Propriety and Effectiveness in English-Hindi Google Translation: A Study of Select Official Texts

[A minor project of Linguistics (Major 1)]

Utkarsh Mishra

BA (NEP), Semester VI

Roll No: 2110011015277

Supervisor

Prof. K. Srikumar

Department of Linguistics, University of Lucknow

May, 20

Table of Contents

Acknowledgments

First and foremost, I extend my heartfelt thanks to my supervisor Prof. K. Srikumar, for his invaluable guidance, encouragement, and insights throughout this project. His expertise and dedication have been a source of great inspiration, and his constructive feedback has been crucial in shaping and adding quality to my work.

I am profoundly grateful to my parents, whose unwavering support and encouragement have been the foundation of my academic journey. Their belief in my abilities and their constant motivation have provided me with the strength to persevere and achieve my goals.

I would also like to acknowledge the support and constructive comments of my classmates. Their presence has enriched my academic experience.

Abstract

This study investigates the propriety and effectiveness of Google Translation of select official texts from English to Hindi. A dataset of 100 sentences collected from the authentic official sources such as the New Education Policy, NCERT History of India, the Constitution of India, the Union Budget, the Gazette of India, and the Rajbhasha Report has been analyzed for lexical, syntactic, semantic and structural errors by comparing Google Translate outputs with their official translation. Results indicate a significant presence of errors in the Google Translate outputs, particularly in the translation of complex documents like the Gazette of India and the Constitution of India. Simpler documents like the New Education Policy document and Rajbhasha Report have fewer errors. The study highlights the inadequacies of google translation in maintaining formality and cultural appropriateness, which is so essential for official texts. Despite its efficiency, Google Translate's limitations suggest the need for human intervention to ensure accuracy and contextual relevance in formal translation. The identification of the nature and types of errors may be found useful for improving the effectiveness of machine translation.

Introduction

Throughout history, translation and interpretation have played a vital role in cross-linguistic communication. Through translation humans could understand culture, civilization and literature of communities speaking different languages. The need for communication and information exchange among people from different countries has become increasingly significant over time. Since most countries have their own languages, the field of translation plays the crucial role of bridging language differences. Translation also functions as a bridge connecting different languages and cultures by making meaningful contribution to cross-cultural communication (Kafipour & Jahanshahi, 2015).

Initially, the skill of translation was in its infancy and limited to very few bilinguals. The last century saw a great advancement in the study of translation as a discipline. Munday (2001) opines that the academic study of translation truly began in the second half of the previous century and this new discipline emerged as Translation Studies.

Translation is an activity that interprets a text from one language to another language without changing the content of the source text. Translation deals with the transfer of meaning from the source language into the target language. However, every translation, in a true sense, is a slight deviation from the original text in terms of meaning, style and effect. An effective reproduction of a text in the target language has the closest equivalent message of the source text. (Nida and Taber; 1996: 12). A truly effective translation also maintains propriety of formality and style of the source text as much as possible.

With the growing need of translation and advancement in the field of information technology human translators are slowly being replaced by machine translation tools. Machine translation is the process of using artificial intelligence to automatically translate text from one language to another without human involvement. Lots of machine translation tools are freely accessible on the web.

However, the most used and popular one is Google Translate (Aiken and Balan, 2011; Korošec, 2011). Based on the research done by Aiken et al (2011), Google Translate has shown the best accuracy among other machine translations. Google Translate is the most used tool for machine translation and has a market share of 70.37%. Currently Google-Translate can translate into more than 100 languages including English and Hindi.

History of Machine Translation

Machine Translation has a rich history. The field originated during the World War II when researchers like Warren Weaver and his team at MIT explored linguistic challenges and feasibility of computational translation. Inspired by Weaver's efforts, Georgetown-IBM experiment was conducted in 1954 where a computer translated 60 Russian sentences into English. The era also saw the first ever machine translation conference in 1952.

