Jurnal Pendidikan dan Sastra Inggris
Volume. 6 Nomor. 1 April 2026
e-ISSN : 2827-8860; p-ISSN : 2827-8852, Hal. 304-318
DOI: https://doi.org/10.55606/jupensi.v6i1.6924
Tersedia: https://journalshub.org/index.php/JUPENSI

Comparing Neural Machine Translation and Human Translation of Dickinson's *Because I Could Not Stop for Death*

Supriyono[1*]
[1]English Literature Study Program, Translation Studies, Universitas Terbuka, Indonesia
**Author Correspondence: priyo.ckp@gmail.com*

Abstract. The development of neural machine translation (NMT) as a branch of artificial intelligence has significantly influenced translation practices, including literary translation. However, the capacity of NMT systems to preserve the aesthetic and stylistic dimensions of poetry remains contested. This study compares the performance of DeepL and a human translation in rendering Emily Dickinson's poem Because I Could Not Stop for Death into Indonesian. The study employs a qualitative line-by-line comparative analysis focusing on equivalence, diction, shifts, and stylistic effects. The machine translation was generated through DeepL using default settings without post-editing. The findings reveal that DeepL demonstrates strong lexical accuracy and structural consistency, but tends to produce literal renderings that reduce metaphorical depth and aesthetic nuance. In contrast, the human translation shows greater flexibility in handling figurative language, rhythm, and emotional tone, resulting in a more poetic and communicative target text. These findings indicate that while DeepL can function as an initial assistive tool in poetry translation, human interpretive intervention remains essential to achieve artistic and stylistic adequacy.

Keywords: DeepL; Human Translation; Neural Machine Translation; Poetry Translation; Stylistic Analysis.

1. INTRODUCTION

Artificial Intelligence (AI), particularly through Neural Machine Translation (NMT), has significantly transformed translation studies by enabling context-sensitive automated translation. Recent research highlights the growing integration of AI systems in academic and professional translation contexts (Ahmed et al., 2025; Dolmaci, 2025). Although neural systems such as DeepL demonstrate high levels of lexical and syntactic accuracy, several studies note that challenges persist when translating stylistically complex texts, especially literary works (Guerberof-Arenas & Toral, 2022; Karabayeva & Kalizhanova, 2023).

Beyond individual system evaluations, broader discussions have examined the structural transformation of translation practice in the age of artificial intelligence. Bellos and Sennrich (2019) argue that machine translation development should be understood not as a replacement of human translators but as a technological evolution that reshapes translational workflows and competence. In the Indonesian academic context, Siregar et al (2023) map the landscape of machine translation error research and highlight persistent challenges in semantic ambiguity and contextual interpretation. Similarly, Kembaren et al (2023). note that while machine translation demonstrates increasing technical sophistication, qualitative dimensions such as cultural nuance and stylistic sensitivity remain critical areas of concern.

In literary translation, the primary difficulty lies not only in transferring semantic meaning but also in reconstructing figurative language, symbolism, and emotional resonance. Empirical findings indicate that while AI systems often achieve grammatical fluency, they may

Naskah Masuk: 09 Desember 2025; Revisi: 11 Januari 2026; Diterima: 17 Februari 2026;
Tersedia: 06 Maret 2026

reduce metaphorical subtlety and interpretative depth (Castilho et al., 2018; Jassim, 2025; Moorkens, 2018). These limitations become particularly visible in poetry, where meaning is condensed and stylistic devices function integrally within the text.

Despite the growing body of research on AI-assisted translation, studies specifically examining the translation of English poetry into Indonesian remain limited. Moreover, existing investigations tend to focus on isolated aspects such as lexical accuracy or general equivalence without integrating equivalence theory, translation techniques, shift analysis, and stylistic criticism within a unified analytical framework. Within the Indonesian context, research on AI-assisted literary translation is still developing, and comprehensive evaluations of stylistic performance are scarce (Pratiwi & Gusthini, 2025; Yuliasri, 2025).

