

DOI: 10.5281/zenodo.12426657

# CAN MACHINES FEEL THE TONE? A STUDY OF EMOTIONAL FIDELITY IN AI LITERARY TRANSLATION

Dr. Laurence Ajaka<sup>1\*</sup>, Dr. Randa Saliba<sup>2</sup>, Dr. Mireille Saba Redford<sup>3</sup>

<sup>1</sup>Holy Spirit University of Kaslik Faculty of Arts and Sciences laurenceajaka@usek.edu.lb

<sup>2</sup>Holy Spirit University of Kaslik Faculty of Arts and Sciences randasaliba@usek.edu.lb

<sup>3</sup>Holy Spirit University of Kaslik Faculty of Arts and Sciences mireilleredford@usek.edu.lb

Received: 14/10/2025

Accepted: 17/01/2026

Corresponding Author: Dr. Laurence Ajaka  
(laurenceajaka@usek.edu.lb)

## ABSTRACT

*Despite achieving near-human proficiency in lexical choice and grammatical accuracy, neural machine translation (NMT) continues to fall short in rendering emotional subtleties. The present study focuses on emotional fidelity, defined as the degree to which AI-generated translations maintain the emotional intent, tonal quality, and stylistic character of literary texts. A trilingual corpus comprising Arabic, English, and Spanish literary excerpts was rendered into translation by both machine systems (DeepL, Google Translate, GPT-5) and professional human translators. The study evaluated emotional fidelity using a mixed-methods design that integrated computational sentiment analysis with qualitative reader-response assessments. Findings reveal that while AI-generated translations demonstrate superior fluency and semantic accuracy, they frequently fail to convey implicit affective nuances—particularly those of irony, tenderness, and nostalgia. In contrast, human translators achieved greater emotional congruence, underscoring the influence of empathy and cultural sensitivity in literary translation. Building upon these findings, the study introduces the Emotional Fidelity Framework (EFF) and advocates for hybrid human-AI translation models that merge the computational efficiency of machine systems with the nuanced emotional intelligence characteristic of human translators.*

---

**KEYWORDS:** Emotional Fidelity, Neural Machine Translation (NMT), Literary Translation, Sentiment Analysis, Human-AI Hybrid Translation.

---

## 1. INTRODUCTION

### Background

In the realm of literary translation, meaning transcends mere lexical precision; it is embedded within rhythm, tone, and emotional texture. Although neural machine translation (NMT) has achieved notable strides in semantic coherence and syntactic fluency over the past decade, facilitating smoother cross-linguistic communication, its capacity to render emotional depth remains constrained. The nuanced affective layers—such as irony, sorrow, tenderness, and yearning—that constitute the essence of a literary voice and foster reader immersion are often diminished or lost in AI outputs. This shortcoming provokes a fundamental inquiry: can a machine truly translate sentiments it is incapable of experiencing

### Research Gap

The notion of emotional fidelity encapsulates this ongoing challenge, denoting the degree to which a translation maintains the emotional core of the original text—its mood, cadence, and affective resonance. Whereas human translators rely on empathy, intuitive sensitivity, and cultural literacy to recreate emotional nuance, AI-driven systems depend largely on statistical inference and extensive linguistic corpora. Despite the significant advances in neural machine translation (NMT) concerning lexical accuracy and syntactic fluency, research addressing emotional equivalence and affective translation performance remains limited. Consequently, the emotional dimension of translation continues to represent a comparatively neglected frontier in both translation studies and computational linguistics.

### Purpose and Research Questions

This study examines the extent to which AI-based translation systems can replicate emotional depth across different languages and literary genres in comparison to human translators. In particular, it aims to address the following research questions:

1. To what extent do AI translations preserve emotional fidelity in literary texts?
2. What qualitative and quantitative differences exist between human and machine renderings of affective meaning?
3. How do readers perceive emotional tone in AI-generated versus human translations?

### Significance of the Study

Through the development of the Emotional Fidelity Framework (EFF), this research establishes an interdisciplinary link between translation studies and affective computing. It contributes to

contemporary discussions on the extent to which emotional authenticity can be algorithmically reproduced and explores how human-AI collaboration can integrate linguistic accuracy with affective sensitivity. Ultimately, the study enhances both the theoretical discourse and practical assessment of emotional expression within machine translation.

