

Context and WSD: Analysing Google Translate’s Sanskrit to English Output of Bhagavadgītā Verses for Word Meaning

Anagha Pradeep

Language Technology Research Centre
International Institute of Information
Technology Hyderabad, India
anagha.pradeep@research.iiit.ac.in

Radhika Mamidi

Language Technology Research Centre
International Institute of Information
Technology Hyderabad, India
radhika.mamidi@iiit.ac.in

Pavankumar Satuluri

Department of Humanities and Social Sciences
Indian Institute of Technology
Roorkee, India
pavankumar.satuluri@hs.iitr.ac.in

Abstract

In addition to innate human intelligence, having access to extensive context and world knowledge is a crucial factor that aids in comprehending natural language, making it smooth and effortless to understand words with multiple meanings for humans. Although machines lack intrinsic intelligence, their capacity to learn language can greatly improve with access to more data, which serves as valuable context. In Natural Language Processing (NLP), the task of identifying and attributing the right sense of a word in a given context is called Word Sense Disambiguation (WSD). WSD, as a sub-task, plays a crucial role in several NLP applications such as Machine Translation. Every language has a set of words that have multiple senses. Sanskrit, one of the ancient and classical languages of the Indian subcontinent is no exception to this. Like many other languages with a rich literary tradition, Sanskrit features a multitude of polysemous words. However, it is essential to acknowledge that the data used to train machine models on Sanskrit is considerably less compared to European and a few other Indian languages. Consequently, the task of disambiguating word senses in Sanskrit presents a highly complex challenge for machines, especially when considering the unique and rich nature of its literary language. The purpose of this paper is to delineate the potential areas where the infusion of additional data can enhance language learning, through a manual error analysis taxonomy focused on the Bhagavadgītā. Our analysis will delve into the translation outcomes produced by Google Translate, which is considered the state-of-the-art tool for handling Sanskrit and other languages with limited available resources.

1 Introduction

Language is the foremost factor that sets apart humans from other beings. Humans possess the innate ability to understand, analyze and express thoughts through languages. This also means that the process of disambiguating a word having multiple senses is fundamentally natural and effortless in humans, specifically considering the access humans have to context and world knowledge. An exponential growth in the domain of Artificial Intelligence (AI) with an aim of having machines that can simulate human behavior in analyzing and interpreting natural languages is evidently witnessed. Natural Language Processing (NLP) applications like Machine Translation (MT) deal with language complexities among which subtasks such as Word Sense Disambiguation (WSD) are also dealt with. WSD has been a long standing problem in the domain of Computational linguistics and NLP. It is necessary to determine the meaning of each

word in a context, in order to make sense of the text (Itankar and Raza 2020). Nonetheless, it must be acknowledged that this pursuit is far from a straightforward undertaking for a machine. Language modeling has been very helpful in advancing machines in language intelligence. Pre-trained Language Models (PLMs) such as Transformers use large scale unlabelled corpora for their training and have shown great progress in various NLP tasks (Zhao et al. 2023). In 2022, Google Translate, which uses the Transformer architecture, expanded its language repertoire to include Sanskrit and 23 other languages bringing the total number of translatable languages to 133¹. Similar to numerous other languages, Sanskrit, the ancient classical language of the Indian subcontinent, includes a wide range of words with multiple senses. The unique nature of the language within the realm of literature adds an additional layer of complexity in disambiguating the word senses. Google enlists languages that have been trained using monolingual data and zero-resource MT as “long-tail languages” among which Sanskrit is one. These are the languages that do not have as much corpora for training the language model as do most European languages and a few Indian languages such as Hindi and Tamil (Bapna et al. 2022). This paper attempts to outline the importance of increased context in the form of training data for Sanskrit, through a manual error analysis of word meanings in the translation output of Bhagavadgītā (BhG). It proposes a taxonomy for word meaning errors and further explores the potential areas where an increase in contextual knowledge could bring about an enhancement in ambiguity resolution.

The paper is structured as follows: We will begin by discussing our rationale for selecting the BhG for translation in the subsequent section. This will be succeeded by a concise review of pertinent prior research. Moving on to the fourth section, we will provide a detailed account of the experiment and its subsequent analysis. Finally, in the fifth section, we will offer insights into additional observations gleaned from extensive experiments.

2 Motivation to choose Bhagavadgītā

BhG is a part of the well-known Indian epic, the Mahabharata. It can be found in the Bhīṣma-parva which is the sixth book of the epic. This text containing seven hundred verses, is a philosophical masterpiece that presents a conversation between Lord Kṛṣṇa and Arjuna, the Pāṇḍava prince on the battlefield of Kurukṣetra. On seeing his own relatives and loved ones on the opposing side of the battle, Arjuna is faced with a moral conflict about participating in the battle. In this moment of crisis, Lord Kṛṣṇa imparts spiritual wisdom to address Arjuna’s concerns and inspires him to fulfill his duty (Mukundananda 2022). Numerous thinkers and scholars, spanning from Adi Shankaracharya (H. K. Goyandka 2015) to modern figures like Alladi Mahadeva Sastry (Sastry 2004), Mahatma Gandhi (Gandhi 2014), Swamy Dayananda Saraswati (Saraswati 2007), Eknath Easwaran (Easwaran 2009), and many more, have found profound fascination with this revered philosophical masterpiece, each offering their own unique interpretations. The impetus for this study arises from our curiosity to witness how a machine would undertake the translation of such a profound text. It is fascinating to observe where the machine excels and where it encounters challenges in the process.

