UNIVLORA SCIENTIFIC JOURNAL

Leotta, P., et al. DOI: 10.63871/unvl.jsuv2.1.26
UniVlora Scientific Journal 2026, no.II, volume I

Original research

Human Translation and AI: Reimagining EAP/EFL Pedagogy in the Language of Emotions.The Cases of *The Diary of Anaïs Nin* and *Gratitude* by Oliver Sacks

Paola Clara Leotta [1], **Vincenza Tutino**[1], **Heidi Littunen**[1].

[1] University of Catania, Italy

Corresponding author: paola.leotta@unict.it

Article Type: Original research
Received: March 2026 | Accepted: May 2026 | Published: June 2026

Abstract

As English for Academic Purposes (EAP) and English as a Foreign Language (EFL) classrooms evolve with technology, this research aims to argue that integrating the practice of AI-powered literary translation not only modernizes pedagogy but also fundamentally enhances EAP/EFL learners' linguistic, cultural, and critical capabilities.

In classroom interventions across a Master's in Psychology, a University Language Centre and a Secondary School, 106 students translated 18 extracts from two literary Diaries (by Anais Nin and Oliver Sacks). They compared their Italian translations with those from ChatGPT, DeepL, or Google Translate to determine human-AI translation differences and assess which better conveys the source author's emotions.Textual similarity was calculated using the TF-IDF (Term Frequency–Inverse Document Frequency) method, a vector-representation technique that assigns a weight to terms based on their frequency in a text and their distribution across the entire corpus. In addition, an Emotional Detection (ED) analysis was conducted on the human translations and on those generated by the artificial intelligence systems, in order to explore how the affective content of texts is reworked in the transition from the original to the translation.

Results show that human translations present significantly greater variability and emotional sensitivity than AI, particularly in metaphorical and affectively charged texts.

Keywords: *English for Academic Purposes (EAP), English as a Foreign Language (EFL), Translation studies, Intercultural Competence, Artificial Intelligence (AI).*

Original research

Introduction

The objective of this research is to present an authentic project accomplished in the academic year 2024/2025, focusing on creative writing as a psychotherapy (Mastela 2020), followed by diary translation (as an affective, emotional practice).

Through a series of classroom-based interventions with students belonging to three different contexts (a Master's degree programme in Psychology, a University Language Centre and a Secondary School- all based in Catania, Italy), this research focuses on the translation of emotions as a pedagogical tool in English language teaching and on the comparison between human translation and AI translation.

The importance of creating an authentic learning environment and engaging students in translation projects has been recognized in recent years by several translation teachers and scholars.

In addition, following Yu (2022), a standard method to assess Machine Translation (and the latest developments, such as AI) success is to compare it to human translation. Although this field has begun to be explored, research remains limited. Thus, this study addresses the gap by asking: Do human translations of emotionally charged literary texts exhibit greater variability than AI-generated translations?

Methods

Measuring translation divergence

A corpus of 18 extracts from two literary Diaries (Gratitude by Oliver Sacks and The Diary of Anaïs Nin), was administered to a total sample of 106 students (enrolled in B2/C1 English language courses, aged between 16 and 60, with different levels of education—from high school to PhD—and of different genders) who were requested to translate them into Italian and to compare their translations to the ones produced by a chosen AI tool among ChatGPT, DeepL, and Google Translate in order to determine how much human translation differs from AI translation and to assess which one conveys the source texts author's emotions.

The data were collected in a structured file in which human translations and those produced by AI were separated and organized by quotation, while keeping information about the translation tool used.

Starting from this corpus, the analysis was developed in several phases.

In the first phase, for each quotation, the human translations were isolated from the translations generated by the three AI systems. This made it possible to perform a controlled comparison on the same source text and among different engines. Distinguishing between the three tools allowed for the evaluation not only of the distance between human and machine translation, but also the degree of convergence or divergence among the solutions provided by the systems themselves.

