
Journal of English Literature and Cultural Studies
jelcsjournal 2025;6(1):45-53

Original Research
doi: 10.26655/JELCS.2025.06.05

A Comparative Study of the Performance of Human and Machine translation in Conveying the Message in "The Remains of the Days"

Sadighi, Shiva[1]
Falahati Qadimi Fumani, Mohammad Reza[2]
Mahboudi, Hamidreza[3]
Hosseininezhad, Seyed Mohammad [4]

[1]Assistant Professor of ELT, Department of English, Faculty of Humanities, Zand Institute of Higher Education, shiraz, *Iran*.

[2]Associate Professor, Computational Linguistics Research Department, Islamic World Science and Technology Monitoring and Citation Institute (ISC), Shiraz, *Iran*.

[3] Assistant Professor of ELT, Department of English, Faculty of Humanities, Zand Institute of Higher Education, Shiraz, *Iran*.

[4] M.A. in Translation Studies, Zand Institute of Higher Education, shiraz, *Iran*.

Corresponding Author: Falahati Qadimi Fumani, Mohammad Reza
e-mail: mrfalahati81@gmail.com

Article citation: Sadighi, Sh., Falahati Qadimi Fumani, M. R., Mahboudi, H., & Hosseininezhad, M. (2025). A Comparative Study of the Performance of Human and Machine Translation in Conveying the Message in "The Remains of the Days", *Journal of English Literature and Cultural Studies*, 6(1): 45-53.

Received Date: December26, 2024
Accepted Date: February 7, 2025
Online Date: March 19, 2025

Publisher: Kare Publishing

E-ISSN: 2501-1183

ABSTRACT

Comparison of human and machine translation – especially in terms of the preservation of contextual meaning – has already attracted the attention of many researchers. Accordingly, the present study intended to compare human (Daryabandari) and machine (Google Translate) translations of The Remains of the Day into Persian to describe translation error types in conveying contextual meaning. In all, 120 errors were extracted using MQM model from each translation. The study identified four main categories of errors: lexical (35%), semantic (30%), syntactic (23%), and pragmatic (12%). These findings highlighted the limitations of current MT systems in handling the nuances of language, cultural references, and literary style present in sophisticated narratives. The presence of syntactic errors (23%) further demonstrated the difficulty machines face in correctly interpreting and reconstructing grammatical structures across languages. Pragmatic errors, while less frequent (12%), underscored the machine's inability to grasp and appropriately translate cultural nuances and context-dependent meanings. Further, the Chi-square test result ($\chi^2 = 81.4$, df= 3, P= .000) revealed a statistically significant difference between the translation type and error categories. This implied that human and machine translations exhibited different error patterns. The findings of this study suggested that translators and translation students should incorporate both traditional human translation techniques and machine translation technologies into their curricula.

Keywords: Machine Translation; Human Translation; Translation Error; Conveying the Meaning; Literary Text.

Introduction

With the rapid advancements in Machine Translation (MT) technologies, particularly with the development of Neural Machine Translation (NMT) systems, there has been growing interest in the efficacy of machines in translating complex literary works (Lopez, 2008).

The increasing reliance on MT systems for quick, automated translation raises questions about whether machines can truly replace human translators in fields where deep cultural understanding, emotional sensitivity, and literary style are paramount (Bowker & Fisher, 2020). Translation, especially of literary works, involves much more than a word-for-word conversion; it requires an interpreter who can appreciate the emotional undertones and cultural subtleties that are often lost in the algorithms of machine translation (Nida & Taber, 1969).

While MT offers undeniable advantages in terms of speed and consistency, it clearly has its own limitations.

Similarly, human translation is not without problems. Errors stemming from subjectivity, misinterpretation, or inconsistency can sometimes distort the original meaning of a text, affecting its emotional and cultural impact (Pym, 2010).

This study intends to identify the translation errors in both human and machine translations of The Remains of the Day, with particular focus on meaning transfer. Using the Multidimensional Quality Metrics (MQM) model, this study compares how effectively each method conveys the novel's nuanced themes, emotional tones, and cultural references. Therefore, these research questions were introduced:

1. What are the main translation error types in the machine translation of The Remains of the Day?
2. What are the main translation error types in the human translation of The Remains of the Day?
3. Is there a significant difference between the errors found in the machine and human translations?

