# Bridging Cultural Gaps in Automated Translation of Brazilian Expressions: A Study on Cultural Adaptation

**Maria Luiza Silva de Oliveira[1], Andressa Andrade Oliveira dos Santos[1]**
**Leandro Jose Silva Andrade[2]**

[1]Federal University of Bahia (UFBA), Department of Executive Secretariat, Salvador, Brazil
[2]Federal University of Bahia (UFBA), Salvador, Brazil

mlso97@outlook.com, andressaandrade1810@gmail.com, leandrojsa@ufba.br

## Abstract

Automated translation systems exhibit a tendency toward cultural drift when processing non-literal language, often favoring standardized outputs that diverge from the original pragmatic intent. Although Large Language Models (LLMs) have introduced more sophisticated context-handling capabilities, the transition from literal decoding to effective cultural adaptation remains inconsistent. This study investigates these linguistic detours by evaluating ChatGPT-4o, Gemini 1.5 Pro, and Google Translate using a corpus of 100 Brazilian Portuguese expressions. To ensure contemporary relevance, the expressions were validated through *Corpus Carolina* and categorized into four groups: classical idioms, lexically localized expressions, metaphors, and intensifiers. Translation quality was assessed using the Multidimensional Quality Metrics (MQM) framework, focusing on adequacy, fluency, and cultural adaptation. The analysis reveals that, even when grammatical accuracy is achieved, automated systems frequently overlook the sociocultural weight embedded in the source language. Such semantic shifts pose significant challenges in high-stakes professional communication, where nuanced mediation is essential. The findings underscore the limitations of current AI systems in cultural competence and reinforce the ongoing necessity of human intervention to bridge the gap between algorithmic processing and regional identity.

## 1 Introduction

The rapid advancement of Artificial Intelligence (AI) has significantly reshaped the landscape of Machine Translation (MT), expanding its role in professional, institutional, and organizational communication. Recent advances in Large Language Models (LLMs), particularly those based on attention mechanisms, have led to substantial improvements in fluency and contextual coherence in automated translation systems (Vaswani et al., 2017). Despite these advances, the translation of non-literal language—particularly idiomatic expressions, sociolinguistically marked expressionss, metaphors, and pragmatic intensifiers—continues to pose substantial challenges. These linguistic elements are deeply embedded in cultural and social contexts, making them especially vulnerable to semantic distortion when processed by automated systems. While contemporary studies emphasize the growing contextual sensitivity of LLM-based translators, a persistent gap remains between grammatical accuracy and effective cultural adaptation. Automated systems often prioritize standardized or neutralized outputs, leading to what can be described as cultural drift: a gradual displacement of the original pragmatic intent and socio-cultural resonance of the source text. Such deviations may appear subtle at the lexical level but can generate significant communicative breakdowns, particularly in high-stakes professional environments where meaning negotiation and cultural awareness are essential. This issue is especially relevant in multilingual organizational settings, where translation mediates not only linguistic exchange but also institutional credibility, interpersonal relationships, and decision-making processes. In professional fields such as Executive Secretariat, translation tasks frequently involve culturally marked language used in negotiations, internal communication, and external representation. In these contexts, inadequately adapted translations may compromise clarity, tone, and cultural appropriateness, reinforcing the need for critical human mediation even in technologically advanced translation workflows. Against this backdrop, the present study investigates the limitations of automated cultural adaptation in MT by conducting a comparative evaluation of three widely used systems: ChatGPT-4o, Gemini 1.5 Pro, and Google Translate. The analysis is based on a corpus of 100 Brazilian Portuguese expressions, validated



*Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026)*
*Vol. 2*, pages 220–227

through the *Corpus Carolina* (Projeto Corpus Carolina, 2020) to ensure contemporary and authentic language use. The expressions are categorized into four linguistic groups—classical idioms, lexically localized expressions, metaphors, and intensifiers—allowing for a systematic examination of how different forms of non-literal language are handled by each system. Translation quality is assessed using the Multidimensional Quality Metrics (MQM) framework, with particular emphasis on adequacy, fluency, and cultural adaptation. By identifying recurrent patterns of semantic shift and cultural neutralization across the evaluated systems, this study aims to delineate the boundaries of current AI-driven translation technologies. The findings contribute to ongoing discussions on the cultural competence of MT systems and underscore the enduring importance of human expertise in ensuring communicative effectiveness across languages and cultures.