Textual similarity was calculated using the TF-IDF (Term Frequency–Inverse Document Frequency) method, a vector-representation technique that assigns a weight to terms based on their frequency in a text and their distribution across the entire corpus, reducing the impact of words that are too common and not very informative. TF-IDF transforms each translation into a weighted vector, where rare but informative terms are emphasized, and frequent, low-information words are downweighed. Cosine similarity was then applied to these vectors, a measure that quantifies the degree of similarity between two texts on a continuous scale from 0, indicating no similarity, to 1, indicating almost complete overlap. This metric reflects the angle between two vectors: values close to 0 mean the vectors are dissimilar (orthogonal), while values close to 1 indicate nearly identical direction and high textual overlap (Vanmassenhove et al., 2019).

Subsequently, internal variability was operationalized as the standard deviation of the similarities calculated between all possible pairs of translations referring to the same quotation. For each quotation, all pairwise similarities were computed either among human translations or among machine-generated translations. The standard deviation of these values was then calculated to quantify internal coherence: a low standard deviation indicates high consistency, whereas a high one reveals divergence in translational choices. This made it possible to measure, on the one hand, how much human translations differ from one another and, on the other, how much the translations produced by the three AI tools differ from one another. Introducing the

variable "translation tool" , therefore, it was possible to assess whether systematic differences exist among ChatGPT, DeepL, and Google Translate in terms of internal consistency and distance from human translational choices (Freitag et al., 2021). The results were then visualized using boxplots and heatmaps, tools that highlighted respectively the distribution of similarities and the patterns of convergence and divergence between human and AI translations, respectively, both at the level of individual quotations and in comparison, among the various translation tools.

Emotional detection analysis

Alongside the analysis of similarity and translational variability, an Emotional Detection (ED) analysis was conducted on the human translations and on those generated by the artificial intelligence systems, in order to explore how the affective content of texts is reworked in the transition from the original to the translation. The analysis was performed using an emotion-detection model specific to the Italian language, based on the MilaNLProc/feel-it-italian-emotion neural network, capable of classifying each text according to four basic emotions: joy, sadness, fear, and anger.

Each translation, human or automatic, was first analyzed individually by the model, which provided a dominant emotional label and a confidence score. This phase was essential to assess not only linguistic similarity but also whether the emotional impact of the source text was preserved or altered in the translated versions, by humans or AI tools (Bąk, 2023; Qian et al., 2023; Liu et al., 2025).

Questionnaire

Finally, a questionnaire was administered to our student sample (70,5% men and 29,5% women). Among them, 55 of the participants use AI tools occasionally for translation, 38 to use it often, and 9 affirmed to never use it with this purpose. Only 4 affirmed to always use it.

The questionnaire included questions about the use of artificial intelligence in translating literature (diaries, in particular) from English into Italian.

Results

Quantitative results: measuring translation divergence

In the analyzed corpus—consisting of translations produced by 106 students from quotations taken from Gratitude by Oliver Sacks and The Diary of Anaïs Nin, compared with translations generated by three AI tools (ChatGPT, DeepL, and Google Translate)—the quantitative analysis of the data reveals a complex picture that makes it possible to meaningfully address the issue of human translational variability versus artificial translation.

The initial measure used was the semantic–lexical similarity between personal translations and AI translations, calculated through cosine similarity applied to TF-IDF vectors. The resulting values, displayed in a heatmap reorganized by blocks of students (groups of ten) and by quotations, show a clearly heterogeneous distribution. Alongside quotations characterized by a high average alignment with the AI output—where values approach 1—there are systematically darker areas corresponding to lower similarity scores, sometimes close to zero.

