherche du temps perdu," a work known for its lively translation debates. We use stylometry and emotional arc leveraging the newest multilingual generative AI models to evaluate loss in translation according to different translation theories. AI analysis reveals previously undertheorized aspects of translation. Notably, we uncover changes in authorial style and the evolution of sentiment language over time. Our study demonstrates that AI-driven approaches leveraging advanced LLMs yield new perspectives on literary translation assessment. These methods offer insight into the creative choices made by translators and open up new avenues for understanding the complexities of translating literary works.

KEYWORDS

Large Language Models (LLMs), translation, Artificial Intelligence, GPT4o, sentiment analysis, stylometry, Mistral, GenAI

1 Introduction

AI opens up new possibilities for analyzing the task of the translator and the role of the author. Translation is highly theorized in the field of comparative literature with competing claims about translatability (Jahan, 2023). Are translators themselves authors, and is translation creative (Venuti, 2012)? Or are translators merely transposing the literary work from one language to another (Benjamin et al., 2004)? Is there an ideal translation, and what would it look like? Competing claims have argued for creativity or fidelity, for the translatability of a text or the untranslatability of key aspects (MacIntyre, 1985). Word embeddings and more recently, the dynamic word embeddings of LLMs, allow us to test these many theories using the latest methods. This paper outlines a new methodology for exploring translatability through the case study of an author whose translations have engendered significant controversy: Marcel Proust's *In Search of Lost Time* (Proust, 2003, 2012a). AI allows us to identify features of the text that have been undertheorized until now. These include the overall emotional arc, changes in the language of sentiment and grade level pacing, and vocabulary and rhythm. These aspects of authorship have not been adequately addressed by more traditional methods of comparative analysis that rely on a close reading of a phrase or sentence. Our results suggest the need for more studies to further explore these aspects and quantify what has been lost in translation.

1.1 Challenges for literary translation evaluation

Advances in machine translation have led to better ways to evaluate these translations. The earliest evaluation metric, BLEU, employs n-gram comparison to human translators (Papineni et al., 2002). BLEURT (Sellam et al., 2020) captures more semantic similarities using pre-trained language models, while BlonDe (Jiang et al., 2021) combines both approaches to leverage the strengths of each. No matter the metric, however, the overall goal remains one of capturing fidelity to a human translator.

In the case of literary translation, however, fairly significant discrepancies between human translators is assumed. There is no agreed-upon human translation to serve as the gold standard, nor is there agreement about the task of the translator. Rather, translation theory proposes quite divergent tasks for the translator depending on the theory.

1.2 The task of the translator

Debate continues to surround the role of the literary translator. Should the literary translator creatively transform the text or strive for maximum fidelity? To what extent should the foreignness of the translation be captured or the cultural idiom be translated from one language to another?

On one extreme is the literalist approach, which seeks fidelity to the original even at the expense of incorporating elements that may seem foreign to the target language (Prendergast et al., 1993). On the other is a naturalizing approach, which seeks to rewrite the original as if the author had written in the target language (Grieve, 2011). This approach can even extend to translating idiomatic expressions and cultural expressions.

There are also questions about whether elements of the original remain untranslatable, and whether the task of the translator is to ensure widespread dissemination of the text, even if it means creating a relatively distinct new work of art (Derrida, 2001).

1.3 An ideal case study

Marcel Proust's *A la recherche du temps perdu* provides an excellent case study since there are numerous translations, and there is not widespread agreement on which is best (Proust, 1989). Translations also highlight divergent theories of literary translation, sometimes even under a single imprint. The newest Penguin translation edited by Christopher Prendergast has a different translator for each volume. Grieve (2011), the translator for volume two of the series, describes the extent to which some of the translators, including Lydia Davis in the first volume, hew to fidelity even when it creates odd sentence structures. Grieve, on the other hand, wished to translate the novel "as if" Proust had written it in English, even down to invoking English writers like Milton. The latter approach was also taken by Proust's very first translator, Scott Moncrieff, whose original title, *Remembrance of Things Past*, is a nod to Shakespeare (Proust, 2012b). This title was criticized by Proust himself as straying too far from the original.

Lydia Davis, on the other hand, originally translated her volume as "The Way by Swann" in an effort to remain more faithful to the original. This was instead of the more usual "Swann's Way." Her decision was also highly criticized, demonstrating that there is no universal agreement about the task of the literary translator. A robust discussion of the comparative strengths and weaknesses of the original Scott Moncrieff edition, later edited by Kilmartin and revised by Enright, vs. the newer translations can be found across a variety of media. Disagreements are many, but discourse revolves around the vexing question of what is inevitably "Lost in Translation" (Carter, 2005).

1.4 Theory and practice of literary translation

To date the ways to capture these elements and judge the success of the translator has been to engage in relatively small scale analysis by selecting a few choice comparisons and generalizing from those to judge the overall translation. Critics have noted, for example, that Lydia Davis decides to preserve the commas in the French, whereas James Grieve insists that those commas should not translate to English if English expressions would typically avoid such punctuation. Thus, the opening sentence of the novel, "Longtemps, je me suis couché de bonne heure" is translated by Davis as "For a long time, I went to bed early" and by Scott Moncrieff as "For a long time I used to go to bed early."

Moncrieff further adds some interpretation to the passage by adding a habitual statement ("used to") that does not exist in the original. Enright's revision modifies this yet again as "I would go to bed early." These very subtle interpretive choices point to yet one more aspect of literary translation, which revolves around how much the translation preserves the ambiguity of the original or whether, as is the case with these two translations, the ambiguity is partially erased with additional language that clarifies the contextual meaning.

Depending on the critic, one can find attention to different details and different excerpts but, aside from more general statements like those about Davis' preservation of punctuation, it can be hard to get a general sense of the translation from these more subtle aspects of translation choice that are typically highlighted using traditional close reading methods.

2 Translating authorial qualities

Recent methods employing AI for narrative analysis give us some clues as to new ways we can assess literary translation. As opposed to close reading methods that generalize from small discrete examples, we seek AI methods that can pick up signals hard to identify in a few words but visible as general patterns in a larger corpus.

Two computational approaches that can help surface patterns not always visible to human evaluation are stylometry and readability. Stylometry has been used to identify unknown authors as, for example, when a novel published pseudonymously by J.K. Rowling was discovered (Juola, 2013). The approach relies on subtle statistical patterns in language that are often quite distinct

from the more obvious stylistic patterns we identify with a well-known author.

Readability offers another method of surfacing patterns that could be helpful for quantifying translation qualities. Readability influences how accessible a text is to a particular audience. Blatt documents a general change in readability levels in bestsellers over time, with increasing readability decade by decade (Blatt, 2018). Others have found that readability changes over the course of a narrative, likely in order to affect readerly pace and engagement (Clipson, 2021).