From 1966 to 1995, machine translation saw exponential growth. Rule Based Machine Translation evolved and systems like SYSTRAN came into being in 1970s. Rule Based Machine Translation (RBMT) systems relied on linguistic rules and grammatical structures to translate text between languages. Systems like SYSTRAN employed extensive linguistic knowledge encoded in rule sets. These rules governed transformation of source language sentences into target language equivalents. The RBMT approach demonstrated success but faced many challenges in handling ambiguity and dealing with complex nature of languages.

From 1996 to 2012 , MT underwent significant advancements on the World Wide Web . Example Based Machine Translation(EBMT) emerged as a paradigm shift in the field during the late 1990s. This system relied on repositories of bilingual sentence pairs, treating translation as a matching problem. By utilizing a database of aligned source and target language examples, EBMT offered flexibility and adaptability in handling diverse linguistic contexts. EBMT played a crucial role in shaping the trajectory of machine translation by emphasizing the importance of exemplar-based learning.

Statistical Machine Learning(SMT) revolutionized language translation between 1990 and 2010. SMT relied on statistical models, analyzing large bilingual corpora to derive translation probabilities. This data driven approach improved the translation quality marking the transformative era before the arrival of neural networks.

The period from 2013 to present has been characterized as the Age of Neural Machine Translation(NMT). NMT revolutionized language translation by employing artificial intelligence to enhance the accuracy and fluency of translations. This approach powered deep learning models, which gave better results than the statistical and example based models.

Since 2016 there has been a massive improvement in the translation engine of Google Translate. With the use of NMT and the recent introduction of Transformer architecture in 2017, Google Translate has started handling complex linguistic structures and expressions quite successfully.

Machine translation has seen significant advancements in becoming increasingly effective due to improvements in artificial intelligence and neural networks. Modern machine translation systems, like Google Translate and DeepL, use sophisticated algorithms and large datasets to provide more accurate translations. These systems can now handle idiomatic expressions, complex sentence structures, and various dialects more proficiently than earlier rule-based or statistical models. Additionally, continuous learning and user feedback loops further enhance their accuracy. This growing effectiveness makes machine translation a valuable tool for global communication, aiding in everything from casual conversation to professional and academic exchanges.

Although machine translation (MT) has made some incredible advances with natural language processing (NLP), it's still not completely accurate. If you use a tool like Google Translate for your translation projects, some errors are still inevitable. Some language pairs have notoriously poor translation quality, even with the latest innovations in AI. Often times in translation from English to Hindi, we face errors

related to cultural expression, idiomatic expressions, syntax and appropriate collocation.

Identifying errors in English-Hindi machine translation of official texts is crucial for improving translation propriety and effectiveness. By looking at these errors, we can see common mistakes, like wrong grammar, incorrect context, and missed cultural details. By organizing these mistakes into categories, we can find out exactly where the translation system needs improvement. This will help us create specific fixes to make translations more precise, ensuring that the translated texts keep their original meaning and formality. This will lead to more reliable translations in official documents, making communication clearer and reducing misunderstandings.

Theoretical Base

Translating is the written transfer of text messages from one language into another language (Hoed; 2006: 5). In this case, the translated text is called the source text and the language is called the source language. With regard to the translation, the text drafted by the translator is called the target text, and the language is called the target language. Catford (1965:20) stated translation is the transfer of the textual material in the source language into the equivalent textual material in the target language.

Types of Translations

Newmark (1988: 45-47) classifies translation into eight kinds, which are as follows:

1. Word for word translation: In word-for-word translation, the translator keeps the same word order as the original language and translates each word one at a time based on their general meaning, without considering the context.

2. Literal Translation: It means using the closest grammatical structures from the source language in the target language. However, each word is translated individually, without considering the context.

3. Faithful Translation: In this type of translation, the exact meaning from the original language is given into the translated language, even if it doesn't fit perfectly with the translated language's grammar.

4. Semantic Translation: It focuses on capturing the meaning and essence of the original language. It allows the translator to be more flexible and creative, using their intuition to convey the meaning effectively.

