International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Issues in Machine Translation with Reference to the Story “Thakur ka kuan” by Munshi Premchand

Punam Silu

PhD Department of Linguistics, Central University of Rajasthan

Abstract:
This paper explores the issues encountered in machine translation, specifically using Google Translate, in translating Munshi Premchand's short story *Thakur Ka Kuan* ("Thakur's Well"). The study identifies key challenges, including the misinterpretation of cultural references, idiomatic expressions, and contextual meaning that are integral to Premchand's narrative style. Machine translation often produces literal translations that lack the nuance and emotional depth present in the original text, leading to misrepresentation of the story’s themes. For instance, socio-cultural terms and metaphors central to the story's message are frequently mistranslated or omitted entirely. The translation also struggles with maintaining consistency in tense and sentence structure, further diluting the narrative. This paper argues that while machine translation offers convenience for basic translations, it remains insufficient for complex literary works, where cultural, emotional, and contextual understanding is crucial. A comparative analysis with a human-translated version highlights these deficiencies and suggests that human input remains essential for accurate literary translations. The findings underscore the need for improved machine learning models to address these linguistic and cultural intricacies.

Keywords: Machine Translation, Munshi Premchand, Google Translate, Cultural Nuances

1. Introduction:

In Today’s world, everything is dependent on Computer Algorithms. Translation is one of these, which is done using machines. Machines make our work easy, but language is not a phenomenon that Machines can easily understand; it requires a human touch, emotions, cultural meanings, etc., and these can not be fed into any code.

Who doesn’t use Google Translate these days? Most of the people throughout the world use it, and It’s a machine translation tool. In this article, we will discuss a very well-known topic: Machine Translation(MT) and the issues of it. The study identifies issues and shortcomings in the translated document produced by automated translation tools.

1.1 Translation

The word Translation has its root in the Latin word “translatio” which means carrying across and bringing across. Translation means converting a text from one language to another, maintaining its original meaning and essence. The language in which the text is the Source Language, and one in which we translate is the Target Language.

Source text (ST) → Target text (TT)

in the source language (SL) in the target language(TL)

"The Russo-American structuralist Roman Jakobson (1896–1982) described three categories of translation in his seminal paper 'On linguistic aspects of translation'. Jakobson's categories are as follows:

(1) Intralingual translation, or 'rewording' – 'an interpretation of verbal signs using other signs of the same language.'

(2) Interlingual translation, or 'translation proper' – 'an interpretation of verbal signs using some other language.'

(3) Intersemiotic translation, or 'transmutation' – 'an interpretation of verbal signs using signs of non-verbal sign systems.'

(Jakobson 1959/2012)"

Some Types of Translation:

1. Word-for-Word Translation:

Here, the primary unit is "word". In this, we give a substitute for a word from one language to another language without changing the word order.

E.g He is a boy

वह हें एक लड़का

2. Literal translation:

All the words are translated from one language to another, and no additions are made. It is done on languages with similar structures and different structural languages when the text is information-oriented.

E.g अनुवाद कि प्रक्रिया में अब मशीनों कि मदद लि जा रही हें |

Help is now being sought from machines in carrying out the activity of translation.

(page 7, Unit 1 Translation: Its nature and types)

3. Free translation:

The translation is free from word-for-word equivalent, and there is hardly any relation between the translated text and the source text. The emphasis is primarily on the sense of the text.

4. Conceptual Translation:

The message or the concept of the text is more important than the translation of each and every word.

5. Elaborated Translation:

Here, the translated text is longer than the original or the source text; it can be because of the different cultural background.

6. Abridged Translation:

The Translated text is shorter than the original text because of the language's structural differences.

7. Back Translation:

Here, the translated text becomes the basic or source text for the translation back into the original language.

E.g., Hindi - English

English - Hindi

8. Machine translation:

The translation is done using machines or computer software.

1.2 Machine Translation

It is a sub-branch of Computational Linguistics, which works with Natural Language Processing (NLP), which works on developing computer-based translation systems. The goal of Machine Translation is an

automated translation of a text from one language to another with the help of a computer algorithm without human involvement.

1.3 Historical Evolution

The history of Machine translation Started with the work of Warren Weaver and Yehoshua Bar Hillel in the 1950s. They translated languages with the help of linguistic rules and dictionaries, known as the Rule-based approach of MT. But this struggled with the complexity and nuances of language and made the translation unnatural and inaccurate.