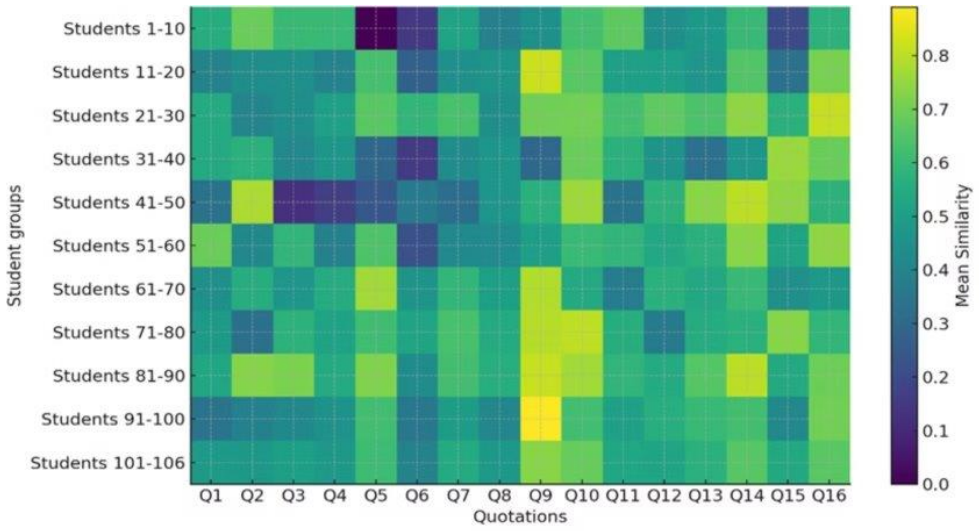
Reorganizing the data into groups makes collective tendencies more visible and highlights how the distance from AI translations is neither uniform nor random; rather, it varies consistently both across different quotations and among different segments of the student sample.

This indicates that in many cases the students did not simply echo or slightly rephrase the AI's output, but instead produced significantly divergent solutions, often restructuring the metaphorical image or altering the semantic orientation of the translation.

Leotta, P., et al. DOI: 10.63871/unvl.jsuv2.1.26
UniVlora Scientific Journal 2026, no.II, volume I

Original research

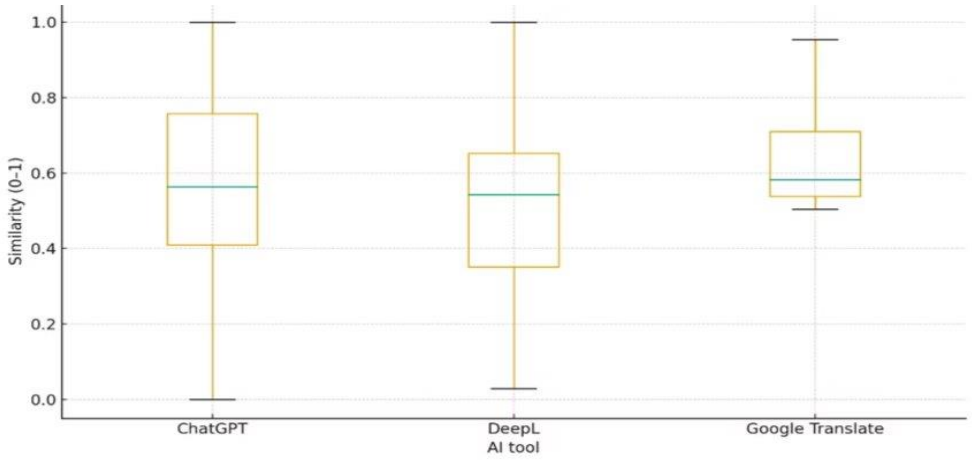
Figure 1 - Heatmap of mean similarity across student groups



The heatmap (Fig. 1), taken as a whole, visually conveys strong inter-individual variability, with students oscillating between high proximity to the AI and highly divergent choices, and with certain quotations proving more "sensitive" than others to this kind of divergence—especially when the original text contains metaphors or complex emotional concepts such as "bugger off," "arrow," "irreversible loss," "rehearsal for separation," or "we are still like animals." This already suggests that interaction with AI does not produce a uniform standardizing effect, but rather a spectrum of responses that depends both on the individual translator and on the complexity of the source text.