3 Related Work

In 2022, Google Translate expanded its language repertoire to include Sanskrit. During the same year, a paper was published, focusing on the semantic and sentiment analysis of BhG translations in English using a language framework based on BERT (Chandra and Kulkarni 2022). This study employed three distinct translations by experts to develop a framework for analyzing the semantic and sentiment aspects of selected verses from the BhG. Subsequently, another paper evaluating the performance of Google Translate in translating the BhG into English was published in 2023. The output from Google Translate was compared to expert translations using sentiment and semantic analysis via BERT-based language models, serving as a continuation of the previous research (Shukla et al. 2023). As far as our knowledge extends, these studies

¹<https://blog.google/products/translate/24-new-languages/> accessed on August 5th 2023

represent the primary investigations into the evaluation of machine translation outputs of the BhG. It is also important to note that the error analysis of semantic relations mentioned in the papers above involves an automated process of measuring semantic similarity between Google’s English translation output and expert human English translations. And is therefore crucial to recognize that this approach significantly differs from the methodology employed in the present study.

In addition to these studies, (Popović 2020) proposes a new method of manual evaluation for Machine Translation output. They adhere to an issue marking strategy rather than a scoring or classifying one. The paper studies Croatian and Serbian outputs of IMDb (Internet Movie Database) movie reviews and Amazon product reviews from three different online MT systems viz: Google Translate², Amazon Translate³ and Bing⁴. It concludes that out of the three MT systems Google generates the most comprehensible translations for both target languages. Usage of MT systems for literary works is not very common yet. This is possibly because of the unique structure of the language that literature carries. The usage of figurative devices, metaphors, idioms, irony and so on, makes it hard for machines to translate, often leading to literal translations that are incorrect. Nevertheless, there have been some studies in this area. (Omar and Gomaa 2020) evaluate the usefulness of applying machine translation systems to literature with a view of identifying challenges that may have negative impacts on the reliability of machine translation systems. The study uses two MT systems Google Translate and Q Translate⁵ to translate two short stories Harry Potter and The Black Cat from English to Arabic. Although they conclude on how both MT systems have performed badly at various levels, we can observe that the error rate in Google Translate was comparatively lesser. In general, the evaluation of MT systems whether automatic or manual is done on the entire translation covering adequacy, comprehensibility, and grammaticality (Popović 2020). However, our work focuses only on errors in word meanings and offers a taxonomy of these errors.

4 Experiment

4.1 Data Collection

The Sanskrit verses of the BhG utilized in this study as the source language were obtained from a Sanskrit computational toolkit named Samsaadhanii⁶. For the translation aspect, the translation outputs were self-acquired through the Google Translate API⁷, a widely employed tool for automated translation. This combination of original Sanskrit verses and translated content forms the foundation of our dataset for analysis⁸.

4.2 Analysis

Words are the fundamental building blocks of any text, and their accurate interpretation plays a key role in understanding the text as a whole as well as preserving its essence. Therefore, the aim of this study is to observe errors in word meanings and categorize them. To achieve this a comprehensive manual examination of each word in the Google translation output of BhG was conducted to ascertain its alignment with the intended meaning of the corresponding Sanskrit word. Two reference translations of the BhG namely Jayadayal Goyandaka’s “Bhagavad Gīta Tattvavivecanī” (J. Goyandka 2011) and Swami Mukundananda’s “Bhagavad Gita The Song of God” (Mukundananda 2022) were employed for this process. Additionally, support was derived from the Monier Williams Sanskrit English dictionary (Monier-Williams 1899) to augment the accuracy of the analysis. Given that the scope of this taxonomy-based analysis was specifically

²<https://translate.google.co.in/>

³<https://ai-service-demos.go-aws.com/translate>

⁴<https://www.bing.com/translator>

⁵<https://qtranslate.en.softonic.com/?ex=CS-1680.3>

⁶<https://sanskrit.uohyd.ac.in/Corpus/>

⁷<https://cloud.google.com/translate>

⁸Translations of all the BhG verses were obtained through Google Translate API on 31st January 2023

confined to word meanings, the errors identified were categorized into four distinct groups, which are:

1. Errors arising from polysemous words.
2. Errors stemming from compounding.
3. Errors originating from words that are knowledge sensitive.
4. Errors springing from incorrect meaning attribution.

We will go through each of these categories of errors in detail with examples for a clearer understanding.

Errors arising from polysemous words:

Words that have multiple senses are known as polysemous words (Agirre and Edmonds 2007). Given such a word, context plays a crucial role in disambiguating which sense of the word is to be taken. For instance, the word हरि(*hari*) in Sanskrit has fourteen meanings such as snake, lion, Vishnu (lord), sun, moon, air, Yama (god of death), Indra (lord of deities), rays, horse, parrot, monkey, frog and the colour yellow⁹. Given the word हरि(*hari*), the machine should be able to aptly choose the right sense of the word based on its context among various possible interpretations. This task is known as Word Sense Disambiguation (Navigli 2009). The outputs received on Google Translate for the following inputs are given below¹⁰.

Sanskrit Sentence	English Translation
हरिः खादति	Hari eats
हरिः कदलीफलं खादति	Hari eats bananas
हरिः वृक्षेऽस्मिन् कदलीफलं खादति	The monkey is eating bananas on this tree

Table 1: Sanskrit Sentences and English Translations

The outputs clearly indicate that when provided with additional contextual information, the machine’s ability to disambiguate the sense of the polysemous word "हरि"(*hari*) becomes more refined. This enhanced disambiguation capability demonstrates the importance of context in NLP tasks. Nonetheless, within the translation outputs we generated, it is important to note that 139 out of the 700 verses contained errors attributable to the incorrect selection of senses for polysemous words. We further categorized this list of verses into two. The former comprised verses in which the error-identified polysemous word’s meaning could potentially be adjusted by altering the context. In contrast, the latter encompassed verses in which the meaning of polysemous words remained unchanged regardless of contextual modifications.