2.1 Methods and materials

Experiments in readability and sentiment analysis (discussed in Section 3) were conducted using an HP Victus Gaming Laptop with a 13th Gen Intel(R) Core(TM) i7-13700HX 2.10 GHz, 32 GB of memory, an NVIDIA GeForce RTX 4060 with CUDA 12.1 and Windows 11 Home Version 23.2 OS.

We experimented with a variety of methods for computing stylistic analysis and readability including a proprietary model (Readable.com, no preprocessing needed)¹, textstat accessed through GPT4o², and Lexical Richness as part of an AI-LIT repository³. In this last case, pysbd (for English) and spacy (for French) were used for text segmentation during preprocessing.

Readable.com offers both a large scale corpus method and a smaller text chunk method with highlighted features. GPT4o now allows for advanced data analysis using the Github repository textstat. One surprising finding using both of these approaches was that Moncrieff was rated as having the highest readability. This finding is significant in that it goes against the common critical reception of that translation, which is typically thought of as the most “Victorian” and ornate. One would also expect increased readability over time in keeping with Blatt’s general findings.

These findings require some qualifications, however. First, GPT4o using textstat and the proprietary model Readable showed considerable divergence in scores, and it was difficult to verify scores over larger corpora. Furthermore, many scores were developed for English, and there is some question as to their reliability for other languages. Lix and Rix scores, both of which were developed in Europe, showed less significant differences between the original French and the translations. Judged by these metrics, all three translations—the Moncrieff, the Enright revision and the Davis—were deemed far more similar to the French original, with a readability aimed at college level. Finally, the most significant divergences identified during the examination of smaller passages often disappeared when evaluating the entire text.

While we are optimistic that GPT4o will eventually yield better results, we opted in the meantime for the AI-LIT Github repository. AI-LIT was developed by Chun (2024) with the express purpose of surfacing qualitative differences between literary texts using quantitative methods. It allows for diachronic stylistic analysis, a

distinct advantage over GPT4o and Readable. It also allows for an ensemble model for surfacing emotional arc that includes both more traditional models like VADER and the latest state-of-the-art models like Mistral.

2.2 Diachronic style in translation

Based on current scholarly evaluations of translations, we would expect to see Moncrieff’s version as the least readable, since Moncrieff’s translation is often cited as being even more ornate and Victorian than Proust’s original French. Davis, on the other hand, is well known for writing “micro” stories that are a single sentence or paragraph long. As a very modern translation, we would expect the Davis version to be the most readable, with the French original to lie somewhere in between these two extremes. Finally, we would expect an increase in readability over time.

The Lexical Richness package offers a variety of metrics for assessing the lexical diversity of the text. Here we broke the text into chunks in order to diachronically visualize the changing lexical diversity and density over time. Some metrics are high when they track low lexical density or diversity while others track in the opposite way. This is the reason why many of the metrics generally show a mirroring trend that revolves around the x-axis.

2.3 Results

Results demonstrate that in the case of Proust’s narrative there is a significant change in style and readability over time. Figure 1 shows the normalized style metrics for the first volume of the French original using a simple moving average. In this case, the narrative begins with medium-high lexical diversity with decreasing diversity partway through the narrative. A somewhat high lexical diversity trends for the first 40 chunks, descending to lower lexical diversity until around chunk 70 before reversing direction.

Comparison with all three translations shows a far noisier pattern, suggesting that the translations fail to capture authorial style in some respects. While some style metrics closely mirror the French original, others do not. We can see this with the most extreme examples, both in the more literal Davis translation shown in Figure 2 and the more interpretive Moncrieff translation shown in Figure 3.