It is clear that the tool used by the students influenced the results. Indeed, the analysis by AI translation tools shows an average similarity that is fairly homogeneous across the different systems: ChatGPT is at 0.5886, while DeepL and Google Translate fall within comparable ranges (0.5936 and 0.5621, respectively), indicating a general tendency toward relatively stable and consistent translations. In terms of frequency within the corpus, ChatGPT is the most widely used tool, with 73 cases, followed by Google Translate with 22 cases and DeepL with 11. This numerical imbalance means that the statistics related to ChatGPT rest on a broader empirical base compared to the other two engines, while the results for DeepL, given the smaller sample size, must be interpreted more cautiously

Figure 2 - Boxplot of similarity values by AI tool



The graphical representation (Figure 2) of the similarity between human and AI translations for the 106 cases analyzed clearly shows that Google Translate displays a more compact distribution, with generally more stable similarity values, while ChatGPT and DeepL show a much wider dispersion, indicating greater fluctuation in the relationship between automatic output and human reformulation. In particular, ChatGPT shows the widest range, with cases of strong convergence and cases of strong divergence, suggesting that its output is sometimes adopted as a model, but other times deeply reworked or rejected by the students.

Regarding the central research question of the study—whether human translations show greater or lesser variability compared to those produced by AI—it is essential to distinguish human–AI similarity from internal variability. Similarity values measure how closely each individual translation aligns with or diverges from the AI output, but they do not directly measure how much students differ from one another, nor how much the automatic systems differ among themselves.

Nevertheless, based on the data obtained, it emerges that the human translations lie along a very wide continuum, with solutions that are very close to AI output and solutions that are profoundly independent. By contrast, AI translations—even when produced by different systems such as ChatGPT, DeepL, or Google Translate—tend to move within a narrower range of variation.

In particular, where the original texts present complex or culturally marked imagery, such as “bugger off” or “rehearsal for separation by death”, human translations show a much broader range of solutions, ranging from semantic softening to radical re-semanticization, while AI tends to converge on more literal formulations with less creative variation. This phenomenon further supports the hypothesis that human translation is not only more variable but also qualitatively more sensitive to the metaphorical, pragmatic, and emotional levels of the source text.

The calculation of internal variability conducted at the level of each individual quotation quantitatively confirms the hypothesis of greater dispersion in human translations compared to those produced by AI systems. The average standard deviation of the students’ translations is 0.2142, while that of the translations generated by ChatGPT, DeepL, and Google Translate is significantly lower, at 0.1178. This indicates that, given the same source text, human solutions show a broader range of variation, reflecting

interpretative differences, individual stylistic choices, and a greater degree of freedom in reformulation. AI translations, although produced by different models, show a tendency toward convergence, with a higher average similarity (0.7012 compared to 0.3608 for human translations) and lower internal variability, suggesting a greater standardization of automatic translation strategies. These results underline that human translation, within the analyzed corpus, is characterized not only by greater distance from artificial output but also by a more pronounced plurality of solutions, especially in metaphorical and semantically complex passages.