An example for the first category is:

अपि चेदसि पापेभ्यः सर्वेभ्यः पापकृत्तमः। सर्वं ज्ञानप्लवेनैव वृजिनं सन्तरिष्यसि ॥4.36॥

Translation: Even if you are the most sinful of all sinners You will cross all troubles by the float of knowledge. 4.36

In this verse, the word प्लव(*plava*) has multiple senses such as float, frog, monkey, boat, sheep, enemy and so on¹¹.

The sense to be taken here is “boat”. However, the machine also recognizes the sense “boat” given an alteration in the context.

रामः प्लवेन नदीं तरति

⁹<https://sanskrit.uohyd.ac.in/scl/amarakosha/frame.html>

¹⁰Translations for all the sample sentences were obtained between 15th August and 8th September 2023

¹¹The dictionary entries for the various words discussed throughout this paper are given in Appendix A

Translation: Rama crosses the river by boat.

Similarly, another example for the first category is a very famous verse from the Gītā.

यदा यदा हि धर्मस्य ग्लानिर्भवति भारत। अभ्युत्थानमधर्मस्य तदात्मानं सृजाम्यहम् ॥4.7॥

Translation: Whenever there is a loss of religion, O Bhārata, When irreligion arises, I create Myself. 4.7

The word धर्म(*dharma*) has been wrongly identified with the sense religion. The appropriate sense to be used in this context is “righteousness”.

Google Translate can also identify other senses of this word given different contexts such as: उष्णत्वम् अग्नेः धर्मः

Translation: Heat is the righteousness of fire.

कुम्भकारस्य धर्मः कुम्भकरणम्

Translation: The duty of the potter is to make pottery.

Let us now go through examples for the second category.

किं कर्म किमकर्मेति कवयोऽप्यत्र मोहिताः। तत्ते कर्म प्रवक्ष्यामि यज्ज्ञात्वा मोक्षसेऽशुभात् ॥4.16॥

Translation: Even the poets are confused here as to what is action and what is inaction I will tell you that action which, knowing it, you will be freed from evil. 4.16

In the above verse, the word कवि(*kavi*) is polysemous. It has several meanings including a wise person, a poet, names of several deities, and so on. The sense that fits well in this context is “a wise person or a thinker”. Although the context was modified, the machine did not recognize the other senses associated with this word.

कविः सर्वदा सन्मार्गे एव चलति

Translation: The poet is always on the right path.

साध्वसाधुविवेचने कविः निपुणः

Translation: The poet is adept at distinguishing between right and wrong.

The same is the case of the word दैव(*daiva*). Despite the multiple senses such as celestial, divine, royal, destiny, and others that the word carries, the translation output irrespective of the context has only been “destiny”.

दैवमेवापरे यज्ञं योगिनः पर्युपासते। ब्रह्माग्नावपरे यज्ञं यज्ञेनैवोपजुहति ॥4.25॥

Translation: Other yogis worship destiny as the sacrifice Others offer the sacrifice in the Brahma-agni by the sacrifice itself. 4.25

निर्मलरात्रौ वयं आकाशे ताराचन्द्रादिकं दैववस्तु द्रष्टुं शक्नुमः

Translation: On a clear night we can see the stars and moon and other objects of destiny in the sky.

दैववृक्षेति सुप्रसिद्धः कल्पवृक्षः प्रार्थितं सर्वं दास्यति

Translation: The Kalpa tree, well known as the tree of destiny, will give everything asked for. In the three example cases that we have seen, the senses of the polysemous word दैव(*daiva*) to be taken in line with the context are celestial and or divine.

Upon examining examples for both scenarios, we observed that in the first case, the machine was able to discern the various senses of polysemous words, only when the context was modified. However, in the second case, even with altered context, the sense remained consistent. As per our observation, we hypothesize that both these cases lack sufficient examples in the training data, that cover a wide range of senses for each of the polysemous words. Thus, inducing additional data through which the machine can learn the distinct senses as well as diverse contexts would be of paramount importance to enhance the performance of the system.

Errors stemming from compounding:

In classical Sanskrit, compounding is a prevalent linguistic feature where multiple words are merged to create a single lexeme. The purpose of this practice is to achieve brevity in language, aiming to express the intended meaning using fewer words. Nevertheless, grasping the meaning of compounds can be intricate due to the fact that they often possess multiple potential interpretations, contingent upon contextual cues (Krishna et al. 2016). In our examination of the translation output of the BhG, we have identified instances where meaning errors arose either from an incorrect interpretation of the compound or from the absence of an interpretation altogether. It is important to note that these errors were relatively infrequent. In the examples that follow, we will delve into instances of both types of errors.

अदृष्टपूर्वं हृषितोऽस्मि दृष्ट्वा भयेन च प्रव्यथितं मनो मे। तदेव मे दर्शय देव रूपं प्रसीद देवेश जगन्निवास
॥11.45॥

Translation: I was delighted to see something I had never seen before and my mind was overwhelmed with fear Show me that very form, O Lord of the gods, O inhabitant of the universe. Have mercy on me. 11.45

The compound जगन्निवास (*jagannivāsa*) has been translated as “inhabitant of the universe”, which is indeed one of the possible interpretations. However, in the given context where Arjuna is addressing Lord Krishna and praising him, a more suitable interpretation would be “the one in whom the world resides” (जगतः निवासः). Referring to Lord Krishna as the one in whom the world resides often carries a deeper and more profound connotation, emphasizing his divine glory, whereas the former interpretation may not capture this essence as effectively. It is important to note that various interpretations of the compound are dependent on the primary derivative suffix.

In the verse,

आब्रह्मभुवनाल्लोकाः पुनरावर्तिनोऽर्जुन। मामुपेत्य तु कौन्तेय पुनर्जन्म न विद्यते ॥ 8.16॥

Translation: O Arjuna, the worlds return to the Abrahma world. But having attained Me, O son of Kunti, there is no rebirth. 8.16

we see that the compound आब्रह्मभुवनात् (*ābrahmabhuvanāt*) has been translated as Abrahma world, while the compound is to be interpreted as “upto the world of Brahma”. This shows that the machine has not been able to interpret the compound.

