Bridging Communication Gaps: Evaluating AI and Google Translate in Tamil Public Signage

S. Pavithra
Department of Translation Studies, University of Jaffna, Sri Lanka
Subapavi1993@gmail.com

Introduction

Public signage is a crucial medium of communication in multilingual societies like Sri Lanka, where Tamil, Sinhala, and English coexist. Traditionally, mistranslations in Tamil signboards have caused confusion, misinterpretation, and occasionally offense. With the rise of AI-assisted translation tools such as ChatGPT, DeepL, and others, there is now an opportunity to automate translations. However, these tools are not flawless, and their use in public signage may introduce new types of errors. This study explores how AI and Machine Translation impact the quality of Tamil translations on public signs, assessing both the benefits and limitations of technology assisted translation in real-world contexts.

Literature Review

Translation studies emphasize the importance of cultural and contextual knowledge in producing accurate translations (Newmark, 1988; Nida, 1964). Past research has documented human mistranslations in Tamil signboards, highlighting spelling errors, literal renderings, grammatical inconsistencies, and semantic mistakes (Davis, 2020; Meyler, 2012). Meanwhile, AI translation tools have become increasingly popular, yet their performance in minority languages like Tamil remains underexplored. Studies on AI translation in global contexts indicate that machine translations often fail to capture idiomatic expressions and contextual nuances (Liu, 2021). By combining insights from translation theory and AI research, this study situates mistranslation issues within both AI and automated frameworks, offering a novel perspective.

Methodology

This study adopts a qualitative comparative approach to evaluate the accuracy of Tamil translations found in public signboards, with a specific focus on identifying and analyzing semantic errors and typographical errors. The research compares translations produced by ChatGPT (AI-generated translation) and Google Translate (machine translation) to assess how effectively each system conveys meaning and maintains linguistic accuracy in Tamil public signage.

Data Collection

Seventeen mistranslated Tamil signboards were purposively collected from urban and semi-urban locations in Jaffna, Colombo, and Kandy. These areas were chosen because they feature multilingual public spaces where English, Sinhala, and Tamil coexist. The signboards were photographed, and both the source text (English or Sinhala) and the Tamil translation were recorded for analysis.

Data Sources and Translation Tools

The primary data for this study consisted of bilingual and trilingual public signboards. To evaluate translation performance:

- The original English or Sinhala text of each signboard was translated into Tamil using ChatGPT (AI-generated translation).
- The same text was also translated using Google Translate (machine translation).
- These AI and machine-generated translations were then compared with the human-translated versions that appeared on the actual public signs.

Analytical Procedure

Each set of translations human, ChatGPT, and Google Translate was compared line by line to identify translation errors. The analysis concentrated on two main error types:

1.Semantic Errors:

These occur when the meaning of the original message is distorted or misrepresented in Tamil.Examples include incorrect word choice, omission of key terms, or mistranslation of culturally specific expressions. Semantic errors were analyzed based on Nida's (1964) concept of dynamic and formal equivalence and Newmark's (1988) semantic vs. communicative translation framework.

2. Typographical Errors:

These involve issues in Tamil script such as misspelled letters, misplaced vowel markers, broken characters, or punctuation errors. They were examined to determine how they affect readability, professional appearance, and public comprehension.

Data Analysis and Discussion

The table 1 presents a comparative analysis of public signage by listing the original English text, the corresponding Sinhala and Tamil translations, the type of errors observed categorized as Semantic Errors (SE) or Typographical Errors (TE) and the translations generated by ChatGPT and Google Translate, highlighting the differences in accuracy and reliability between AI-assisted translations and human translations(translations on the sign boards).

The collected dataset comprises 17 public signboards with original English or Sinhala texts, human-translated Tamil versions, and AI-generated translations using

Google Translate. Each signboard was analyzed for error types, categorized as Semantic Errors (SE) and Typographical Errors (TE).

