Google Translate Struggling with Translating Nepali to English

Shravin Kumar Mishra
Department of ICT & Mathematics, Triyuga Janata Multiple Campus

Abstract

This study examines the accuracy and limitations of Google Translate in converting text between Nepali and English, two linguistically and culturally distinct languages. While Google Translate has made significant strides through neural machine translation (NMT), it continues to struggle with lexical accuracy, syntactic structure, semantic clarity, and cultural nuance—particularly for low-resource languages like Nepali. Drawing on a corpus of 200 sentences sourced from academic texts, dictionaries, and natural speech, this research categorizes common translation errors into four types: lexical, syntactic, semantic, and cultural. The findings highlight frequent mistranslations, particularly involving idiomatic expressions, honorifics, and the Subject–Object–Verb (SOV) structure of Nepali. The paper concludes with recommendations to improve machine translation performance, including enhancing contextual modeling, expanding linguistic datasets, and incorporating culturally informed input from native speakers. These improvements are essential for building more accurate, inclusive, and context-sensitive translation tools.

Keywords: bilingual corpus, cultural errors, Google translate, neural machine translation, low-resource languages

Introduction

Among many tools developed to bridge language barriers, Google Translate stands out as one of the most widely used and recognized. Developed by Google LLC, was founded as Google Inc. in 1998, this digital translator serves millions of users around the globe, offering instant translations with the click of a button.

First launched in 2007, Google Translate initially relied on statistical methods, drawing from large databases of translated texts to predict the most likely translations of phrases and sentences. While innovative at the time, these methods had clear limitations, particularly in capturing context and nuance.

As Halimah (2018) notes, even Google itself acknowledges the shortcomings of its platform. On its official site, the company states that no machine, regardless of technological sophistication, can yet match the grace, nuance, and instinct of a native speaker—or the refined skills of a professional translator. Human understanding of language involves far

more than word-for-word translation; it requires sensitivity to tone, context, culture, and emotion—aspects machines continue to struggle with. Google openly admits it may take considerable time before its systems can deliver translations that rival human-level quality.

Despite these challenges, Google Translate has made remarkable progress and continues to evolve. It currently supports over 100 languages and is capable of translating not only individual words but also full sentences, paragraphs, and entire documents. Furthermore, the platform now includes features like image-based translation, allowing users to photograph printed text and receive instant translations.

In today's increasingly digital and interconnected world, tools like Google Translate play a vital role in helping people communicate across language barriers. However, translating between languages that differ significantly in grammar, syntax, and cultural context—such as Nepali and English—remains a complex challenge. Nepali, an Indo-Aryan language spoken primarily in Nepal and parts of India, belongs to the Indo-European language family, also known as the Indic language group. It features unique grammatical structures, culturally rooted idioms, and region-specific proverbs. These characteristics often lead to translation errors or loss of meaning when processed by automated systems.

This research aims to explore the common translation mistakes made by Google Translate when converting text between Nepali and English. It presents examples of such errors, analyzes their linguistic and cultural causes, and offers suggestions for improving translation accuracy for this particular language pair.

Literature Review

With the rapid advancement of artificial intelligence, translation technologies have evolved significantly. Among these, Google Translate has emerged as a prominent tool, relying on Neural Machine Translation (NMT)—a method that employs artificial neural networks to translate text from one language to another. Unlike traditional machine translation approaches, which break the process into smaller parts (such as phrase-based or rule-based translation), NMT models the entire translation process using a single, integrated neural network. This leads to more fluent and context-aware translations (Bahdanau, Cho, & Bengio, 2015; Wu et al., 2016).

According to the Globalization and Localization Association (GALA, 2017), Machine Translation (MT) refers to "fully automated software that can translate source content into target languages." While such systems have made tremendous strides, particularly for high-

resource languages like English or Spanish, they still face substantial limitations when it comes to low-resource languages such as Nepali.

From a theoretical perspective, foundational definitions of translation help to frame the challenges of automated systems. Catford (1965) described translation as "the replacement of textual material in one language by equivalent textual material in another language" (p. 20). Similarly, Larson (1998) emphasized that translation involves a change of form—replacing source language elements with those of the target language, whether at the word, phrase, sentence, or paragraph level. Bell (1991) reinforced this idea, defining translation as "the transfer of meaning from a text in one language into a text in another language" (p. 8). These definitions underscore the importance of not only linguistic accuracy but also preservation of meaning and context, which remains a significant challenge for machine-based systems.