In an increasingly globalized world, where cross-cultural communication is essential, understanding the comparative strengths of human and machine translation is crucial. The present study is particularly relevant as it explores how each method handles the translation of nuanced literary works that rely on emotional subtlety and cultural context. This research contributes to ongoing discussions about the future of translation in the digital age, where machine learning and artificial intelligence are playing a more significant role in traditionally human-dominated fields (Doherty, 2016). By examining the translation of The Remains of the Day, this study provides a critical assessment of how well both human and machine methods perform in preserving literary integrity and whether hybrid approaches may offer a balanced solution.

Literature Review

In their study, Aslerasouli and Abbasian (2015) compared the efficiency of Google Translation (GT) and Human Translation (HT) in relation to Soft and Hard Science texts. Employing Waddington's (2001) model of Translation Quality Assessment (TQA), quality of five translation pieces of Soft and Hard Science texts translated both by GT and HT was assessed in an effort to test three research hypotheses. The findings revealed that HT's quality was still higher than that of the GT; however, GT's efficiency was to some extent comparable to that of the HT as far as the texts in Physics and Politics were concerned. So, it was concluded that in certain fields, GT could produce roughly reliable translation if not totally as efficient as Human Translator.

By the same vein, Vasheghani Farahani (2020) conducted a comparative study to analyze and compare translation adequacy in machine vs. human translation from English into Persian. In other words, this study sought to investigate whether machine translation product was adequate enough as compared to human translation. For this purpose, seven various texts were selected and translated from Persian into English by two translation students as well as machine translation services including Google Translate and Bing Translation. In addition, a reference translation by two expert translators was prepared against which human and machine translations were compared and analyzed. Moreover, the Sketch Engine software was exploited to analyze the translations. The results of the study revealed that there was no statistically significant difference between human and machine translations as compared to the reference translation and that machine translations were adequately appropriate translations.

Abu Saaleek (2016) ran a comparative study on the adequacy and acceptability of machine translation in translating Islamic texts. To do so, he selected four different machine translations including World lingo, Babylon translation, Google Translate, and Bing. His corpus of the study included five Islamic text-genres covered supplication, part of a sermon, pillars of Islam, Hadith

(sayings of the prophet of Islam) from Arabic into English. The findings of this study showed that from among the four various MTs, Google Translate functioned better in terms of adequacy and acceptability. The results also indicated that Google Translate had better output in terms of (accuracy, suitability, and well-formedness) and sub-characteristics (syntax, terminology, reliability and fidelity).

Farahsani, Rini and Jaya (2020) found the translation accuracy of Google Translate in English while translating Indonesian, especially technical terms into engineering terms. The results of their descriptive qualitative survey showed that when translating Indonesian, especially technical terms into English engineering terms, Google Translate was accurate for word translations, but not for phrases. This statement is based on the exact appearance rate results of Google Translate. This was 70.73% at the word level and 46.37% at the phrase level.

Methodology

Data collection

This study aimed to compare the machine and human translations of the book The Remains of the Day. This book was selected because it was one of the most famous short stories in the world. Further, its author, Kazuo Ishiguro, was a nobble prize winner in literature. As the data of the study, the original English text of *The Remains of the Day Kazuo Ishiguro* was obtained through the Web. Further, as a human translation, the existing and well-known Farsi translation of the book by Najaf Daryabandari was sourced since it was the only Farsi translation of the book. Google Translate was used as the machine translation tool to produce a Farsi translation of the entire text. Google Translate was selected because it is one of the most popular machine translation engines used consistently by people for translation tasks.

The original book had 245 pages and 50 pages (20% of the book) were chosen as the data of the study. Based on *The Handbook of Corpus Linguistics and Natural Language Processing* ([HCLNLP], 2010), the minimum requirement for sampling is 10% of the main source. Simple randomization was avoided since there was a need for the sentences to be parallel.

The final dataset consisted of the aligned passages from the original English text, the human translation, and the machine translation, totaling 240 cases. This dataset was then used to undertake the data analysis. The MQM model was employed to assess and categorize the translation errors in these two types of translation.