Upon observation, we noticed that the occurrence of errors related to compounding in the verses was relatively infrequent. This is likely because the training data treated compounds as single words and therefore identified and translated them accurately. To substantiate our observation, we quote the following examples:

Sanskrit Sentence	English Translation
अहं रामालयं गच्छामि	I am going to Ramalaya
अहं विद्यालयं गच्छामि	I am going to school
गणेशः लम्बोदरः अस्ति	Ganesha is tall
भीमः वृकोदरः अस्ति	Bhima is a wolf-belly

Table 2: Sanskrit Sentences and English Translations

Albeit, even in cases where the machine is trained on various interpretations of a compound, determining which interpretation is appropriate for a specific context can still be a challenging task.

Errors originating from words that are knowledge sensitive:

From our observation errors stemming from words that are sensitive to cultural, societal, or philosophical nuances are quite common in machine-generated content. Machines often struggle to recognize the senses of such words, particularly when those words hinge on a deep grasp of human culture, history, or philosophy. These nuances can be challenging for machines to capture accurately. BhG being a profound philosophical text, that is a part of ऐतिहासिक (*aitihāsika*) text is susceptible to such errors. Here are a couple of instances that we may look through.

पाञ्चजन्यं हृषीकेशो देवदत्तं धनञ्जयः। पौण्ड्रं दध्मौ महाशङ्खं भीमकर्मा वृकोदरः ॥1.15॥

Translation: Hrishiksha married Panchajanya and Arjuna married Devadatta Bhimakarma blew the great conch of Paundra 1.15

From the translation, it is apparent that the words पाञ्चजन्य, देवदत्त, पौण्ड्र (*pāñcajanya, devadatta, paunḍra*) have been incorrectly recognized as names of entities and place, instead of being correctly identified as the names of the conches used by Kṛṣṇa, Arjuna, and Bhīma. This misinterpretation occurs due to the machine’s limitation in historical and contextual knowledge. Another observation regarding the translation is that the verse does not contain a verb related to “marriage”, yet the translation includes this verb twice. This discrepancy may be attributed to instances of similar sentences in the training data.

Similarly, in the verse:

प्रकृतिं पुरुषं चैव क्षेत्रं क्षेत्रज्ञमेव च। एतद्वेदितुमिच्छामि ज्ञानं ज्ञेयं च केशव ॥13.1॥

Translation: Nature, the person, the field, and the knower of the field. I wish to know this knowledge and the knowable, O Kesava. 13.1

Although the words have been translated to their corresponding senses, the philosophical nuances they carry often remain unrepresented in these. Take, for example, the word प्रकृति (*prakṛti*); it does not refer to nature in a general sense but is intricately linked to the specific concept of material nature discussed in the Sāṅkhya philosophy. The machine frequently struggles to convey these intricacies due to its inherent limitations.

Errors springing from incorrect meaning attribution:

Out of the 700 verses in the BhG, it was found that 25 verses had word translations that were completely unrelated to their actual word meanings. Unlike polysemous words where at least one of the listed senses was considered, in these instances, the translations provided did not align with any of the listed meanings for the respective words. Let us take a look at a few sample cases.

अयनेषु च सर्वेषु यथाभागमवस्थिताः। भीष्ममेवाभिरक्षन्तु भवन्तः सर्व एव हि ॥1.11॥

Translation: Situated in their proper places in all the moons May you all protect Bhishma alone. 1.11

The word अयन (*ayana*) refers to path in this context. It is wrongly translated as “moon”, which does not correspond to any of the listed senses of the word.

Similarly, the word परित्राण (*paritrāṇa*), which means “to protect”, was incorrectly translated in the following verse.

परित्राणाय साधूनां विनाशाय च दुष्कृताम्। धर्मसंस्थापनार्थाय सम्भवामि युगे युगे ॥4.8॥

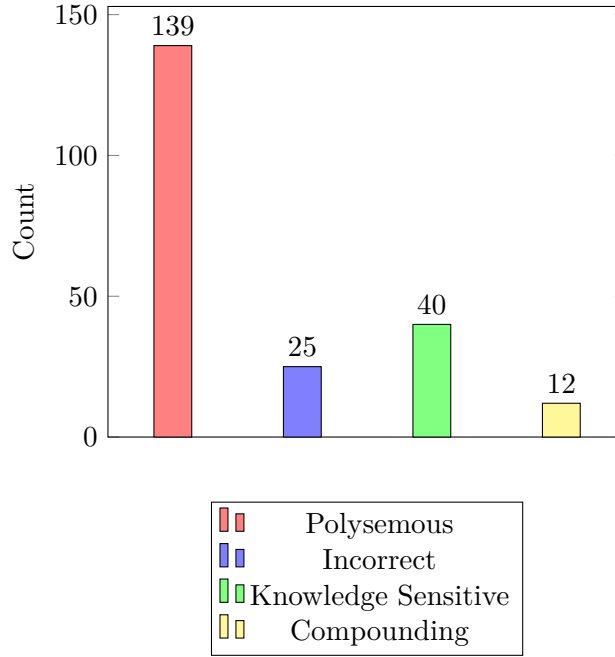


Figure 1: Bar chart depicting the number of verses in each category

Translation: For the salvation of the righteous and the destruction of the wicked I am able to establish righteousness in every age. 4.8

The sense “salvation” has not been enlisted as one of the senses of the word परित्राण (*paritrāṇa*). We could not draw any conclusions regarding the reasons for these incorrect translations, except for the possibility that the training data might have contained errors.