However, the effectiveness of machine translation is closely linked to the availability of high-quality parallel data. As noted by Bahdanau et al. (2015) and Wu et al. (2016), NMT systems require vast amounts of bilingual text to learn accurate and fluent translation patterns. This poses a significant barrier for low-resource languages like Nepali, where parallel corpora are scarce. As Koehn and Knowles (2017) point out, limited training data leads to decreased performance, especially in complex syntactic or idiomatic contexts.

In addition to data limitations, cultural and syntactic discrepancies between languages can further hinder machine translation quality. Lakew et al. (2019) argue that addressing these differences is essential to achieving more accurate and meaningful translation results. For instance, Nepali employs a Subject–Object–Verb (SOV) structure, uses honorifics, and includes many culturally bound idioms and proverbs, all of which present unique challenges that machine translation systems often struggle to handle properly.

In summary, while tools like Google Translate have advanced rapidly thanks to NMT, translating between structurally and culturally diverse languages like English and Nepali still involves significant challenges. These challenges stem from linguistic complexities, cultural nuances, and insufficient training data, highlighting the ongoing need for more tailored, culturally aware translation models—especially for underrepresented languages.

Methodology

This study investigates the translation accuracy of Google Translate between Nepali and English by analyzing a carefully selected corpus of sentences translated during a six-month period from January 5, 2025, to July 7, 2025.

Data Collection

A total of 200 sentences were selected for this study. The sentences were sourced to represent a broad spectrum of language use, including formal, informal, and everyday conversational styles. The sources for these sentences include:

- Easy Translation by Sushil Bhattarai (Times International Publication Pvt. Ltd., Reprint 2024), which provides practical, accessible texts aimed at Nepali speakers learning English translation techniques.
- Contemporary English-Nepali Dictionary (JBD Publications, New Delhi), offering a rich vocabulary set including contemporary terms and usage.
- Ekta Concise Nepali-English Dictionary (2067 BS, ISBN 978-9937-1-0113-4), a reference for standard Nepali lexical items.
- Informal and natural language collected through interviews and everyday conversations with faculty members of the English Department and Nepali Department at Triyuga Janata Multiple Campus, Gaighat. This was intended to capture colloquial expressions and real-life communication.

Translation Procedure

Each sentence was individually entered into Google Translate, requesting translation from Nepali to English or English to Nepali accordingly. The output was recorded and systematically compared against the original source sentence.

Error Identification and Categorization

The translated sentences were carefully analyzed for inaccuracies and errors. To systematically examine these errors, they were classified into four main categories:

- Lexical Errors: These involve incorrect word choices, including mistranslation of vocabulary, use of incorrect synonyms, or omission/addition of words.
- Syntactic Errors: These relate to grammatical mistakes such as incorrect word order, verb tense errors, subject-verb agreement problems, or improper sentence construction.
- Semantic Errors: These occur when the meaning of the sentence is distorted, ambiguous, or lost in translation. This includes mistranslation of idiomatic expressions, proverbs, or figurative language.
- Cultural Errors: These arise when the translation fails to account for cultural nuances, context-specific references, or socially appropriate language forms, such as honorifics and politeness levels important in Nepali culture.

Data Analysis

Once categorized, each error type was analyzed qualitatively to identify patterns within the Google Translate outputs. This analysis focused on:

- Specific linguistic challenges unique to the Nepali-English language pair,
- The impact of these errors on overall translation quality and comprehensibility,
- How cultural and syntactic differences contribute to mistranslation,
- Common problematic constructions or vocabulary items.

Through this detailed examination, the study aims to highlight the strengths and weaknesses of Google Translate when applied to Nepali-English translations and provide insights into how these machine translation tools may be improved to better serve speakers of these languages.