5. Adaptation Translation: It is the most flexible type of translation, often used for drama and poetry. It allows for significant changes to better fit the cultural and artistic context of the translated language.

6. **Free Translation:** It focuses on conveying the message and meaning from the original language without sticking to its exact form or structure. The content is translated, but the original shape and wording are not closely followed.

7. Idiomatic translation: It involves conveying the message or meaning into the target language, but sometimes the exact shade of meaning might change because idioms are used that weren't present in the original language.

8. Communicative Translation: In this type of translation, the meaning from the original language is adapted to ensure that the message can be received and understood by the audience of the translated language.

Major Theories in Translation

In order to understand the process of translation at some depth, a few important theories in translation have been reviewed here. The focus is mainly on three theories- Skopos Theory, Equivalence Theory and Domestication and Foreignization Theory.

Skopos Theory, proposed by Hans J. Vermeer (1989), emphasizes the purpose (skopos) of a translation as the primary determinant of the translation methods and strategies. According to this theory, the end goal of the translation dictates how it should be carried out, rather than adhering strictly to the source text.

Eugene Nida's (1964) **equivalence theory**, focuses on the degree to which the translation matches the source text in terms of both form and meaning. Nida talks of two types of equivalence namely formal equivalence and dynamic equivalence. Formal equivalence seeks a close match to the source text structure, while dynamic equivalence aims at naturalness and the same impact on the target audience as the source text had on its audience.

Lawrence Venuti's (1995) concept of **domestication and foreignization** deals with the translator's approach to the cultural elements in the source text. Domestication involves adapting the text to fit the cultural context of the target audience, making it more familiar and accessible. Foreignization, on the other hand, retains the foreign elements to preserve the original culture's flavor and promote cross-cultural understanding.

However, despite the fact that many novel theories were proposed by translation scholars, the concept has always emerged out of the two traditional approaches of “word vs. sense” or “literal vs. free” translation (Newmark, 1998). Whereas a word-for-word translation attempts to maintain the meaning of source text in the target language, its sense-for-sense counterpart permits a translator significant scope for negotiation and interpretation of meaning.

These foundational translation theories are essential for examining machine translation errors. Google Translate, as an automated tool still struggles to match the accuracy and naturalness achieved by skilled human translators. Consequently, preliminary investigations indicate that common errors in Google Translate arise from the divergence between the practicality of word-for-word and sense-for-sense translation approaches. These theories will guide my analysis of errors in formality, lexical choices, syntax, stylistics, propriety, and overall effectiveness.

The translation of scientific texts, official texts, and government documents gives as high importance to formal equivalence as to capturing its semantic essence. This approach prioritizes accuracy, precision, and adherence to the source text's structure, terminology, and meaning. The mechanism of formal equivalence is employed in this study because the data for this paper is based on texts having high formality and are downloaded from government sources.

Errors in Translation: Causes and Effects

Translation is not an easy work as it is not simply the substitution of words in one language by words in another language, but the transfer of meaning in the most natural way. Therefore, proper training of translators is required. A translator needs to know different types of errors. An error in translation refers to the production of incorrect forms in speech and writing by a non-native speaker of the target language, due to the incomplete knowledge of the rules of that target language.

The errors in translation have been classified as follows:

1. Linguistic Errors : These are mistakes in the basic language elements like grammar, vocabulary, or spelling. It takes place when you accidentally use the wrong word or mess up the sentence structure in the translated text.
2. Cultural Errors: These mistakes happen when the translation doesn't fit the cultural context of the audience. Something that makes sense in one culture gets lost or misunderstood in another.
3. Stylistic Error: These are errors in the tone or style of the translation. It happens when the translated text doesn't match the formality or casualness of the source text.

Silalahi, Rafli, and Rasyid (2018) found some more types of errors that appeared in the scientific text translation from English into Indonesian. These were lexical , morphological and syntactic errors.