5 Additional Experiments

In addition to the previously mentioned error categories, we also identified a few other factors that influenced the translation of word meanings. These factors are listed below, along with examples.

5.1 Sandhi

Sandhi encompasses sound or form alterations occurring either at morpheme or word boundaries (Hyman 2007). In Table 3, we have listed a few examples illustrating how the presence or absence of sandhi influenced the translation of word meanings.

Sanskrit Sentence/Verse	English Translation
हरिः वृक्षेऽस्मिन् कदलीफलं खादति	The monkey is eating bananas on this tree
हरिः वृक्षे अस्मिन् कदलीफलं खादति	Hari is eating bananas on this tree
न च मां तानि कर्माणि निबध्नन्ति धनञ्जय। उदासीनवदासीनमसक्तं तेषु कर्मसु ॥9.9॥	Nor do those actions bind Me, O Arjuna Indifferent as if indifferent , unattached to those actions.
न च मां तानि कर्माणि निबध्नन्ति धनञ्जय। उदासीनवत् आसीनम् असक्तं तेषु कर्मसु ॥9.9॥	Nor do those actions bind Me, O Arjuna Seated as if indifferent , unattached to those actions.
चातुर्वर्ण्यं मया सृष्टं गुणकर्मविभागशः। तस्य कर्तारमपि मां विद्ध्यकर्तारमव्ययम्॥4.13॥	I have created the four varnas according to the divisions of virtue and action Know Me, the inexhaustible doer, to be the doer of it
चातुर्वर्ण्यं मया सृष्टं गुणकर्मविभागशः। तस्य कर्तारम् अपि मां विद्ध्य अकर्तारम् अव्ययम्॥4.13॥	I have created the four varnas according to the divisions of virtue and action Know Me also as the doer of it, the inexhaustible non-doer .

Table 3: Sanskrit Sentences and English Translations

5.2 Playing with words

In some cases, we also observed how the absence and replacement of words in a sentence has rendered changes in the meanings of the polysemous words in the sentence. Table 4 lists instances of a few examples.

ये यथा मां प्रपद्यन्ते तांस्तथैव भजाम्यहम्। मम वर्त्मानुवर्तन्ते मनुष्याः पार्थ सर्वशः ॥4.11॥

Translation: I reward those who worship Me in the same way that they worship Me. Men follow My path in every way, O Arjuna.

In the case of the word प्रपद्यन्ते (*prapadyante*), we needed its specific sense of “surrender” in the given context. To explore if we could achieve the desired meaning through some modifications, we initially transformed the metrical form into a sentence: ये यथा मां प्रपद्यन्ते तान् तथा अहं भजामि इति भगवान् कृष्णः भगवद्गीतायाम् अवदत्। This sentence was translated as “Lord Krishna said in the Bhagavad Gita that I worship those who worship Me in the same way as they worship Me”.

To experiment, we decided to remove certain words from the sentence to see if we could obtain the intended sense. Surprisingly, when we removed the word अहम् (*aham*), the desired output was achieved: भगवान् कृष्णः भगवद्गीतायां ये यथा मां प्रपद्यन्ते तान् तथा भजामि इति अवदत्।

Translation: Lord Krishna said in the Bhagavad Gita that I worship those who surrender to Me in the same way.

However, it is to be noted that when the verse was transformed into prose, the meaning of the word भजामि (*bhajāmi*) was also inadvertently altered.

Although this was not an example from the Gītā, we encountered something similar with the word हरि (*hari*) as well.

Sanskrit Sentence	English Translation
हरिः वृक्षेऽस्मिन् स्थित्वा खादति	The monkey is standing on this tree and eating
हरिः वृक्षेऽस्मिन् स्थित्वा पश्यति	The monkey stands in this tree and watches
हरिः वृक्षेऽस्मिन् उपविश्य खादति	The monkey is sitting in this tree and eating
हरिः वृक्षेऽस्मिन् उपविश्य पश्यति	Hari is sitting in this tree and watching
हरिः वृक्षेऽस्मिन् स्थित्वा पश्यति	The monkey stands in this tree and watches
हरिः वृक्षेऽस्मिन् गत्वा खादति	The monkey goes to this tree and eats
हरिः वृक्षेऽस्मिन् गत्वा पश्यति	The monkey goes to this tree and looks

Table 4: Sanskrit Sentences and English Translations

It was interesting to observe how making seemingly random changes to words could bring out sense modifications. Nonetheless, despite our efforts, we could not discern a logical pattern or inference point to explain why these alterations were occurring.

6 Conclusion and Future Work

The objective of this study was to underscore the significance of context, in the form of extensive data, in improving the performance of language models for translating word meanings in Sanskrit. To achieve this goal, we assessed Google Translate’s output of BhG from Sanskrit to English and developed a taxonomy of errors. Notably, our taxonomy revealed that the most frequent errors arose from polysemous words, highlighting the critical role of WSD in NLP tasks like machine translation.

However, we also observed a significant limitation: the state-of-the-art PLMs lack access to substantial Sanskrit-language data, unlike many European languages and a few other Indian languages. To address this, we illustrated how more context could potentially reduce errors stemming from polysemous words.

In addition to polysemy, we identified three other error categories: compounding, knowledge-sensitive words, and incorrect meaning attribution. While increased contextual learning might

mitigate these errors to some extent, they remain complex due to the inherent challenge of instilling machines with world knowledge.

Our study further delved into the functioning of Google Translate through experiments. However, we were unable to derive logical inferences regarding why and how the mere presence or absence of certain factors influenced the machine’s translations of word meanings.

In conclusion, this study provides a broad exploration based on our observations and experiments with Google Translate. With the disclosure of the data used to train language models, future research can yield more specific linguistic-based solutions to enhance performance in Sanskrit translation tasks.