More advanced techniques were introduced in the 1960s with the development of computers and computational linguistics. In one of the early MT demonstrations, approximately 60 Russian lines were translated into English in the Georgetown IBM project 1954. Despite its limited success, it generated curiosity in the field.

In the 1970s and 1980s, Researchers started experimenting with statistical methods. Large bilingual corpora were employed to create probabilistic translation models rather than depending exclusively on the rules of linguistics.

Examples from this period include the IBM Candide system and European SYSTRAN systems. Although these statistical models enhanced translation quality, they were still limited by context and ambiguity.

In the late 1990s, the development of example-based and hybrid methodologies—which mix statistical and rule-based approaches began. Systems such as METEO and SMT aimed to solve linguistic complexity through data-driven methods. The internet enabled more training data to be collected, allowing these algorithms to perform better.

In the 2010s, the introduction of Neural Machine Translation was a game changer. Deep learning made this method possible, which completely changed the field. NMT employed neural networks to directly learn patterns from multilingual text pairs rather than depending on established rules or statistical models. NMT improved translation quality and naturalness significantly. Translation became more accurate and fluid as a result of its successful context and idiomatic expression capture.

With the incorporation of transformer types, NMT has continued to improve in the 2020s—these models, like BERT and GPT, enhanced translation by considering broader context windows and comprehending subtleties. To reduce the requirement for enormous bilingual datasets, unsupervised and semi-supervised learning approaches became more popular at this time.

In recent years, the emphasis has switched to multilingual and low-resource languages. Researchers are focused on creating systems that can efficiently translate across many languages and assisting languages with little training data to benefit from transfer learning.

1.4 Research question

- What are the issues of machine translation in Hindi to English translation.

1.5 Aims

- To understand Machine Translation.
- To find out the issues we face while doing machine translation.

2. Literature Review

This section deals with some published research on “Machine translation”. Much research has been done on Machine translation in different languages worldwide, and there are many translation systems in different languages.

There is some debate over who first came up with the idea of machine translation, and it is attributed to

Andrew D. Booth and Warren Weaver of the Rockefeller Foundation in 1947 and, in particular, to a memo written by Weaver that contained the following two sentences.

*"I have a text in front of me that is written in Russian, but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the text."(*Arnold, Douglas, et al., *page 12,2001)*

This memorandum prompted a huge amount of interest and research, and by the early 1950s, there were a large number of research groups operating in Europe and the United States, representing a significant financial investment.

Still, after this much success and the fact that many questions were raised and remained important till now, there was disappointment in the funding authorities.

The uncertainty of funding authorities was brought up in a report authorized by the US National Academy of Sciences in 1964 when it established the Automatic Language Processing Advisory Committee (ALPAC) to report on the state of play with respect to MT in terms of quality, cost, and prospects, as opposed to the existing cost of, and need for translation. The so-called ALPAC Report was devastating, concluding that there were plenty of human translators and that there was no imminent possibility for machine translation to provide accurate translations of general scientific publications. Almost all government support in the USA ended due to this report. Even worse, it resulted in a general decline in morale on the battlefield since early hopes were thought to be unfounded.

The philosopher Bar-Hillel expressed the theoretical concern in a 1959 report, contending that fully automatic, high-quality machine translation was both theoretically and practically impossible. He talked about context-sensitive sentences.

During the 1930s, two researchers created mechanical systems that were focused on semiautomatic translation and multilingual dictionaries (Hutchins, 2004).

The first effort was created by Georges Artsrouni, a French engineer of Armenian descent who had finished his education in Russia before immigrating to France in 1922. He submitted a patent application in July 1933 for what he called a "mechanical brain"—a device that could automatically store and retrieve different kinds of information rather than being a forerunner of contemporary computers. Two prototypes were created (perhaps between 1932 and 1935), and during public displays, they generated a lot of curiosity. The device even won a "Grand Prix" during the 1937 Universal Exposition in Paris (a second prototype was started but never finished; the two still in existence are kept at the Paris Museum of Arts and Crafts). According to Artsrouni, his device might automate the use of telephone directories, dictionaries, and railway schedules. His system was not dedicated to Translation, but from the beginning, he said that this field was one of the most interesting fields. The device could keep simple words in several languages as linguistic data (i.e., on a strip of paper). Using a series of holes along the paper strip in a manner similar to punch cards, each word was uniquely encoded. The system was instructed to find the corresponding translations from the coding strip using a keyboard. Artsrouni was unable to proceed further due to the system. Although he was not a linguist and never discussed the challenges of machine translation, his records showed that he was among the pioneers of the fully automatic system based on multilingual dictionaries.