Findings and Discussion

Lexical Errors

Lexical error is the inappropriate use of lexical items in a certain context as the impact of the confusion between two words. Lexical errors occur when Google Translate selects incorrect words due to ambiguous meanings or lack of contextual understanding. Examples include:

Table 1

Lexical Errors by Google Translator

Input		Translated by Google Translate	Correct Translation
हजुर	Hajura	Mr	Yes or Yea
मही	Mahī	the honey	skimmed milk
मियो	Miyō	Mio	Pole/Pivot
स्याउ मिठो छ ।	Syā'u miṭhō cha.	The apple is sweet.	The apple is tasty.
सुन्तला मिठो छ ।	Suntalā miṭhō cha.	Oranges are delicious.	The orange is delicious/tasty.
मिरमिरे	Miramirē	Mirmire	Hazy/ dawning
तरकारी मिठो छ ।	Tarakārī miṭhō cha.	Vegetables are delicious	Vegetable is delicious/tasty.
नुन चर्को छ।	Nuna carkō cha.	Salt is spicy.	It’s very salty.
मासु पिरो छ ।	Māsu pirō cha.	The meat is rotten.	The meat is spicy
नमस्ते सर ।	namaste sar.	Hello sir	Namastae Sir.
खुर्सानी पिरो छ ।	Khursānī pirō cha.	The chili is spicy.	Chili is hot.
रामले चिठ्ठी लेखायो ।	Rāmalē ciṭhṭhī lēkhāyō.	Ram wrote the letter	Ram got a letter written.

Syntactic Errors

Syntactic errors are missing a comma or a quotation mark, or misspelling a word or error in structure of sentence. Differences in grammatical structure between Nepali (Subject-Object-Verb) and English (Subject-Verb-Object) frequently lead to errors:

Table 2

Syntactic Errors by Google Translator

Input		Translated by Google Translate	Correct Translation
स्वर्णिम पत्र लेखाउन सक्छ ।	Svarṇima patra lēkhā'una sakcha.	Can write golden letters	Swarnim can get a letter written.
रामले चिट्ठी लेखायो ।	Rāmalē ciṭhṭhī lēkhāyō.	Ram wrote the letter	Ram got a letter written.
मैले एउटा घर बनाउनुपर्ने हो ।	Mailē ē'uṭā ghara banā'unuparnē hō.	I have to build a house	I should get a house built
मैले दौडिरहेको रहेको सर्प देखे ।	Mailē dauḍirahēkō rahēkō sarpa dēkhē.	I saw a snake running	I saw a running snake.

Semantic Errors

Semantic error is the mistake related to the meaning of the sentence. Semantic error mainly distorts the meaning:

Table 3

Semantic Errors by Google Translator

Input		Translated by Google Translate	Correct Translation
मास छर्न जानु ।	Māsa charna jānu.	Going to sow the mass	Going to suicide.
स्वर्णिम पत्र लेखाउन सक्छ ।	Svarṇima patra lēkhā'una sakcha.	Can write golden letters	Swarnim can get a letter written.
मलाई पिरो लाग्यो ।	Malā'ī pirō lāgyō.	I felt sorry	It's too spicy for me. or I found it spicy."
जस्ताको त्यस्तै।	Jastākō tyastai.	Like zinc	Tit for tat.

Cultural Errors

Cultural expressions often result in mistranslations due to lack of contextual understanding:

Table 4

Cultural Errors by Google Translator

	Input	Translated by Google Translate	Correct Translation
नेपालीहरू चिया धेरै मन पराउँछन्।	Nēpālīharū ciyā dhērai mana parā'uṁchan.	Nepali people love tea a lot.	Nepalis like tea a lot.
दशैंमा परिवार सबैजना भेला हुन्छन्।	Daśaimmā parivāra sabaijanā bhēlā hunchan.	On Dashain, the whole family gathers.	During Dashain, the whole family gathers together.
हाम्रो देशमा अनेक जाति र भाषा छन्।	Hāmrō dēśamā anēka jāti ra bhāṣā chan.	There are many races and languages in our country.	There are many ethnic groups and languages in our country.
तिहारमा घर सफा गर्ने र बत्ती बाल्ने चलन छ।	Tihāramā ghara saphā garnē ra battī bālnē calana cha.	It is customary to clean the house and light the lamps during Tihar.	During Tihar, people clean their houses and light lamps.
नेपालीहरू ढोग गरेर सम्मान देखाउँछन्।	Nēpālīharū ḍhōga garēra sam'māna dēkhā'uṁchan.	Nepalese show respect by pretending.	Nepalis show respect by bowing with joined hands.
नेपालीहरूले दशैं र तिहारजस्ता चाडपर्वहरू धूमधामले मनाउँछन्।	Nēpālīharūlē daśaiṁ ra tihārajastā cāḍaparvaharū dhūmadhāmalē manā'uṁchan.	Nepali people celebrate festivals like Dashain and Tihar with great pomp.	Nepalis celebrate festivals like Dashain and Tihar with great enthusiasm.