Table 1 - Human vs. AI Translation Variability and Similarity

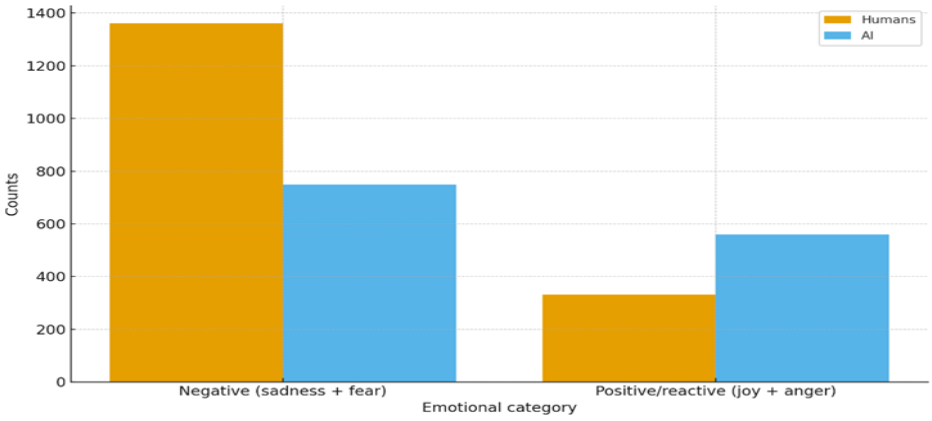
Metric	Value	Description
Human standard deviation	0.21	Average variability in human translations
AI standard deviation	0.12	Average variability in AI-generated translations
AI internal similarity	0.70	Consistency across ChatGPT, DeepL, and Google Translate
Human internal similarity	0.36	Consistency across student translations

Quantitative results: Emotional detection analysis

Regarding the analysis of emotions, since the corpus examined consists of literary texts that are strongly marked on the existential and thematic level—centering on experiences of loss, death, separation, and memory—the four categories returned by the model were subsequently reorganized into two broader interpretative areas consistent with the nature of the material. On the one hand, there is a pole of deep negativity, comprising the emotions of sadness and fear (linked to grief, vulnerability, and anguish). On the other hand, there is a pole of positive/reactive re-elaboration, comprising joy and anger (interpretable respectively as signals of vital affirmation and of emotional distancing).

This reorganization made it possible to avoid a mechanical reading of the automatic categories, reintegrating the analysis within an interpretative framework more appropriate to the type of texts under examination.

Figure 3 - Emotional polarity distribution



The results (Fig. 3) show a clear divergence between human translations and AI translations. Human translations are strongly concentrated in the deep negativity pole, with 1,367 occurrences

(out of 1,696 total translations) falling into this macro-category, compared to only 329 in the positive/reactive pole. This suggests that students tend to preserve, and in some cases intensify, the emotional weight related to loss, pain, and vulnerability present in the original texts. AI-generated translations, on the other hand, show an overall more balanced emotional profile compared to human translations, with 750 occurrences in the deep negativity pole and 554 in the positive/reactive pole. Notably, within this latter category, there is a relatively high presence of the emotion of joy, interpreted not as simple happiness but as a signal of a distanced or rebalancing reformulation of the traumatic and memorial content of the source texts. The remaining 392 AI translations, characterized by marked emotional neutrality or scores too low for reliable classification, were not included in the graphical representation in order to preserve readability and interpretive coherence. This gap should not be interpreted as an "emotional deficiency" of AI, but rather as the result of its text generation strategies, which favor more balanced and less affectively polarized linguistic solutions.

Table 2 - Emotional Distribution in AI and Human Translations

Emotion	AI Translations	Human Translations
Anger	130	73
Fear	122	1117
Joy	424	256
Sadness	628	250
Neutral	392	-
Total classified	1304	1696

As detailed in Tab. 2, human translations show a strongly skewed concentration on the emotion of fear (1,117 occurrences out of 1,696), which can be interpreted as a greater emotional involvement with themes of loss, vulnerability, and existential anguish present in the texts. The fear detected by the model should not be understood strictly in psychological terms but as a linguistic index of a strong proximity to pain and the experience of grief, while sadness is less represented (250 occurrences), suggesting that students are less likely to "cool down" the emotional content into a form of detached sadness and more likely to maintain it in a dimension of immediate affective exposure.

In AI translations, instead, a more nuanced emotional distribution is observed: the categories of sadness (628 occurrences) and fear (122 occurrences) are both present but less dominant than in human translations, with the prevalence of sadness indicating a tendency to redistribute the tragic content of the text between processing and attenuation of emotional tension. The relatively high presence of joy in AI translations (424 occurrences out of the 1,304 classified) does not indicate a naive distortion of the original content but rather a systematic tendency to rebalance negative emotional load through less dramatically connoted linguistic formulations aimed at semantic reworking. In this sense, the "joy" detected by the model should not be interpreted in hedonic or consolatory terms but as a marker of a process of attenuating the tragic intensity present in the sources.