After Georges Artsrouni, Russian scientist Petr Petrovitch Smirnov-Trojanskij (1894–1951) submitted a patent for a machine that could select and encode words for translation between different languages. Probably, a working prototype of the device was never built. This machine was close to Georges Artsrouni, but Smirnov-Trojanskij's invention only concerned translation. Instead of concentrating on a single

device, Smirnov-Trojanskij's system was created so that a translator may initially utilize it to hunt for translation components at the word level. The text was then edited and corrected stylistically at the very end by a professional text editor or translator. His proposal does not go into great length regarding the challenges of machine translation, but what makes this idea noteworthy is that Trokanskij planned an environment for assisted translation rather than a fully automated procedure (Hutchins, 1986, chapter 2 "Precursors and pioneers").

There are many different approaches, which are broadly categorized into five groups: Direct Machine Translation (DMT), Rule-Based MT (RBMT), Corpus-Based MT (CBMT), Knowledge-Based MT (KBMT), and Hybrid Based MT (HBMT).

RBMT is further divided into 2=two types: Transfer Based MT (TBMT) and Interlingua Based MT (IBMT).

CBMT is also divided into two types: Statistical MT (SMT) and Example-Based MT (EBMT). Neural MT is a type of SMT (Dorr et al. 2004; Seasly 2003).

Direct Machine Translation (DMT):

Only word-to-word matching is used for translation; there is no intermediate representation of the source and target languages. The system may also include pre-processing and post-processing: paring phases for input sentence morphological analysis and target sentence reordering, respectively. The method matches SL words with TL terms using a multilingual dictionary.

Rule-based MT(RBMT):

It translates sentences by applying grammar rules to the source and target languages and then doing a grammatical analysis of both. However, it needs extensive editing, and given how heavily it relies on dictionaries, competency is only attained after a considerable amount of time.

Transfer Based MT (TBMT):

According to this method, a parse tree is produced after the input text has undergone morphological analysis in order to determine its grammar structure. The system uses a bilingual source-target language dictionary and a set of transfer rules to translate an SL parse tree into TL. Using syntactic and semantic generator modules, the target language dictionary, and other tools, the TL text is generated according to the TL grammar.

Interlingua Based MT (IBMT):

In this method, SL text is analyzed to produce an intermediate language-independent code, which is then translated into TL text. The intermediate code format can be utilized for multilingual machine translation because it is independent of SL and TL. The target language generator depends on the specific target language, while the language analyzer depends on SL for the input process.

Statistical MT (SMT):

It operates by referring to statistical models that are based on the analysis of massive amounts of multilingual information. It anticipates determining the relationship between a word from the source language and a word from the objective/source language. This approach has two main parts: the language model and the translation model. The language model generates the likelihood that a string of words will appear in both the source and the target languages, as well as the conditional likelihood that a word in the target language will appear after a word in the source language. The frequency of source and destination word pairings occurring in the corpus accessible for translation is determined by multiplying the likelihood of occurrence of a word in SL by the conditional probability of occurrence of a word corresponding to this word in TL. Machine translation can be performed based on a word, phrase, sentence, or hierarchical

phrase. The N-gram model is typically employed by the translation model. Using the text's preceding words as a guide, the N-gram model forecasts the appearance of the following word.

Neural MT(NMT):

It uses neural networks to construct statistical models with translation as the ultimate goal. As a result, it doesn't rely on specialized systems that are typical of other machine translation systems, like SMT. It provides a single system that can be prepared to unravel the source and target text.

Example-Based MT(EBMT):

The above approach uses analogy as the fundamental translation concept. This method doesn't need a large corpus; it just needs a bilingual corpus of samples that have been stored, and it uses one of the matching algorithms to find the translation which is equivalent to the sentence in the source language. In general, EBMT simply uses the stored examples and the matching algorithm to determine the closest match relating to the input sentence; it does not require any detailed grammatical rule foundation.