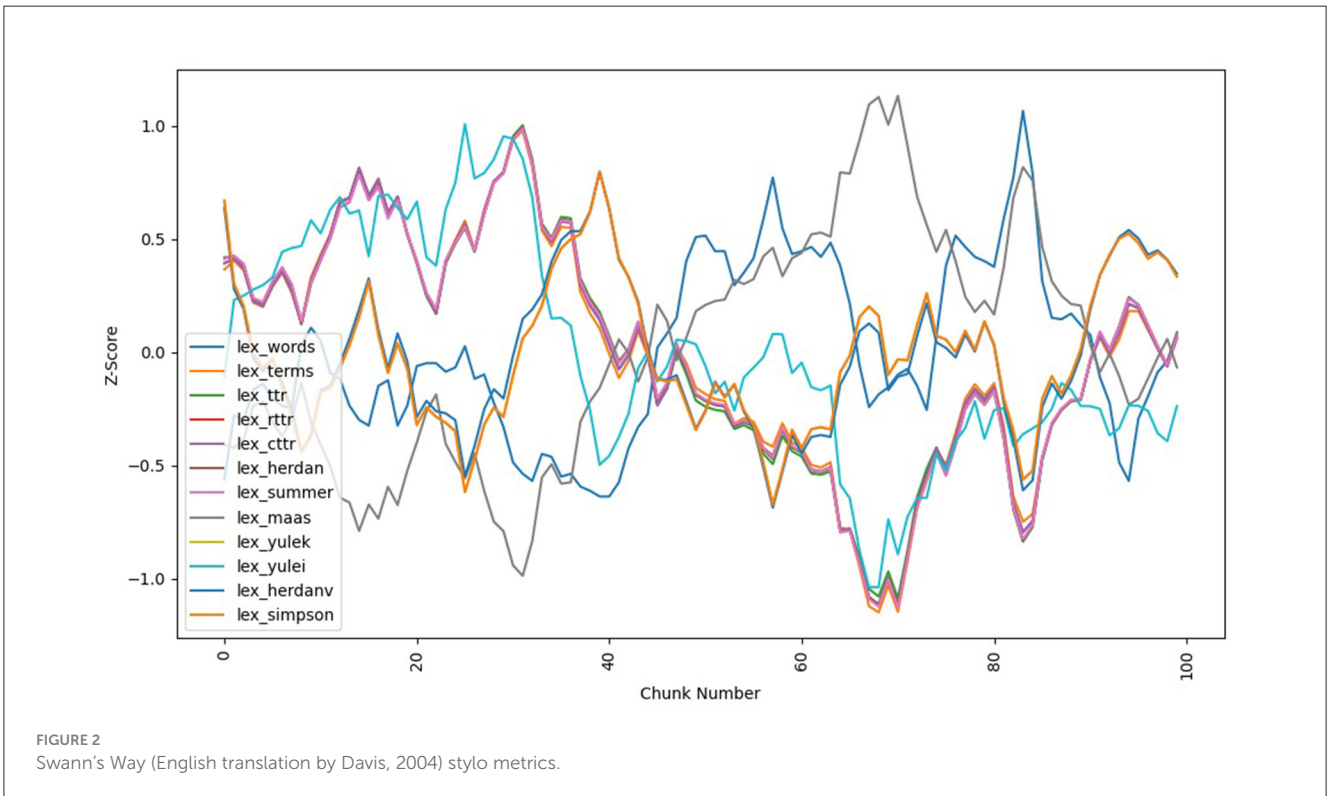
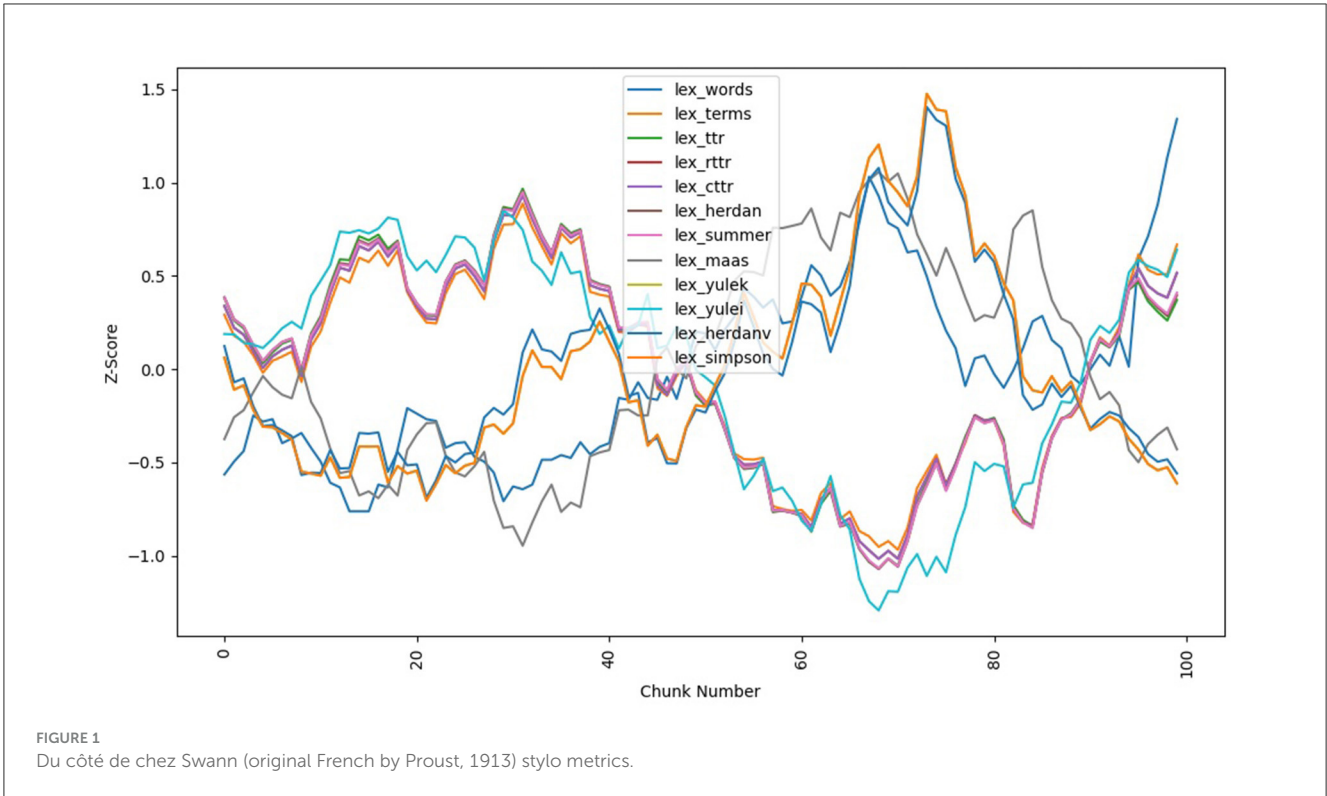
Comparing translations and the French original along a single metric further illuminates these nuances. The number of unique words track fairly well across all versions as shown in Figure 4. Both the Davis translation and Proust’s original French alternate between diverging from this common trend slightly. They also both demonstrate a gradual decrease in lexical diversity. However the Davis evinces the greatest divergence from the original near the beginning. The French, on the other hand, shows the greatest divergence in lexical diversity near the end. These relatively minor differences are mirrored in other metrics like the Type Token Ratio.

By contrast, there is considerably more divergence between the French and the three translations using more complex algorithms like Yule’s K, Herdan’s Vm and Simpson’s D. These

