



Is machine translation the greatest thing since sliced bread? Human and machine translation of food-related idioms

Rozane R. Rebechi

Federal University of Rio Grande do Sul
Porto Alegre, Rio Grande do Sul, Brazil

rozane.rebechi@ufrgs.br 

<https://orcid.org/0000-0002-1878-7548> 

Isabela Moreira de Oliveira

Federal University of Rio Grande do Sul
Porto Alegre, Rio Grande do Sul, Brazil

demoreirapraoliveira@gmail.com 

<https://orcid.org/0009-0003-3796-8787> 

Maria Augusta Tedesco

Federal University of Rio Grande do Sul
Porto Alegre, Rio Grande do Sul, Brazil

maria.tedesco@ufrgs.br 

<https://orcid.org/0009-0000-2930-5615> 

Isabela Freitas Cardoso

Federal University of Rio Grande do Sul
Porto Alegre, Rio Grande do Sul, Brazil

isabelafcardoso10@gmail.com 

<https://orcid.org/0009-0007-9084-6673> 

Abstract: This study examines the translation of food-related idiomatic expressions (IEs) from English into Brazilian Portuguese, by comparing the performance of trainee translators and machine translation (MT) systems operating under two distinct paradigms: neural machine translation (NMT) and large language models (LLMs). Contexts containing IEs are extracted from the Corpus of Contemporary American English and translated by undergraduates, and by four AI-based systems: Google Translate and DeepL (NMT-based), and ChatGPT-4 and Gemini 2.5 Pro (LLM-based). All translations are qualitatively evaluated for idiomaticity and conventionality using corpus evidence and classified into four categories: idiomatic and conventional, non-idiomatic but conventional, idiomatic but non-conventional, and inaccurate. When human translation (HT) and MT are compared overall, HT produces a higher proportion of idiomatic and conventional equivalents. While both approaches show similar proportions of non-idiomatic but conventional and idiomatic but non-conventional renderings, inaccurate translations are almost twice as frequent in MT. When the performance of HT is compared with that of the LLM-based systems, the results indicate that the latter, largely driven by Gemini's performance, slightly outperform students in producing appropriate translations. A more fine-grained analysis, however, reveals important differences within MT: NMT systems perform markedly worse, generating few idiomatic and conventional equivalents and a high proportion of inaccurate translations. Statistical analyses confirmed a significant

association between translator type (human, NMT, or LLM) and translation quality (idiomatic and conventional, non-idiomatic but conventional, idiomatic but non-conventional, or inaccurate) across the four evaluation categories. This association is driven primarily by the poorer performance of NMT systems, whereas no significant difference is found between human translators and LLMs. These results highlight the importance of integrating human expertise with advanced MT tools in idiom-rich translation tasks. They also highlight the need for translator training that emphasizes idiomaticity, conventional usage, and critical evaluation of machine-generated output, preparing translators to work effectively in increasingly hybrid translation environments.

Keywords: food idioms; machine translation; large language models; neural machine translation; human translation.

I. Introduction

The automation of translation processes was initiated several decades ago with the development of computer-assisted translation (CAT) tools and machine translation (MT) systems¹. Early attempts date back to the post-World War II period, when researchers developed computers capable of performing translation tasks (Rothwell et al., 2023). Although full automation proved elusive, the enthusiasm surrounding MT in the early 1960s inspired continuous research and technological advances aimed at using computational systems to replicate aspects of human translation (HT) (Hutchins, 1986). Over time, MT has evolved from rule-based and statistical approaches to neural architectures, such as Google Translate and DeepL (Jurafsky & Martin, 2026), which can produce more fluent and contextually appropriate output than the previous systems (Rothwell et al., 2023).

More recently, translation has emerged as one of the most visible and widely discussed applications of artificial intelligence (AI), particularly with the rise of large language models (LLMs) based on the generative pre-trained transformer (GPT) architecture. Based on prompt-driven interaction, models such as ChatGPT-4 and Gemini 2.5 Pro employ deep neural networks trained on massive text corpora to generate fluent and contextually appropriate language, enabling the models to perform translation, among other linguistic tasks (Jiao et al., 2023; Rothwell et al., 2023).

Recent studies indicate that LLMs perform particularly well in translation tasks, achieving results comparable to traditional neural machine translation (NMT) systems for high-resource languages² and spoken or conversational registers (Jiao et al., 2023). However, they still face limitations when translating highly creative texts. Testing ChatGPT in the English-German translation of a travel blog text, Puppel and Borg (2025) conclude that, while the system produces translations that are generally coherent, fluent, and communicatively effective, it fails to consistently render figurative language and stylistic nuance. Rather, it frequently resorts to literal and non-idiomatic solutions that weaken aesthetic and rhetorical effects, highlighting the continued need for human

¹ In this article, machine translation is understood broadly as any system employed to automatically generate translations, whether originally designed specifically for translation tasks – such as dedicated machine translation systems – or more general-purpose models, including large language models, when used for translational purposes.

² In Natural Language Processing, the term high-resource refers to languages for which large amounts of digital linguistic data and tools are available. The classification refers to data availability and infrastructure, not the number of speakers or the language's social status (Joshi et al., 2020).

post-editing to preserve the appealing characteristics of this genre. Mammadova (2025) also notes that, although NMT has indeed surpassed previous paradigms in overall quality, it still struggles to handle idiomatic language and maintain consistency across longer passages.

Hence, despite its advances, MT still faces considerable challenges in rendering texts that rely heavily on creativity, cultural references, and figurative language (Rebecchi et al., 2024, 2025). Idiomatic expressions (IEs) are particularly problematic due to their non-literal meaning and cultural specificity, which often prevents direct word-for-word translation (Donthi et al., 2025).

Food-related IEs occupy a particularly expressive space in both everyday communication and literary discourse, as they reflect cultural practices, social values, and shared ways of thinking (Batista et al., 2026; Moreira de Oliveira, 2022). In Brazilian Portuguese (PTBR), for example, the expression *preço de banana* (lit. “banana price”) is used to refer to something extremely cheap, since bananas are abundant and therefore inexpensive in tropical countries. In different varieties of English, a similar figurative meaning is conveyed through expressions such as “for peanuts” (usually in the United States), “chicken change” (India), and “cheap as chips” (Australia), which also associate low monetary value with common, easily accessible items. Such expressions encode experiences that are simultaneously linguistic and cultural, and their accurate translation requires more than lexical equivalence. It demands pragmatic awareness, cultural competence, and sensitivity to idiomatic conventions in the target language (Moreira de Oliveira, 2022; Rebecchi & Trindade, 2021).

The present study aims to investigate the extent to which MT, on the one hand, and trainee translators, on the other hand, can handle idiomaticity in authentic contexts from English (EN) into PTBR. More specifically, this study focuses on food-related IEs. 75 expressions were selected out of an initial list of approximately 1,000, based on their occurrences in the Corpus of Contemporary American English (COCA) (Davies, 2008–). Contextualized sentences containing these expressions were extracted from COCA and subsequently submitted to human, Google Translate and DeepL (NMT-based systems), and ChatGPT-4 and Gemini 2.5 Pro (LLM-based systems) translations.

The analysis combined qualitative and quantitative approaches. First, the EN > PTBR human, NMT-based, and LLM-based translations of the 75 food-related IEs were manually classified into four categories: idiomatic and conventional, non-idiomatic but conventional, idiomatic but non-conventional, and inaccurate. Next, the resulting data were subjected to statistical tests to assess the significance of the differences between human translation, NMT-based translations, and LLM-based translations, as well as between NMT-based translations and LLM-based translations themselves.

This article is structured as follows. After this introduction, we present a brief review of the literature on MT, translation of food-related idioms, and conventionality. Next, we outline the methodology adopted. Then, we analyze the results obtained, and the article concludes with some final remarks.

2. Literature review

2.1 Machine translation

The growing use of technology to support translation is not new and can be traced back to the development of CAT tools, which incorporated translation memories and terminology databases



to enhance productivity and consistency. More recently, translation has become one of the most visible applications of AI, particularly with the consolidation of NMT systems and the emergence of LLMs (Rothwell et al., 2023).

Recent research has increasingly focused on improving the ability of MT systems, particularly LLMs, to handle idiomatic and culturally embedded language. Jiang et al. (2026), for example, propose a reasoning framework specifically designed for the English–Chinese translation of IEs. By combining idiom detection with external retrieval and alternative translation strategies, their approach improved the performance of different LLMs. Their findings suggest that explicitly identifying idiomatic language and providing models with additional contextual information can reduce overly literal translations. However, their evaluation relies primarily on automated metrics, which may not fully account for the cultural and phraseological dimensions involved in idiom translation.

The distinction between semantic adequacy and cultural or phraseological appropriateness has also emerged as an important issue in recent research. Oliveira et al. (2026), while investigating the translation of culturally embedded PTBR expressions, found that improvements in fluency and grammatical accuracy did not necessarily result in successful cultural adaptation. Although LLM-based systems such as ChatGPT-4o and Gemini 1.5 Pro generally outperformed conventional MT systems in fluency and overall adequacy, they frequently relied on paraphrasing and semantic normalization, which could attenuate pragmatic meaning, idiomatic force, and cultural specificity.