## 2 Methodology

This study adopts an applied, descriptive, and comparative research design aimed at evaluating the capacity of automated translation systems to handle culturally and sociolinguistically embedded, non-literal language. The analysis focuses on identifying patterns of semantic shift and cultural neutralization produced by different Machine Translation (MT) architectures when translating Brazilian Portuguese expressions into English.

### 2.1 Corpus Selection and Validation

The dataset consists of 100 non-literal expressions originally formulated in Brazilian Portuguese. These expressions were selected to represent recurrent forms of culturally and socially embedded language commonly used in both professional and informal communication contexts. To ensure linguistic authenticity and contemporary relevance, all expressions were validated through the *Corpus Carolina*, a large-scale corpus of Brazilian Portuguese that reflects current language usage across multiple registers. This validation step was employed to confirm that the selected expressions are attested in real communicative contexts rather than artificially constructed examples. Following validation, the expressions were organized into four analytically grounded linguistic categories:

(i) Idiomatic expressions, defined as multiword units whose meanings are not compositionally derived from their individual lexical components (Tagnin, 2013);

(ii) Sociolinguistically marked expressions, referring to expressions associated with specific social groups, communicative contexts, or identities, reflecting structured linguistic variation rather than deviation from a standard norm (Bagno, 2007);

(iii) Metaphorical expressions, understood as linguistic realizations of conceptual mappings between domains, in line with Conceptual Metaphor Theory (Lakoff and Johnson, 1980);

(iv) Intensifiers, defined as lexical or phrasal elements that amplify or modulate the degree of a predicate.

Although some overlap between categories is possible, particularly between idiomatic and sociolinguistically marked expressions, each item was classified according to its dominant linguistic function to ensure analytical consistency. This categorization adopts a descriptive sociolinguistic perspective, in which linguistic variation is treated as a legitimate and meaningful component of language use, rather than as deviation from a standardized variety.

### 2.2 Evaluated Translation Systems

Three widely used MT systems were selected for comparative evaluation: ChatGPT-4o, Gemini 1.5 Pro, and Google Translate. ChatGPT-4o and Gemini 1.5 Pro represent Large Language Model (LLM) based systems, characterized by their ability to generate context-aware and fluent output through neural generative mechanisms. Google Translate was included as a representative of a conventional MT engine, serving as a baseline for comparison. All translations were generated under default system settings, without user intervention or post-editing, in order to reflect typical real-world usage scenarios. Each source expression was translated independently into English by all three systems.

### 2.3 Machine Translation Systems and Prompting Strategy

To ensure methodological consistency and avoid bias across translation systems, a minimal and uniform interaction strategy was adopted. For Large Language Models (ChatGPT and Gemini), translations were generated using a direct and uncontextualized prompt equivalent to the instruction "Translate this sentence," followed by the source expression. No additional contextual information, examples, or clarifications were provided. This decision was motivated by the need to maintain comparability with Google Translate, which does not allow prompt-based interaction beyond direct input. By restricting LLM usage to a basic translation command, the study sought to minimize potential advantages associated with prompt engineering and to ensure that observed differences in output could be attributed to the underlying system architectures rather than to interaction design. All translations were generated under default system settings, without manual intervention or post-editing. This approach aligns with the study's objective of evaluating how automated translation systems handle culturally marked expressions under typical user conditions.

### 2.4 Evaluation Framework

Translation quality was assessed using the Multidimensional Quality Metrics (MQM) framework (Lommel et al., 2014), which enables fine-grained evaluation across multiple linguistic dimensions. Building on the methodological procedures established in the original undergraduate thesis, the present study operationalized MQM through a structured scoring scheme designed to capture both linguistic accuracy and cultural adequacy. For analytical purposes, the evaluation focused on four dimensions: adequacy, fluency, coherence and cohesion, and cultural adaptation. Adequacy refers to the extent to which the translated output preserves the semantic content of the source expression. Fluency assesses grammatical correctness and naturalness in the target language. Coherence and cohesion examine internal textual consistency, while cultural adaptation evaluates whether the translation conveys the pragmatic intent and sociocultural meaning embedded in the original expression. Each dimension was scored on a scale ranging from 0 to 3, where 0 indicates a complete failure to meet the criterion and 3 represents optimal performance. After the evaluation stage, all scores were systematically organized into comparative spreadsheets, allowing both quantitative aggregation and qualitative inspection of recurrent error patterns across categories and systems. The final score of each translated expression was calculated using a simple arithmetic mean, as shown in Equation (1):