1 Readable. Available at: <https://readable.com/>.

2 Textstat. Available at: <https://github.com/textstat>.

3 Lexical, R. Available at: <https://lexicalrichness.readthedocs.io/en/latest/README.html>.



more complex algorithms were designed to compensate for the limitations of simpler metrics and, in the case of the Yule equations, stem from a consideration of literary corpora. With more complex metrics, we see more significant divergences from

the French in translation. For example, Figure 5 shows that using Herdan's metric, the French and its translations are quite different between chunks 10 and 20 and again from around chunk 10 to chunk 80.

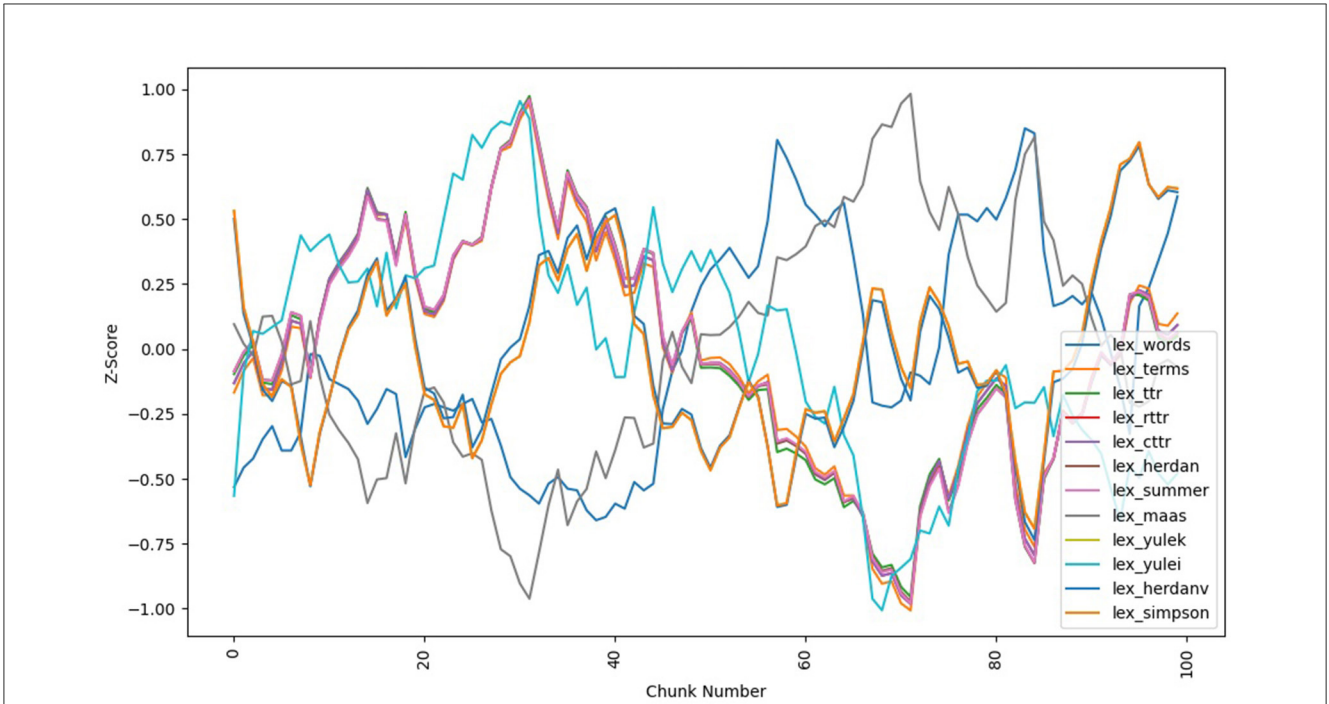


FIGURE 3 Swann's Way (English translation by Moncrieff, 1922) stylo metrics.