Emily Dickinson's *Because I Could Not Stop for Death* provides a relevant case for such examination due to its compact structure, layered metaphors, and symbolic density. Therefore, this study aims to analyze DeepL's translation of Dickinson's poem in comparison with a human translation by examining equivalence, translation techniques, shifts, and stylistic effects. By combining quantitative and qualitative approaches, the study seeks to provide a comprehensive evaluation of AI performance in literary translation and to clarify the extent to which machine systems approximate human interpretative competence.

2. LITERATURE REVIEW

Translation studies generally focus on the issue of equivalence as the core of transferring meaning from the source language to the target language. Eugene Nida (1964) distinguishes between formal equivalence and dynamic equivalence. Formal equivalence emphasizes fidelity to the form and structure of the source text, whereas dynamic equivalence prioritizes the response of the target readers. In poetry translation, dynamic equivalence is particularly relevant, as success is measured not only by lexical accuracy but also by the ability to recreate an equivalent aesthetic experience.

Peter Newmark (1988) differentiates between semantic translation and communicative translation. Semantic translation attempts to preserve contextual and cultural nuances of the source text, while communicative translation prioritizes clarity and acceptability for the target audience. Literary translators often negotiate between these two approaches. Neural machine translation systems such as DeepL tend to produce outputs closer to semantic translation by maintaining structural and lexical correspondences, whereas human translators have greater flexibility to apply communicative strategies to preserve artistic effects.

The concept of translation shifts proposed by John C. Catford (1965) is also fundamental to this study. Shifts occur when formal correspondence between source and target structures is absent, requiring changes at the linguistic level or category. In poetry translation, shifts frequently appear in syntax, word class, and metaphorical structure. Analyzing these shifts enables identification of systematic differences between DeepL and human translations, particularly in handling figurative and symbolic language.

Methodologically, Lucía Molina and Amparo Hurtado Albir (2002) propose a classification of translation techniques, including literal translation, modulation, transposition, amplification, reduction, and adaptation. This framework is employed to identify strategies in both translation types. Neural machine translation systems like DeepL often rely on literal and pattern-based transposition techniques, while human translators are more likely to apply modulation and adaptation to maintain stylistic effects.

In addition to translation theory, this research adopts a stylistic approach. Geoffrey Leech and Mick Short (2007) explain that stylistics examines the relationship between linguistic form and artistic function in literary texts. Elements such as metaphor, personification, imagery, diction, tone, and rhythm construct poetic meaning. In this study, stylistic analysis evaluates the extent to which DeepL represents not only denotative meaning but also connotative meaning and emotional effect in the poetry of Emily Dickinson.

Recent developments in neural machine translation (NMT) demonstrate that systems such as DeepL employ deep learning models based on artificial neural networks to predict word sequences in the target language using global contextual information. This approach improves fluency and coherence compared to earlier rule-based or statistical machine translation systems. However, studies indicate that despite lexical and syntactic improvements, NMT systems still face limitations in handling implicit meaning, metaphorical ambiguity, and symbolic density in literary texts.

Comparative investigations between AI and human translation further reveal nuanced distinctions in stylistic handling. Castilho et al. (2018) demonstrate that although neural machine translation improves fluency compared to phrase-based systems, professional translators consistently outperform NMT in stylistic refinement. Moorkens (2018) similarly notes that neural outputs often appear grammatically coherent but may lack pragmatic sensitivity. In the domain of literary translation, Karabayeva and Kalizhanova (2023) emphasize that rhetorical and figurative elements remain particularly challenging for machine systems. Likewise, Jassim (2025) observes that AI-generated poetry translations tend to preserve surface meaning while reducing emotional intensity and metaphorical subtlety. These

studies collectively reinforce the argument that stylistic and interpretative dimensions remain critical benchmarks in evaluating AI-based literary translation.

Comparative studies in different linguistic and cultural settings further confirm these tendencies. Anuar and Sabri (2025) examining AI and human translations in Malaysian literary texts, report that AI systems tend to prioritize structural correspondence, whereas human translators demonstrate greater interpretative flexibility. Batubara et al. (2025) in their systematic review of artificial intelligence in translation, identify literary translation as one of the most complex domains for AI application due to its high density of figurative and contextual meaning.