1. Lexical Errors: Lexical errors occur when a word or phrase is incorrectly translated, often due to a misunderstanding of the word's meaning or context. These errors

typically involve choosing an incorrect word in the target language that does not convey the intended meaning of the source text.

2. Morphological Errors: Morphological errors happen when the internal structure of words is not properly translated. This includes errors in affixes (prefixes and suffixes), word formation, and inflectional forms (such as tense, number, or gender).

3. Syntactic Errors: Syntax errors involve incorrect sentence structure, which can lead to confusion or a complete change in meaning. These errors occur when the grammatical rules of the target language are not followed, affecting how words and phrases are arranged.

Architecture of Machine Translation Systems

Image 1 **:** *Figure showing architecture of LSTM-RNN Translation System*

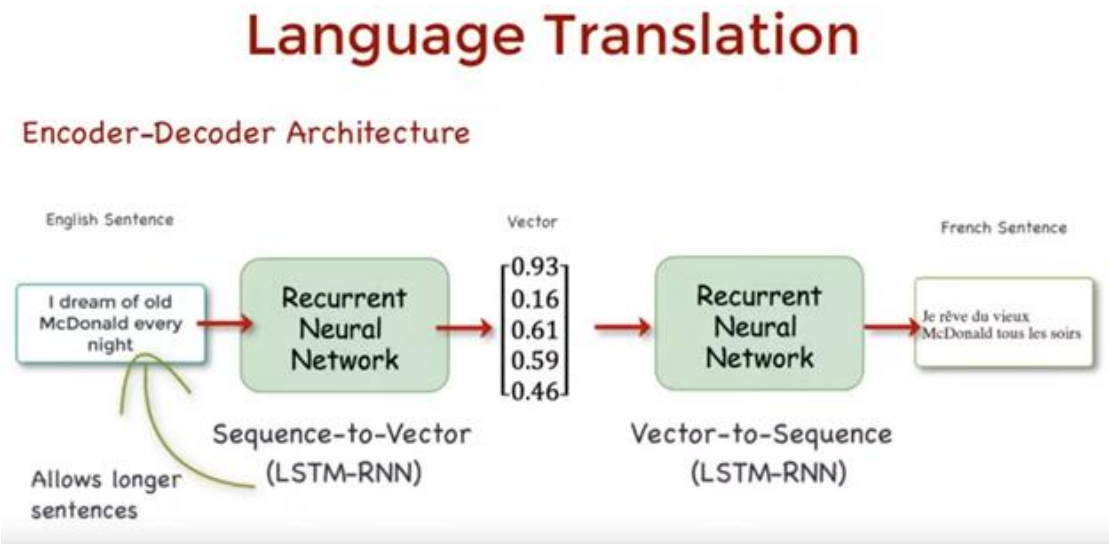


Image-1 illustrates the Encoder-Decoder Architecture used in machine translation systems, specifically using Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN).

The Encoder processes the sentence word by word and converts the entire sequence into a fixed-length vector representation. This vector captures the meaning of the sentence.This fixed-length vector serves as an intermediate representation of the input sentence, encapsulating its semantic meaning.The vector is then passed to a Decoder, which is another RNN using LSTM units.The Decoder generates the target language sentence (in this case, French) from the vector.

This method is powerful for translating sentences of varying lengths while maintaining the contextual and semantic integrity of the original sentence. While the Encoder-Decoder architecture with LSTM-RNNs was a crucial development in the evolution of Google Translate, the current systems have advanced to use Transformer models with attention mechanisms, providing even more accurate and efficient translations. This shift represents the state-of-the-art in neural machine translation technology.

Errors in Machine Translation

In machine translated texts, translation errors are classified into several types. (Vilar 2006: 698) divide the errors into five big classes such as “Missing Words”, “Word Order”, “Incorrect Words”, “Unknown Words” and “Punctuation” errors.