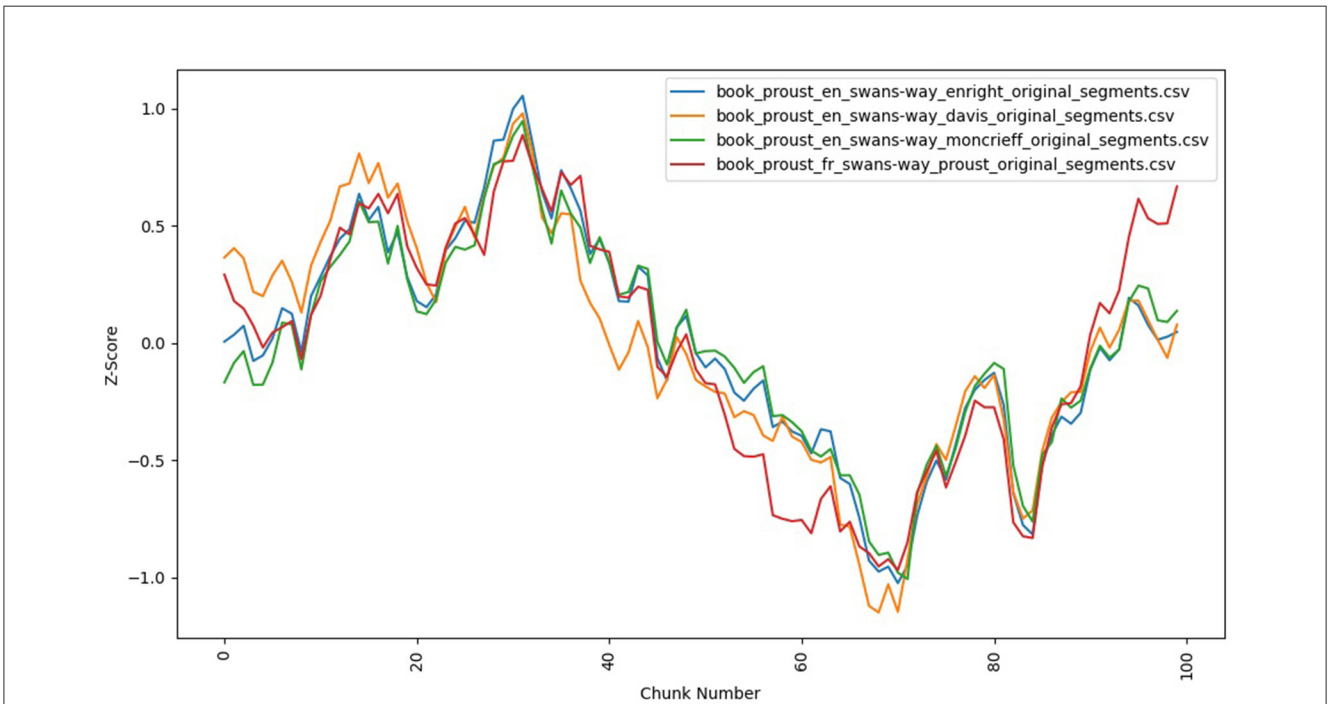
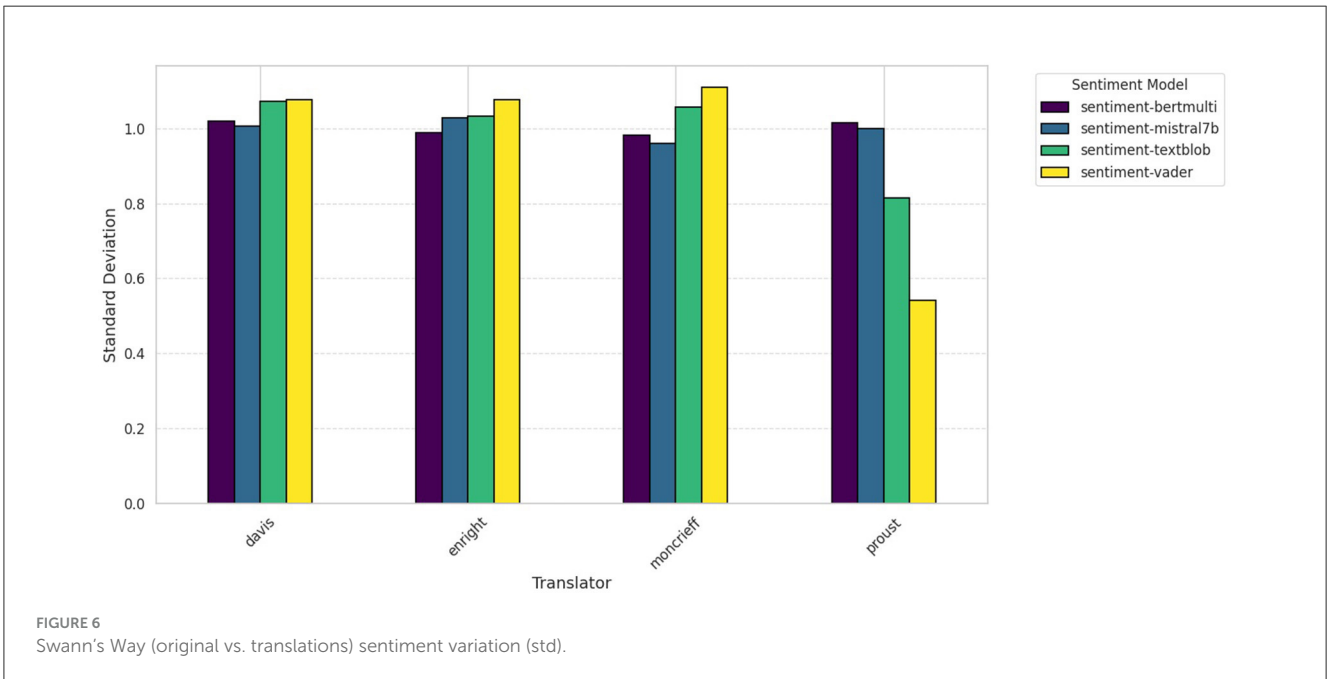
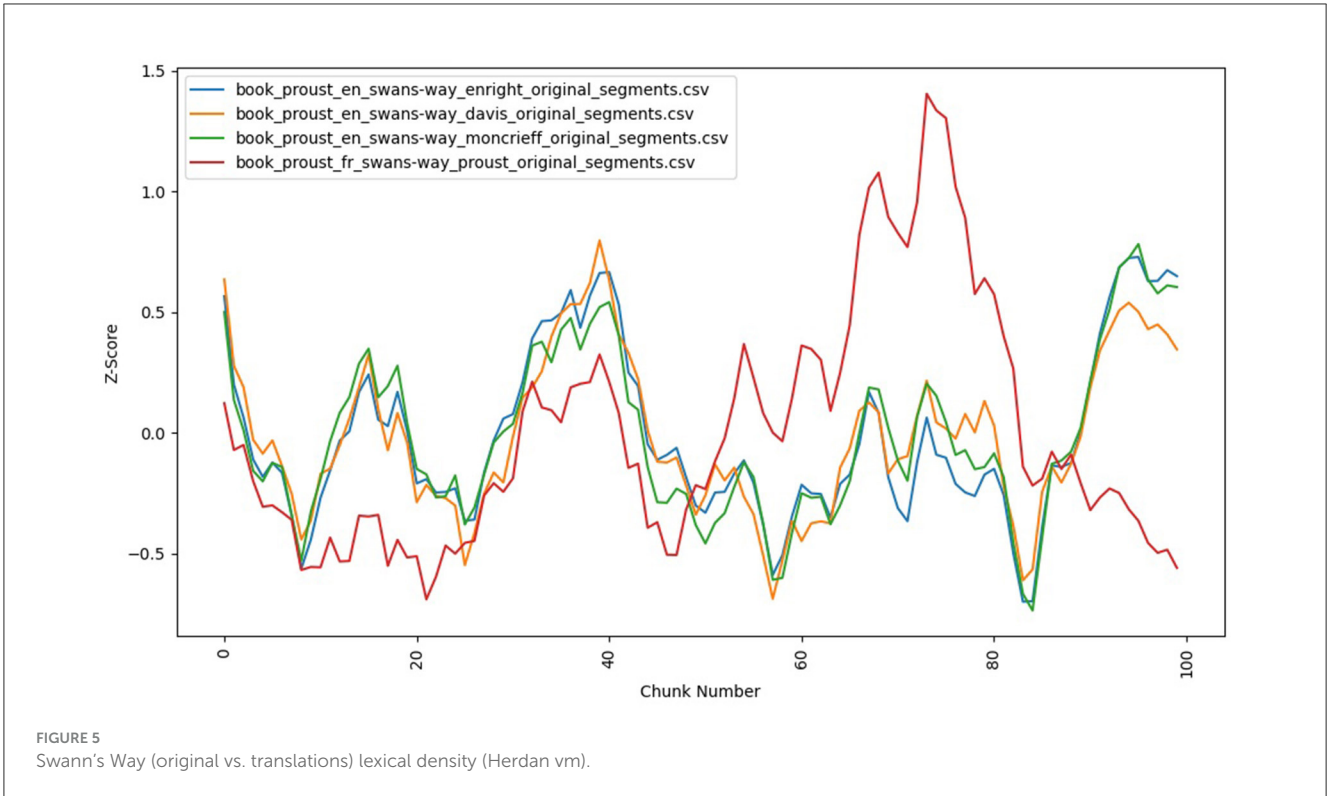