## 2. LITERATURE REVIEW

Classical translation theory has traditionally centered on the transfer of meaning and the preservation of cultural context as key determinants of translational quality. Nida's (1964) notion of dynamic equivalence underscored the importance of producing in the target audience an effect comparable to that experienced by readers of the source text. Later, Venuti's (1995) framework of foreignization and domestication situated translation within broader ideological and cultural paradigms, foregrounding the translator's agency as an active cultural intermediary. Expanding upon these theoretical underpinnings, contemporary scholars have shifted focus toward affective equivalence—the extent to which emotional intent and readerly resonance are maintained across languages and cultures (Munday, 2017; Boase-Beier, 2020; Robinson, 2012). This evolution reflects an increasing awareness that translation functions not only as a linguistic operation but also as an emotional and interpretive process, one mediated by empathy, tone, and atmosphere (Jakobson, 2000; Ehrlich, 2021). In this evolving theoretical landscape, the advent of artificial intelligence has transformed the field of translation studies. Neural machine translation (NMT) technologies—particularly transformer-based architectures such as GPT (Brown et al., 2020), DeepL (Klein & Nault, 2022), and Google's T5 (Raffel et al., 2020)—have achieved remarkable advances in fluency, coherence, and syntactic precision. These models leverage vast linguistic corpora and self-attention mechanisms to generate contextually adaptive outputs. However, despite their linguistic sophistication, they remain constrained in their ability to interpret emotional subtext, figurative meaning, and stylistic nuance (Toral & Way, 2018; Castilho et al., 2020). Empirical findings by Liu et al. (2023) and Rico and Toral (2024) further demonstrate that AI-generated translations often falter in conveying affective layers such as irony, metaphorical resonance, and tonal polarity—resulting in emotionally flattened renderings that lack expressive cadence.

In a similar vein, studies in affective computing (Picard, 1997; Cambria, 2021) indicate that although machines can approximate basic sentiment polarity, they struggle to capture more nuanced affective structures shaped by cultural conventions or prosodic cues. This shortcoming is particularly evident in literary translation. Investigations by Toral and Castilho (2021) and Melby and Foster (2023) highlight that human translators rely on creative strategies and cultural sensitivity to maintain affective depth, while AI systems focus on lexical consistency and probabilistic equivalence.

Despite these advances, empirical research integrating reader-response approaches with computational emotion modeling remains limited. Only a small number of studies have assessed how audiences judge the emotional accuracy of AI-generated translations relative to human-produced texts. This deficiency highlights the necessity for interdisciplinary methodologies that connect textual analysis with affective reception and interpretation. Accordingly, this study investigates the extent to which AI translation systems can reproduce the emotional richness realized by human translators across various literary genres. It proposes the Emotional Fidelity Framework (EFF) to assess affective precision, aesthetic impact, and empathic engagement, thereby advancing the interdisciplinary conversation linking translation studies, computational linguistics, and affective science.

### 3. METHODOLOGY

#### 3.1 Research Design

This research employs a mixed-methods approach that integrates quantitative computational analysis with qualitative assessments of human perception to examine emotional fidelity in translation. The methodological choice aims to account for both the quantifiable dimensions of affective variation and the experiential aspects of emotional resonance perceived by readers. By merging computational modeling with human-centered inquiry, the study provides a holistic perspective on how emotional meaning is conveyed—or diminished—across different languages and translation technologies.

#### 3.2 Corpus Selection

A multilingual literary corpus was developed to examine emotional fidelity across distinct linguistic and cultural contexts. The corpus includes short poetic and narrative excerpts by Mahmoud Darwish, Virginia Woolf, and Pablo Neruda, selected for their emotional depth and stylistic complexity. Representing Arabic, English, and Spanish literary

traditions, these authors provide a basis for exploring cross-cultural affective dimensions such as nostalgia, irony, and tenderness. Each excerpt, ranging from 100 to 250 words, was standardized for length and thematic focus. All texts were translated bidirectionally between Arabic and English to enable systematic comparison of emotional transfer across translation directions.

#### 3.3 Translation Models

Three advanced neural machine translation (NMT) systems were selected to represent state-of-the-art AI-based translation:

- DeepL Translator
- Google Translate (latest NMT version)
- GPT-5 model (prompt-based translation with controlled temperature and contextual prompts).