Another line of research has explored whether idiom-specific training data can improve MT performance. Montesinos et al. (2026) used LLMs to expand a manually aligned Spanish–Galician dataset into a larger synthetic parallel corpus of IEs, which was subsequently used to retrain a sequence-to-sequence translation model. The inclusion of idiom-focused data significantly improved the translation of these expressions while also producing a modest improvement in overall translation performance. Although the original dataset comprised only 66 core idiom pairs, the findings indicate that targeted phraseological data can enhance MT systems' ability to handle idiomatic language.

The present study distinguishes between NMT, represented by systems such as Google Translate and DeepL, and LLM-based translation, exemplified by models such as ChatGPT and Gemini. Although both rely on AI and neural architectures, they embody distinct paradigms of translation technology. NMT systems are specialized engines trained on large parallel corpora and optimized to map source sentences to target sentences more efficiently than previous systems. Their performance, however, is strongly conditioned by the availability and quality of parallel data, which tends to favor high-resource language pairs and frequently translated genres. In low-resource languages and creative texts, however, they tend to have limited capacity in terms of accuracy and fluency. In contrast, LLMs are general-purpose generative models trained on vast multilingual corpora to perform a wide range of language tasks. While translation is not their primary objective, it emerges from their capacity for probabilistic text generation and contextual reasoning (Naveen & Trojovsky, 2024).

As LLMs encode broader contextual, cross-linguistic, and world-knowledge signals during pretraining, they may demonstrate greater adaptability when translating between languages with uneven data representation, potentially mitigating some of the constraints traditionally associated with low-resource settings. At the same time, their translation quality remains uneven across

languages, as performance is still closely tied to the distributional patterns present in the training data, often privileging high-resource languages and dominant varieties (Pang et al., 2025).

Empirical studies comparing human translators, NMT systems, and LLMs consistently show weaker performance of MT for culture-bound expressions. For instance, Jibreel (2024) found that both Google Translate and ChatGPT performed poorly in the translation of English–Arabic allusions, achieving substantially lower success rates than human translations produced through machine translation post-editing (MTPE).

A study by Rizki and Masykuroh (2025) examined how well ChatGPT handled cultural nuances, idiomaticity, and narrative style inherent to literary texts. It investigated the techniques employed by ChatGPT in translating J.K. Rowling’s *Harry Potter and the Sorcerer’s Stone*, focusing on the accuracy and acceptability of the translations produced. Utilizing a qualitative approach, the study analyzed data from the original English text, the Indonesian translation generated by ChatGPT, and the feedback from five evaluators based on predetermined criteria. The data analysis process involved comparing the source and target texts to categorize translation strategies and calculating average scores from raters. The findings revealed that ChatGPT effectively utilized various translation strategies, with “translation by paraphrase using a related word” being the most used (62%). Other strategies included “translation by a more neutral or less expressive word” (24%) and “translation by a more general word” (8%). The overall accuracy score was relatively high. However, challenges arose in maintaining acceptability, particularly regarding expressive tones and cultural nuances.

In a study comparing the performance of Google Translate and ChatGPT 3.5 in translating partially compositional and non-compositional multiword expressions (cf. Ramisch et al., 2016) from EN into PTBR, Rebechi et al. (2024) found that both systems performed statistically similarly, both in terms of their correspondence to gold standard equivalents – high-quality human-produced translations that serve as a benchmark against which machine-generated translations are evaluated – and in the adequacy of their proposed translations that did not match the gold standard options, regardless of the expressions’ degree of compositionality. MT performance did not exceed 48% correspondence with gold standard equivalents for non-compositional expressions and 64% for partially compositional ones. Regarding adequacy, MT achieved 63.5% for partially compositional expressions and 70% for non-compositional ones.

When the performance of trainee translators was compared to that of MT in rendering contexts with partially and non-compositional multiword expressions, Rebechi et al. (2025) found that both Google Translate and ChatGPT 3.5 were able to generate plausible translations for a range of multiword expressions, generally recognizing their idiomatic nature and producing adequate equivalents. However, the systems tended to favor literal renderings more often than human translators, who were more successful in capturing idiomaticity and conventionality, as well as providing a wider range of equivalent expressions. Statistical tests indicated no significant difference between the two systems in either accuracy or adequacy. Qualitative results revealed that, although both MT systems sometimes offered creative or contextually acceptable solutions, they also produced inconsistent or culturally inappropriate renderings, including European Portuguese variants. In contrast, human translators showed greater sensitivity to context and figurative meaning,

confirming that idiomatic translation still benefits from human interpretative and cultural competence.

2.2 Translator training in the age of GenAI

The growing integration of GenAI into translation practice has prompted renewed discussion of the competences required of future translators and, consequently, of how translator training should respond to these technologies. The recent volume *Teaching Translation in the Age of Generative AI: New Paradigm, New Learning?*, edited by Penet et al. (2026), is entirely devoted to this question, bringing together different perspectives on the pedagogical implications of GenAI for translator education. Rather than approaching AI exclusively in terms of automation or replacement, several contributions to the volume examine how it can be integrated into translator training while preserving human agency, critical judgment, and established translation competences. Of particular relevance to the present study are proposals concerning human–AI collaboration, critical assessment of AI-generated translations, and the use of GenAI to support the development of translation expertise.

As the capabilities of LLMs expand, however, the relationship between human and machine translation need not be understood exclusively in terms of competition or replacement. Massey and Ehrensberger-Dow (2026) conceptualize GenAI as a collaborative technology capable of complementing human agency, provided that translators bring their own knowledge, experience, and transferable skills to the interaction. From this perspective, translation competence increasingly involves combining established textual and interlingual skills with the ability to interact effectively with GenAI through prompting and reprompting, while retaining human agency over the translation process.

From a pedagogical perspective, GenAI may also contribute to translator training by supporting deliberate practice rather than simply generating translations. Angelone (2026) proposes using conversational GenAI as a personalized training resource through which learners can receive immediate and informative feedback, identify recurring translation problems, and engage in further practice targeting specific areas of difficulty. In this approach, learners remain active participants: through strategic prompting, they can request feedback on their translations and generate additional tasks focusing on particular error patterns, including overly literal translation. GenAI can thus support self-directed learning and reflection on translation performance, although the author emphasizes that learners remain responsible for monitoring their own progress and critically engaging with the feedback provided.

This perspective is particularly relevant to the translation of IEs, where apparently fluent output may nevertheless reflect literal interpretation or non-conventional target-language choices. The use of GenAI for deliberate practice could therefore provide trainee translators with opportunities not only to generate possible equivalents but also to compare alternatives, identify problems of idiomaticity and conventionality, and reflect on the reasons why particular solutions are more appropriate in a certain context. In this sense, GenAI may function not simply as an alternative translation agent, but as a resource for developing the critical translation competence required to evaluate machine-generated output.



2.3 Translating food idioms

Following Alisoy (2025), IEs are conventionalized multiword expressions that function as single semantic units whose overall meaning cannot be straightforwardly derived from their individual lexical components. Their translation is particularly challenging because these expressions combine semantic opacity with conventionality (Tagnin, 2013) and are often deeply rooted in the source culture, limiting the availability of direct equivalents in the target language.

Beyond these intrinsic properties, IEs are also highly context-dependent and require careful disambiguation, since their idiomatic meaning may only emerge in specific co-textual or discourse contexts. These combined challenges pose significant difficulties for both human and machine translators (Rebechi et al., 2025). For example, the expression “bring home the bacon” illustrates how disambiguation can become particularly complex. The phrase may be interpreted literally, referring to the act of bringing home a cured pork product, or idiomatically, meaning “to earn a living” or “to provide financially for one’s family”. To further complicate matters, speakers may deliberately exploit this ambiguity to create wordplay, as in “Are you getting paid for this thing? Because you’re supposed to bring home the bacon, Dad, and I eat a lot of bacon”. (COCA). In this case, the sarcasm relies precisely on the coexistence of the literal and idiomatic readings, both of which are activated by the surrounding context.

The extent to which an idiom integrates with its surrounding context varies according to its degree of idiomaticity, which directly affects how it is interpreted and translated. In highly idiomatic expressions, such as “not know beans about”, a literal, word-by-word interpretation is incompatible with the surrounding linguistic context. This lack of semantic compatibility between the compositional reading and the surrounding context signals a high degree of opacity and idiomaticity (Moreira de Oliveira & Teixeira, 2025). By contrast, in idioms with lower levels of idiomaticity, such as “eat like a bird”, the figurative meaning is more easily recoverable, as it builds on widely shared background knowledge – namely, that birds eat small quantities of food. Similarly, the expression “as different as chalk from cheese” draws on culturally shared knowledge that, despite a superficial resemblance, the two elements differ fundamentally in texture, function, and use, enabling the idiom to express a strong contrast.

Recognizing and interpreting these varying degrees of semantic opacity requires sensitivity to contextual, cultural, and conventional cues, which explains why the translation of IEs remains a persistent challenge for both HT (Rebechi & Trindade, 2021; Xatara et al., 2001) and MT (Baziotis et al., 2023; Musaad & Al Towity, 2023). Concerning the language pair English–Italian, Castaldo and Monti (2024, p. 32) conclude that “[t]raditional neural machine translation (NMT) systems often falter in accurately capturing the essence of idiomatic expressions, tending towards translations that are either overly literal or misinterpret the intended meaning”. The authors nevertheless acknowledge that more recent GPT-based models exhibit greater flexibility in handling idiomaticity by leveraging broader contextual and semantic information.