Acknowledgement

I would like to thank Prof. Amba Kulkarni for the Sanskrit verses of the Bhagavadgītā (BhG) utilized in this study and Mr. Sriram Krishnan for his valuable input and guidance.

References

- Agirre, Eneko and Philip Edmonds (2007). *Word sense disambiguation: Algorithms and applications*. Vol. 33. Springer Science & Business Media.
- Bapna, Ankur et al. (2022). “Building machine translation systems for the next thousand languages”. In: *arXiv preprint arXiv:2205.03983*.
- Chandra, Rohitash and Venkatesh Kulkarni (2022). “Semantic and sentiment analysis of selected Bhagavad Gita translations using BERT-based language framework”. In: *IEEE Access*.
- Easwaran, Eknath (2009). *The Bhagavad Gita*. Jaico Publishing House.
- Gandhi, Mahatma (2014). *The Bhagavad Gita*. Jaico Publishing House.
- Goyandka, Hari Krishnadas (2015). *Srimad Bhagavad Gita with Shankara Bhashya*. Gita Press Gorakhpur.
- Goyandka, Jayadayal (2011). *Śrīmadbhagavadgītā Tattvavivecanī (English Commentary)*. Gita Press, Gorakhpur, India.
- Hyman, Malcolm D (2007). “From pāṇinian sandhi to finite state calculus”. In: *International Sanskrit Computational Linguistics Symposium*. Springer, pp. 253–265.
- Itankar, Prashant Y and Nikhat Raza (2020). “Ambiguity Resolution: An Analytical Study”. In: Krishna, Amrith et al. (2016). “Compound type identification in sanskrit: what roles do the corpus and grammar play?” In: *Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP2016)*, pp. 1–10.
- Monier-Williams (1899). *A Sanskrit-English Dictionary: Etymologically and Philologically Arranged with Special Reference to Cognate Indo-European Languages*. Motilal Banarsidass.
- Mukundananda, Swami (2022). *BHAGAVAD GITA: THE SONG OF GOD*. Rupa & Company.
- Navigli, Roberto (2009). “Word sense disambiguation: A survey”. In: *ACM computing surveys (CSUR)* 41, pp. 1–69.
- Omar, Abdulfattah and Yasser Gomaa (2020). “The machine translation of literature: Implications for translation pedagogy”. In: *International Journal of Emerging Technologies in Learning (iJET)* 15, pp. 228–235.
- Popović, Maja (2020). “Informative manual evaluation of machine translation output”. In: Saraswati, Swamy Dayananda (2007). *Śrīmad Bhagavad Gītā*. Arsha Vidya Research and Publication Trust.
- Sastry, Alladi Mallesh (2004). *The Bhagavad Gita*. Samata Books.
- Shukla, Akshat et al. (2023). “An evaluation of Google Translate for Sanskrit to English translation via sentiment and semantic analysis”. In: *arXiv preprint arXiv:2303.07201*.
- Zhao, Wayne Xin et al. (2023). “A survey of large language models”. In: *arXiv preprint arXiv:2303.18223*.

Appendix

The appendix is structured as follows: In Sec. A we present details about the dictionary entry¹² for words.

A Dictionary

1. **plava** (H1) [Printed book page 715,1] plava a plavaka &c. See col. 2. [ID=141442] (H2) [Printed book page 715,2] plava b mf(ā)n. swimming, floating, ŚāṅkhGr. ; Suśr. [ID=141469] sloping towards, inclined, Hariv. ; Var. ; Hcat. (in astrol. applied to a constellation situated in the quarter ruled by its planetary regent, Var. , Sch.) [ID=141470] transient, MuṇḍUp. [ID=141471] plava mn. (ifc. f(ā).) a float, raft, boat, small ship, RV. &c. &c. [ID=141472] plava m. a kind of aquatic bird (= gātra-samplava, kāraṇḍava, jala-vāyasa, jala-kāka or jala-kukkuṭa, L.), VS. &c. &c. [ID=141473] a frog, L. [ID=141474] a monkey, L. [ID=141475] & sheep, L. an arm, L. [ID=141476] a Caṇḍāla, L. [ID=141477] an enemy, L. [ID=141478] Ficus Infectoria, L. [ID=141479] a snare or basket of wicker-work for catching fish, L. [ID=141480] the 35th (or 9th) year in a cycle of Jupiter, VarBr̥S. [ID=141481] plava m. swimming, bathing (ifc. f(ā).), MBh. ; R. ; Kathās. [ID=141482] plava m. flooding, a flood, the swelling of a river, MBh. ; Mār̥kP. [ID=141483] the prolated utterance of a vowel (= pluti), L. [ID=141484] protraction of a sentence through 3 or more Ślokas (= kulaka), L. [ID=141485] sloping down or towards, proclivity, inclination, L. [ID=141486] (in astrol.) = plava-tva, VarBr̥S. , Sch. [ID=141487] a kind of metre, Col. [ID=141488] N. of a Sāman (also with vasiṣṭhasya), Ār̥ṣBr. [ID=141489] jumping, leaping, plunging, going by leaps or plunges, R. (cf. comp. below) [ID=141490] returning, L. [ID=141491] urging on L. [ID=141492] plava n. Cyperus Rotundus or a species of fragrant grass, Suśr. [ID=141493]
2. **dharma** (H2) [Printed book page 510,3] 1. dharma m. (rarely n. g. ardharcādi; the older form of the RV. is dharman, q.v.) that which is established or firm, steadfast decree, statute, ordinance, law [ID=99903] usage, practice, customary observance or prescribed conduct, duty [ID=99904] right, justice (often as a synonym of punishment) [ID=99905] virtue, morality, religion, religious merit, good works (dharmaṇa ind. or °māt ind. according to right or rule, rightly, justly, according to the nature of anything; cf. below; °mesthita mfn. holding to the law, doing one's duty), AV. &c. &c. [ID=99906] Law or Justice personified (as Indra, ŚBr. &c.; as Yama, MBh. ; as born from the right breast of Yama and father of Śama, Kāma and Harṣa, ib. ; as Viṣṇu, Hariv. ; as Prajā-pati and son-in-law of Dakṣa, Hariv. ; Mn. &c.; as one of the attendants of the Sun, L. ; as a Bull, Mn. viii, 16 ; as a Dove, Kathās. vii, 89 , &c.) [ID=99907] the law or doctrine of Buddhism (as distinguished from the saṅgha or monastic order, MWB. 70) [ID=99908] the ethical precepts of Buddhism (or the principal dharma called sūtra, as distinguished from the abhi-dharma or 'further dharma' and from the vinaya or 'discipline', these three constituting the canon of Southern B°, MWB. 61) [ID=99909] the law of Northern B° (in 9 canonical scriptures, viz. Prajñā-pāramitā, Gaṇḍavyūha, Daśa-bhūmīśvara, Samādhirāja, Laṅkāvatāra, Saddharma-puṇḍarīka, Tathā-gataguhya, Lalita-vistara, Suvarṇa-prabhāsa, ib. 69) [ID=99910] nature, character, peculiar condition or essential quality, property, mark, peculiarity (= sva-bhāva, L. ; cf. daśadh°arma-gata, ŚBr. &c. &c.; upamānopameyayor dh°, the tertium comparationis, Pāṇ. ii, 1, 55, Sch.) [ID=99911] a partic. ceremony, MBh. xiv, 2623 [ID=99912] sacrifice, L. [ID=99913] the ninth mansion, Var. [ID=99914] an Upaniṣad, L. [ID=99915] associating with the virtuous, L. [ID=99916] religious abstraction, devotion, L. [ID=99917] = upamā, L. (cf. above) [ID=99918] a bow, Dharmāś. [ID=99919] [Printed book page 1329,2] a thing, Sukh. i [ID=99919.1] [Printed book page 510,3] a Soma-drinker, L. [ID=99920] N. of the 15th Arhat of the present Ava-sarpiṇī, L. [ID=99921] of a son of Anu and father