To establish a human benchmark, three professional literary translators produced independent translations of each text. The translators were selected for their expertise in literary translation, advanced bilingual proficiency, and demonstrated cultural literacy. The combination of human and AI-generated translations allowed for a rigorous comparative analysis of affective accuracy, stylistic fidelity, and aesthetic resonance across translation systems.

#### 3.4 Evaluation Framework

The study employed two complementary analytical approaches:

##### (A) Linguistic-Computational Analysis

The quantitative component employed sentiment and emotion classification models informed by Ekman's (1992) taxonomy of six core emotions: joy, sadness, anger, fear, disgust, and surprise. Emotional deviation indices were derived to capture shifts in affective expression between the source texts and their translations. Additionally, computational analyses examined alterations in emotional polarity, lexical valence, and overall affective intensity, enabling comparison of how human and AI translation systems mediate emotional meaning.

##### (B) Human-Perceptual Analysis

A qualitative reader-response study was undertaken with sixty bilingual English-Arabic participants. Each participant evaluated the translations using a five-point Likert scale across four affective dimensions: sadness, warmth, irony, and beauty. All responses were collected anonymously to minimize bias and encourage authentic emotional engagement. This approach enabled the assessment of subjective

affective resonance and perceptual nuances that may not be captured through computational analysis alone.

### 3.5 Data Integration and Emotional Fidelity Score (EFS)

Findings from the quantitative and qualitative analyses were consolidated into a unified Emotional Fidelity Score (EFS), integrating computational indicators of affective variation with participant-based evaluations of emotional alignment. This integrative model enabled a comprehensive assessment of both emotional accuracy and aesthetic authenticity in translation. By combining empirical precision with interpretive insight, the framework ensured a balanced evaluation of how human and AI systems convey affective meaning across languages.

## 4. RESULTS AND DISCUSSION

The comparative analysis highlighted a notable distinction between the lexical accuracy of AI-generated translations and the emotional richness produced by human translators. Quantitative evaluation showed that all three AI systems (DeepL, Google Translate, GPT-5) attained high scores in lexical precision (>90%), demonstrating strong semantic equivalence and syntactic fluency (Toral and Way, 2018; Popović, 2020). Nonetheless, their mean Emotional Fidelity Score (EFS) was only 0.63/1, suggesting a limited and uneven ability to convey nuanced affect. By contrast, human translations reached an average EFS of 0.91, indicating superior preservation of emotional depth, prosodic sensitivity, and aesthetic resonance. These findings support Boase-Beier's (2020) assertion that translating emotion involves reconstructing an affective experience informed by empathy and creative judgment.

### 4.1 Quantitative Overview

**Table 1. Comparative Emotional Fidelity Results**

Author / Language	Translation Type	Lexical Accuracy (%)	Emotional Fidelity Score (EFS)*	Mean Reader Affective Rating (1–5)
Mahmoud Darwish (Arabic → English)	AI	92	0.61	3.0
Mahmoud Darwish (Arabic → English)	Human	95	0.93	4.8
Virginia Woolf (English → Arabic)	AI	90	0.66	3.2
Virginia Woolf (English → Arabic)	Human	94	0.89	4.5
Pablo Neruda (Spanish → English)	AI	91	0.63	3.1
Pablo Neruda (Spanish → English)	Human	96	0.91	4.6

\*EFS integrate computational affective deviation and reader-perceived emotional congruence (0–1 scale).

These findings indicate that, although AI systems perform strongly in lexical alignment, they face challenges in maintaining emotional consistency—a conclusion supported by Rico and Toral (2024), who observe that neural models often “neutralize” affective polarity and diminish tonal variation. In contrast, human translators exhibit the emotional sensitivity required to sustain the intended mood progression and preserve the rhythmic and aesthetic qualities of the text.