$$M = \frac{A + F + C + AC}{4} \tag{1}$$

where $M$ represents the final mean score of the translation, $A$ denotes adequacy, $F$ fluency, $C$ coherence and cohesion, and $AC$ cultural adaptation.

To obtain the overall performance score of each machine translation system (ChatGPT, Gemini, and Google Translate), the individual mean scores of all evaluated expressions were summed and divided by the total number of expressions analyzed, as expressed in Equation (2):

$$M_G = \frac{\sum M_i}{n} \tag{2}$$

where $M_G$ corresponds to the global mean score of the system, $M_i$ refers to the mean score of each individual translation, $n$ represents the total number of evaluated expressions, and $\sum$ indicates summation. This quantitative procedure enabled the identification of numerical performance trends across systems and categories, which were subsequently interpreted through qualitative analysis in the Results and Discussion section, with particular attention to patterns of cultural drift and pragmatic loss.

### 2.5 Analytical Procedure

The analysis was conducted by comparing source expressions and their respective translations into English within each category. Particular attention was given to cases in which grammatical accuracy was achieved at the expense of cultural or pragmatic equivalence. Examples illustrating successful and unsuccessful adaptations in both target languages were selected to support the discussion of system behavior. This approach enables a nuanced understanding of how different MT architectures handle culturally marked language and highlights the boundaries of automated cultural competence in multilingual professional communication contexts.

### 2.6 Related Work

Research on Machine Translation (MT) has evolved significantly over recent decades, transitioning from rule-based systems to statistical and neural approaches. While early MT models relied on explicit linguistic rules and bilingual lexicons, limiting their capacity to address ambiguity and contextual variation (Hutchins, 2001), statistical models introduced corpus-driven learning, improving scalability but still facing challenges related to fluency and long-distance dependencies (Koehn, 2010). The emergence of Neural Machine Translation (NMT), particularly through encoder–decoder architectures and attention mechanisms, marked a turning point in the field by enabling more context-sensitive translations (Bahdanau et al., 2015). The Transformer architecture further advanced this paradigm by relying on self-attention to model linguistic relationships more effectively across languages (Vaswani et al., 2017). More recently, Large Language Models (LLMs) have incorporated translation as part of broader generative frameworks trained on massive multilingual datasets. Although these systems often demonstrate substantial improvements in grammaticality and surface-level coherence, growing evidence suggests that such gains do not necessarily extend to pragmatic adequacy or cultural fidelity. Neural and LLM-based systems have been shown to normalize culturally and sociolinguistically marked expressions, favoring standardized renderings that may obscure pragmatic intent and cultural specificity (Toral et al., 2018). From a sociolinguistic perspective, such normalization may contribute to the attenuation of linguistic diversity and the underrepresentation of socially situated language use. These limitations are particularly salient in the translation of idiomatic expressions, metaphors, and sociolinguistically marked expressions, which rely on shared cultural knowledge rather than compositional meaning. Idiomatic expressions, in particular, have been widely studied in phraseology and are characterized by their non-compositional meaning, as discussed by Tagnin (2013). Within Translation Studies, meaning has long been understood as inseparable from its social and cultural context. From this perspective, translation quality cannot be evaluated solely through linguistic accuracy, but must also consider pragmatic intent, cultural resonance, and communicative purpose (Nida, 1964; House, 2015). Automated translation systems, however, tend to operationalize quality through formal equivalence and fluency-based metrics, often overlooking context-dependent and culturally embedded meanings (Pym, 2010). From a sociolinguistic perspective, linguistic variation is understood as a structured and meaningful phenomenon rather than deviation from a standard norm. In the Brazilian context, studies by Bagno (2007) and Bortoni-Ricardo (2004) emphasize that language reflects social identity, communicative practices, and power relations. This perspective is particularly relevant for the analysis of non-literal and socially marked expressions, which rely on shared cultural knowledge and contextual interpretation. Recent Brazilian scholarship has contributed important empirical and theoretical insights into the use, perception, and evaluation of machine translation in real communicative contexts. Nouatin and Parreiras (Nouatin and Parreiras, 2021) investigate the role of machine translation in the teaching and learning of non-native languages, highlighting teachers' perceptions and attitudes toward automated translation tools. Their findings emphasize that, while MT systems are increasingly present in educational and professional environments, their outputs frequently require critical mediation due to limitations in pragmatic and contextual adequacy. Similarly, Esqueda (Esqueda, 2021) discusses pedagogical and cognitive challenges associated with the use of machine translation, underscoring the need for human interpretive competence to contextualize and refine automated outputs. These Brazilian studies reinforce broader concerns regarding the gap between linguistic fluency and communicative effectiveness in automated translation systems. They also foreground the importance of qualitative evaluation approaches that account for discourse, pragmatics, and cultural meaning, dimensions that are often underrepresented in automatic metrics. In this regard, the Multidimensional Quality Metrics (MQM) framework has emerged as a robust alternative, offering a fine-grained taxonomy of error categories across linguistic, semantic, and pragmatic dimensions (Lommel et al., 2014). MQM enables systematic qualitative analysis of translation behavior, making it particularly suitable for studies concerned with culturally marked language. Corpus-based approaches have likewise played a central role in advancing MT research and evaluation. The use of curated and linguistically validated corpora allows researchers to ground their analyses in authentic language use and contemporary discourse patterns. In the Brazilian context, the *Corpus Carolina* has been employed as a reference for contemporary Brazilian Portuguese usage, providing a reliable basis for the selection and validation of idiomatic and culturally situated expressions (Projeto Corpus Carolina, 2020). Such resources are especially relevant for studies examining non-literal language, as they help ensure that analyzed expressions reflect current linguistic practices rather than prescriptive or outdated forms. Despite growing interest in culturally informed evaluation and the increasing availability of neural and LLM-based translation systems, comparative studies that systematically examine how different MT architectures handle non-literal language across multiple categories remain limited. Existing re-