It is also important to emphasize that all 1,696 human translations were classified by the model into one of the four emotional categories, whereas 392 AI translations were excluded from the analysis due to emotional neutrality or classification scores too low to be considered reliable. This asymmetry seems to indicate that

automatic translations tend toward reorganizing and partially neutralizing emotional load rather than erasing it, while human translations are overall more sensitive to the affective and symbolic layering of the original texts, maintaining a closer adherence to their emotional complexity.

Qualitative results

From a qualitative point of view, it is noteworthy how human and AI translations diverge, not only in lexical choice but, more importantly, in their handling of tone, register, and cultural resonance. By examining a colloquial, irreverent sentence: "He thought he would live to eighty and then bugger off" from Oliver Sacks, alongside a metaphorically charged expression "pierced by an arrow" from Anaïs Nin, the analyses bring to the fore the different pressures exerted by style-sensitive prose. Together, they serve to illustrate that AI systems tend toward safe, literal, semantically stable solutions, whereas human translator, through creativity, intuition, and subjective interpretation, produce a far wider spectrum of possibilities, running the range from highly effective recreations of the original voice to unintended tonal shifts. In both examples, the contrast brings to light the core dilemma at the heart of translation, balancing accuracy with expressiveness, and fidelity of meaning with fidelity of effect.

The different human translations of the sentence "He thought he would live to eighty and then bugger off" show a whole gamut of strategies, which indicate how differently translators interpret tone, register, and cultural nuance. On one extreme, very neutral interpretations such as "Lui pensava che avrebbe vissuto fino agli ottant'anni e che poi ne sarebbe andato", retain the literal meaning but flatten Sacks's playful irreverence into something bland and almost bureaucratic type of language. This is a tendency that is often seen in AI output. More expressive ones, such as "Egli pensava che non avrebbe vissuto più di 80 anni e che dopo si sarebbe levato dai piedi" or "poi si sarebbe levato di torno", get much closer to "bugger off", by adopting colloquial Italian idioms; despite small semantic shifts, they reproduce more effectively Sacks's light-hearted and self-mocking tone. Others fall into tonal mismatches: "passato a miglior vita", "si sarebbe spento", or "andarsene in pace" are correct, yet they introduce solemnity or tenderness absent from the original, while "tirato le corde" tries creativity but uses a non-standard idiom which may raise readers' eyebrows. Taken altogether, the comparison shows, in a nutshell, the dilemma between semantic accuracy, tonal fidelity, and cultural equivalence faced by any translation. Neutral, safe translations-the ones typically produced by AI-retain meaning but lose voice, while human translators, through colloquial choices, euphemisms, or experimental idioms, can either succeed in recreating the author's irreverence or, at times, inadvertently distort it. All in all, the versions using "levarsi di torno" or "levarsi dai piedi" turn out to be the closest reproductions to Sacks's intended tone, with the best balance of clarity, informality, and humor.

The AIs' translations of the expression "pierced by an arrow" from The Diary of Anaïs Nin, compared with human renderings, show clear differences in strategy, nuance, and stylistic sensitivity. In this case, DeepL and ChatGPT converged on "trafitto da una freccia", a solution that is lexically precise, slightly poetic, and fully appropriate to Nin's literary tone. But here, too, this choice reflects limits that are proper to AI systems, which always head for the most conventional dictionary equivalent, avoiding risks in style and failing to capture possible emotional, symbolic, or sensuous nuances proper to Nin's writing. Quite another story, however, is told by human translators, who offer a much broader and more heterogeneous range of possibilities. Some alternatives, like "trafitto da una freccia", or the more dramatic "trapassato da una freccia", stay faithful to the original while modulating intensity and register, whereas others deviate decidedly in tone: "perforato da una freccia" introduces an overly clinical