Contemporary research comparing AI and human translation often focuses on academic or informative texts, while studies on poetry remain limited and rarely integrate systematic stylistic analysis. Furthermore, the English–Indonesian language pair has received relatively little attention in evaluating DeepL's performance in poetic texts. Therefore, this study situates itself at the intersection of equivalence theory, translation techniques, shift analysis, and stylistics to provide a comprehensive evaluation of DeepL's performance in translating poetry into Indonesian.

By integrating these theoretical frameworks, this research not only compares semantic accuracy but also examines how poetic effects, metaphor, imagery, and tone are realized in translation. This multidimensional approach strengthens the analysis and contributes more specifically to AI-based literary translation studies within the Indonesian context.

Within the Indonesian context, recent studies have begun to explore AI-assisted literary translation. Pratiwi and Gusthini (2025) compare ChatGPT and human translations of Dickinson's poetry, concluding that AI demonstrates lexical adequacy but struggles with poetic tone and symbolic resonance. Yuliasri (2025) further report that student translations assisted by AI tools show improved structural accuracy yet require substantial revision to achieve stylistic naturalness. These findings indicate that while AI tools are increasingly integrated into translation education in Indonesia, their role remains complementary rather than substitutive in literary translation practice.

3. RESEARCH METHOD

This study employs a descriptive qualitative approach with a comparative design to examine differences between machine translation and human translation of a poetic text. The research focuses on analyzing equivalence of meaning, diction choices, structural shifts,

translation techniques, and stylistic realization in the translations of the poem *Because I Could Not Stop for Death* by Emily Dickinson.

The study does not involve reader-response testing; therefore, the stylistic evaluation is conducted through analytical and descriptive textual analysis.

Data and Data Sources

The data of this study consist of three types of texts forming a comparative corpus.

The source text is taken from Dickinson's *Because I Could Not Stop for Death*, included in the editorial edition Selected Poems published by Gramedia. The version in Selected Poems is an editorial edition that has undergone standardization of punctuation, capitalization, and layout by modern editors; thus, it is not a diplomatic reproduction of Dickinson's original manuscript. Accordingly, the source text in this study represents a modern published version commonly used in educational contexts and general readership.

The use of this editorial version carries textual implications, as Dickinson's poems are known for their variations in punctuation and capitalization in the original manuscripts. However, since the focus of this study is the comparison of translation strategies (machine and human) based on a widely circulated printed text, the use of an editorial edition is considered methodologically consistent with the research objectives.

The machine translation was generated using the platform DeepL (https://www.deepl.com) with the English–Indonesian language pair under default settings, without any post-editing. The poem was input in its entirety according to the version in Selected Poems, and the output was collected in its original form on February 10, 2026. This version is treated as the machine translation data (TT1). Since DeepL is continuously updated, the findings are limited to the version accessed on February 10, 2026 (DeepL, 2026).

The human translation refers to the version published in BSE Bahasa Indonesia XII untuk SMA dan MA Program Bahasa by Nurita Bayu Kusmayati and Eka Trianingsih, issued by Mediatama in 2009. This version is treated as the human translation data (TT2).

This study limits the analysis to 20 lines contained in the Selected Poems version used as the source text and that have corresponding equivalents in the human translation found in the BSE textbook. This limitation is intended to maintain corpus equivalence and consistency in the comparative analysis.

The unit of analysis in this study is the poetic line as the primary unit, with particular attention to metaphorical phrases, diction choices, and syntactic structures containing stylistic elements such as personification, imagery, symbolism, and tone. The analysis is conducted at

the line level to identify translation techniques and to compare strategic tendencies between machine translation and human translation.

Data Collection

Data were collected using the documentation method. The source text and the two translated versions were aligned in a line-by-line comparative table to facilitate the identification of differences. Each line was numbered and categorized according to the stylistic elements it contains, such as metaphor, personification, visual imagery, and temporal expressions.

Analytical Framework and Procedure

The analysis was conducted through the following systematic stages:

Identification of Meaning Equivalence

Comparing lexical and contextual meaning between the source text, DeepL translation, and human translation based on the concepts of formal and dynamic equivalence.