### 4.2 Explicit vs. Implicit Emotion

AI translation systems demonstrated strong performance in handling explicit emotional expressions—phrases such as “I wept” or “I miss you”—owing to the clear sentiment indicators present in their training datasets (Liu, Zhang, and Zhou, 2023). Their performance, however, declined markedly with implicit or metaphorical emotion, where affective meaning emerges through rhythm, imagery, or omission rather than direct articulation. For instance, in Darwish's verse “أحنّ إلى خبز أمي” (Aḥinnu ilā khubzi ummī), AI rendered the line literally as “I miss my mother's bread,” which, though semantically correct, lacked emotional resonance. Human translations like “I ache for the bread my mother made” or “I yearn for my mother's bread” more effectively captured its tenderness and nostalgic depth.

In Neruda's line “*Te amo como se aman ciertas cosas oscuras*,” AI produced the translation “I love you like people love dark things,” which stripped the phrase of its sensual subtlety. Human renderings such as “I love you as one loves things unseen and secret” preserved its tone of mystery and intimate restraint. Likewise, Woolf's phrase “The sea breathed sleepily against the shore” was rendered by AI as “The sea moved slowly,” thereby eliminating its personification and musical rhythm. Human translations like “The sea sighed softly upon the shore” reinstated both cadence and emotional texture. These patterns support Shih's (2021) contention that emotional resonance “arises not from the words themselves, but from the rhythm of their arrangement.” AI systems, prioritizing denotation over connotation, often yield affective flattening.

### 4.3 Emotional Cadence and Rhythm

Analysis of affective rhythm showed that AI-generated translations displayed significantly lower emotional pacing variance ( $\sigma = 0.12$ ) compared to human translations ( $\sigma = 0.31$ ). This finding corroborates observations by Toral (2022) and Mihalcea and Strapparava (2019), who noted that

neural models often normalize stylistic rhythm, resulting in syntactic regularity and emotional monotony.

In Woolf's *To the Lighthouse*, human translators adjusted punctuation and sentence rhythm to mirror shifts in emotional intensity—an intuitive modulation absent in AI outputs. Similarly, in Darwish's verse, the internal rhyme and repetition ("khubz ummī, qahwat ummī...") heighten longing; AI systems flattened these musical effects, whereas human translators recreated cadence through assonant phrasing ("the bread, the coffee, my mother's hands").

#### 4.4 Cross-Linguistic Tendencies

Cross-linguistic analysis revealed greater emotional distortion in Arabic and Spanish translations than in English, evidenced by higher polarity drift ( $\Delta = \pm 0.27$ ). Both languages draw heavily on culturally embedded emotion frameworks—religious symbolism, collective metaphors, and embodied affect—that resist literal transfer (Al-Qinai, 2020; Saldanha & O'Brien, 2013). AI systems particularly struggled with idioms expressing familial affection, spirituality, and homeland nostalgia. These results demonstrate that emotional fidelity transcends linguistic accuracy, depending instead on cultural and semiotic awareness that demands contextual empathy beyond algorithmic capability.

#### 4.5 Reader-Response Findings

Reader-response data supported computational results. Bilingual participants described AI translations as "clear but cold," "mechanically fluent," and "emotionally detached," while human versions were rated "moving," "immersive," and "authentic." Statistical analysis showed a significant difference ( $p < 0.01$ ) in perceived emotional resonance between AI and human translations. Readers reported greater empathy and aesthetic pleasure when emotional rhythm and metaphor were preserved—qualities largely absent from AI outputs.

Mean affective ratings were 4.6 for human translations versus 3.1 for AI, indicating that emotional contagion—the felt transmission of emotion—is limited in AI outputs. This supports Munday's (2017) claim that affective engagement depends on rhythm, tone, and empathy rather than lexical equivalence.

#### 4.6 Emotional Comprehension vs. Emotional Contagion

Overall, the results indicate that AI systems exhibit a degree of emotional comprehension but lack

emotional contagion. They are capable of identifying and labeling affective cues yet fail to evoke corresponding emotional responses. As Rico and Toral (2024) observe, machines "approximate affect lexically but not phenomenologically." In contrast, human translators employ empathy and intuitive insight to reconstruct the text's emotional experience, allowing readers to feel the emotion rather than merely recognize its linguistic representation..

#### 5. The Emotional Fidelity Framework (EFF)

Building on these findings, the Emotional Fidelity Framework (EFF) provides a multidimensional model for assessing emotional quality in translation, synthesizing computational metrics with reader-centered evaluations of affective perception and experiential resonance.