English	Sinhala	Tamil	Type of error	AI translation	Google Translate
Reserved for Pregnant Mothers	ගැබිණි කාන්තාවන් සඳහා වෙන්කර ඇත	கர்ப்பிணி நாய்மார்களுக்காக ஒதுக்கப்பட்டுள்ளது.	SE	கர்ப்பமாகிய மாக்களுக்காக ஒதுக்கப்படடுள்ளது	கர்ப்பிணித் தாய்மார்களுக்கு ஒதுக்கப்பட்டுள்ளது.
Galle National Museum	ගාල්ල ජාතික කෞතුකාගාරය	காலி தசேிய நூதனசாலை	SE	காலி தேசிய அருங்காட்சியகம்	காலி தேசிய அருங்காட்சியகம்
Library	පුස්තකාලය	நூலகும்	TE	நூலகம்	நூலகம்
Ministry of National Coexistence Dialogue and Official Languages	ජාතිකය සහජීවන, සංවාදය හා රාජ්‍ය භාෂා අමාත්‍යාංශය	தேகிய சகவாழ்வு இகலந்தலைபாடல் மறறும் அரசகரும மொழீகன அனமக்சு	TE	தேசிய ஒற்றுமை உரையாடல் மற்றும் அதிகாரபூர்வ மொழிகள் அமைச்சகம்	தேசிய சகவாழ்வு உரையாடல் மற்றும் உத்தியோகபூர்வ மொழிகள் அமைச்சகம்
Kelaniya Kohuwala	කැලණිය වකාහුවල	முகடுசவலரலக முகசநவலரலகு	TE	கௌலணியா கொகுவளா	களனி கொகுவல
Dharga Town	දර්ගා නගරය	தர்கா நரகம்	SE	தர்கா டவுன்	தர்கா நகரம்
Entrance	ඇතුල්වීම	உல்லேவான்க	SE	நுழைவு	நுழைவாயில்

Table 1: Sample Findings of Mistranslated sign boards (SE –Semantic Errors,TE:Typographical Errors)

Analysis shows that typographical errors were the most frequent, accounting for approximately 51% of the sample, followed by semantic errors at 39% . Typographical errors mainly involved garbled characters, incorrect encoding, or misplaced punctuation, as observed in translations of the Ministry of National Coexistence Dialogue and Official Languages. Semantic errors, which pose a significant risk to comprehension, were found in critical notices such as

“Reserved for Pregnant Mothers,” where mistranslation altered meaning entirely. Spelling errors were typically minor but affected readability and overall presentation quality.

AI translations were generally effective in correcting typographical and spelling errors. For instance, garbled Tamil text from the human translation of “Library” was accurately rendered by AI. However, semantic errors persisted, particularly in idiomatic expressions and culturally nuanced phrases, demonstrating AI’s limitations in contextual understanding. Furthermore, place names were often rendered incorrectly by AI: “Kelaniya Kohuwala” appeared as மகளலணியா, மகாகுவளா, and “Dharga Town” appeared as தரக்கா டவுன், instead of accurate transliteration or usage names. This inconsistency highlights the challenge of standardizing proper nouns in Tamil public signage when relying solely on AI translation tools.

In contrast, Google Translate provided more accurate translations for these cases, producing களனி , கொஹுவல for Kelaniya and Kohuwala, தர்கா நகரம for Dharga Town, and Eiothapy; for Entrance. This demonstrates that AI (ChatGPT) and conventional machine translation systems function differently, with each having strengths and limitations.

Semantic errors remain the most critical, potentially causing confusion, while typographical and spelling errors primarily affect readability and credibility. Proper nouns and culturally sensitive terms require careful attention, as AI alone may not provide accurate results.

Using standard, formal Tamil in public signage is essential to ensure clarity, accuracy, and professionalism. While spoken or colloquial Tamil may feel relatable, it can lead to confusion or misinterpretation, especially in formal public spaces such as bus stations, airports, and government offices. For example:

Entrance: நுழைவாயில் (formal, correct) vs. உல்லேவான்க (colloquial)

Departure: புறப்பாடு (formal, correct) vs. வெளியே போங்க (colloquial)

These examples demonstrate that formal Tamil provides consistency and avoids ambiguity, making instructions clear to all Tamil-speaking audiences, including those less familiar with colloquial expressions.