Identification of Translation Techniques

Classifying the translation techniques applied using the framework proposed by Molina and Albir including literal translation, modulation, transposition, adaptation, amplification, and reduction (Molina & Albir, 2002).

Analysis of Translation Shifts

Identifying structural and linguistic category shifts based on Catford's theory (Catford, 1965).

Stylistic Analysis

Evaluating how metaphor, imagery, tone, and rhythm are realized in the target language using a stylistic approach.

Comparative Interpretation

Interpreting the differences between the DeepL translation and the human translation to determine their respective translational tendencies.

To enhance analytical rigor, the frequency of translation techniques was recorded and classified to identify dominant patterns in each translation version. This quantification was not intended as a full statistical measurement but rather as supportive evidence for qualitative interpretation.

Data Validity

Data validity was ensured through theoretical triangulation by integrating the frameworks of Nida, Newmark, Catford, Molina and Albir, and stylistics in the analysis. Interpretations

were conducted consistently with reference to the overall context of the poem to maintain analytical coherence.

The classification of translation techniques was carried out by the researcher based on the operational definitions proposed by Molina and Albir (Molina & Albir, 2002). Each line was examined to identify the dominant technique applied. Although inter-rater validation was not employed, consistency was maintained through the systematic application of established theoretical criteria.

This study does not aim to generalize all AI-based translation systems; rather, it evaluates DeepL as a representation of neural machine translation in translating English poetry into Indonesian.

4. RESULTS AND DISCUSSION

General Patterns of DeepL and Human Translation

The analysis of the 20 lines of *Because I Could Not Stop for Death* reveals consistent differences between the DeepL translation and the human translation. In general, DeepL preserves the syntactic structure and lexical relations of the source text. Clause arrangement, parallelism, and diction are rendered directly, resulting in relatively stable formal equivalence. Although grammatically accurate, the translation in several lines appears less idiomatic and less capable of conveying the poem's nuanced poetic tone.

In contrast, the human translation demonstrates greater flexibility in restructuring both form and meaning. Shifts in diction, adjustments of metaphor, and expansion or condensation of phrases are employed to maintain rhythm and symbolic resonance. Rather than strictly preserving the original form, the translator seeks to reconstruct the aesthetic experience in the target language. This contrast indicates a clear distinction between a form-oriented approach and an effect-oriented approach to poetry translation.

Overall Line-by-Line Comparison

Table 1. Comparison of Poetic Line Translations in *Because I Could Not Stop for Death* by Emily Dickinson.

No.	Source Text (Gramedia)	DeepL Translation	Human Translation (Mediatama)	Brief Analysis
1.	Because I could not stop for Death	Karena aku tidak bisa berhenti untuk Kematian	Karena aku tak bisa berhenti untuk Sang Maut	DeepL closely follows the source wording; Human strengthens personification through lexical choice.
2.	He kindly stopped for me	Dia dengan baik hati berhenti untukku	Dialah yang berbaik hati dan berhenti	DeepL keeps the original structure; Human adds subject emphasis.
3.	The carriage held but just ourselves	Kereta itu hanya membawa kami berdua	Kereta yang distop adalah diri kita sendiri	DeepL maintains the clause pattern; Human reformulates the structure.
4.	And Immortality	Dan Keabadian	Dan kelanggengan	Both show lexical variation with similar meaning.
5.	We slowly drove, he knew no haste	Kami melaju perlahan, dia tidak terburu-buru	Kita gerakkan kereta kuda perlahan saja Dia mana kenal sikap tergesa	DeepL preserves the two-clause structure; Human uses a more idiomatic expression.
6.	And I had put away	Dan aku telah menyisihkan	Dan aku sisihkan semua kerja	DeepL retains the original line structure; Human adds clarification through expansion.
7.	My labor and my leisure too	Pekerjaanku dan waktu luangku pula	Serta kesenggangan masa	DeepL keeps the parallel terms; Human reduces lexical elements.
8.	For His Civility	Untuk Kesopanannya	Untuk mengimbangi budi bahasanya	DeepL renders the phrase concisely; Human provides a more elaborated expression.
9.	We passed the School where Children played	Kami melewati sekolah tempat anak-anak bermain	Kita laluilah sekolah itu yang murid-muridnya sedang bermain	DeepL translates directly; Human adds descriptive detail.
10.	Their lessons scarcely done	Pelajaran mereka baru saja selesai	Gulat di atas pentas	DeepL remains close to the source meaning; Human shows significant adaptation.
11.	We passed the Fields of Gazing Grain	Kami melewati ladang gandum yang memandang	Kita lalui padang ternak memamahbiak	DeepL maintains the original imagery; Human changes the imagery.
12.	We passed the Setting Sun	Kami melewati matahari terbenam	Kita lalui kesibukan matahari Membagi cahaya pada bumi	DeepL follows the original expression; Human expands the interpretation.
13.	We paused before a House that seemed	Kami berhenti di depan sebuah	Kita berhentilah depan rumah yang seakan	Both versions are relatively equivalent with minor structural differences.