Instruments

Google Translate is a free multilingual neural machine translation service developed by Google. It can translate texts, documents, and websites between over 100 languages. It uses neural machine translation technology, which allows it to translate whole sentences at a time rather than piece by piece. It can be accessed via website, mobile apps, or through an API for developers. It offers features like real-time camera translation, conversation mode for bilingual chats, and the ability to save translations for offline use. While generally quite accurate, it may struggle with nuanced language idioms, or highly technical content. Google continually improves the service using machine learning techniques and user feedback. It is widely used for quick translations but is not considered a substitute for professional human translation for critical or official documents.

Framework

The primary focus of this study was to compare the accuracy of human and machine translations in conveying the original message of the novel. This involved measuring translation errors and analyzing differences in error rates between the two translation methods. The MQM model was used to assess various aspects of translation quality, including linguistic accuracy, fidelity to the source text, stylistic appropriateness, and compliance with client requirements (Lommel et al., 2014). This multifaceted evaluation is particularly crucial for literary works like *The Remains of the Day*, where the artistic style, thematic richness, and the author's original voice are as significant as the factual correctness of the translation (Drugan, 2013). The MQM model's flexibility and adaptability make it

an ideal choice for this study, as it can be tailored to address the unique challenges of translating nuanced literary texts and ensuring that all relevant quality dimensions are accounted for (Castilho et al., 2014). By employing the MQM framework, the study can deliver a more holistic and in-depth analysis of translation quality, setting it apart from research that relies on more limited or one-dimensional evaluation criteria.

Based on the MQM model, errors could be classified into the following categories:

1- Lexical Errors: These errors involve incorrect word choice or mistranslations at the word level. In MQM, lexical errors can be categorized under "Terminology", and "Lexical Consistency." For instance, if a character's name or a specific term is translated inconsistently or incorrectly, it would fall under this category.

Original English: "Mr. Stevens embarked on his journey."

MT (Farsi): "آقای اســتیونز سفر خود را سوار شد "

HT: "آقای استیونز سفر خود را آغاز کرد."

Explanation: The MT incorrectly translated "embarked" as "سوار شد", which means "getting on" instead of the appropriate term for "embarked" in Farsi, which is "started".

2- Syntactic Errors: These errors relate to the structure of sentences and the rules of grammar. In MQM, they are often classified under "Grammar" and "Syntax." For example, if a sentence in the translated text is structurally incorrect or follows Farsi syntax incorrectly, it would be noted as a syntactic error.

Original English: "The butler meticulously performed his duties."

MT (Farsi): "پیشخدمت وظایف خود را با دقت انجام می‌داد."

HT: "پیشخدمت وظایف خود را با دقت انجام داد"

Explanation: The MT failed to translate the verb tense correctly, and hence the grammatical problem.

3- Semantic Errors: These errors involve misunderstandings or mistranslations that alter the meaning of the text. Within MQM, semantic errors can be categorized under "Mistranslation" and "Accuracy." If a passage in the translation changes the original meaning or fails to convey the intended sense of the original text, it would be considered a semantic error.

Original English: "The subtle nuances of the conversation were lost on him."

MT (Farsi): "او جزئیات فراوان مکالمه را از دست داد"

Human Translation: "نکات ظریف مکالمه از نظر او پنهان مانده بود."

Explanation: The MT failed to capture the intended meaning of "subtle nuances" and translated it as "abundant details," altering the original sense of the sentence.

4- Pragmatic Errors: These errors relate to the contextual or cultural appropriateness of the translation. In MQM, they are categorized under "Register" and "Internationalization/Localization." For instance, if a cultural reference is not adapted appropriately for the Farsi audience or if the tone and style are not suitable for the target readership, it would be noted as a pragmatic error.

Original English: "His humor fell flat in the formal setting."

MT (Farsi): "شوخی او در تنظیم رسمی خنثی شد"

HT: "شوخ طبعی او در محیط رسمی بیروح بود"

Explanation: The MT mistranslated "fell flat" as "خنثی شد" which does not convey the appropriate cultural or contextual implications of the original phrase.

Data analysis

To answer the first two research questions, "What are the main translation errors that occur in machine translation of the book *The Remains of the Day* in the context and meaning transfer?", and "What are the main translation errors that occur in human translation of the book *The Remains of the Day* in the context and meaning transfer?", the MQM framework, along with descriptive statistics, was used to categorize types of errors (lexical, syntactic, semantic, and pragmatic) for the MT and the HT. The frequency of each error type was counted and reported.