Missing words error refers to the missing items which ought to be present in a sentence. This error is distinguished into content and filler words. Content word concerns with missing words which are important to express the meaning of a sentence whereas filler word concerns with missing words which are merely needed to form a grammatically correct sentence.

Word order error directs to the wrong structure of the sentence. This error is distinguished into word and phrase based orderings, and within each of these categories is more specified into local and long range reordering. The distinction between local or long range is that local reorders the words in a same syntactic chunk of the source text, whereas long range moves the words into another chunk.

Incorrect words error refers to the incorrect translation. This error is distinguished into wrong lexical choice, incorrect disambiguation, incorrect form, extra words, style, and idiom errors.

Unknown words error directs to untranslated words. This error is distinguished into unknown stem and unseen form. Unknown stem is a condition when MT cannot find the right translation due to the source lexeme while unseen

form is a condition when MT cannot recognize the morphological complex words of SL.

Punctuation error refers to the errors related to marks used in writing.

These types of errors are prominent in Google Translate. Hence, Google Translate alone cannot be relied upon for the translation of official texts. Once the text is translated, we need to proofread it and then approve it for higher authorities.

A Brief Review of Available Research

Machine translation (MT) has become an indispensable tool in the era of globalization, enabling real-time translation of text and speech across diverse languages. Despite significant advancements in MT technologies, such as neural machine translation (NMT) and transformer models, the accuracy and reliability of translations remain imperfect.

This review aims to explore the prevalent types of errors encountered in machine translation systems, analyzing the underlying causes and examining the research dedicated to addressing these challenges. By understanding the common pitfalls and the efforts to mitigate them, we can better appreciate the current state and future directions of machine translation technology.

Anggaira and Hadi (2017) in their study examine the linguistic errors in narrative text translations from English to Indonesian using Google Translate. The research identifies and categorizes errors into morphological, syntactic, and semantic aspects, using the narrative text "Snow White" as a case study. The results indicate that morphological errors are the most frequent, followed by syntactic and semantic errors. The study underscores that while Google Translate can serve as a preliminary tool for translation, it often fails to capture the nuances of context and cultural meaning, resulting in significant inaccuracies. Consequently, translations produced by Google Translate require careful post-editing to ensure accuracy and appropriateness, especially for educational purposes. The authors advocate for caution among

educators who use machine translation tools, emphasizing the need for thorough proofreading and correction of errors to avoid misleading translations (Anggaira & Hadi, 2017).

The research paper "Error Analysis of Abstract Translation in Scientific Writing by Using Google Translate" by Nyayu Yayu Suryani and Tira Nur Fitria investigates the types of errors that occur when using Google Translate for translating scientific abstracts from Indonesian to English. The study involves a qualitative analysis of six abstracts from various undergraduate papers, highlighting frequent errors in lexico-semantics, tense, prepositions, word order, verb group usage, and voice (active/passive). The research finds that Google Translate often produces inaccurate translations, particularly with respect to word choice, syntactic structure, and semantic accuracy. The authors emphasize the need for professional translation services, especially in academic contexts, to ensure the quality and reliability of translated texts. They suggest that future studies should explore similar issues with a broader range of sources and recommend involving human translators to enhance translation quality (Suryani & Fitria, 2021).

Muzaffar and Behera (2021) provide a comprehensive qualitative evaluation of Google Translate's English-Urdu machine translation systems, comparing Phrase-Based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT). Utilizing a model corpus of 100 English sentences, the authors employ the Inter-translator Agreement (IA) method, leveraging Fleiss' Kappa statistical measure to assess comprehensibility and grammaticality. The results indicate that PBSMT performs poorly, with Kappa scores of 0.24 and 0.22 for comprehensibility and grammaticality, respectively. This system also exhibited high word error rates (WER) of 21.11% and sentence error rates (SER) of 72.39%. Conversely, the NMT system demonstrated superior performance, with Kappa scores of 0.61 for comprehensibility and 1 for grammaticality, alongside improved WER and SER of 32.58% and 28%, respectively. The study underscores the limitations of current MT systems, including the restricted scope of computational errors and general-domain data. Despite the superior performance of NMT, the authors caution that it does not fully resolve all MT challenges. They advocate for further research, particularly extending evaluations

to other Indian languages to enhance MT reliability and accuracy (Muzaffar & Behera, 2021).