2.4 Conventionality

From a corpus-based perspective (McEnery & Hardie, 2012), idiomaticity is grounded in the recurrent and conventional nature of authentic language use. In this sense, conventionality refers to



the establishment of expressions through repeated use within a speech community, rather than through the free creation of novel combinations by individual speakers (Tagnin, 2013). In other words, language users rarely produce or interpret texts by freely combining words at will. As Sinclair (1991) proposes, fluent communication relies less on an “open-choice principle”, where words are selected one by one, and more on what he termed the idiom principle – the tendency to use preconstructed patterns and recurrent combinations. These prefabricated units form the backbone of conventionality in language, enabling speakers to process and produce meaning efficiently.

Concerning IEs, conventionality operates as a shared cognitive and cultural resource. Idioms, as well as other multi-word combinations, are conventional not only because they recur frequently, but also because their form-meaning pairing has become cognitively stabilized through repeated use within a speech community. This stability means that even minor alterations – such as changing a lexical component or modifying syntactic order – can disrupt recognition and comprehension (Erman & Warren, 2000). For instance, the English idiom “spill the beans” conveys the act of revealing a secret. Substituting one of its elements (e.g. “pour the beans”, “spill the peas”) or translating it literally into another language would likely obscure its idiomatic meaning, since the expression’s conventionality resides not in its individual words, but in its fixed combination and figurative sense (Moreira de Oliveira & Teixeira, 2025).

Conventionality can manifest at different levels, including syntactic patterning (e.g., fixed word order), semantic meaning (e.g., figurative or non-literal interpretations), and pragmatic use (e.g., contextual appropriateness and cultural salience) (Tagnin, 2013). In food-related idioms, these dimensions frequently overlap, as expressions draw on culturally shared knowledge while occurring in relatively fixed linguistic patterns. For instance, expressions such as “the cream of the crop” and “bite off more than one can chew” derive their figurative meanings from culturally established associations with food. Their idiomatic interpretation, therefore, depends not only on understanding the individual words but also on recognizing the expression as an established form–meaning pairing within a speech community.

Predictability poses an important challenge to MT systems. Unlike human translators, who activate contextual, encyclopedic, and cultural knowledge to identify conventionalized meaning, MT systems often rely on statistical or probabilistic associations derived from training data. As a result, deviations from canonical forms – pluralization, tense variation, or the insertion of modifiers – may lead to literal renderings or (partial) misinterpretations (Schmid, 2020). For example, while a human translator is expected to primarily recognize “bring home the bacon” as an idiom, an MT system might produce a literal translation, unless it has encountered sufficient contextual examples of the idiomatic use in its training data.

In his usage-based socio-cognitive model, Schmid (2020) describes conventionalization as a dynamic, multilayered process that underlies linguistic stability and change. Conventions arise when members of a community repeatedly align their linguistic behavior, through recurrent patterns of use and their social diffusion, until these patterns become collectively recognized and expected. The degree of conventionality of an expression, such as an idiom, depends on the extent to which speakers conform to shared regularities across formal, semantic, contextual, and social dimensions. Once established, these conventions guide comprehension and production, licensing familiar utterances while delimiting the space for innovation. In translation, viewing idioms as products of

conventionalization helps explain why human translators can draw on shared socio-cultural knowledge to recognize – and render – figurative meanings, whereas MT systems often struggle when these conventional associations are underrepresented in their training data.

3. Methodology

3.1 The corpus

Corpus-based research shows that idioms are unevenly distributed across genres and registers. Biber et al. (1999, p. 1025-1026) argue that IEs, such as “how do you do”, “a piece of cake”, or other formulaic phrases, occur only occasionally in fiction – approximately less than five instances per million words – and are even rarer in other written genres, such as news, academic prose, or official documents. To the authors’ surprise, idioms are not as frequent in spontaneous conversation as one might expect; instead, they tend to occur more often in fictional dialogue, where they are used to evoke informality or to construct stereotypical speech. They also note that idioms sporadically appear in news writing, where they serve stylistic and rhetorical purposes, such as to inject vividness, humor, or a touch of colloquialism into otherwise formal reporting.

Hence, identifying an ideal, ready-to-use comparable corpus from which to extract contexts containing IEs in English and Portuguese is far from straightforward. For this study, we therefore opted to work with the Corpus of Contemporary American English (COCA) (Davies, 2008-), a large, balanced corpus comprising more than one billion words drawn from a wide range of genres, including spoken language, fiction, popular magazines, newspapers, academic texts, TV and movie subtitles, blogs, and other web-based materials. COCA was used to identify, from an extensive list of English idioms, those expressions that occur at least five times in the corpus, ensuring that the items selected for analysis are attested in contemporary usage.

To assess the use and conventionality of the translated IEs in Portuguese, this study draws on two complementary sources: the Corpus do Português: NOW (Davies, 2012-2019) and the Corpus do Português: Web/Dialects (Davies, 2016-). NOW is a large monitor corpus composed of online news texts published between 2012 and 2019, with approximately one billion words, and provides access to edited, semi-formal language across different Portuguese-speaking countries. The Web/Dialects corpus is a web-based collection of informal texts drawn from blogs, forums, and personal webpages, also comprising around one billion words, and is designed to capture more spontaneous, non-edited, and regionally varied language use.

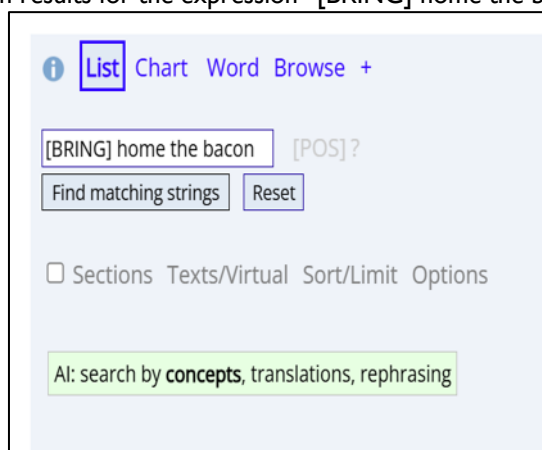
Although these corpora differ in composition and genre coverage, their large size and broad coverage allow for systematic searches of idiomatic patterns in authentic contexts. To mitigate the limitation regarding genre comparability, we established specific criteria to identify recurrent IEs in EN and to search for their functional or conventional equivalents in PTBR, as detailed in the following section.



3.2 Selection of food-related idiomatic expressions

75 expressions were selected out of an initial list of approximately 1,000 food-related idioms in English compiled by Batista et al. (2026)³. The selection followed the alphabetical order of the content words in the expressions and was restricted to idioms that occurred at least five times in COCA. The adoption of this relatively low cutoff point in such a large corpus was motivated by previously discussed corpus-based evidence showing that IEs are generally infrequent across genres and registers (Biber et al., 1999). First, all occurrences of each IE were retrieved from COCA. Figure 1 illustrates this procedure using the expression “bring home the bacon”, which was searched by means of a lemma-based query on the verb.

Figure 1: Search results for the expression “[BRING] home the bacon” in COCA



The screenshot shows the COCA search interface. At the top, there are tabs for 'List', 'Chart', 'Word', and 'Browse'. The 'List' tab is selected. Below the tabs, there is a search bar containing the query '[BRING] home the bacon' and a placeholder '[POS]?'. To the right of the search bar are buttons for 'Find matching strings' and 'Reset'. Below the search bar, there is a checkbox for 'Sections' and a label 'Texts/Virtual Sort/Limit Options'. At the bottom, there is a green box with the text 'AI: search by concepts, translations, rephrasing'.

Source: The authors

COCA retrieved 175 instances of this expression, with the verb occurring in various inflected forms (Figure 2):

Figure 2: Results of the lemma-based search for the expression “bring home the bacon” in COCA

ALL FORMS (SAMPLE): 100	FREQ
BRING HOME THE BACON	86
BRINGING HOME THE BACON	51
BROUGHT HOME THE BACON	21
BRINGS HOME THE BACON	17
TOTAL	175

Source: The authors

An analysis of the contexts in which the expression appeared revealed that it mostly conveys the meaning “to earn money to provide for one’s family” (Cambridge University Press, n.d.), as shown in Figure 3:

³ The list was compiled as part of the EITA project (Expressões Idiomáticas com a Temática da Alimentação [Food-Related Idiomatic Expressions]), coordinated by Professor Elisa D. Teixeira at the University of Brasília (UnB). The project investigates multilingual food-related IEs from a corpus-based perspective.

Figure 3: Excerpts of the contexts with the expression “bring home the bacon” in COCA

further evidence that there is more than one way to bring home the bacon and fry it up in a pan. " The pre ad to a dead heat by stressing his ability to bring home the bacon. Beginning his long tenure in the Senat through a crisis, Jack. I have to bring home the bacon now. Sally's pregnant. Damn, man. Why didn't you t aid for this thing? Because you're supposed to bring home the bacon, Dad, and I eat a lot of bacon. For y d it. Well, somebody's got ta bring home the bacon. - And the bacon and the bacon. - Downtime is one th doctorate in physical therapy. I guess she'll bring home the bacon in our family. " # " When I was going t iona is who I want in Sacramento to bring home the bacon, " said San Francisco Board of Supervisors Pr ear and brag about how we've been able to bring home the bacon. Or, so that we can be sure that the sp mbers. # " Luckily I'm not just here to bring home the bacon, " he said. # Parents were apolitical # Born i ea. I got ta go to work so I can bring home the bacon. - And a porterhouse steak. - Hey, I'm not kidding.
--

Source: The authors

A context in which the idiomatic meaning of the expression could be inferred was then selected, as in: “He was content to be merely the man of the house, **bringing home the bacon** and letting his wife handle their home”. (COCA, our emphasis). Preference was given to contexts that did not contain excessive specialized terminology but used more general vocabulary. The selected contexts range from 9 to 43 words in length.