		rumah yang tampak		
14.	A Swelling of the Ground	Seperti gundukan tanah	Sebuah pembengkakan bumi	Both preserve the metaphor with different word choices.
15.	The Roof was scarcely visible	Atapnya hampir tak terlihat	Atapnya nyaris tak begitu jelas	Both are semantically equivalent with slight idiomatic variation.
16.	The Cornice but a mound	Balok atapnya hanya gundukan	Tepinya cuma tanah yang meninggi	DeepL keeps the original form; Human adjusts lexical choice.
17.	Since then 'tis Centuries; but each	Sejak itu sudah berabad-abad; tetapi setiap	Sejak abad timbul dan tenggelam	DeepL retains the temporal contrast; Human uses a more figurative expression.
18.	Feels shorter than the Day	Rasanya lebih singkat dari hari	Namun satuannya serasa sehari saja	DeepL expresses the idea directly; Human restructures it idiomatically.
19.	I first surmised the Horses' Heads	Ketika aku pertama kali menyadari kepala kuda-kuda	Sudah kuduga bahwa kuda-kuda kereta	DeepL maintains the original inference; Human modifies the verb choice.
20.	Were toward Eternity	Menghadap ke kekekalan	Mengarahkan langkah ke titik yang abadi	DeepL renders the line concisely; Human expands the symbolic meaning.

Source texts: Dickinson (2020); DeepL (2026); Mediatama translation (2009).

Distribution and Orientation of Translation Techniques

To examine translational orientation, the techniques applied in both versions were classified according to the framework of Molina and Albir (Molina & Albir, 2002). The frequency distribution is presented in Table 2.

Table 2. Frequency Distribution of Translation Techniques.

Technique	DeepL	%	Human Translation	%
Literal	13	65%	5	25%
Modulation	2	10%	7	35%
Transposition	2	10%	3	15%
Adaptation	1	5%	2	10%
Amplification	2	10%	3	15%
Total	**20**	**100%**	**20**	**100%**

As shown in Table 2, literal translation dominates the DeepL output, accounting for 65% of the total techniques identified. This indicates a strong source-text orientation in which lexical and syntactic structures are preserved with minimal alteration. Such consistency contributes to semantic clarity and grammatical accuracy; however, it may limit stylistic flexibility, particularly in metaphorical or symbolically dense lines.

In contrast, the human translation demonstrates greater variation, with modulation emerging as the most frequent technique (35%). Adaptation and amplification are also applied more frequently than in the DeepL version. This pattern suggests a target-text orientation, where adjustments are made to maintain poetic resonance, cultural appropriateness, and emotional nuance in Indonesian.

The divergence in technique distribution reflects broader theoretical distinctions in translation studies. DeepL's reliance on literal rendering aligns with formal equivalence as proposed by Eugene Nida (1964), emphasizing structural correspondence. Meanwhile, the human translator's frequent use of modulation resonates more closely with dynamic equivalence, prioritizing equivalent effect over structural similarity.