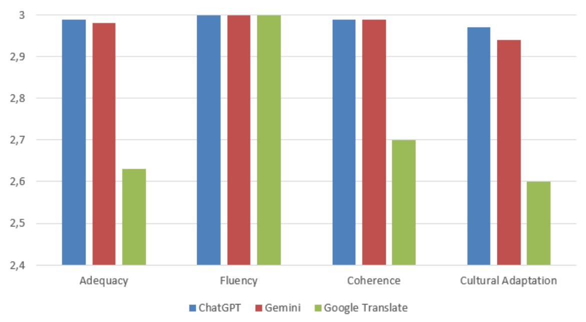


Figure 1: Average dimension scores per tool (Scale 1-3) across the 100-item corpus.

search often prioritizes sentence-level accuracy or post-editing effort, leaving open questions regarding broader patterns of cultural drift in automated translation. This study addresses this gap by analyzing how distinct MT systems process idioms, sociolinguistically marked expressions, metaphors, and intensifiers in Brazilian Portuguese, combining a corpus-based methodology with a qualitatively grounded evaluation framework.

# 3 RESULTS AND DISCUSSION

## 3.1 Overall Patterns Observed

This section presents and discusses the results obtained from the MQM-based evaluation of automated translations of Brazilian Portuguese expressions into English. To ensure analytical focus and conciseness, the analysis concentrates on a selected subset of the corpus, consisting of representatives expressions per category (idioms, lexically localized expressions, metaphors, and intensifiers/comparative expressions). While all selected expressions were evaluated according to the same criteria, the presentation of results prioritizes the most illustrative cases in each category, as is common practice in full-length conference papers to avoid excessive extension.