The 75 selected contexts were randomly distributed among the 25 students enrolled in the undergraduate course English Translation II at a Brazilian federal university⁴. Although the IEs were evenly distributed, with each student receiving 30 contexts, the number of completed translations per student ranged from 15 to 30. Each IE was translated between one and eight times by different students. This variation is explained by two factors: only 13 students were present on the day of data collection, and some of them did not complete the task within the two-hour time limit. As a result, 338 human translation outputs were collected.

The same set of contexts was subsequently translated by four MT systems – two NMT systems (Google Translate and DeepL) and two LLMs (ChatGPT-4 and Gemini 2.5 Pro) –, resulting in 300 machine-generated translations⁵.

4. Analyses

This section presents a qualitative analysis of the translation outputs, comparing human and machine translations in terms of how IEs were rendered and classified, as well as contrasting the performance of the MT systems with one another. Statistical metrics were then applied to assess whether the observed differences are significant.

⁴ The human translation task was conducted by the first author during an onsite class held at the Federal University of Rio Grande do Sul (UFRGS) on May 28, 2025.

⁵ The machine-generated translations were obtained in May 2025.

4.1 Qualitative analysis

Both HT and MT outputs were organized into separate tabs of a Google spreadsheet. Each translation was independently assessed by four members of the research team with expertise in Translation Studies, corpus linguistics, and phraseology: the project coordinator, one master's student, one PhD student, and one undergraduate research assistant supervised by the project coordinator. Each evaluator identified the target-language equivalent used for the IE and assigned it to one of the categories in the annotation scheme described below. The annotations were subsequently compared and discussed to resolve any discrepancies. Whenever disagreements arose, the evaluators sought consensus through discussion. In the few cases where consensus could not be reached, the majority decision was adopted. Although no formal inter-rater reliability coefficient was calculated, the use of multiple independent evaluators followed by consensus meetings was intended to enhance the consistency and reliability of the qualitative analysis. The annotation scheme comprised four categories, defined as follows:

- i) Idiomatic and conventional (green): the equivalent provided is idiomatic and occurs at least five times with an idiomatic meaning in CdP;
- ii) Non-idiomatic but conventional (orange): the expression is not rendered idiomatically in PTBR, but the proposed equivalent is considered appropriate in the context where it appears;
- iii) Idiomatic but non-conventional (blue): a creative idiomatic equivalent is presented with no attested recurrence in CdP;
- iv) Inaccurate (red): the idiomatic meaning is not preserved in the translation.

Figure 4 illustrates the annotated spreadsheet, showing the IE, its context, the students' translations, and the highlighted PTBR equivalent.

Figure 4: Annotation procedure

Idiomatic Expression (IE)	Context to be translated	Student 1	IE 1	Student 5	IE 5	Student 7	IE 7
appetite, whet someone's	The trailer is there to whet your appetite for more and to get you	0				O trailer serve pi	dar só um gostir
ate	When you watch her work, you can see why Aretha is who she	"Quando você v	engoliu			"Quando você vi	devorou
bacon, bring home the	He was content to be merely the man of the house, bringing ho	0		Ele estava conte	trazer o pão de c	Ele se contentav	fazer o trabalho
bacon, save one's	Of course a few months later he was our golden boy, the guy th	0					
barrel, scrape (the bottom of) the	The supersonic burn rate has left few qualified people in Washi	0					
beans about (someone or something), not know	The trouble with women is they always think they know what's t	0					
beans, spill the	He's not gonna kill me for it. Which, I can't say the same about	0					
beef, (have/to)	I think it's pretty obvious what happened. I read somewhere tha	0					
beefcake, to be a	I'd met him before but hadn't seen him in over five years. Dr. G	Eu conheci ele a	ser um pedaço d	Já o tinha visto	a ser um colírio pa		
beet-red, go	As soon as she sits down, dude just goes beet-red. He's so em	Assim que ela se	ficar com o rosto			No minuto em q	ficar vermelho c

Source: The authors

Each category is presented below and illustrated using the following source text: "Oh, Sara, I'm sorry the visit took so long. Mr. Campbell was **chewing my ear off**" (COCA, our emphasis).



Idiomatic and conventional

The idiomatic and conventional category includes translations in which the source idiom is rendered by a PTBR IE that is both semantically equivalent and attested as conventional usage. An example of this category is presented below:

Student 10: “Ah, Sara, peço desculpas pela visita ter se estendido tanto. A sra. Campbell fala pelos cotovelos”.

To verify the conventionality of the proposed equivalent, we searched CdP for the expression *[FALAR] pelos cotovelos* [lit. speak through the elbows]. A total of 323 occurrences were found in CdP with the idiomatic meaning, confirming its established use in PTBR.

Non-idiomatic but conventional

The non-idiomatic but conventional category comprises translations in which the idiomatic meaning of the source expression is correctly interpreted but rendered through a non-idiomatic yet conventional expression in PTBR. The example below illustrates this category:

Student 19: “Ah, Sara, desculpa pela visita ter demorado tanto. O sr. Campbell não parava de falar”.

In this case, the student accurately captured the intended meaning of the source idiom, but opted for a literal rendering – *não parar de falar* [lit. not stop talking/speaking]. A search in the CdP returned 144 occurrences of this expression, confirming its conventional use.

Idiomatic but non-conventional

This category includes translations in which the source idiom is rendered by an IE in PTBR that is semantically appropriate, but lacks attested conventionality in reference corpora. The example below illustrates this category:

Student 23: “Ai, Sara, desculpe a visita ter levado tanto tempo. O Sr. Campbell ficou de palestra comigo”.

In this translation, the student successfully interpreted the idiomatic meaning of the source expression and employed the term *palestrinha* [lit. mini-lecture], an informal and often humorous or mildly pejorative expression used to describe someone who talks excessively or adopts an unsolicited, expert-like explanatory stance. In contemporary PTBR, the term circulates widely in social media discourse and is frequently associated with gendered uses, as in *macho palestrinha*, functioning in ways comparable to the notion of ‘mansplaining.’ Although the expression *ficar de*



palestrinha does not yield occurrences in CdP, it was classified as a creative and idiomatic equivalent, as it effectively conveys the figurative meaning of the original expression.

Inaccurate

The inaccurate category comprises translations in which the meaning of the source expression is not preserved in the target. The following example illustrates this category:

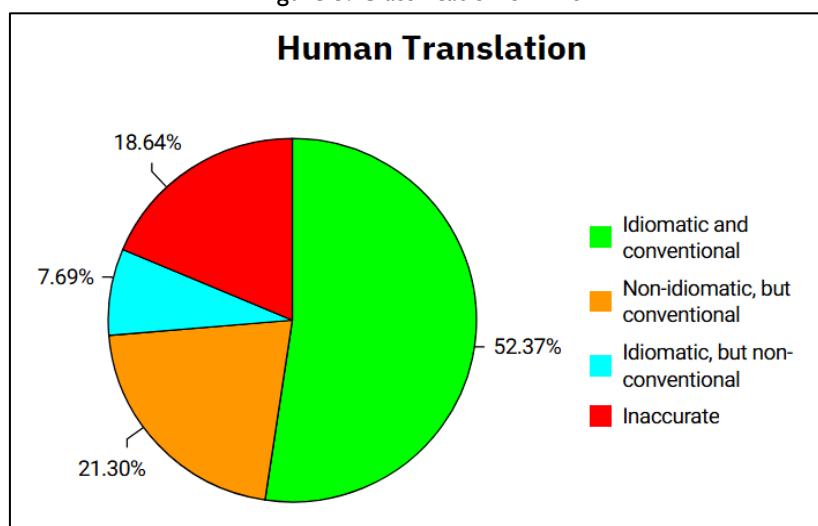
Student 5: “Ah, Sara, desculpa a visita ter demorado tanto. O sr. Campbell estava descendo a lenha em mim”.

The expression *descer a lenha* [lit. to strike with firewood] is a colloquial PTBR idiom meaning ‘to criticize harshly’ or ‘to attack verbally’, which does not correspond to the source idiom’s meaning.