¹²<https://www.sanskrit-lexicon.uni-koeln.de/scans/MWScan/2020/web/webtc/indexcaller.php>

of Ghṛta, Hariv. [ID=99922] of a s° of Gāndhāra and f° of Dhṛta, Pur. [ID=99923] of a s° of Haihaya and f° of Netra, BhP. [ID=99924] of a s° of Pṛthu-śravas and of Uśanas, ib. [ID=99925] of a s° of Su-vrata, VP. (cf. dharma-sūtra) [ID=99926] of a s° of Dīrgha-tapas, VāyuP. [ID=99927] of a king of Kaśmīra, Rāj. iv, 678 [ID=99928] of another man, ib. vii, 85 [ID=99929] of a lexicographer &c. (also -paṇḍita, -bhaṭṭa and -śāstrin), Cat. [ID=99930] dharma [cf. Lat. firmus, Lith. dermé.] [ID=99930.05] (H2C) [Printed book page 510,3] dharmeṇa ind., according to right or rule, rightly, justly, according to the nature of anything [ID=99930.1] dharmāt ind., according to right or rule, rightly, justly, according to the nature of anything [ID=99930.2] (H2) [Printed book page 512,3] 2. dharma Nom. P. °mati, to become, law, Vop. [ID=100471] 3. dharma in comp. for °man, q.v. [ID=100472] (H1) [Printed book page 513,1] dharma a See p. 510, col. 3. [ID=100518]

3. **kavi** (H1) [Printed book page 264,2] kavi mfn. (√1. kū cf. 2. kava, ākūta, ākūti, kāvya, Naigh. iii, 15 ; Nir. xii, 13 ; Uṇ. iv, 138) gifted with insight, intelligent, knowing, enlightened, wise, sensible, prudent, skilful, cunning [ID=46509] kavi (is), m. a thinker, intelligent man, man of understanding, leader [ID=46510] a wise man, sage, seer, prophet [ID=46511] a singer, bard, poet (but in this sense without any technical application in the Veda), RV. ; VS. ; TS. ; AV. ; ŚBr. i, 4, 2, 8 ; KaṭhUp. iii, 14 ; MBh. ; Bhag. ; BhāgP. ; Mn. vii, 49 ; R. ; Ragh. [ID=46512] N. of several gods, (esp.) of Agni, RV. ii, 23, 1; x, 5, 3; iii, 5, 1; i, 31, 2; 76, 5 [ID=46513] of Varuṇa, Indra, the Aśvins, Maruts, Ādityas [ID=46514] of the Soma [ID=46515] of the Soma priest and other sacrificers [ID=46516] (probably) N. of a particular poet [ID=46517] cf. aṅgiras (Mn. ii, 151) and uśanas (Bhag. x, 37) [ID=46518] of the ancient sages or patriarchs (as spirits now surrounding the sun) [ID=46519] of the R̥bhus (as skilful in contrivance) [ID=46520] of Pūṣan (as leader or guider) [ID=46521] N. of a son of Brahmā, MBh. xiii, 4123, 4142-4150 [ID=46522] of Brahmā, W. [ID=46523] of a son of Bhṛgu and father of Śukra, MBh. i, 2606 (cf. 3204; BhāgP. iv, 1, 45 and Kull. on Mn. iii, 198) [ID=46524] that of Śukra (regent of the planet Venus and preceptor of the demons), Rājat. iv, 495 [ID=46525] of the planet Venus, NBD. [ID=46526] of the sons of several Manus, Hariv. ; BhāgP. ; VP. [ID=46527] of a son of Kauśika and pupil of Garga, Hariv. [ID=46528] of a son of R̥ṣabha, BhāgP. [ID=46529] of Vālmiki, L. [ID=46530] a keeper or herd, RV. vii, 18, 8 [ID=46531] (fig.) N. of the gates of the sacrificial enclosure, TS. v, 11, 1, 2 (cf. kavaṣ) [ID=46532] the sun, W. [ID=46533] of various men [ID=46534] the soul in the Sāṃkhya philosophy Comm. [ID=46535] a cunning fighter, L. [ID=46536] an owl, L. [ID=46537] kavi (is or ī, W.), f. the bit of a bridle, L. [ID=46538] the reins (cf. kavikā), W. [ID=46539] a ladle (cf. kambi), L. [ID=46540]