Therefore, the distribution of techniques reveals not merely methodological differences but contrasting translational philosophies: one privileges semantic preservation through structural proximity, while the other foregrounds interpretative engagement and stylistic adaptation in poetic translation.

Analysis of Meaning Equivalence

In terms of semantic equivalence, both versions generally preserve the basic meaning of the source text, particularly in descriptive lines. However, differences in technique affect the degree of precision and scope of meaning.

DeepL tends to produce stable and direct semantic equivalence, allowing the denotative meaning to remain relatively intact. Nevertheless, in several lines, this approach results in renderings that are less idiomatic or less sensitive to implicit connotations.

In contrast, the human translation shows a tendency toward expansion and semantic shift. In a number of lines, there is semantic elaboration, strengthened connotation, or even imagery deviation. In some cases, these shifts enrich the interpretative dimension of the poem; however, in certain lines such as the change of imagery in the school journey section the deviation moves noticeably away from the original meaning.

Thus, DeepL demonstrates greater consistency at the denotative level, whereas the human translation is more dynamic but carries a higher risk of interpretative deviation.

Structural Shift

The shift analysis indicates that DeepL generally performs minor structural and level shifts while preserving the original syntactic pattern. Clause structure, word order, and grammatical relations are largely maintained. In contrast, the human translation exhibits more noticeable structural changes, including structure shifts, unit shifts (line expansion), and class shifts.

For instance, in “The carriage held but just ourselves,” DeepL maintains the simple verbal clause structure, whereas the human version reformulates it into a copulative construction, reflecting a structural and semantic shift. Similarly, in “And I had put away,” DeepL preserves the perfective form and enjambment, while the human translation adds an explicit object and omits the perfective marker, resulting in structural simplification and semantic expansion. In “Feels shorter than the Day,” DeepL retains the comparative structure, whereas the human version restructures the expression and shifts the temporal focus.

These findings suggest that DeepL tends to maintain the grammatical configuration of the source text, while the human translator applies greater flexibility to adjust structure for fluency and expressive effect in the target language.

Stylistic and Semantic Analysis by Stanza

In the first stanza, Dickinson constructs the personification of death through the capitalization of “Death” and the phrase “He kindly stopped for me.” This personification is ambiguous: death is not portrayed as a threat, but rather as a polite figure. DeepL preserves this structure directly in its rendering, maintaining the basic semantic meaning. However, in Indonesian, the structure becomes more declarative and loses some of its subtle existential ambiguity. The human translator, on the other hand, chooses the expression “Sang Maut,” which strengthens the symbolic and mythological dimension of death. While this choice enriches the cultural resonance, it also shifts Dickinson’s intimate tone toward a more dramatic nuance.

In the second stanza, the phrase “For His Civility” functions as a subtle irony, positioning death as a courteous and respectable figure. DeepL translates it concisely, preserving the density and minimalism characteristic of Dickinson’s style. In contrast, the human translator expands the phrase into a more elaborated expression. This expansion results in semantic amplification but reduces the stylistic compactness typical of Dickinson’s poetic voice. Thus, DeepL tends to maintain poetic density, whereas the human translation introduces interpretative elaboration.

In the third stanza, the repetition of the phrase “We passed” creates syntactic parallelism that produces a processional rhythm, suggesting that the journey toward death unfolds slowly and inevitably. DeepL consistently preserves this repetition, maintaining the stylistic effect of foregrounding through parallel structure. Although the human translation retains the general meaning of the journey, it varies the sentence structure, resulting in a softer parallel effect. From a stylistic perspective, DeepL remains closer to the rhythmic structure of the source text.

In the fourth stanza, the metaphor of the grave as a house is expressed through the phrase "A Swelling of the Ground." Both translations preserve this core metaphor, though they differ in diction intensity. The human version employs a more figurative lexical choice, while DeepL remains more structurally aligned with the source expression. Conceptually, however, both maintain the underlying metaphor of DEATH AS A HOUSE.