4.1.1 Human translation

The analysis of the HTs revealed that 177 out of 338 translations were idiomatic and conventional, indicating that most students were able to recognize the figurative meaning of the expressions and render them with recurrent PTBR idioms. 72 were classified as non-idiomatic but conventional, which means that although these renderings were acceptable in PTBR, they did not preserve the figurative nature of the source idioms. A smaller portion – 26 translations – was identified as idiomatic but non-conventional, reflecting creative solutions that, while figurative, lacked recurrence in the corpus. The remaining 63 translation choices were considered inaccurate, showing that nearly one-fifth of the translations failed to convey the intended meaning of the source IE. Figure 5 summarizes these results:

Figure 5: Classification of HTs



Source: The authors

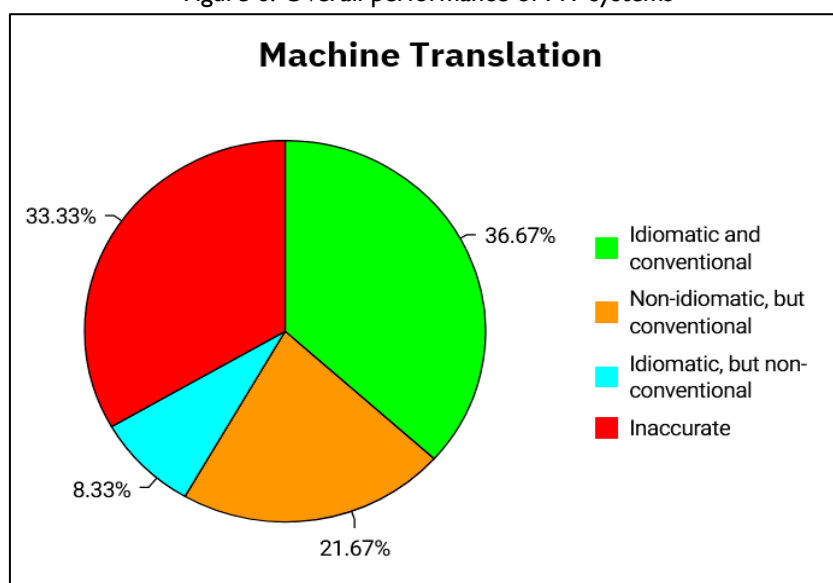
4.1.2 Machine translation

For this stage, the same 75 sentences translated by the HTs were submitted to MT. Four AI-based systems were selected, representing the two technological paradigms outlined earlier: LLM-based translation, represented by Gemini 2.5 Pro and ChatGPT 3.5; and NMT-based translation, represented by Google Translate and DeepL.

When considered jointly, the 300 translations produced by the four systems reveal substantial variation in how idiomatic meaning is handled across automatic translation paradigms. A total of 110 translations were classified as idiomatic and conventional, indicating that just over one third of the outputs successfully conveyed the figurative meaning of the source expressions through recurrent and natural equivalents in PTBR. 65 translations were categorized as non-idiomatic but conventional, meaning that although the target-language output was acceptable and fluent, the figurative meaning of the source idiom was neutralized or paraphrased.

A smaller subset of 25 translations fell into the idiomatic but non-conventional category, reflecting creative renderings that preserved figurative meaning, but relied on expressions not attested in the CdP. Finally, 100 translations (one third) were deemed inaccurate, as they not only failed to preserve the idiomatic meaning of the source expressions, but also resulted in semantic distortion or misinterpretation. Figure 6 summarizes the overall performance of the MT systems in percentage terms.

Figure 6: Overall performance of MT systems



Source: The authors

The following sections examine each system individually and then compare their performance.

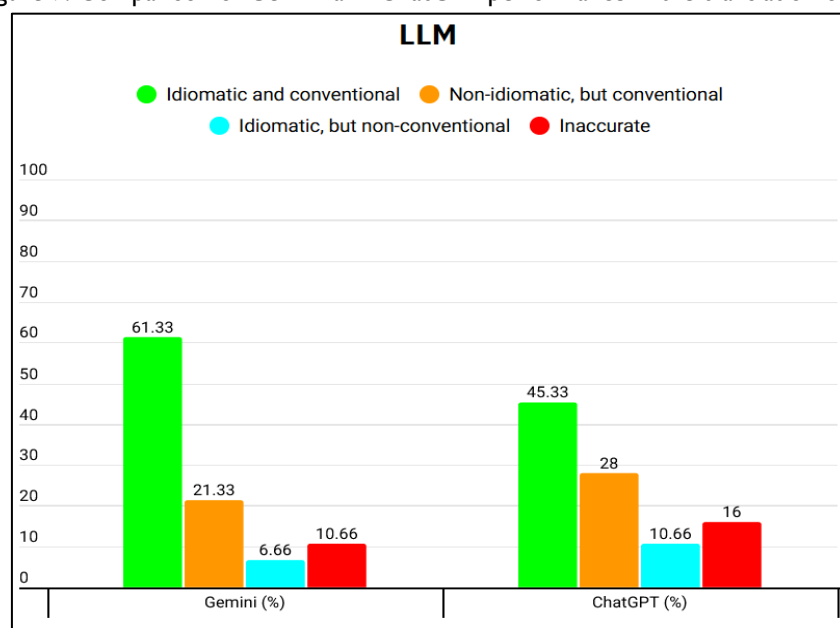
4.1.2.1 Machine translation by LLMs

When designing the prompts for the LLM-based translation, particular attention was paid to the segmentation of contexts containing IEs. Many of the selected idioms occurred in multi-sentence contexts, as in “I’m starving. Why don’t we go **grab a bite** at Kellogg’s?” (our emphasis). To ensure that the models correctly interpreted such sequences as a single context, sentences were grouped accordingly and each group was assigned a numerical identifier. This procedure prevented contextual overlap across different items and ensured that each idiom was processed independently. The LLMs were prompted as follows:

Translate the following groups of English sentences, each containing at least one food-related idiomatic expression, into Brazilian Portuguese. Ensure that the idiomatic meaning of the expressions is maintained by adapting them naturally to Brazilian Portuguese usage. Each group is numbered and should be treated independently; sentences within the same group are related to one another, but have no connection to the other groups. Use only the context provided within each group to guide your translation.

When the two LLMs are compared, Gemini outperforms ChatGPT in the translation of IEs, producing a higher proportion of idiomatic and conventional equivalents (46 versus 34). This result suggests that Gemini was more consistent in identifying idiomatic meaning and selecting established idiomatic equivalents in PTBR. ChatGPT, in turn, yielded a higher number of non-idiomatic but conventional translations (21 versus 16) and idiomatic but non-conventional renderings (8 versus 5), indicating a greater tendency toward creative solutions. Conversely, the proportion of inaccurate translations was higher for ChatGPT (12) than for Gemini (8). Figure 7 summarizes these results in percentage form.

Figure 7: Comparison of Gemini and ChatGPT performance in the translation of IEs

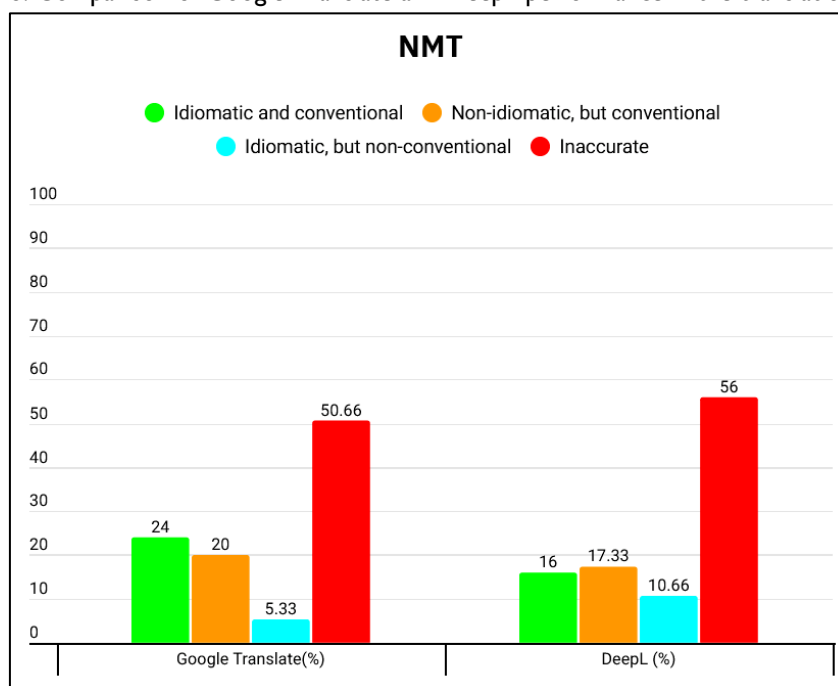


Source: The authors

4.1.2.2 Machine translation by NMT

Of 75 IEs translated by each system, Google Translate outperformed DeepL in producing (i) a higher proportion of idiomatic and conventional equivalents (18 versus 12), and (ii) a slightly greater number of non-idiomatic but conventional translations (15 versus 13). DeepL, by contrast, generated more idiomatic but non-conventional renderings (8 versus 4), suggesting a stronger tendency toward creative solutions. However, regarding translation inaccuracies, DeepL produced a higher number of errors than Google Translate (42 versus 38). Figure 8 presents these results in percentage form.