Figure 1 2 presents the mean scores for cultural adaptation obtained by each translation system across all evaluated expressions. The results indicate systematic differences in how the systems handle culturally embedded language. LLM-based systems (ChatGPT-4o and Gemini 1.5 Pro) achieved higher average scores in fluency and adequacy, reflecting their ability to generate grammatically natural and contextually coherent output. However, gains in linguistic quality did not consistently translate into effective cultural adaptation. Google Translate, while exhibiting lower fluency scores overall, demonstrated more stable behavior in literal equivalence, albeit frequently failing to convey pragmatic intent in non-literal expressions. These findings suggest that improvements in surface-level coherence do not necessarily correspond to better handling of cultural meaning, reinforcing concerns raised in recent MT research regarding semantic normalization in neural systems.

## 3.2 Idiomatic Expressions

Idiomatic expressions constitute one of the most challenging categories for automated translation, as their meanings are not compositionally derived from individual lexical items. Table 1 presents representative examples of idiomatic expressions and their corresponding MQM-based scores.

| Expression | System | Translation | Score |
|---|---|---|---|
| Chutar o balde | ChatGPT | give up completely | 2.25 |
| | Gemini | lose patience | 2.00 |
| | Google Translate | kick the bucket | 0.75 |
| Acabar em pizza | ChatGPT | end with no consequences | 2.50 |
| | Gemini | end in nothing | 1.75 |
| | Google Translate | end in pizza | 0.50 |
| Engolir sapo | ChatGPT | put up with something unfair | 2.25 |
| | Gemini | swallow an insult | 2.00 |
| | Google Translate | swallow a frog | 0.75 |

Table 1: MQM scores for selected idiomatic expressions

Table 2 summarizes the translation results for the selected idiomatic expressions (*Chutar o balde, Acabar em pizza, Engolir sapo, Amigo da onça, Quebrar o galho*). Across systems, idioms posed substantial challenges, particularly with respect to cultural adaptation. Literal translations were frequent, especially in cases where idiomatic equivalents exist in the target language but require pragmatic inference rather than compositional decoding. ChatGPT-4o demonstrated relatively higher adequacy scores for idioms with close functional equivalents, such as *Quebrar o galho*. Nevertheless, even in these cases, translations often favored explanatory paraphrases over idiomatic substitutions, reducing pragmatic force. Gemini 1.5 Pro exhibited similar tendencies, while Google Translate consistently produced literal renderings, resulting in lower cultural adaptation scores. These results align with previous findings that idiomaticity remains a persistent weakness in automated translation systems.

### 3.3 Sociolinguistically Marked Expressions

Sociolinguistically marked expressions reflect patterns of linguistic variation associated with specific social groups, communicative contexts, and cultural practices. From a sociolinguistic perspective, such variation is understood as a structured and meaningful component of language use rather than deviation from a standardized norm (Bagno, 2007; Bortoni-Ricardo, 2004). These expressions are therefore especially sensitive to normalization strategies in automated translation. Table 2 summarizes representative translations of Brazilian Portuguese expressions exhibiting sociolinguistic marking. In this study, these expressions are identified based on their distribution across communicative contexts and their attestation in corpus data, rather than on prescriptive or geographically restrictive criteria. In this study, the term “regionalism” and "sociolinguistically marked expressions" is not employed in a normative or hierarchical sense; instead, it functions as an operational label for lexical items with predominantly localized distribution, supported by corpus-based evidence.

| Expression | System | Translation | MQM Mean |
|---|---|---|---|
| Oxente | ChatGPT | wow/really? | 2.00 |
| | Gemini | what? | 1.75 |
| | Google Translate | oxente | 0.50 |
| Arretado | ChatGPT | impressive | 2.25 |
| | Gemini | intense | 1.75 |
| | Google Translate | angry | 1.00 |
| Migué | ChatGPT | an excuse/deception | 2.00 |
| | Gemini | trick | 1.75 |
| | Google Translate | migué | 0.50 |

Table 2: MQM mean scores for selected Brazilian Portuguese regionalisms.

The analysis reveals a strong tendency toward cultural neutralization. While LLM-based systems attempted contextual approximation, they frequently diluted regional identity. Google Translate systematically failed to interpret sociolinguistically marked expressions, often leaving terms untranslated or assigning semantically unrelated meanings.