In the final stanza, Dickinson plays with the concept of temporality through the paradox "'tis Centuries – and yet / Feels shorter than the Day." DeepL preserves this paradoxical structure with relative accuracy. The human translation adopts a more poetic expression, but shifts the focus from subjective temporal relativity toward an impression of historical cycles. This shift illustrates how human interpretation can enrich expressive nuance, while also potentially altering the original conceptual framework.

Implications of the Findings in the Age of Artificial Intelligence

The findings of this study indicate that neural machine translation systems such as DeepL have achieved a high level of semantic accuracy in translating poetry. However, in the dimensions of stylistics and artistic effect, human interpretation continues to demonstrate a more varied capacity for stylistic reconstruction. Literary translation is not merely the transfer of meaning, but an aesthetic re-creation that requires sensitivity to symbolism, tone, and musicality.

This study contributes to the development of AI-based literary translation studies, particularly through an empirical evaluation of DeepL's performance in translating a classical English poem into Indonesian. By integrating translation technique analysis, shift examination, and stylistic reading, the study underscores the aesthetic dimension as a primary parameter in assessing the quality of literary translation.

The findings further reveal that although neural machine translation shows strong semantic precision, the interpretative dimension and the reconstruction of poetic effects still require human intervention. Therefore, this research enriches the evaluative framework of AI-based translation by positioning stylistics not as a supplementary aspect, but as the central axis in assessing the quality of poetry translation.

The present findings align with broader discussions on AI creativity in translation. Guerberof-Arenas and Toral (2022) argue that neural systems demonstrate controlled linguistic creativity but remain limited in producing culturally and emotionally grounded interpretations. Similarly, Manapbayeva et al. (2024) observe that AI-generated literary translations often require human post-editing to achieve expressive authenticity. The current study confirms this tendency within the English–Indonesian poetic context.

5. CONCLUSION AND RECOMMENDATIONS

Conclusion

This study identifies a clear difference in translational orientation between DeepL and human translation in rendering Dickinson's Because I Could Not Stop for Death into Indonesian. DeepL predominantly applies literal techniques and preserves syntactic structure, resulting in high lexical accuracy but limited stylistic flexibility. The human translator more frequently employs modulation and structural shifts to reconstruct metaphor, tone, and poetic resonance.

The findings confirm that semantic precision does not guarantee stylistic adequacy in poetry translation. While DeepL performs reliably at the denotative level, aesthetic reconstruction and interpretative depth remain dependent on human intervention. By integrating technique distribution, shift analysis, and stylistic evaluation, this study contributes to AI-based literary translation research within the English–Indonesian context.

AI may serve as an assistive tool in literary translation; however, it has not replaced the interpretative dimension of human translators. As the analysis is limited to one poem and one system version, broader generalizations should be approached cautiously.

Recommendations

Based on these findings, future research is encouraged to expand the object of study to other literary genres, such as drama or narrative prose, in order to examine whether similar AI translation tendencies persist across different textual contexts. Comparative analysis involving more than one neural machine translation system would also provide a more comprehensive understanding of the performance and limitations of each system in literary translation.

From a technological perspective, the integration of stylistic modeling, cultural sensitivity, and pragmatic context awareness should become key considerations in the development of AI-based translation systems. Such efforts are expected to enhance AI's capacity to reproduce the aesthetic and symbolic dimensions of literary texts, which remain a primary limitation.

In translation practice and education, AI may be utilized as a supportive instrument at the initial stage of the translation process. However, human involvement remains essential for interpretation, evaluation, and final editing to ensure stylistic quality and depth of meaning.

Furthermore, the evaluation of AI-based literary translation should not rely solely on lexical accuracy and structural equivalence. Stylistic parameters, emotional resonance, and aesthetic effect must be positioned as primary indicators in assessing the success of literary translation in the era of artificial intelligence.