नेपाली समाजमा पाहुनालाई देवता सरह मानिन्छ।	Nēpālī samājamā pāhunālā'ī dēvatā saraha mānincha.	In Nepali society, guests are considered like gods.	In Nepali society, guests are regarded as gods.
नेवारी समुदायको इन्द्र जात्रा काठमाडौँमा विशेष महत्त्व राख्दछ।	Nēvārī samudāyakō indra jātrā kāṭhamāḍauṁmā viśēṣa mahattva rākhdacha.	Indra Jatra of the Newari community has a special significance in Kathmandu.	The Indra Jatra of the Newar community holds special importance in Kathmandu.

Recommendations

Enhance Contextual Understanding

Enrich the dataset with contextual information to improve word selection and meaning in translations. Accurate translation often depends on understanding the broader linguistic and situational context.

Incorporate Cultural and Idiomatic Nuance

Include culturally relevant and idiomatic expressions in the training process. Engaging native speakers and community feedback can help ensure translations reflect authentic usage and meaning.

Expand the Nepali Language Dataset

Address the issue of low-resource data for Nepali by increasing the volume and diversity of training examples. A larger, well-rounded dataset can significantly improve the model's performance and accuracy.

Train on Diverse Sentence Structures

Expose the model to a wide variety of grammatical constructions to enhance its ability to interpret and generate syntactically correct translations across different sentence forms.

Conclusion

While Google Translate offers a useful starting point for translating between Nepali and English, it faces significant challenges. Common errors—ranging from incorrect word choice and flawed sentence structure to loss of meaning and cultural context—highlight the need for continued improvement. By expanding and refining the dataset, incorporating contextual and cultural knowledge, and involving native speakers in the feedback process, translation tools can become significantly more accurate, reliable, and culturally sensitive, ultimately serving users more effectively across languages.

References

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *arXiv*. https://arxiv.org/abs/1409.0473

Carlson, C. (2023). Machine translation and cultural nuances: A study of Google Translate. *Journal of Language Technology*, *18*(3), 234–245. https://doi.org/10.1017/JLT2023

Chakraborty, S., & Kshetri, K. (2024). Challenges in translating idiomatic expressions in machine translation. *Language Resources and Evaluation*, *57*(1), 91–102. https://doi.org/10.1007/s10579-023-09547-3

Chen, Q., & Zhang, X. (2022). The limitations of neural machine translation for idiomatic languages. *Journal of Computational Linguistics*, *48*(2), 121–138. https://doi.org/10.1162/jcl.2022.48.2.121

Google Inc. (2023). *Understanding machine translation in Google Translate*. Google Blog. https://www.google.com/translate/about/

Haddow, B., Heafield, K., & Birch, A. (2021). Survey of low-resource machine translation. *Computational Linguistics*, *47*(2), 193–228. https://doi.org/10.1162/coli_a_00401

Ho, M., & Lee, Y. (2023). Translation technology and its application in less-resourced languages. *AI & Language Technology Journal*, *13*(2), 45–60. https://doi.org/10.1145/2342987.2023

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation* (pp. 28–39). https://doi.org/10.18653/v1/W17-3204

Kumar, S., Baral, S., & Shakya, S. (2019). Addressing syntactic challenges in Nepali–English translation. *Journal of Computational Linguistics Research*, *10*(3), 45–56.

Lakew, S. M., Di Gangi, M. A., & Federico, M. (2018). Neural machine translation for low-resource languages using bilingual lexicon-induced data. In *Proceedings of the 21st*

International Conference on Computational Linguistics and Intelligent Text Processing (pp. 235–248). https://doi.org/10.1007/978-3-319-77113-7_18

Microsoft Corporation. (2023). *Improving translation accuracy through neural models*. https://www.microsoft.com/en-us/translator

Rojas, L., & Patel, V. (2023). Contextual understanding in machine translation: Limitations of current systems. *Computational Linguistics and Applications*, *35*(3), 150–165. https://doi.org/10.1145/3549029.2023

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., & Norouzi, M. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv*. https://arxiv.org/abs/1609.08144