Knowledge-based MT(KBMT):

This method takes the linguistic data from SL and stores it in the knowledge base that will be utilized for translation. Bilingual dictionaries, linguistic structure, pre-stored translation data, domain-specific information dictionaries, etc., are all used to extract information.

One of the prominent names in Machine Translation is Bonim J. Dorr. She talked about divergence and gave two types of classification for divergence. Many times, the naturally occurring conversion of one language into another takes on a different form than the original. Divergence describes this phenomenon of translation. Divergences between languages are complicated by cross-cultural differences. Depending on the unique language structures, many divergences between pairs of languages can be found.

Dorr's Classification of Divergence

There are two types of divergence:

1. Syntactic Divergence

It's further divided into five types:

1. Constituent Order Divergence: When the constituent order in the two languages is different, as in the case of verb phrases and inflectional phrases, which are head-final in German but head-initial in English and Spanish, there is a divergence.
2. Preposition Stranding Divergence: Differences in proper governor explain preposition stranding differences between languages translated.
3. Long Distance Movement Divergence: Divergence happens when the choice of bounding node is allowed in one language (such as Spanish) but not in the other (such as English and German), and the distance between co-referring elements does not allow for more than one bounding node.
4. Null-Subject Divergence: Null-subject languages have a feature that subjects can freely invert into post-verbal positions. For instance, a null element in the subject position can be pro in Spanish but not in English or German.
5. Dative divergence: Divergence explains differences in how dative construction is used. Alternating dative construction is permitted in English and Spanish but not in German.
6. She also discussed Adjunction divergence, Movement divergence, Verb movement divergence, and Pleonastic divergence.

2. Lexical-Semantic Divergence

The methods used to resolve lexical-semantic divergences are based on the extended version of the lexical conceptual structure called the interlingua representation.

1. Thematic divergence occurs when the main verb in one language becomes the verbal object in another language (argument structure changes).
2. Promotional divergence: The modifier is realized as the primary verb in one language but as an adverbial phrase in another (head flipping).
3. Structural divergence: In one language, the verbal object is realized as a noun phrase; in the other, it is realized as a prepositional phrase (changes in argument structure).
4. Conflational divergence: where the meaning of a single word in one language necessitates using at least two words in the other language(N-to-1 and 1-to-N lexical gaps, default/exception).
5. Categorial divergence: A shift in the category (a change in category).
6. Lexical divergence: The event is lexically realized as the primary verb in one language but as a distinct verb in another (N-to-1 lexical gaps, support verbs).
7. Demotional divergence: A primary verb in one language is realized as an adverbial modifier in the other (head flipping).

3. Research Methodology:

In this paper, we have used the Descriptive and Empirical methods. Here, we have selected a story written by Munshi Premchand, "Thakur ka Kuan" written in Hindi language and translated this story into English using the Google Docs Translate tool.

After translating the story, we put both documents side by side and did a close introspection to find if the translation was done correctly or not, whether the translation was accurate or not, whether the translation was credible or not, or whether the translation sounded natural or not, we tried to check all the shortcomings and issues which we face while translating any text using automated translating tools.

4. Analysis

Translation of any text demands a very deep understanding of grammar and culture here we need to know the grammatical rules of the languages. Even people who are professional translators make mistakes, so we can't assume that machine translation will be error-free and accurate.

While examining both documents, we have found many issues in the English-translated text of the story. The most occurring problem was the Accuracy and naturalness of the text, and the problem in finding the equivalent of any word, ignoring functional words, at someplace over-translation and at someplace in-translation was done.

4.1 Lack of Accuracy

One of the most common issues is Accuracy while doing Machine Translation. If we look into the translated document, most of the paragraphs are Inaccurate, for example:-

Original Text:-

जोखू ने लोटा मंहु से लगाया तो पानी में सख्त बदबू आई।

Translated Text:-

When Jokhu applied Lota to Manhu, the water had a strong smell.

By looking at the text, we can say the translated text is not accurate the right translation for this will be, "As Jokhu brought the *lota* to his lips to drink, the water gave off a nasty smell."