FIGURE 4 Swann's Way (original vs. translations) lexical density (Lexical Terms).

2.4 Discussion

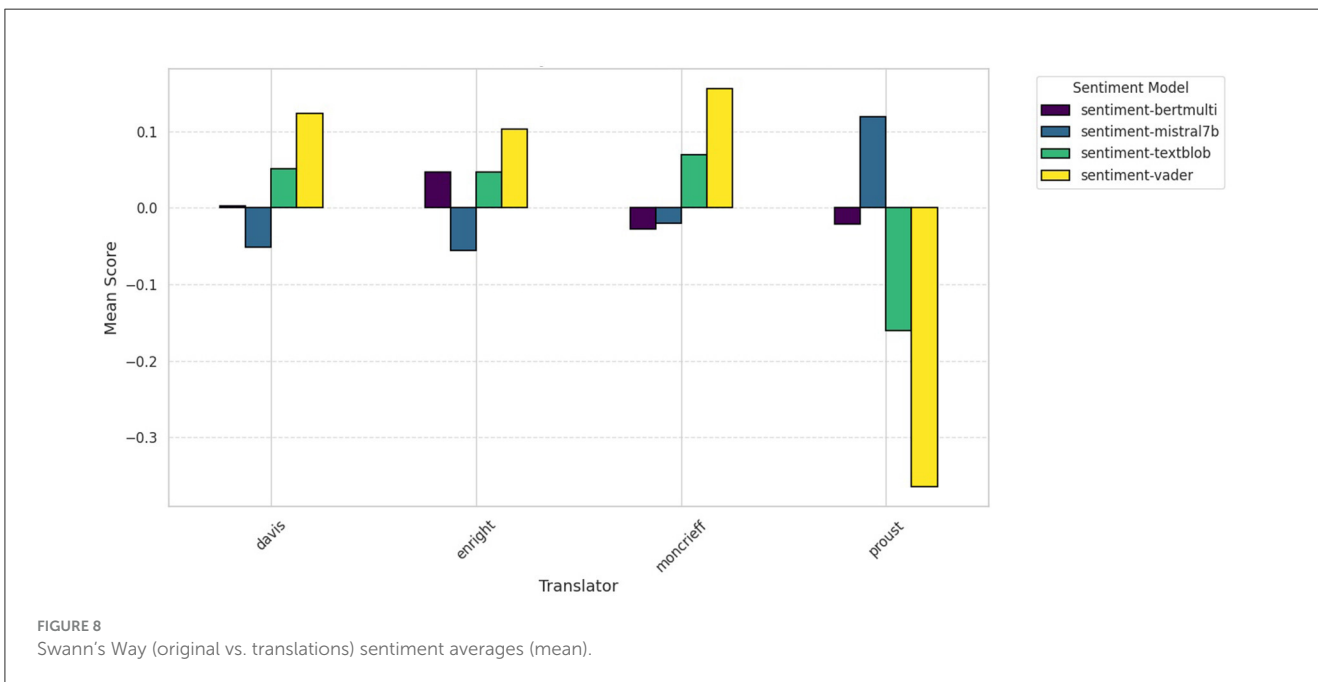
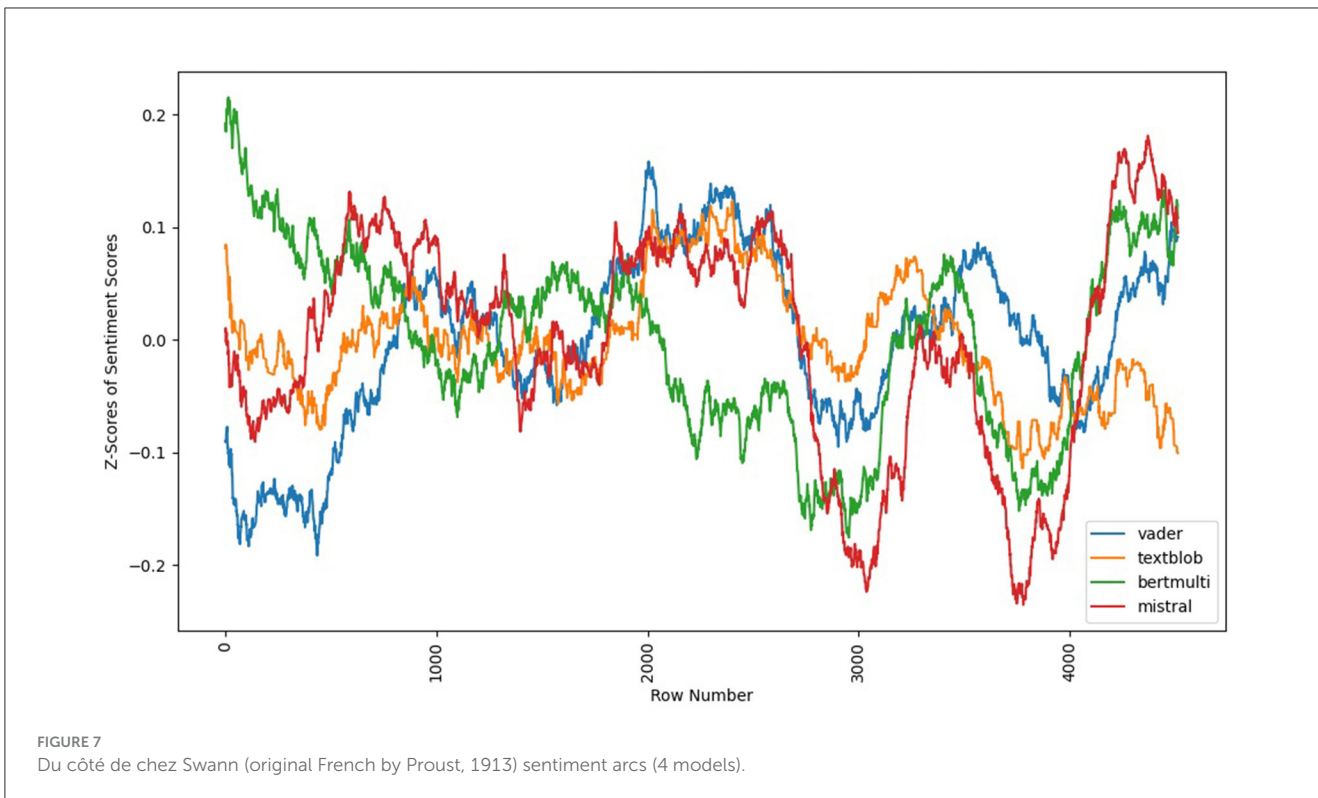
Several conclusions can be drawn from these findings. First, they confirm preliminary research suggesting that certain stylistic

features show distinctive diachronic patterns in literary texts. More research is needed to determine whether these are general patterns or whether they change depending on different periods, genres, and cultural traditions.



In this case study, lexical density and diversity decrease over time, suggesting that the narrative begins with a fairly complex style before moving toward simpler language over time. This pattern loosely follows the narrative structure of the novel in which the opening follows a more modern impressionist style before ceding to the third-person story of Swann, often considered a narrative style and structure more typical of a 19th-century novel.

Finally, it would seem that some stylistic aspects of the original are only partially captured by the translations. Unlike the original, all three translations show a flattening of lexical richness in a way that distinguishes them from the original. The Davis translation, often considered the most “literal,” comes closer to capturing this aspect than the other two translations.



3 Emotional arc in translation

Like stylistic analyses, sentiment analysis surfaces an emotional arc that can determine key differences in the way a translation is experienced by a reader. It surfaces a background signal, a series of emotional ups and downs, that are harder to analyze using a more typical comparative approach that isolates a few key words or phrases. Sentiment analysis therefore offers another

potentially useful method for comparing a literary translation to an original.

3.1 Background

Research has demonstrated emotional arc as a key component that distinguishes bestselling narratives (Archer and Jockers,

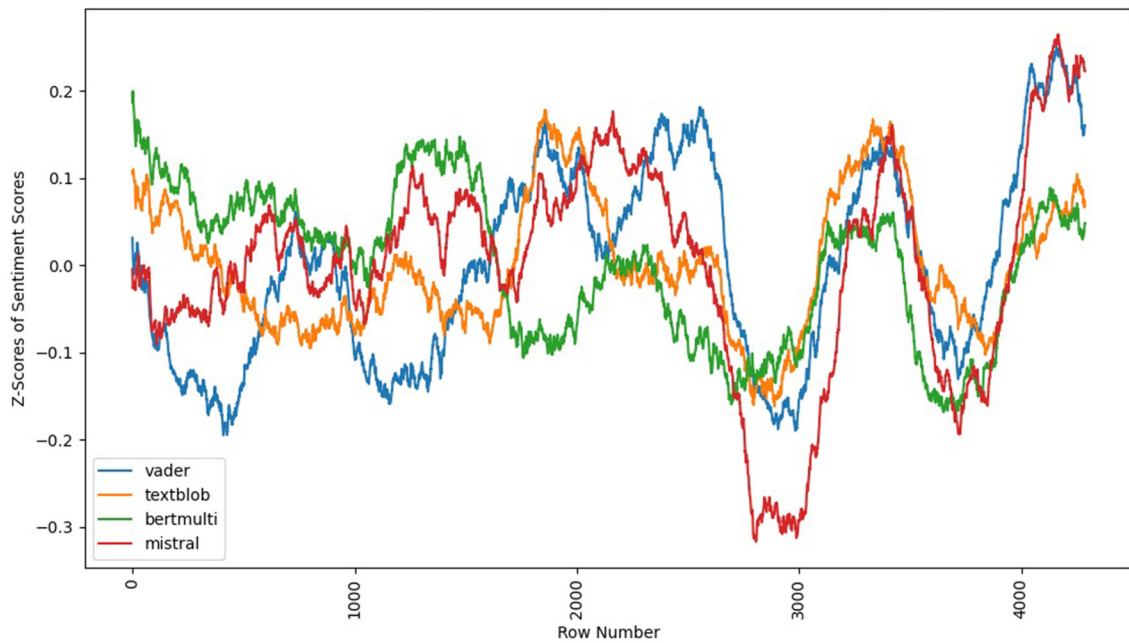


FIGURE 9 Swann's Way (English translation by Moncrieff, 1922) sentiment arcs (4 models).