REFERENCES

Ahmed, O. F., Dewi, I. K., & Anis, M. Y. (2025). Evaluation of machine translation systems: A literature review on ChatGPT and Google Translate. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 13(1). https://doi.org/10.24256/ideas.v13i1.6236

Anuar, A. F., & Sabri, M. A. Z. (2025). Artificial intelligence and Malaysian literature: A comparative study of AI and human translation. International Journal of Academic Research in Progressive Education and Development, 14(4), 67–87. https://doi.org/10.6007/IJARPED/v14-i4/26790

Batubara, N. H. (2025). A systematic literature review on the application of artificial intelligence in translation: Challenges, innovations, and impact across diverse fields. Journal of Linguistics, Literature and Language Teaching, 5(1), 67–87. https://doi.org/10.37249/jlllt.v5i1.934

Bellos, D., & Sennrich, R. (2019). Machine translation: Between myth and reality. Translation Spaces, 8(1), 1–18. https://doi.org/10.1075/ts.18026.bel

Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, P., Lohar, P., Way, A., Barone, A. V. M., & Gialama, M. (2018). A comparative quality evaluation of PBSMT and NMT using professional translators. Machine Translation, 32(3), 255–275. https://doi.org/10.1007/s10590-018-9211-x

Catford, J. C. (1965). A linguistic theory of translation. Oxford University Press.

DeepL SE. (2026). DeepL Translator (Computer software). DeepL. https://www.deepl.com/id/translator

Dolmaci, M. (2025). Linguistic and cultural fidelity in AI-driven translation: A comparative analysis of DeepL and human translations in English–Turkish literary short stories. International Journal of Language and Translation Studies, 5(1), 92–109. https://doi.org/10.63673/Lotus.1662868

Guerberof-Arenas, A., & Toral, A. (2022). Creativity in neural machine translation: A case study on literary text. Target, 34(3), 415–438. https://doi.org/10.1075/target.21057

Jassim, H. A. (2025). Human vs. posthuman author: A comparative analysis of human and AI-generated poetry translations. Journal of Arts, Literature, Humanities and Social Sciences, 13. https://doi.org/10.33193/JALHSS.121.2025.1455

Karabayeva, I., & Kalizhanova, A. (2023). Evaluating machine translation of literature through rhetorical analysis. Absorbent Mind, 3(2), 169–183. https://doi.org/10.37680/absorbent_mind.v3i2.4486

Kembaren, F. R. W., Hasibuan, A. K., & Natasya, A. (2023). Technology trends in translation: A comparative analysis of machine and human translation. Absorbent Mind, 3(2), 169–183. https://doi.org/10.37680/absorbent_mind.v3i2.4486

Leech, G., & Short, M. (2007). Style in fiction: A linguistic introduction to English fictional prose (2nd ed.). Pearson Education.

Manapbayeva, Z., Zaurbekova, G., K, A., A, K., & Pirmanova, K. (2024). AI in literary translation: ChatGPT-4 vs. professional human translation of Abai's poem "Spring." Procedia Computer Science, 251, 526–531. https://doi.org/10.1016/j.procs.2024.11.143

Molina, L., & Albir, A. H. (2002). Translation techniques revisited: A dynamic and functionalist approach. Meta, 47(4), 498–512.

Moorkens, J. (2018). What to expect from neural machine translation: A practical in-context assessment. Translation Spaces, 7(1), 93–112. https://doi.org/10.1075/ts.18011.moo

Newmark, P. (1988). A textbook of translation. Prentice Hall International.

Nida, E. A. (1964). Toward a science of translating. E. J. Brill.

Pratiwi, R. I., & Gusthini, M. (2025). ChatGPT sebagai alat penerjemahan AI untuk puisi "Hope Is the Thing with Feathers": Perbandingan dengan penerjemahan manusia. JISHUM (Jurnal Ilmu Sosial Dan Humaniora), 3(4), 513–526. https://doi.org/10.57248/jishum.v3i4.589

Siregar, S. D. P. Z. A.-F., & Naser, A. D. M. (2023). Mapping the landscape of machine translation error research: A bibliometric review (1980–2023). Buletin Al-Turas, 30(1). https://doi.org/10.15408/bat.v30i1.32655

Yuliasri, I. (2025). Student poetry translations with and without the use of AI: An evaluation of translation quality. LITE: Jurnal Bahasa, Sastra, Dan Budaya, 21(2). https://doi.org/10.33633/lite.v21i2.14336