Original Text:-

ठाकुर के कुएं पर कौन चढ़ने देगा ? दूर से लोग डाँट बताएगे। साहू का कुआँ गाँव के उस सिरे पर है, परन्तु वहाँ कौन पानी भरने देगा ? चौथा कुआँ गाँव में नहीं है।

Translated Text:-

Who will allow people to climb Thakur's well? People will scold from a distance. Sahu's well is at the other end of the village, but who will allow water to be filled there? The fourth well is not in the village.

All most in every paragraph, we can find the issue of Accuracy.

4.2 Lack of Naturalness

It is obvious that a machine can not sound natural because it has precoded vocabulary; translation of words is possible, but machines can not give the human touch.

For example:-

Original Text:-

गंगी ने पानी न दिया। खराब पानी से बीमारी बढ़ जाएगी इतना जानती थी, परंतु यह न जानती थी की पानी को उबाल देने से उसकी खराबी जाती रहती हैं। बोली, "यह पानी कैसे पियोंगे? न जाने कौन जानवर मरा हैं। कुएँ से मै दसूरा पानी लाए देती हूँ।"

Translated Text:-

Gangee didn't give water. She knew that diseases would increase due to bad water, but she did not know that by boiling water, its bad qualities go away. She said, "How will you drink this water? Don't know which animal has died. I will bring other water from the well."

By looking at the translated text, we can say it seems a bit weird, it doesn't sound natural.

4.3 Wrong Translation

Sometimes, it happens that the machine translates a sentence wrong, maybe because the words that are used in the sentences are not encoded in their system. To understand this issue, we can see the examples given below:

Original Text:-

कोई पचास मांगता, कोई सौ। यहाँ बेपैसे-कौडी नकल उडा दी। काम करने का ढ़ग चाहिए।

Translated Text:-

Some ask for fifty, some a hundred. Here, every penny was copied. Need a way to work.

Here, in the 2nd sentence of the paragraph, we can see that the translation of "यहाँ बेपैसे-कौडी नकल उडा दी।" is wrong, it means something else, it should be He obtained the copy without paying a single cowrie or paisa. The machine translated the sentence wrong, maybe because of the different languages of the languages. And this issue is also very common.

4.4 Lack of Discourse

Understanding the discourse of the sentence is very difficult for a machine as we know; translating an idiom is so difficult for a machine because the machine cannot understand the meaning behind the sentence. It is just the translation of sentences according to the data fed into the machine. That's why machine translation shows an issue if the meaning is unclear.

In 4.3's example in the last line, " काम करने का ढ़ग चाहिए।", the translation is "Need a way to work." but if we think thoroughly, it should be "One should know the art of manipulation."

4.5 Intranslation

Intranstion happens when words are omitted, or the translation is reduced. To understand this, we can look at some examples:

In the first paragraph " जोखू ने लोटा मंहु से लगाया तो पानी में सख्त बदबू आई। गंगी से बोला "यह कैसा पानी है ? मारे बास के पिया नहीं जाता। गला सुखा जा रहा है और तू सडा पानी पिलाए देती है !"

Translated text:

When Jokhu applied Lota to Manhu, the water had a strong smell. Said to Gangi, “What kind of water is this? Can't drink due to bad smell. My throat is going dry and you give me water to drink!”

In the translated text we can see in the second sentence subject is dropped it should be He said to Gangi and in the last sentence,”गला सुखा जा रहा है और तू सड़ा पानी पिलाए देती है !”, here in the translation is reduced, it is “My throat is going dry and you give me water to drink!”, the translation should be like this “My throat’s dry, and you’re making me drink this foul water.”

4.6 Overtranslation

Overtranslation happens when the translated text has more than enough words making the translated text long. For example

Original Text:-

गंगी क्या जवाब देती, किन्तु उसने वह बदबुदार पानी पीने को न दिया।

Translated text:-

What answer could Gangi give, but she did not allow him to drink that stinking water.

In the translated text, there is nothing wrong with the translation, but it would have been better if the first sentence were a little short, like in this: “Gangi kept quiet, but she did not let him drink the stinking water.” It’s easy to understand and short.

4.7 Lack of Word Equivalent and Cognate Transfer

Sometimes, it's hard to find an equivalent word in another language, and it becomes an issue because, in this case two things happen the word will stay the same in the translated text, or the wrong translation will happen.