Standard Tamil also preserves the linguistic integrity and cultural correctness of public signage. While AI translation tools can assist in producing translations and reducing basic errors, they are not sufficient on their own. Human oversight remains essential to verify semantic accuracy and cultural appropriateness.

A hybrid approach combining AI assistance with professional human translators is recommended. Developing standardized guidelines, translation memory in CAT tools, and specialized glossaries for frequently used terms, idioms, and place names can ensure consistency, clarity, and long-term reliability in public signage. Community feedback can further enhance inclusivity, but the use of colloquial or spoken Tamil should be avoided in official public signs to maintain professionalism and accuracy.

It is important to note that not all mistranslations we come across in public spaces are genuine. With the rise of social media, there are also instances of fake or exaggerated posts that highlight supposed mistranslations, which may not actually exist in the original signage. In such cases, it becomes crucial to verify the actual truth before drawing conclusions or criticizing authorities.

At the same time, there are many positive examples where institutions and public spaces follow the trilingual language policy accurately, ensuring that names and information appear correctly in all three languages without errors. These efforts deserve recognition and appreciation, as they demonstrate both linguistic inclusivity and administrative responsibility.

However, in certain cases where mistranslations do occur particularly when one language is rendered inaccurately, it is essential to immediately address the issue with the responsible official and ensure corrections are made without delay. To make this process more efficient, officers can rely on tools such as Google Translate, which, despite its limitations, often performs better than some AI translation systems. Still, the use of such tools should be complemented by the expertise of professional Tamil translators, who can ensure semantic accuracy and cultural appropriateness.

This combined approach of technological assistance and human expertise can significantly reduce mistranslation errors and contribute to clearer, more reliable multilingual communication in public signage.

Recommendations

Future studies should explore a hybrid approach that combines AI-assisted translation with human expertise to ensure both accuracy and cultural appropriateness. One effective strategy would be to create a Translation Memory (TM) within Computer-Assisted Translation (CAT) tools, which would store verified translations of public signage and official terms. This would ensure that translations remain consistent over time, reducing recurring errors and maintaining standardization across locations.In addition, the development of related software tools could help automate error detection, flagging potential semantic or grammatical mistakes before signs are finalized. The use of specialized glossaries, particularly for technical, cultural, or place names, would further support translators and officials in producing accurate and reliable translations. By integrating AI, CAT-based translation memory, glossaries, and human review, public authorities can achieve sustainable, high-quality multilingual communication that is clear, inclusive, and culturally appropriate for Tamil-speaking populations.

Conclusion

While AI-assisted translation offers some support for Tamil public signage, this study finds that Google Translate performs better in terms of accuracy and reliability. Nevertheless, human expertise remains essential to correct semantic and cultural errors. Combining Google Translate with professional translators, along with Translation Memory in CAT tools and specialized glossaries, can ensure consistent, accurate, and culturally appropriate translations, enhancing clarity and inclusivity in Sri Lanka's multilingual public spaces.

References

Davis, D. (2020). Mistranslations in public signage: A case study of South Asian languages. *Journal of Translation Studies, 15(*2), 45–60.

Liu, Y. (2021). Machine translation and cultural nuances: Challenges for minority languages. *Language Resources and Evaluation, 55*(1), 1–19.

Meyler, M. (2012). *The language of public signs in multilingual societies.* Multilingual Matters.

Mohomed Sakeef, S. N. (2022). *Translation of news reports in foreign newspapers to Sinhala and Tamil languages and misinterpretation of democracy in Sri Lankan media in reference to selected newspapers* (Unpublished thesis). University of Kelaniya.

Newmark, P. (1988). *A textbook of translation*. Prentiuce Hall.

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