Figure 8: Comparison of Google Translate and DeepL performance in the translation of IEs

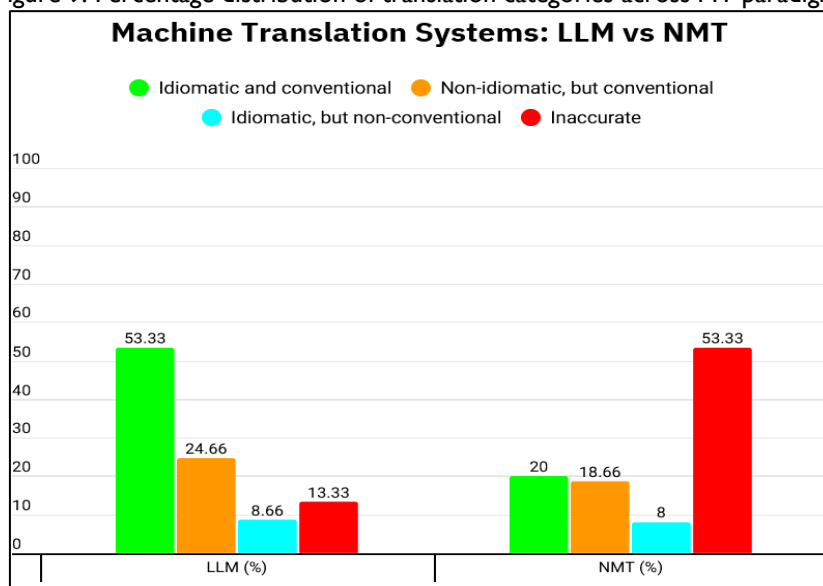


Source: The authors

4.1.2.3 LLM versus NMT

When the performance of the two system types – LLM-based and NMT-based translations – is examined separately, the results indicate that LLMs outperform NMT systems across most categories. LLMs produced a higher number of idiomatic and conventional equivalents (80 versus 30), as well as more non-idiomatic but conventional translations (37 versus 28) and idiomatic but non-conventional renderings (13 versus 12). In contrast, NMT systems generated four times as many inaccurate translations as LLMs (80 versus 20). Figure 9 presents a comparative overview of these distributions across MT systems in percentage terms, allowing for a more direct examination of how each paradigm performs with respect to idiomaticity, conventionality, and (in)accuracy.

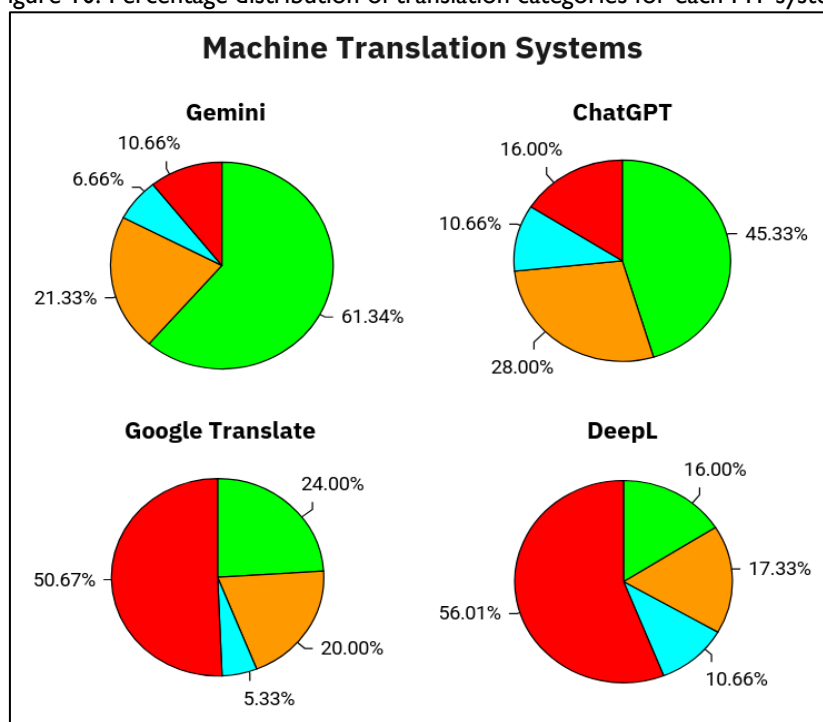
Figure 9: Percentage distribution of translation categories across MT paradigms



Source: The authors

To facilitate a more fine-grained comparison across MT systems, Figure 10 presents the distribution of translation categories for each system separately. By visualizing the results in percentage terms, it becomes possible to contrast system-specific patterns with respect to idiomaticity, conventionality, and (in)accuracy.

Figure 10: Percentage distribution of translation categories for each MT system



Source: The authors

Next, we compare human and machine translations, drawing attention to systematic differences in the strategies employed to translate the IEs.

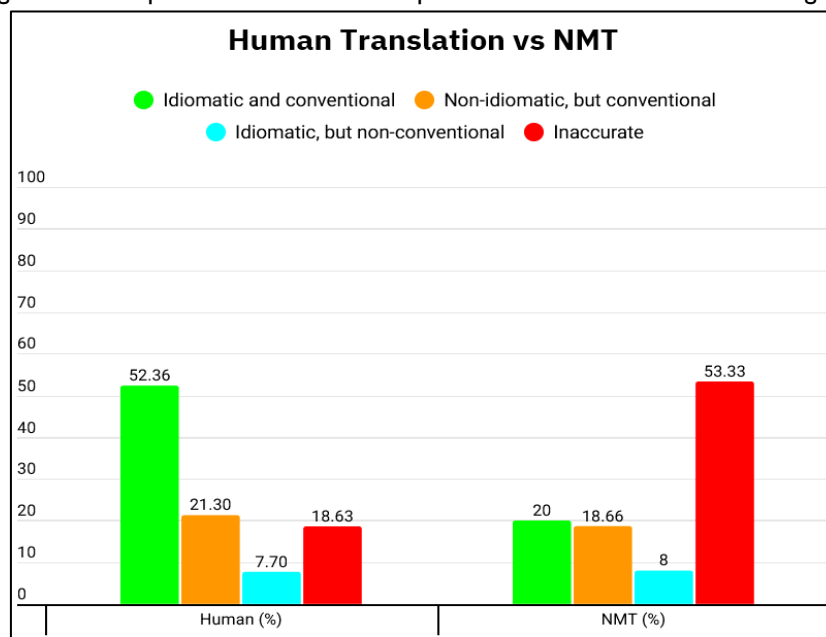
4.1.3 HT versus MT

When HT (Figure 5) is compared to MT altogether (Figure 6), clear differences emerge across translation categories. Human translations show a higher proportion of idiomatic and conventional renderings than MT (52.37% vs. 36.6%). Both approaches display similar proportions of non-idiomatic but conventional translations (21.3% for HT and 21.67% for MT) and idiomatic but non-conventional renderings (7.69% and 8.33%, respectively). In contrast, inaccurate translations are almost twice as frequent in MT as in HT (33.33% vs. 18.64%).

4.1.3.1 HT versus NMT systems

In the comparison between HT and NMT systems, trainee translators achieved considerably higher performance in producing idiomatic and conventional equivalents, while showing performance similar to Google Translate and DeepL in both non-idiomatic but conventional and idiomatic but non-conventional renderings. On the other hand, NMT systems produced almost three times as many inaccurate translations as HT. These results are summarized in Figure 11.

Figure 11: Comparison of HT and NMT performance across translation categories

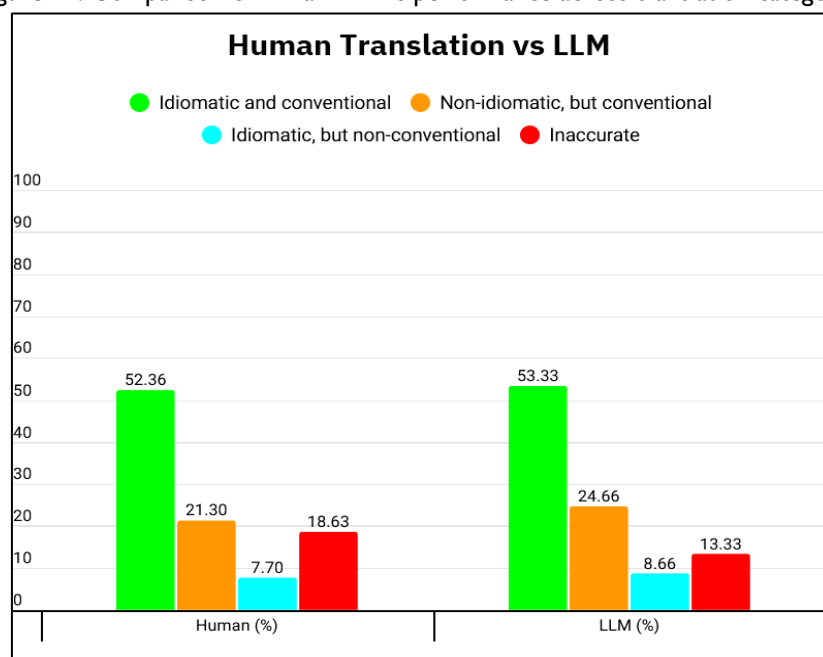


Source: The authors

4.1.3.2 HT versus LLMs

When the performance of HT is compared with that of the LLMs, the results indicate that the latter – largely driven by Gemini’s performance – slightly outperformed trainee translators in producing idiomatic and conventional equivalents, non-idiomatic but conventional translations, and idiomatic, but non-conventional equivalents. In addition, LLMs produced fewer inaccurate translations. These results are illustrated in Figure 12.

Figure 12: Comparison of HT and LLMs performance across translation categories



Source: The authors

In the following section, we examine whether the differences observed between HT and MT were statistically significant.

4.2 Quantitative analysis

To determine whether a statistically significant association exists between the type of translator and the quality of the translation, a chi-squared (χ^2) test of independence was conducted. This method is appropriate as the collected data consist of frequency counts for categorical variables. The two variables analyzed were:

1. Translator Type, with two levels: HT and MT; and
2. Translation Quality, with four levels based on the established criteria.