To answer the third research question, "Are there any significant differences between the errors of human and machine Farsi translations of the book The Remains of the Day based on MQM framework?", Chi-square test was conducted in SPSS Version 26 to determine whether there was a significant difference in the error rates between HT and the MT.

After analysis of the collected data from the MT performance, the following results were found:

1. Lexical errors, accounting for 35% of the total errors, emerged as the most prevalent type. These errors reflect the MT's struggles in accurately translating individual words and phrases, often resulting in mistranslations or inappropriate word choices.

2. Syntactic errors, responsible for 23% of the errors, indicate the MT's difficulty in correctly processing and translating the grammatical structures of English sentences into Farsi.

3. Semantic errors, accounting for 30% of the errors, demonstrate the MT's inability to grasp the deeper meaning and context of the source text.

4. Pragmatic errors, constituting 12% of the errors, highlight the MT's inability to grasp the social and cultural context of the novel.

Table 1. Error Distribution in the MT.

Error Category	Number of Errors	Percentage of Total Errors
Lexical Errors	42	35%
Syntactic Errors	28	23%
Semantic Errors	36	30%
Pragmatic Errors	14	12%
Total	120	100%

Analysis of the collected data from the HT performance revealed that:

1. Lexical and pragmatic errors, each accounting for 21% of the errors, highlight the HT's occasional struggles with choosing the most appropriate words and adapting cultural references for the target audience.

2. Syntactic errors, comprising only 10% of the errors, indicate that the HT generally maintains a high level of grammatical accuracy in the translation.

3. Semantic errors, accounting for 48% of the errors, demonstrate the HT's inability to grasp the deeper meaning and context of the source text.

Table 2. Error Distribution in the HT.

Error Category	Number of Errors	Percentage of Total Errors
Lexical Errors	25	21%
Syntactic Errors	12	10%
Semantic Errors	58	48%
Pragmatic Errors	25	21%
Total	120	100%

The Chi-square test was used to check whether there was a significant difference in the errors that HT and MT made in each category. As indicated in Table 5, the result of the Chi-square test (χ^2 = 81.4, df= 3, P= .000) revealed a statistically significant difference between the translation type and error categories. This implies that human and machine translations exhibit different error patterns, which may have important implications for improving translation quality and understanding the translation. Based on the analysis of error types and the results obtained from the Chi-Square test, a significant difference was found between the errors of the machine and human translations of The Remains of the Day into Persian. Based on the available data, MT demonstrates a higher frequency of lexical errors (35%) and syntactic errors (23%), while HT exhibits more semantic errors (48%). Pragmatic errors occur at similar rates in both translation types (12% in MT and 21% in HT).

Table 3. Comparative Error Analysis of Machine and Human Translations

Error Category	Machine Translation (Frequency/ Percentage)	Human Translation (Frequency/ Percentage)
Lexical Errors	42 / 35%	25 / 21%
Syntactic Errors	28 / 23%	12 / 10 %
Semantic Errors	36 / 30%	58 / 48%
Pragmatic Errors	14 / 12%	25 / 21%
Total	120 / 100%	120 / 100%

Table 4. Cross Tabulation Table

Error Type	MT			HT			Total
	Observed	Expected	Total	Observed	Expected	Total	
Lexical	42	33	75	25	42	67	142
Syntactic	28	12	40	12	28	40	80
Semantic	36	58	94	58	36	94	188
Pragmatic	14	17	31	25	14	39	70
Total	120	120	240	120	120	240	480

Table 5. Chi-square Test

Test Statistics	Value	df	Asymp. Sig. (2-sided)
	81.4	3	0.000
Likelihood Ratio	84.1	3	0.000
Linear-by-Linear Association	76.8	1	0.000
No of Valid Cases	480		

Discussion and Conclusion

The analysis of the MT errors in Kazuo Ishiguro's The Remains of the Day reveals significant challenges in accurately translating complex literary works. The study identified four main categories of errors: lexical (35%), semantic (30%), syntactic (23%), and pragmatic (12%). These findings highlight the limitations of current MT systems in handling the nuances of language, cultural references, and literary style present in sophisticated narratives.