precision; "infilzato da una freccia" risks sounding crude, or grotesque; and "bucato da una freccia" trivializes the metaphor through colloquial language. Further illustrative of how human creativity may also result in distortion or typographical mistakes is the error "pugnalato da una feccia". Overall, while AI offers a safe, neutral, and stylistically consistent option, human translations exhibit both greater expressive potential and greater variability, oscillating between poetic refinement, dramatic intensity, and mismatched or inappropriate choices. The best-balanced, faithful rendering, shared by both AI and one human version, remains "trafitto da una freccia", which maintains Nin's literary tone without excess or distortion.

Discussion

A discussion on the role of AI in text translation, with advantages and disadvantages, is now in order.

The research conducted revealed a fairly clear and shared perception by the students of the role of artificial intelligence in translation work. Students acknowledged that AI represents a useful tool, capable of speeding up the process and providing generally correct translations, often offering synonyms, alternative phrasing, and explanations of the meaning or origin of certain expressions. Many highlighted how AI can help enrich one's vocabulary and quickly provide a general understanding of a text.

However, almost all respondents emphasized the limitations of AI compared to a human translator. In particular, automatic translation often appears too literal, mechanical, and lacking in sensitivity. AI struggles to capture emotions, cultural nuances, irony, and metaphors—elements that are essential when producing a natural translation faithful to the spirit of the original text. Although AI may be more precise grammatically, it cannot reproduce interpretative richness or human "feeling."

Another recurring point concerns the importance of human oversight: AI can suggest good solutions, but its translations always need to be checked, corrected, and adapted. Revision is considered essential, as is the ability to compare different versions and critically reflect on linguistic choices. Some students even noted that this process helped them improve their own linguistic awareness.

In conclusion, students agreed that AI is an effective support tool, provided it is used with moderation and critical judgment. It can facilitate, complement, and enrich translation work, but it cannot replace the sensitivity, deep contextual understanding, and creativity typical of a human translator.

As for advantages, the responses clearly indicate that the primary advantage of using artificial intelligence in translation is speed. Students repeatedly emphasize that AI allows them to obtain a translation in a matter of seconds, drastically reducing the time required compared to manual translation. This speed is seen as particularly useful not only in academic study but also in professional contexts, where large volumes of text must be managed, or quick responses to messages and communications are needed.

In addition to speed, many appreciate the convenience and accessibility of the tool: AI is always available, easy to use, and allows users to quickly generate a first version of a text or gain a general understanding of its content. In several cases, it is also considered an effective tool for checking the accuracy of one's own translation or for comparing alternative ways of expressing the same idea.

Another highlighted advantage concerns AI's lexical knowledge, which allows users to quickly learn new words, synonyms, stylistic variations, and correct grammatical constructions. AI also offers alternative translations depending on the context, helping users choose the most suitable formulation. For those who struggle to understand individual terms or expressions, AI provides immediate and reliable support to clarify meanings and linguistic structures.

Finally, many students recognize that AI is useful for producing a first draft of a text to be further refined: a base that is accurate, fluent, and often

well-structured, which can then be adapted to stylistic or communicative needs.

In summary, the use of AI in translation is perceived as fast, practical, efficient, and educational, capable of facilitating comprehension and offering multiple linguistic suggestions. Despite its limitations, it represents a valuable support tool for anyone wishing to translate quickly and more easily access the meaning of words and texts.

As for drawbacks, the responses reveal a strong awareness of the risks associated with using artificial intelligence for translation. The first limitation highlighted concerns the possibility of errors: AI can translate too literally, misinterpret the meaning of words, or completely miss cultural nuances, metaphors, idioms, and emotions. This can result in translations that are grammatically correct but unnatural, nonsensical, or unable to convey the tone and depth of the original text.