4. **daiva** (H1) [Printed book page 497,2] 1. daiva or daiva mf(ī)n. (fr. deva) belonging to or coming from the gods, divine, celestial, AV. ; Br. ; Mn. ; MBh. &c. [ID=96788] sacred to the gods (-tīrtha n. the tips of the fingers, Mn. ii, 59 ; cf. s.v.; °vīdik f. the north, L. ; cf. 2. diś) [ID=96788.05] royal (vāc), Rājat. v, 205 [ID=96788.1] depending on fate, fatal, Kāv. [ID=96788.15] daiva m. (with or without vivāha) a form of marriage, the gift of a daughter at a sacrifice to the officiating priest, Mn. iii, 21 ; 28 [ID=96788.2] the knowledge of portents, Śaṃk. [ID=96788.25] patr. of Atharvan, ŚBr. [ID=96788.3] pl. the attendants of a deity, TāṇḍBr. xvii, 1, 1 [ID=96788.35] (H1B) [Printed book page 497,2] daivī (ī), f. a woman married according to the Daiva rite, Viṣṇ. xxiv, 30 [ID=96788.4] a division of medicine, the medical use of charms, prayers &c., W. [ID=96788.45] (H1B) [Printed book page 497,2] daiva n. a deity (cf. kula-), BhP. iii, 1, 35 &c. [ID=96788.5] (scil. karman, kārya &c.) a religious offering or rite, Yājñ. ; MBh. [ID=96788.55] daiva n. divine power or will, destiny, fate, chance (°vāt ind. by chance, accidentally), AV. ; Mn. ; MBh. &c. [ID=96788.6] (H1C) [Printed book page 497,2] daivāt (°vāt), ind., by chance, accidentally [ID=96788.7] (H2) [Printed book page 497,3] 2. daiva Vṛddhi form of deva in comp. [ID=96855]

5. **ayana** (H2) [Printed book page 84,2] ayana a mfn. going, VS. xxii, 7 ; Nir. [ID=14674] ayana n. walking, a road, a path, RV. iii, 33, 7 &c. (often ifc. cf. naimiṣāyana, puruṣāyana, praśamāyana, samudrāyana, svedāyana), (in astron.) advancing, precession, Sūryas. [ID=14675] (with gen. [e.g. angirasām, ādityānām, gavām, &c.] or ifc.) ‘course, circulation’, N. of various periodical sacrificial rites, AV. ; ŚBr. &c. the sun’s road north and south of the equator, the half year, Mn. &c., the equinoctial and solstitial points, Var-BṛS. &c. [ID=14676] way, progress, manner, ŚBr. [ID=14677] place of refuge, Mn. i, 10 [ID=14678] a treatise (śāstra cf. jyotiṣām-ayana), L. [ID=14679] (H1) [Printed book page 84,3] ayana b See √ay, col. 2. [ID=14744.1]
6. **paritrāṇa** (H3) [Printed book page 595,3] pari-trāṇa n. rescue, preservation, deliverance from (abl.), protection or means of protection, refuge, retreat, Mn. ; MBh. &c. [ID=117701] self-defence, L. [ID=117702] the hair of the body, L. [ID=117703] moustaches, Gal. [ID=117704]
7. **prapad** (H1) [Printed book page 682,1] 1. pra-√2. pad Ā. -padyate (ep. also P.), to fall or drop down from (abl.), throw one’s self down (at a person’s feet), MBh. ; to go forwards set out for, resort to, arrive at, attain, enter (with acc., rarely loc.), AV. &c. &c.; to fly to for succour, take refuge with (acc.), TS. &c. &c.; to fall upon, attack, assail, RV. ; AV. ; to come to a partic. state or condition, incur, undergo (acc.), MBh. ; Kāv. &c.; (with an adv. in sāt), to become, e.g. sarpaśāt pra-√pad, to bec^o a serpent, Bhaṭṭ. ; to obtain, gain (patini, ‘as husband’), partake of, share in (acc.), ib. ; to adopt or embrace (a doctrine), Rājat. ; to undertake, commence, begin, do, MBh. ; Kāv. ; to form (a judgement), MBh. ; to assume (a form), Kathās. ; to enjoy (pleasure), R. ; to take to (dat.), Hariv. ; to come on, approach, appear, AV. ; R. ; Hariv. ; to take effect, succeed, MBh. ; to turn out (anyathā, ‘differently’ i.e. without any effect or consequence), Hariv. ; to admit (a claim), R. : Caus. -pādayati, °te, to cause to enter, introduce into (acc. or loc.), Br. : Desid. P. pitsati, to wish to enter, ŚBr. ; Ā. -pitsate (cf. Pāṇ. vii, 4, 54), to be going to incur or undertake, Daś. [ID=135115] (H2) [Printed book page 682,1] 2. pra-pad f. away, AitBr. [ID=135118] N. of partic. sacred texts, Br. ; GṛŚrS. [ID=135119] (H1) [Printed book page 682,2] 3. pra-pad f. (fr. 3. pad) the fore part of the foot, AV. [ID=135140]