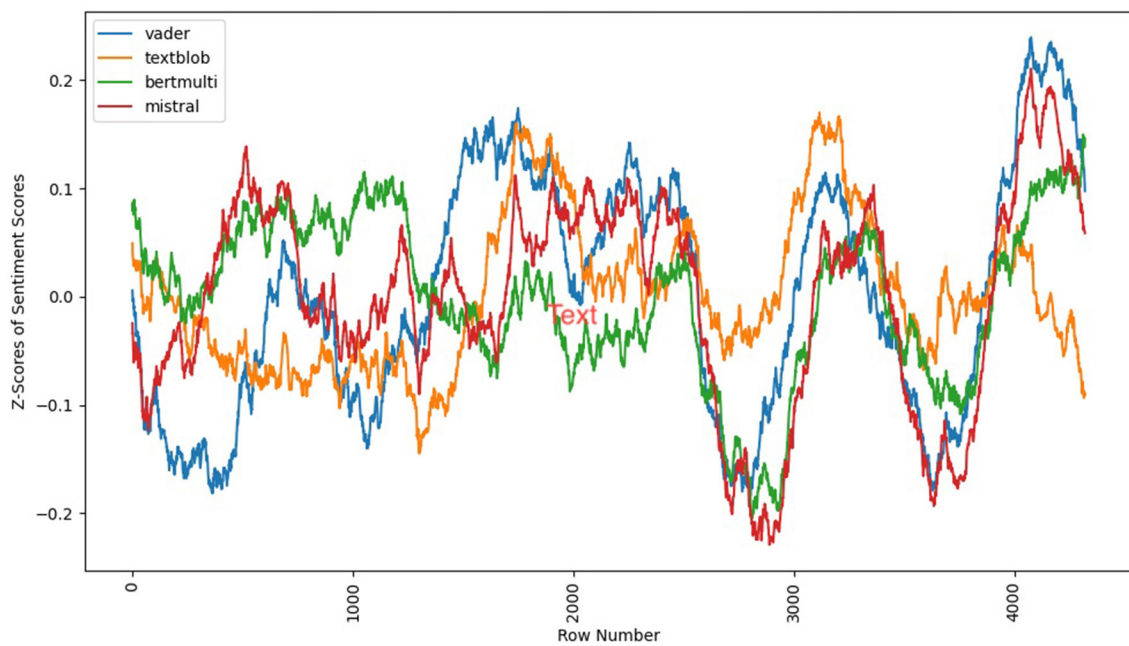


FIGURE 10 Swann's Way (English translation by Davis, 2004) sentiment arcs (4 models).

2016). Scholars have also demonstrated that some translations exhibit subtle differences in emotional arc that are not easily explained by time period or distinctive linguistic styles (Elkins,

2022). Other studies have found more significant variations in emotional arc depending on the target language of the translation (Strain, 2022).

3.2 Large Language Models vs. traditional computational approaches

Up until recently multilingual sentiment analysis has not always proven reliable across languages, and early research suggested BERT models might not always perform well (Elkins, 2022). When comparing emotional arcs of stories across languages, therefore, any significant differences could be attributed to the performance of the model rather than the object of study.

Recently, however, Large Language Models have demonstrated significantly better performance across languages. While earlier approaches to narrative analysis necessitated an ensemble of models, LLMs make it more and more likely that we can begin to rely on a single model that performs well across languages.

The performance of Mistral offers one good test case to assess performance against older models given its known reliability in both English and French. Our first goal was therefore to compare model performance. Unlike the smaller transformer models like BERT, larger language models like Mistral offer better context awareness. We would expect the larger language models to be able to evaluate ambiguity and edge cases more reliably across languages.

3.2.1 Methods

An ensemble of models of increasing complexity was chosen, including a simple lexical model (VADER), a more complex model (Textblob), a multilingual Transformer (BERTmulti available on huggingface) and a state-of-the-art 7 billion parameter model (Mistral) using an ollama library. Raw sentiment polarity was extracted from each model according and z-score normalized (set mean = 0 and y-axis transformed to [-1.0 to 1.0]) to enable direct comparisons. VADER uses a lexical approach with additional heuristic rules, Textblob uses a Naive Bayes classifier, and BERT-Multilingual is fine-tuned to excel at multilingual tasks. Mistral is a general LLM that renders sentiment via a prompting. Default Mistral ollama hyperparameters (temp = 0.8, top p = 0.9) were used along with the following prompt:

```
prompt=f"###SENTENCE:\n{sentence}\n\n\n###INSTRUCTIONS:\nGiven the above ###SENTENCE, estimate the sentiment as a float number from -1.0 (most negative) to 0.0 (neutral) to 1.0 (most positive). Return only one float number between -1.0 and 1.0 for sentiment polarity and nothing else, no header, explanation, introduction, summary, conclusion. Only return a single float number for the sentiment polarity"
```

More details about each model can be found in specific library documentation. A good overview as well as justification for an ensemble comparative approach is made by Elkins (2022).

3.3 Results

Sentiment scores varied across models and translations as can be seen in Figure 6. The largest difference in scores can be seen with the French original as assessed by different models, confirming

that multilingual sentiment analysis poses a challenge in terms of cross-model coherence. VADER and Textblob exhibited fairly different ranges of sentiment from the Transformer models even when normalized.

Further exploration of Proust's French original suggests that Mistral best captures the subtleties of the original. This was confirmed using a peak crux detection and analysis method first developed by Elkins (2022). As seen in Figure 7, BERTmulti showed the most divergence outside of the strongest emotional arc area, which forms the "W" in the second half of the narrative. Textblob and VADER tracked only loosely with Mistral, which performed the best according to human evaluation. Textblob surfaced the least strong emotional arc, moreover, hovering closer to neutral for large swaths of narrative time. Mistral surfaced the highest intensity low points of the "W," alongside a more neutral midpoint between both "V"s. These intensities were also confirmed by human evaluation.

There was less agreement between models for the opening of the narrative. Mistral offered the clearest curvy shape as would be expected in a highly popular novel. As yet one further checkpoint, mean sentiment scores were evaluated across models and translations as seen in Figure 8. Here, there is a surprising lack of agreement between models and translations. Notably, Mistral assessed translations as overall negative but the original as overall positive.

Next, a comparison of the emotional arc between translations and the original suggests striking differences as seen in Figures 9, 10. Mistral identified the same "W" shape, first identified as bestseller curves by Archer and Jockers, in all three translations. However the beginning of the narrative proved more of a challenge, and none of the translations were able to accurately capture the same slightly positive curves in the beginning chapters of the novel. This beginning shape was difficult to surface for the simpler models as well, suggesting a more subtle pattern. Given that it also fails to appear strongly with the Mistral in translations, it may also be less noticeable to translators, and thus less well captured in translation.