For example:-

मार्के, मोहतिमिम and जगत in these three words for मार्के the translation was marque's, but it should be landmark court judgement.

For मोहतिमिम the word was the same in the translated text, but it should be Court Officials.

Original Text:-

“कितनी अकलमंदी से एक मार्के के मुकदमे कि नक़ल ले आए |” नाजिर और मोहतिमिम, सभी कहते थें, नकल नहीं मिल सकती। कोई पचास मांगता, कोई सौ।

Translated Text:-

“How wisely you brought a copy of a marque's case.” Nazir and Mohtimim, everyone used to say that imitation cannot be found.

Correct Translation:-

“How cleverly he had obtained the copy of a landmark court judgement. The *nazir* and other court officials had said that a copy couldn’t be given.”

For जगत, in some places, it doesn’t change, and in someplace, it is world, but it should be platform.

Original Text:-

गंगी जगत की आड मे बैठी मौके का इंतजार करने लगी।

Translated Text:-

Gangi sat under the cover of Jagat and started waiting for the opportunity.

Correct Translation:-

Gangi came and sat close to the platform around the well, sheltering herself, and waited for an opportunity.

Original Text:-

दानों पानी भरकर चली गई, तो गंगी वक्षृ की छाया से निकली और कुएँ की जगत के पास आयी।

Translated Text:-
When she filled the grains with water and went away, Gangi emerged from the shadow of the tree and came to the world of wells.
Correct Translation:-
Both the women walked away after drawing water. Gangi came out of the tree's shadow and walked towards the well.
These examples tell us about both the issues of finding word equivalents and doing cognate transfer.

4.8 Lack of Punctuation Marks

While doing the translation machine, most of the time omits Punctuation marks and makes the translated text hard to understand.
Original Text:-
जरुर कोई जानवर कुएं में गिरकर मर गया होगा, मगर दूसरा पानी आवे कहां से ?
Translated Text:-
Surely some animal must have died after falling into the well, but where would the other water come from?
Here, after some there should be a coma, but it's not there the same type of problems are in many paragraphs.
Besides these issues sometimes there are issues in finding the correct pronoun and omitting functional words like determiners, conjunctions, etc.

5. Summary

In this paper, we have looked at the challenges or issues faced by Machine Translation systems, We have discussed what is Translation and Types, In particular, about Machine Translation's historical background of Machine translation and the different approaches of MT. We have also discussed Dorr's classification. In particular, this paper aimed to find issues while translating a text using automated translating tools with reference to a story, and by this, we have identified many issues such as Accuracy, Naturalness, Intranslation, Overtranslation, Word Equivalent, Cognate transfer, Discourse, Wrong translation, Punctuation marks, etc.

References:

1. Ameur, Mohamed Seghir Hadj, Farid Meziane, and Ahmed Guessoum. "Arabic machine translation: A survey of the latest trends and challenges." *Computer Science Review* 38 (2020): 100305.
2. Anju, E. S., and K. V. Manoj Kumar. "Malayalam to English machine translation: An EBMT system." IOSR Journal of Engineering (IOSRJEN) 4.1 (2014): 18-23.
3. Arnold, Douglas, et al. "An Introductory Guide." (2001).
4. Godase, Amruta, and Sharvari Govilkar. "Machine translation development for Indian languages and its approaches." International Journal on Natural Language Computing 4.2 (2015): 55-74.
5. Hutchins, William John. Machine translation: past, present, future. Chichester: Ellis Horwood, 1986.
6. Jakobson, Roman. "On linguistic aspects of translation." On translation. Harvard University Press, 1959. 232-239.
7. Munday, Jeremy, Sara Ramos Pinto, and Jacob Blakesley. *Introducing translation studies: Theories and applications*. Routledge, 2022.
8. Muzaffar, Sharmin, Pitambar Behera, and Girish N. Jha. "Classification and resolution of linguistic divergences in English-Urdu machine translation." WILDRE: LREC (2016).

9. Sitender, et al. "A comprehensive survey on machine translation for English, Hindi and Sanskrit languages." Journal of Ambient Intelligence and Humanized Computing (2021): 1-34.
10. Machine Translation 2, EpgPathsala.