### 3.4 Metaphorical Expressions

Metaphorical expressions displayed intermediate levels of difficulty. As shown in Table 3, some metaphors were rendered through literal transfer, while others were paraphrased, affecting expressive intensity.

| Expression | System | Translation | MQM Mean |
|---|---|---|---|
| Luz no fim do túnel | ChatGPT | light at the end of the tunnel | 2.75 |
| | Gemini | hope ahead | 2.25 |
| | Google Translate | light at the end of the tunnel | 2.50 |
| Fogo de palha | ChatGPT | short-lived enthusiasm | 2.25 |
| | Gemini | something temporary | 2.00 |
| | Google Translate | straw fire | 1.00 |

Table 3: MQM mean scores for selected metaphorical expressions.

The final category revealed marked divergence across systems. Highly figurative comparisons posed particular difficulties, frequently resulting in either literal translations or neutral paraphrases. ChatGPT-4o occasionally generates culturally adapted equivalents, especially for expressions with approximate analogs in English. However, these adaptations were inconsistent and often depended on implicit inference rather than systematic cultural mapping. Google Translate, by contrast, largely failed to capture the evaluative force of intensifiers, producing translations that conveyed factual meaning without pragmatic emphasis.

### 3.5 Cross-System Comparison and Implications

Overall, ChatGPT-4o and Gemini demonstrated greater contextual flexibility, achieving higher adequacy and fluency scores. However, both systems exhibited a tendency toward cultural neutralization, which reduced region-specific and idiomatic features. Google Translate, while more consistent in literal transfer, showed limited capacity for pragmatic interpretation. These patterns highlight the limitations of current automated translation systems in high-stakes communicative contexts, where cultural mediation plays a central role. The findings reinforce the need for human intervention in

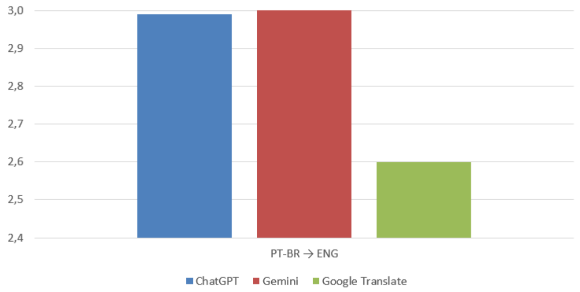


Figure 2: Mean Cultural Adaptation scores per MT system

professional domains, where accurate interpretation of culturally embedded language is essential for effective communication.

### 3.6 Examples of Cultural Mismatch

MT tools often fail to preserve the cultural nuances of source texts, leading to misinterpretations or loss of intended meaning. Table 2 illustrates some examples of Brazilian Portuguese expressions translated into English, highlighting cases of cultural mismatch and inadequate adaptation. These examples demonstrate how literal translations or lack of context can distort the original message.

| Original Expression | Tool | Translation | Justification |
|---|---|---|---|
| Moscou | ChatGPT | Moscow | Confuses Brazilian slang (“distraiu / vacilou”) with proper noun “Moscow.” Ideal: “He/She zoned out / slipped up.” |
| Chutar o balde | ChatGPT | Kick the bucket | Literal translation misleads; English idiom refers to dying instead of “giving up.” |
| Ficar de molho | Google Translate | Stay in sauce | Literal translation loses the meaning of resting or recovering. |
| Pagar o pato | ChatGPT | Pay the duck | Literal translation loses idiomatic meaning (“take the blame”). |
| Dar com os burros n’água | Gemini | Give the donkeys in the water | Literal translation misinterprets the expression meaning (“fail”). |