3.4 Discussion

The Mistral arc best captures the emotional nuance of the original French as well as the English translations. That a similar distinctive pattern has been surfaced in another novel from this period (Elkins and Chun, 2019) lends further confirmation to this conclusion. The French is more accurate than the translations, however, since human evaluation confirms the more negative dips of the French "W," which occur during the protagonist Swann's disillusionment with love. The rise between the two halves of the "W" is also more neutral than other models or translations rated it. This passage corresponds to an explicitly referenced neutral emotional state of the protagonist Swann.

Both the Enright and the Moncrieff surface fairly similar emotional arcs to each other, which is not surprising given that the Enright is an edited revision of the Moncrieff. In neither translation does the Mistral arc dip as low during the second drop of the "W." The earlier half of the narrative also evinces less of a clear wave than the original.

By contrast, the Davis shows an arc that is closest to the original overall. The beginning of the narrative somewhat conveys the curves of the original using Mistral, albeit not as clearly, and the

ending “W” lies somewhere between the French original and two other English translations in terms of intensity and neutrality.

4 Conclusion

It is notable that two strong patterns emerge in the original French. One is a striking change in lexical richness over time. The second is a stronger pattern of emotional arc in the second half of the narrative. Taken together, one can deduce that the pace of the second half of the novel, both in terms of easy readability and in terms of a strong narrative arc, typify this particular story. It is not always the case that a reader will experience the emotional arc that the language of sentiment traces. Nonetheless, emotional arc gives us one good way to assess translation differences that may be fairly subtle.

More research is needed, but this study suggests that AI can help surface elements of a translation that may not be visible using a traditional close reading method. Patterns of lexical richness alongside differences in emotional arc are likely to give rise to a significantly different reading experience. In this case, only one translation was able to partially capture both phenomena, and it did so only imperfectly. It is possible that the most successful translator according to these metrics may have been more attuned to these aspects due to her deep experience as a published and successful story writer. While one can never truly replicate the way it feels to read a literary work in the original, both of these methods reveal subtle aspects of the reading experience that can easily be lost in translation.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: Available at: <https://github.com/jon-chun>.

References

- Archer, J., and Jockers, M. L. (2016). *The Bestseller Code: Anatomy of the Blockbuster Novel*. New York: St. Martin's Press.
- Benjamin, W. W., Bullock, M., and Jennings, M. W. (2004). “The task of the translator,” in *Walter Benjamin: Selected Writings. Vol. 1 1913-1926*. Cambridge, MA: Belknap Press of Harvard Univ. Press, 253–263.
- Blatt, B. (2018). *Nabokov's Favorite Word Is Mauve: What the Numbers Reveal about the Classics, Bestsellers, and Our Own Writing*. New York; London; Toronto; Sydney; New Delhi: Simon Et Schuster.
- Carter, W. C. (2005). Lost in translation. *Moder. Modern.* 12, 695–704. doi: 10.1353/mod.2005.0106
- Chun, J. (2024). *AI-LIT. Github Repository*. Available at: <https://github.com/jon-chun?tab=repositories> (accessed August 2, 2024).
- Clipson, S. (2021). *Analyzing the Reading Levels of Fifty Shades of Gray and the DaVinci Code*. Digital Kenyon, 2021. Available at: https://digital.kenyon.edu/dh_ips_prog/47/ (accessed August 2, 2024).
- Derrida, J. (2001). What is a relevant translation? *Trans. Venuti. Crit. Inquiry* 27, 174–200. doi: 10.1086/449005
- Elkins, J., and Chun, J. (2019). Can sentiment analysis reveal structure in a plotless novel? *ArXiv abs/1910.01441*.
- Elkins, K. (2022). *The Shapes of Stories*. Cambridge: Cambridge University Press. doi: 10.1017/9781009270403
- Grieve, J. (2011). *Working with the Demented: Penguin's Proust, an Experiment in Collaborative Foreignization*. The Australian National University Open Research Repository.
- Jahan, D. (2023). Comparative literature and translation studies: approaching an understanding between the two. *Int. Soc. Sci. Hum. Res.* 06:32. doi: 10.47191/ijsshr/v6-i3-32
- Jiang, Y. E., Liu, T., Ma, S., Zhang, D., Yang, J., Huang, H., et al. (2021). *BlonDe: An automatic evaluation metric for document-level machine translation. arXiv preprint arXiv:2103.11878*.
- Juola, P. (2013). “How a Computer Program Helped Show J.K. Rowling Write a Cuckoo's Calling.” *Scientific American*. Available at: <https://www.scientificamerican.com/article/how-a-computer-program-helped-show-jk-rowling-write-a-cuckoos-calling/> (accessed August 2, 2024).
- MacIntyre, A. (1985). Relativism, power and philosophy. *Proc. Addr. Am. Philosop. Assoc.* 59:5. doi: 10.2307/3131644
- Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. (2002). “Bleu: a method for automatic evaluation of machine translation,” in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (Philadelphia, PA, USA: Association for Computational Linguistics), 311–318. doi: 10.3115/1073083.1073135
- Prendergast, C., Proust, M., Scott Moncrieff, C. K., Kilmartin, T., and Enright, D. J. (1993). “*English Proust*.” *London Review of Books*. Available at: <https://www.lrb.co.uk/the-paper/v15/n13/christopher-prendergast/english-proust> (accessed July 8, 1993).
- Proust, M. (1989). *A La Recherche Du Temps Perdu*. Amazon. Paris: Gallimard.

Author contributions

KE: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The author would like to thank Kenyon College for supporting this project with a Faculty Research Development Grant.

Acknowledgments

The author wishes to thank Jon Chun for his work in creating AI-LIT, an open-source repo on Github that was used for the text processing and visualizations in this paper.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Proust, M. (2003). *In Search of Lost Time. I, The Way by Swann's*. Trans. Davis. London: Penguin.
- Proust, M. (2012a). *In Search of Lost Time, Volume I*. Trans. Moncrieff and Kilmartin, Revised by Enright. Penguin Random House.
- Proust, M. (2012b). *Remembrance of Things Past, Volume I*. Trans. Scott Moncrieff. Vintage.
- Sellam, T., Das, D., and Parikh, A. P. (2020). "BLEURT: learning robust metrics for text generation," in *Annual Meeting of the Association for Computational Linguistics*. doi: 10.18653/v1/2020.acl-main.704
- Strain, F. (2022). *The trials of translation: a cross-linguistic survey of sentiment analysis on Kafka's the trial*. Available at: https://digital.kenyon.edu/cgi/viewcontent.cgi?article=1056&context=dh_iphs_prog (accessed August 2, 2024).
- Venuti, L. (2012). *The Translator's Invisibility: A History of Translation*. London: Routledge. doi: 10.4324/9780203553190