The prevalence of lexical and semantic errors (totally 65%) suggests that MT systems struggle most with accurately conveying the meaning and context of individual words and phrases within a literary context. This indicates a fundamental challenge in translating works that rely heavily on nuanced language and complex themes. The presence of syntactic errors (23%) further demonstrates the difficulty machines face in correctly interpreting and reconstructing grammatical structures across languages. Pragmatic errors, while less frequent (12%), underscore the machine's inability to grasp and appropriately translate cultural nuances and context-dependent meanings.

Lexical: Pym (2018), for instance, found that lexical errors were more prevalent in his analysis of literary translations, contrasting with our findings where semantic errors dominated. This difference might be attributed to the specific nature of *The Remains of the Day* and its heavy reliance on subtle cultural and historical contexts.

These results imply that while MT can provide a basic understanding of the text, it falls short in capturing the depth, subtlety, and artistic quality of literary works. The high percentage of errors across all categories suggests that human intervention remains crucial in the translation of complex literature.

The findings of this study align with previous research on machine translation of literature. For instance, Toral and Way (2018) found that neural MT systems struggled with literary texts, particularly in preserving style and capturing context-dependent meanings. Their study reported error rates similar to those found in our analysis, supporting the notion that machine translation faces significant challenges in literary contexts.

However, some studies have shown more promising results in specific areas. Moorkens et al. (2018) reported improvements in machine translation of certain types of figurative language, suggesting that advancements are being made in handling some aspects of literary translation. This contrasts somewhat with our findings, which show persistent difficulties across all error categories. The high prevalence of lexical and semantic errors in our study is consistent with research by Besacier and Schwartz (2015), who identified vocabulary and meaning preservation as major challenges in machine translation of literature. They argued that these issues stem from the machine's inability to understand context and cultural references, which aligns with our observations.

Our findings on pragmatic errors are particularly significant in the context of literary translation. This aligns with research by Sánchez-Gijón and Torres-Hostench (2014), who emphasized the importance of cultural adaptation in literary translation and the current limitations of MT in this area. Bassnett (2013), also emphasizes the importance of cultural adaptation in literary translation. Bassnett argues that the translator's role extends beyond linguistic transfer to cultural mediation, which aligns with our observations on the challenges of adapting cultural references.

Semantic: Nord (2018) stresses the importance of understanding the source text's intended

meaning and function, which correlates with the high semantic error rate in our study.Munday (2016) discusses the challenges translators face in interpreting and conveying the intended meaning in literary texts, particularly when dealing with culturally specific concepts. This supports our observation of high semantic error rates. However, some studies have shown varying results in error distribution.

The syntactic errors observed in our study are consistent with findings by Burchardt et al. (2017), who noted that even advanced neural MT systems struggle with complex sentence structures and long-range dependencies, which are common in literary texts.

In conclusion, while MT continues to improve, our results indicate that it still faces significant challenges in accurately translating complex literary works. The nuanced nature of literature, with its reliance on context, cultural understanding, and stylistic elements, remains a formidable challenge for automated translation systems.

The relatively low rate of syntactic errors in our study is consistent with findings by Chesterman (2016), who notes that professional translators generally have a good grasp of grammatical structures in both source and target languages. However, Chesterman also points out that even experienced translators can struggle with complex sentence structures, which aligns with the syntactic errors we observed. The analysis of human translation errors in The Remains of the Day underscores the complexity of literary translation. While human translators surpass machines in understanding and conveying cultural nuances, they are not infallible, particularly in areas requiring deep cultural immersion and interpretation of subtle literary devices.

This study reaffirms the need for translators to engage deeply with both source and target cultures, and perhaps for readers to approach translated literature with an understanding of these inherent challenges. It also highlights areas where even human translators could benefit from further training or tools to aid in capturing the full essence of literary works.

Interestingly, our results show a different pattern compared to MT errors. While machines struggle more with lexical and syntactic aspects, human translators in our study faced greater challenges with semantics and pragmatics.

Toral and Way (2018) found that neural MT systems could produce translations of literary texts that were sometimes indistinguishable from HTs in blind evaluations.