A second recurring risk is dependency. Many students worry that frequent use of AI could reduce motivation to fully engage in language learning, weaken critical thinking, and diminish the ability to translate independently. There is concern about a gradual "mental laziness," leading to underuse of memory, creativity, and interpretative skills, ultimately compromising personal linguistic development.

Many responses also emphasize the danger of losing originality: AI-generated translations tend to standardize style, erase the author's personality, and neutralize emotions. Some note that, if not carefully reviewed, a text produced by AI can feel cold, artificial, or stylistically anonymous.

Finally, broader risks were noted, such as the potential replacement of human professionals, with resulting employment consequences, and even indirect effects related to the environmental impact of large computational systems.

Conclusion

In conclusion, students recognize that AI can be a useful support tool, but they stress that excessive or uncritical use can lead to errors, misunderstandings, loss of skills, and diminished quality of the translated text. Human intervention remains essential to verify, interpret, and preserve meaning, creativity, and emotion.

In addition, in a world where English for Academic Purposes/English as a Foreign Language classrooms evolve with technology, integrating the practice of AI-powered literary translation not only modernizes pedagogy but also fundamentally enhances EAP/EFL learners' linguistic, cultural, and critical capabilities.

Human translations show significantly greater variability and emotional sensitivity than AI, particularly in metaphorical and affectively charged texts. The future lies not in replacing human translators but in harnessing AI's strengths while safeguarding human creativity, judgment, and the irreplaceable capacity to translate the language of emotions.

As a follow-up, this research will further compare the aforementioned results to the literary translation from Italian authors (respectively, Diario V 1947>1955 by Delfina Vezzoli, 2016, Gratitudine by Isabella C. Blum, 2016), reflecting on the role of the translator's "affect" (Koskinen 2020) in language choices. This will prompt discussions among students regarding 'foreignizing' versus 'domesticating' translation choices (Venuti 1995), and will encourage them to consider how cultural nuances and affect/emotions are either preserved or erased by AI.

Acknowledgment

The authors acknowledge that the quantitative analysis was conducted using a Python script, developed by Gabriele Caruso (https://orcid.org/0009-0009-9040-7296), designed to process the dataset and generate the visualizations.

References

Bąk, H. (2023). Issues in the translation equivalence of basic emotion terms. *Ampersand,*

11,100128. https://doi.org/10.1016/j.amper.2023.100128

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics, 9*, 1460–1474. https://doi.org/10.1162/tacl_a_00437

Koskinen, K. (2020). *Translation and affect: Essays on sticky affects and translational affective labour*. John Benjamins.

Liu, Y., Lee, S. Y. M., & Li, D. (2025). Examining emotions in English and translated Chinese children's literature: A bilingual emotion detection model based on LLMs. *Language Resources and Evaluation*. https://doi.org/10.1007/s10579-025-09846-z

Mastela, O. (2020). Retelling legends and folk tales: A transcreative approach in the collaborative translation classroom. *Research in Language, 18*(2), 151–171.

Nin, A. (1975). *The diary of Anaïs Nin: Volume 5, 1947–1955*. Harvest/HBJ Book.

Qian, S., Orasan, C., do Carmo, F., & Kanojia, D. (2023). Challenges of human vs machine translation of emotion-loaded Chinese microblog texts. In *Proceedings of Machine Translation Summit XIX: Users Track* (Vol. 2, pp. 217–236). Asia-Pacific Association for Machine Translation. https://aclanthology.org/2023.mtsummit-users.21.pdf

Sacks, O. (2015). *Gratitude*. Alfred A. Knopf.

Vanmassenhove, E., Shterionov, D., & Way, A. (2019). Lost in translation: Loss and decay of linguistic richness in machine translation. In *Proceedings of Machine Translation Summit XVII: Research Track* (pp. 222–232). European Association for Machine Translation. https://aclanthology.org/W19-6622/

Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.

Yu, C. (2022). *Online collaborative translation in China and beyond: Community, practice, and identity*. Routledge.