The analysis was performed under the following statistical hypotheses:

- Null Hypothesis (H_0): Translation quality is independent of the translator type. This implies that the proportion of translations in each quality category is the same for both HT and MT; and
- Alternative Hypothesis (H_a): Translation quality is dependent on the translator type. This implies that the distribution of translations across the quality categories differs between HT and MT.

The significance level (α) was set at 0.05, meaning that a difference was considered significant if there was less than a 5% likelihood that it occurred by chance, which is the standard threshold used in comparative studies. The observed frequencies from the sample were organized into a 2×4 contingency table, as shown in Table 1.

Table 1: Observed frequencies (X_{ij})

Translator type	Idiomatic and conventional	Non-idiomatic but conventional	Idiomatic but non-conventional	Inaccurate	Row total
HT	177	72	26	63	338
MT	110	65	25	100	300
Column Total	287	137	51	163	

Source: The authors

Based on the null hypothesis, the expected frequencies (E_{ij}) were calculated for each cell using the formula:

$$E_{ij} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

The resulting expected frequencies are shown in Table 2.

Table 2: Expected frequencies (E_{ij})

Translator type	Idiomatic and conventional	Non-idiomatic but conventional	Idiomatic but non-conventional	Inaccurate
HT	152.0470219	72.5799373	27.01880878	86.35423197
MT	134.9529781	64.4200627	23.98119122	76.64576803

Source: The authors

The Chi-Square test statistic (Q) was then computed to measure the discrepancy between observed and expected frequencies, according to the formula:

$$Q = \sum \frac{(X_{ij} - E_{ij})^2}{E_{ij}}$$

The calculated test statistic was $Q \approx 22.23$. With 3 degrees of freedom ($v = (\text{rows} - 1) \times (\text{columns} - 1) = (2 - 1) \times (4 - 1) = 3$) and a significance level of $\alpha = 0.05$, the critical value from the χ^2 distribution is 7.81.

Since the calculated value (22.23) is greater than the critical value (7.81), the null hypothesis was rejected. This indicates that there is a statistically significant association between the translator type and the quality of the translation. To further investigate the source of this dependency, two additional Chi-square analyses were conducted: one comparing HT with LLMs, and another comparing HT with NMT systems. These follow-up tests revealed no statistically significant difference between HTs and LLMs, but confirmed a substantial discrepancy between HTs and NMT systems.

An examination of the observed versus expected frequencies clarifies the nature of this difference. For NMT systems, the observed count in category 1 (30) was significantly lower than expected (~63), while the count in category 4 (80) was much higher than expected (~44). Conversely, for HT, the observed count in category 1 (177) exceeded the expected value (~143), while the count in category 4 (63) was well below the expected value (~99). This disparity indicates that HT demonstrates greater adequacy, with a significantly higher number of translations classified in the top category and a lower number in the 'Inaccurate' category compared to NMT systems.

4.3 Discussion

The findings of the present study are broadly consistent with previous research reporting that recent LLMs outperform conventional NMT systems in the translation of figurative language (Jiao et al., 2023; Mammadova, 2025; Rebecchi et al., 2024), while still falling short of human translators in certain aspects (Puppel & Borg, 2025; Rebecchi et al., 2025, among others). In particular, the substantially lower proportion of inaccurate translations produced by ChatGPT 3.5 and Gemini 2.5 Pro compared with Google Translate and DeepL suggests that the broader contextual representations learned by LLMs enable them to better recognize and interpret the non-compositional meaning of IEs. However, our results are also in consonance with previous studies by demonstrating that high translation accuracy does not necessarily preserve idiomaticity. Although LLMs frequently generated semantically appropriate translations, they were less likely than human translators to produce idiomatic and conventional renderings, often opting for acceptable, but more literal or less conventional alternatives. This distinction is especially relevant because translation quality involves not only conveying meaning correctly but also producing expressions that conform to target-language phraseological norms. By explicitly evaluating both idiomaticity and conventionality, the present study provides a more in-depth assessment of translation quality than studies relying primarily on correctness or fluency metrics.

5. Concluding remarks

This study set out to examine how IEs are handled by different translation agents – human translation (HT), represented by trainee translators, and machine translation (MT), represented by neural machine translation (NMT) systems and large language models (LLMs) – with particular attention to idiomaticity, conventionality, and accuracy. The results reveal clear differences across translation paradigms. Overall, LLMs demonstrated greater ability to preserve idiomatic meaning and select conventional equivalents in Brazilian Portuguese than NMT systems, while HT consistently outperformed NMT across all quality categories. The quantitative analysis confirmed that translation quality is significantly associated with translator type, with the most pronounced contrast observed between HT and NMT systems. No statistically significant difference was found between HT and LLMs, suggesting that recent generative models have substantially narrowed the gap traditionally observed between human and machine performance in the translation of idiomatic language. These findings reinforce the view that idiomaticity remains a demanding testing ground for trainee



translators as well as automatic translation technologies, while also highlighting the growing performance of LLMs in addressing figurative and culture-bound language.

Despite these contributions, the study presents some limitations. First, the human participants were trainee translators, which constrained the interpretation of the results, as their performance cannot be directly equated with that of professional translators whose experience may lead to different strategies and outcomes. Moreover, given their enrollment in the same undergraduate course, all trainee translators were treated as having comparable proficiency and exposure to translation practice. Individual differences in linguistic competence, translation experience, and familiarity with idiomatic language were not measured and may have influenced the results. Second, the relatively small number of participants limited the generalizability of the findings.

Future research should also examine linguistic factors that may influence the translation of IEs, but were not treated as independent variables in the present study. These include idiom frequency, degree of semantic opacity, conventionality, contextual support, and the availability of established target-language equivalents. By investigating these variables systematically, we may help explain why some IEs are successfully rendered by both human and machine translators, whereas others remain particularly challenging, and may provide a more fine-grained understanding of the conditions under which different translation paradigms succeed or fail.

Another direction for future research is to extend the investigation beyond the translation of pre-selected IEs to their automatic identification and translation in naturally occurring parallel data. Ongoing research by our group is pursuing this direction using an EN–PTBR parallel corpus of subtitles to investigate the ability of LLMs to identify IEs in context and retrieve their corresponding translations from target-language subtitles (Rebechi & Alves, 2026). This work is also being extended to translator training, with trainee translators evaluating and post-editing translations from different human and machine sources. Such studies may contribute to a better understanding not only of the extent to which current technologies can handle idiomatic language, but also of when and how human intervention is required and how critical engagement with machine-generated output can be incorporated into translator education.

From a pedagogical perspective, these findings stress the continued importance of translator training that foregrounds idiomaticity, conventional usage, and critical evaluation of machine-generated output. In an increasingly AI-mediated translation environment, developing translators' ability to work reflexively and collaboratively with advanced translation technologies emerges as a key component of professional competence.

References

- Alisoy, H. (2025). A Taxonomic Approach to Structural and Semantic Dimensions in English Phraseology. *Porta Universorum*, 1(4), 72–79. <https://doi.org/10.69760/portuni.0104007>
- Angelone, E. (2026). Generative AI as a facilitator of deliberate practice in translator training. In J. C. Penet, J. Moorkens & M. Yamada (Eds.), *Teaching translation in the age of generative AI: New paradigm, new learning?* (pp. 27–47). Language Science Press.
- Batista, J. S., Moreira de Oliveira, I., Tagnin, S. E. O., Teixeira, E. D., & Rebechi, R. R. (2026). Classificadores semânticos na identificação de equivalentes interlinguais para expressões



- idiomáticas: descascando esse abacaxi. *Letras & Letras*, 42(suppl.), 1–26. <https://doi.org/10.14393/LL63-v42S-20226-0>
- Baziotis, C., Mathur, P., & Hasler, Eva. (2023). Automatic Evaluation and Analysis of Idioms in Neural Machine Translation. In A. Vlachos & I. Augenstein (Eds.), *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 3682–3700). Association for Computational Linguistics.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Pearson Education Limited.
- Cambridge University Press. (n.d.). Bring home the bacon. In *Cambridge Dictionary*. <https://dictionary.cambridge.org/dictionary/english/bring-home-the-bacon>
- Castaldo, A., & Monti, J. (2024). Prompting Large Language Models for Idiomatic Translation. In B. Vanroy, M.-A. Lefer, L. Macken & P. Ruffo (Eds.), *Proceedings of the 1st Workshop on Creative-text Translation and Technology* (pp. 32–39). European Association for Machine Translation.
- Davies, M. (2008-). *The Corpus of Contemporary American English (COCA)*. <https://www.english-corpora.org/coca/>
- Davies, M. (2016-). *Corpus do Português: Web/Dialects*. <http://www.corpusdoportugues.org/web-dial/>
- Davies, M. (2012-2019). *Corpus do Português: NOW*. <https://www.corpusdoportugues.org/now/>
- Donthi, S., Spencer, M., Patel, O. B., Doh, J. Y., Rodan, E., Zhu, K., & O'Brien, S. (2025). Improving LLM abilities in idiomatic translation. In H. Hettiarachchi, T. Ranasinghe, P. Rayson, R. Mitkov, M. Gaber, D. Premasiri, F. A. Tan & U. Lasitha (Eds.), *Proceedings of the First Workshop on Language Models for Low-Resource Languages* (pp. 175–181).
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29–62. <https://doi.org/10.1515/text.1.2000.20.1.29>
- Hutchins, W. J. (1986). *Machine translation: Past, present, future*. Ellis Horwood.
- Jiang, R., Liu, X., Liu S., Wang, Y., Zhang, M., Tao, S., Wei, D., & Zhang, M. (2026). DeReA: Improving Idiom Translation with Detect-Retrieve-Arbitrate Reasoning. In M. Liakata, V. Moreira, J. Zhang & D. Jurgens (Eds.), *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics [Long Papers]* (pp. 6603–6621). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2026.acl-long.299>
- Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT a Good Translator? Yes with GPT-4 as the Engine. *arXiv:2301.08745v4*, 1–10. <https://doi.org/10.48550/arXiv.2301.08745>
- Jibreel, I. (2024). Translation Quality of Artificial Intelligence and Machine Translation Vs. Human Translation Utilizing MTPE Skills (An Empirical Study on Allusion Translation). *Journal of Social Studies*, 30(3), 46–72, <https://doi.org/10.20428/jss.v30i3.2545>
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In D. Jurafsky, J. Chai, N. Schluter & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6282–6293). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.560>