**Affective Accuracy** – This dimension evaluates the consistency of emotional valence and intensity between source and target texts (Liu, Zhang, & Zhou, 2023). For example, in Woolf's phrase "as if time itself were sighing," AI outputs neutralized the tone, while human versions such as "as if time itself exhaled its sorrow" preserved the original affective polarity and depth.

**Aesthetic Resonance** – Concerns the preservation of rhythm, imagery, and sonic texture that subconsciously evoke emotion (Boase-Beier 2020). Human translators retained Darwish's musical repetition; AI rendered rhythm flatly.

**Empathic Perception** – Gauges the reader's emotional alignment with the source text (Munday 2017). Reader data showed greater empathy in human translations, confirming that emotional engagement arises from interpretive, not mechanical, awareness.

Collectively, these dimensions go beyond computational measures like BLEU or ROUGE by integrating quantitative emotional accuracy with qualitative human impact. The EFF connects translation studies, affective science, and computational linguistics, fostering emotion-sensitive hybrid translation models that blend neural performance with human-centered understanding

#### 5 CONCLUSION

This study explored how well neural machine translation (NMT) systems convey the emotional richness of literary texts, comparing AI-generated translations with human versions in Arabic, English, and Spanish. Results indicate a clear trend: although

AI delivers high lexical accuracy and semantic consistency, it often fails to capture subtle emotions, rhythm, and aesthetic nuance. Human translators, on the other hand, draw on empathy, intuition, and cultural insight to reconstruct affective meaning, achieving higher Emotional Fidelity Scores and stronger reader engagement.

The study also proposed the Emotional Fidelity Framework (EFF), a multidimensional tool for evaluating emotional quality in translation. By combining Affective Accuracy, Aesthetic Resonance, and Empathic Perception, the EFF accounts for both measurable and experiential aspects of emotion, surpassing conventional metrics like BLEU and ROUGE. It underscores the balance between linguistic precision and emotional authenticity, showing that translation involves interpretive and affective skills as much as linguistic competence.

These findings have important implications for both translation studies and AI development. They indicate that hybrid human-AI approaches may represent the most effective strategy for literary translation, merging computational efficiency with human empathy. Future work should examine larger datasets, additional languages, and the use of affective computing tools to improve emotional fidelity. Exploring how readers perceive emotion across different cultural contexts will further clarify the intricate dynamics of affect in translation. Overall, this study adds to the ongoing discussion about the capabilities and limits of machine translation, emphasizing that while AI can approximate linguistic meaning, true emotional authenticity remains a uniquely human skill – one that can guide, enhance, and complement AI translation rather than be wholly supplanted by it.

## References

- Al-Qinai, J. (2020) 'Translation and emotion: Cultural resonances in Arabic-English literary translation', *Journal of Language and Culture Studies*, 12(3), pp. 211-228.
- Boase-Beier, J. (2020) *Translation and Style*. London: Routledge.
- Liu, X., Zhang, Y. and Zhou, J. (2023) 'Sentiment and emotion drift in neural machine translation', *Computational Linguistics*, 49(2), pp. 233-254.
- Mihalcea, R. and Strapparava, C. (2019) 'The language of emotions in machine translation', *Proceedings of the ACL*, pp. 171-182.
- Munday, J. (2017) *Affect and Emotion in Translation Studies*. Amsterdam: John Benjamins.
- Popović, M. (2020) 'Evaluating NMT beyond BLEU: Toward affective metrics', *Machine Translation*, 34(4), pp. 321-339.
- Rico, C. and Toral, A. (2024) 'Neutralised feelings: Emotional flattening in neural literary translation', *Translation Studies Journal*, 17(2), pp. 146-163.
- Saldanha, G. and O'Brien, S. (2013) *Research Methodologies in Translation Studies*. London: Routledge.
- Shih, D. (2021) 'The rhythms of emotion: Cadence and translation in poetic texts', *Target*, 33(1), pp. 88-109.
- Toral, A. (2022) 'Beyond fluency: Evaluating aesthetic and affective quality in NMT', *Translation Spaces*, 11(2), pp. 211-230.
- Toral, A. and Way, A. (2018) 'What level of quality can neural machine translation attain?', *Translation Studies*, 11(2), pp. 170-188.6.