Table 4: Examples of cultural mismatch in machine translation

### 3.7 Professional and Organizational Implications

Although artificial intelligence models demonstrate superior performance in terms of fluency and processing speed, the findings of this study indicate that they do not fully replace human judgment in tasks involving culturally nuanced language. It is important to emphasize that this outcome is directly related to the methodological choices adopted, as translations were generated from isolated expressions, without sentential context, and through minimal prompts, with no iterative refinement or interaction. These controlled methodological decisions highlight how the absence of linguistic and pragmatic context can significantly affect the cultural interpretation of non-literal expressions by automated systems. These results support the argument proposed by Amini et al. (2024), who frame machine translation as a collaborative rather than a substitutive technology. In this perspective, automated systems function as productivity-enhancing tools whose outputs still require human mediation, particularly in communicative situations where pragmatic intent and cultural meaning are central. The implications of these findings are especially relevant for professionals engaged in multilingual information management and intercultural communication, including translators, interpreters, language educators, and, notably, Executive Secretariat professionals. As emphasized by Florido (2021), executive secretaries act as cultural mediators and strategic information managers within organizational contexts, playing a key role in companies that operate internationally. Consequently, developing critical awareness of the limitations and affordances of machine translation systems becomes an essential professional competence for ensuring communicative accuracy and cultural adequacy in documents, meetings, negotiations, and corporate correspondence. From an organizational perspective, the results suggest that companies operating in multicultural environments may benefit from adopting structured evaluation frameworks such as MQM to assess the quality of machine-translated content used in reports, emails, contracts, and institutional materials. Such practices contribute to the standardization of intercultural communication and reduce risks associated with misinterpretation. Furthermore, the strategic use of well-designed prompts in AI-based translation tools can improve semantic and pragmatic alignment with the intended communicative context. Nevertheless, even when productivity gains and cost reductions are achieved, effective deployment of these technologies still depends on continuous monitoring and qualified human mediation to ensure that culturally embedded meanings are preserved throughout the translation process.

## 4 Conclusion

This study investigated how contemporary automated translation systems handle culturally embedded Brazilian Portuguese expressions, focusing on idioms, sociolinguistically marked expressions, metaphors, and intensifiers. Using a corpus-based selection validated through the *Corpus Carolina* and an MQM-inspired evaluation framework, the analysis revealed that advances in fluency and grammatical accuracy do not necessarily translate into effective cultural adaptation. Even when translations were semantically adequate, pragmatic intent and socio-cultural resonance were frequently attenuated. The results demonstrate that LLM-based systems such as ChatGPT-4o and Gemini 1.5 Pro generally outperform conventional MT engines in terms of fluency and overall adequacy. However, these systems often favor paraphrasing and semantic normalization, which reduces idiomatic force and cultural specificity. Google Translate, while more consistent in literal equivalence, systematically failed to convey non-literal meaning, particularly in expressions that depend on shared cultural knowledge. From a professional perspective, especially in fields that require intercultural mediation, these findings highlight the continued necessity of human intervention in high-stakes communication. Automated translation tools function more effectively as support technologies rather than substitutes for culturally informed mediation. Future research may expand the corpus size, explore additional language pairs, and investigate prompt-sensitive evaluation strategies to further examine the boundaries of automated cultural adaptation.

## References


Marcos Bagno. 2007. *Preconceito linguístico: o que é, como se faz*. Loyola.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations*.

Stella Maris Bortoni-Ricardo. 2004. *Educação em língua materna: a sociolinguística na sala de aula*. Parábola.

Marileide Dias Esqueda. 2021. Machine translation: teaching and learning issues. *Trabalhos em Linguística Aplicada*, 60(1):282–299.

Juliane House. 2015. *Translation Quality Assessment: Past and Present*. Routledge.

John Hutchins. 2001. Machine translation and human translation: In competition or cooperation? *International Journal of Translation*, 13(1).

Philipp Koehn. 2010. *Statistical Machine Translation*. Cambridge University Press.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press.

Arle Lommel, Aljoscha Burchardt, and Hans Uszkoreit. 2014. Multidimensional quality metrics (mqm): A framework for declaring and describing translation quality metrics. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*.

Eugene A. Nida. 1964. *Toward a Science of Translating*. Brill.

Gbènoukpo Gérard Nouatin and Vicente Aguimar Parreiras. 2021. Tradução automática no ensino e na aprendizagem de línguas não maternas: percepções, atitudes e opiniões de professores. *Trabalhos em Linguística Aplicada*, 60(3):841–852.

Projeto Corpus Carolina. 2020. Corpus carolina: um corpus de referência do português brasileiro contemporâneo. Disponível para pesquisa linguística e validação de uso contemporâneo.

Anthony Pym. 2010. *Exploring Translation Theories*. Routledge.

Stella Esther Ortweiler Tagnin. 2013. *O jeito que a gente diz: combinações consagradas em inglês e português*. Disal.

Antonio Toral, Sheila Castilho, Ke Hu, and Andy Way. 2018. Post-editing effort of neural machine translation: A study of human factors. *Machine Translation*, 32(1–2).

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*.