- Jurafsky, D., & Martin, J. H. (2026). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed). [Online manuscript]. <https://web.stanford.edu/~jurafsky/slp3/>
- Mammadova, I. (2025). Machine Translation vs. Human Translation: A Linguistic Analysis. *Porta Universorum*, 1(1), 26-31. <https://doi.org/10.69760/vhrq8s76>
- Massey, G., & Ehrensberger-Dow, M. (2026). Translation competence in the age of generative AI: Debates, dilemmas, directions. In J. C. Penet, J. Moorkens & M. Yamada (Eds.), *Teaching translation in the age of generative AI: New paradigm, new learning?* (pp. 3–25). Language Science Press.
- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics*. Cambridge University Press.
- Montesinos, L. S., Buján, S., Bardanca D., & Gamallo, P. (2026). Improving Machine Translation of Idioms: A Spanish–Galician Parallel Dataset and Synthetic Augmentation Approach. In M. Souza, I. de-Dios-Flores, D. Santos, L. Freitas, J. W. C. Souza & E. Ribeiro (Eds.), *Proceedings of the 17th International Conference on Computational Processing of Portuguese* (Vol. 1; pp. 980–987). Association for Computational Linguistics.
- Moreira de Oliveira, I. (2022). Expressões idiomáticas com a temática alimentação: uma proposta de glossário Português – Inglês. [Dissertação de Mestrado]. Universidade de Brasília.
- Moreira de Oliveira, I., & Teixeira, E. (2025). Expressões idiomáticas com a temática alimentação: um estudo contrastivo português inglês. In A. Husky, S. Mejri & C. S. N. Salvador (Orgs.), *A Fraseologia: O paradoxo do universal e do idiomático* (pp. 195–236). Pontes Editores.
- Musaad, M. M. A. M., & Al Towity, D. A. A. (2023). Translation Evaluation of Three Machine Translation Systems, with Special References to Idiomatic Expressions. *Humanities and Educational Sciences Journal*, 29, 678–708. <https://doi.org/10.55074/hesj.vi29.700>
- Naveen, P., & Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. *iScience*, 27(10), 1–25. <https://doi.org/10.1016/j.isci.2024.110878>
- Oliveira, M. L. S., Santos, A. A. O., & Andrade, L. J. S. (2026). Bridging Cultural Gaps in Automated Translation of Brazilian Expressions: A Study on Cultural Adaptation. In M. Souza, I. de-Dios-Flores, D. Santos, L. Freitas, J. W. C. Souza & E. Ribeiro (Eds.), *Proceedings of the 17th International Conference on Computational Processing of Portuguese* (Vol. 2; pp. 220–227). Association for Computational Linguistics.
- Pang, J., Ye, F., Wong, D. F., Yu, D., Shi, S., Tu, Z., & Wang, L. (2025). Salute the classic: Revisiting challenges of machine translation in the age of large language models. *Transactions of the Association for Computational Linguistics*, 13, 73–95. <https://doi.org/10.48550/arXiv.2401.08350>
- Penet, J. C., Moorkens, J., & Yamada, M. (Eds.). (2026). *Teaching translation in the age of generative AI: New paradigm, new learning?* Language Science Press. <https://doi.org/10.5281/zenodo.17580856>
- Puppel, M., & Borg, C. (2025). Evaluating ChatGPT's performance in creative text translation for communication: A case study from English into German. *Media and Intercultural Communication*, 3(1), 1–27. <https://doi.org/10.22034/mic.2024.480506.1023>
- Ramisch, C., Cordeiro, S., Zilio, L., Idiart, M., Villavicencio, A. (2016) How Naked is the Naked Truth? A Multilingual Lexicon of Nominal Compound Compositionality. In K. Erk & N. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*

(Vol. 2; pp. 156–161). Association for Computational Linguistics.
<https://doi.org/10.18653/v1/P16-2026>

- Rebechi, R., & Alves, D. A. (2026). A Comparative Study of LLMs for Idiomatic Expression Identification and Translation Equivalent Extraction in an English–Portuguese Subtitle Corpus. In M. Escribe, A. P. Izquierdo, C. Orăsan, T. Ranasinghe, G. C. Pastor, M. Tadic & R. Mitkov (Eds.), *Proceedings of the 3rd International Conference New Trends in Translation and Interpreting Technology 2026* (Vol. 1; pp. 54–62). Incoma.
- Rebechi, R. R., Marcon, N. O., & Faller, G. A. (2025). Tradução automática e tradução humana de expressões multipalavras: Descascando esse abacaxi. *ReVEL*, 23(44), 346–380.
- Rebechi, R. R., Marcon, N. O., Faller, G. A., & Villavicencio, A. (2024). Machine Translation of Multiword Expressions: Double Dutch or Crystal Clear? In C. Orăsan, T. Ranasinghe, G. C. Pastor, R. Mitkov, M. Kunilovskaya, V. Sosoni & M. Escribe (Eds.), *Proceedings of the Conference New Trends in Translation and Technology 2024: Translation in the AI age* (pp. 116–137). Incoma.
- Rebechi, R. R., & Trindade, E. (2021). Traduzir metáfora não é mamão com açúcar: A busca por equivalentes de botanomorfismos. *Polissema*, 21, 110–132.
- Rizki, K., & Masykuroh, Q. (2025). Evaluating ChatGPT's Translation of Harry Potter: A Qualitative Study of Translation Techniques, Accuracy, and Acceptability. *Journal of English Language Teaching and Literature*, 6(1), 181–192.
- Rothwell, A., Moorkens, J., Fernández-Parra, M., Drugan, J., & Austermuehl, F. (2023). *Translation tools and technologies*. Routledge.
- Schmid, H.-J. (2020). *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford University Press.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Tagnin, S. E. O. (2013). *O jeito que a gente diz: Combinações consagradas em inglês e português*. Disal.
- Xatara, C., Riva, H. C., & Rios, T. H. C. (2001). As dificuldades na tradução de idiomatismos. *Cadernos de Tradução*, 2(8), 183–194.

Editorial notes

Authorship contribution

Conceptualization: R. R. Rebechi, I. Moreira de Oliveira, M. A. Tedesco, I. F. Cardoso

Data collection: R. R. Rebechi, I. Moreira de Oliveira, M. A. Tedesco, I. F. Cardoso

Data analysis: R. R. Rebechi, I. Moreira de Oliveira, M. A. Tedesco, I. F. Cardoso

Results and discussion: R. R. Rebechi, I. Moreira de Oliveira, M. A. Tedesco, I. F. Cardoso

Writing – review and editing: R. R. Rebechi, I. Moreira de Oliveira, M. A. Tedesco

Research dataset

The research data were collected as part of the project “Machine Translation of Multiword Expressions in the English–Portuguese Language Pair”, conducted at the Federal University of Rio Grande do Sul (UFRGS).

Funding

This research was initiated during the first author’s postdoctoral fellowship at the University of Sheffield in 2023, funded by a CAPES-PrInt scholarship (grant no. 88887.838833/2023-00).



Image copyright

Not applicable.

Approval by ethics committee

This research was approved by the Research Ethics Committee of the Federal University of Rio Grande do Sul (CAAE: 83967624.8.0000.5347).

Conflict of interest

Not applicable.

Data availability statement

The data generated in this study that are not included in this article may be made available by the authors upon reasonable request.

License

The authors grant *Cadernos de Tradução* exclusive rights for first publication, while simultaneously licensing the work under the [Creative Commons Attribution](#) (CC BY) 4.0 International license. This license enables third parties to remix, adapt, and create from the published work, while giving proper credit to the authors and acknowledging the initial publication in this journal. Authors are permitted to enter into additional agreements separately for the non-exclusive distribution of the published version of the work in this journal. This may include publishing it in an institutional repository, on a personal website, on academic social networks, publishing a translation, or republishing the work as a book chapter, all with due recognition of authorship and first publication in this journal.

Publisher

Cadernos de Tradução is a publication of the Graduate Program in Translation Studies at the Federal University of Santa Catarina. The journal *Cadernos de Tradução* is hosted by the [Portal de Periódicos UFSC](#). The ideas expressed in this paper are the responsibility of its authors and do not necessarily represent the views of the editors or the university.

Section editors

Andréia Guerini – Willian Moura

Style editors

Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

Article history

Received: 27-01-2026

Approved: 07-08-2026

Revised: 03-09-2026

Published: 09-2026