At the end, it is important to note that our study focused on a single, highly complex literary work. The results may not be generalizable to all types of literary translations. As Besacier (2014) points out, the effectiveness of MT can vary significantly depending on the specific language pair and text genre. In conclusion, our results suggest that MT has made significant strides in literary translation. The differing error patterns between machine and human translations suggest that a hybrid approach, combining the strengths of both methods, might be most effective for translating complex literary works.

References

Abu Saaleek, A. O. (2016). The adequacy and acceptability of machine translation in translating the Islamic texts. Innovational Journal of English Linguistics, 6(3), 1-85.https://doi.org/10.5539/ijel.v6n3p185.

Aslerasouli, P., & Abbasian, G. R. (2015). Comparison of google online translation and human translation with regard to soft vs. hard science texts. Journal of Applied Linguistics and Language Research, 2(3), 169-184.

Bassnett, S. (2013). Translation studies (4th ed.). Routledge.

Besacier, L. (2014). Machine translation for literature: A pilot study. In Proceedings of TALN 2014 (Volume 2: Short Papers).

Besacier, L., & Schwwartz, L. (2015). Automated translation of literary work: A pilot study. In Proceedings of the 4th Workshop on Computational Linguistics for Literature (pp. 114-122). Association for Computational Linguistics.

Bowker, G. C., & Fisher, S. G. (2020). The role of MT in the literary translation process. The Journal of Specialized Translation, 32, 9-26.

Burchardt, N., Schutze, H., & Stede, M. (2017). Neural machine translation in the wild: Evaluation of Google translate. In Proceedings of the 55th Annual Meeting of Association for Computational Linguistics (ACL), 1, 914-924.

Castilho, S, O'Brien, S., Alves, F., & O'Brien, M. (2014). Does post-editing increase usability? A study with Brazilian Portuguese as target language. In Proceedings of the 17th Annual Conference of the European Association for Machine Translation. 16-18 June 2014, Dubrovnik, pp. 183-190.

Chesterman, A. (2016). Memes of translation: The spread of ideas in translation theory (revised edition), Vol.123. Amsterdam: Benjamins Publishing Company.

Doherty, S. (2016). The impact of translation technologies on the process and product of translation. International Journal of Communication, 10, 947-969.

Drugan, J. (2013). Quality in professional translation: Assessment and improvement. Bloomsbury, London.

Farahsani, Y., Rini, I. P., & Jaya, P. H. (2020). Google translate accuracy in translating specialized language from English to Bahasa Indonesia: A case study on mechanical engineering terminology. In Proceedings of the 4th International Conference on Sustainable Innovation 2020 – Social, Humanity, and Education (ICoSIHESS 2020). Atlantis Press.

Lommel, A., Uszkoreit, H., Burchardt, A. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. Revista Tradumatica 12, 455-463.

Lopez, A. (2008). Translating literature in the age of machine translation. New Voices in Translation Studies, 5(1), 33-47.

Moorkens, J., Toral, A., Castilho, S., & Way, A. (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. Translation Spaces, 7(2), 240-262.

Munday, J. (2016). Introducing translation studies. Routledge.

Nida, E. A., & Taber, C. R. (1969). The theory and practice of translation. Brill.

Nord, C. (2018). Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis (2nd ed.). Rodopi.

Pym, A. (2018). Fault types in translation: Lexical vs. grammatical errors. Translation Spaces, 2(2), 166-187.

Pym, A. (2010). Can human translation be objective? The Translator, 16(2), 145-161.

The handbook of corpus linguistics and natural language processing (2010) (Clark, A., Fox, Ch., & Lappin, Sh. Ed.). Wiley Blackwell.

Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text? arXiv preprint arXiv:1801.04962.

Sanchez-Gijon, E., & Torres-ostench, M. (2014). Literary translation and cultural adaptation: The need for human intervention in audiovisual translation. The Translator, 20(2), 207-226.

Vasheghani Farahani, M. (2020). Adequacy in machines vs. human translation: A comparative study of translation from English into Persian. Iranian Journal of Translation Studies, 3(6), 51-78.

Waddington, C. (2001). Different methods of evaluating students' translations: The question of validity. Retrieved Jan. 12, 2026 from http://id.erudit.org/iderudit/004583a.