Malik and Baghel (2021) present a study on enhancing the performance of machine translation systems for the English-Hindi language pair using Natural Language Processing (NLP) techniques. The research focuses on translating English verses into Hindi and evaluates the effectiveness of their proposed NLP-based translation system against traditional tools like Google Translate and Bing. The authors utilized various NLP metrics, including tokenization and ATEC scores, to measure translation precision and recall. The results indicate that the NLP-based translator significantly outperforms the existing systems, with an ATEC score of 0.508 compared to 0.402 and 0.342 for Google and Bing, respectively. This improvement is attributed to the detailed tokenization and probability calculations inherent in the NLP approach. The study concludes that the NLP tool provides a more accurate translation, though it acknowledges that machine translation cannot yet match human translation accuracy. The authors recommend further refinement and expansion of this NLP-based system to enhance its applicability and reliability in translating between English and Hindi (Malik & Baghel, 2021).

Despite these advancements, none of the existing studies have comprehensively analyzed the translation of official government texts between English and Hindi. Official texts are critical for clear communication in governance and require a high level of formality and propriety, which are often challenging for machine translation systems to maintain. This study will try to fill this gap by evaluating the propriety and effectiveness of Google Translate in translating official texts from English to Hindi.

The Aim of Study and Research Questions

The aim of this research is to evaluate the propriety and effectiveness of Google Translate in translating official texts from English to Hindi. Specifically, the study will analyze translation errors, assess the maintenance of formality, and examine the cultural and contextual appropriateness of the translations. By identifying

and categorizing common issues, the research aims to provide actionable insights for improving machine translation systems, ensuring they meet the high standards required for official documents. Ultimately, this study seeks to enhance the reliability and accuracy of machine translation tools in formal and governmental contexts.

The research questions guiding this study are:

1. What types of errors are most common in Google Translate's English to Hindi translations of official texts?
2. To what extent does Google Translate maintain the necessary level of formality in translating official documents?
3. How contextually appropriate are the translations produced by Google Translate for official use?

These questions aim to identify specific challenges and areas for improvement, ultimately enhancing the effectiveness and reliability of machine translation tools in formal settings.

Methodology

The data for this study was meticulously collected from various government sources, ensuring a comprehensive representation of bilingual data to assess formality, appropriateness, and propriety in official texts. The sources included the Ministry of Education, Ministry of Finance, legislative documents, and other advisory bodies. The primary goal was to analyze English-to-Hindi translations, specifically focusing on their accuracy when performed by Google Translate compared to professional government translations.

Initially, the sentences were extracted from the PDFs, cleaned, and all pointers and bullets were removed to maintain uniformity. Each sentence was meticulously selected to ensure a balanced representation from each source. To streamline the process, a Python script was employed to calculate the average number of words per sentence, given the impracticality of manual calculations for large datasets.

The organized data was then structured into a table comprising 100 rows and three columns. The first column contained the source sentences in English, while the second column housed their correct translations in Hindi, provided by highly qualified government translators, thus ensuring authenticity and credibility. The third column contained the translations generated by Google Translate for the same sentences.

Subsequently, the Google Translate outputs were compared against the government-provided translations to identify errors. These errors were meticulously classified and defined based on their nature and impact on the translated text's accuracy. The classification process involved identifying common error types such as grammatical inaccuracies, contextual misinterpretations, and inappropriate formality levels.

Finally, the study concluded with a comprehensive analysis of the findings, highlighting the specific areas where Google Translate falters and the implications of these errors on the accuracy and appropriateness of translated official documents. The research underscored the need for improved machine translation systems for official use, ensuring that translated texts maintain the required level of formality and propriety.

Analysis

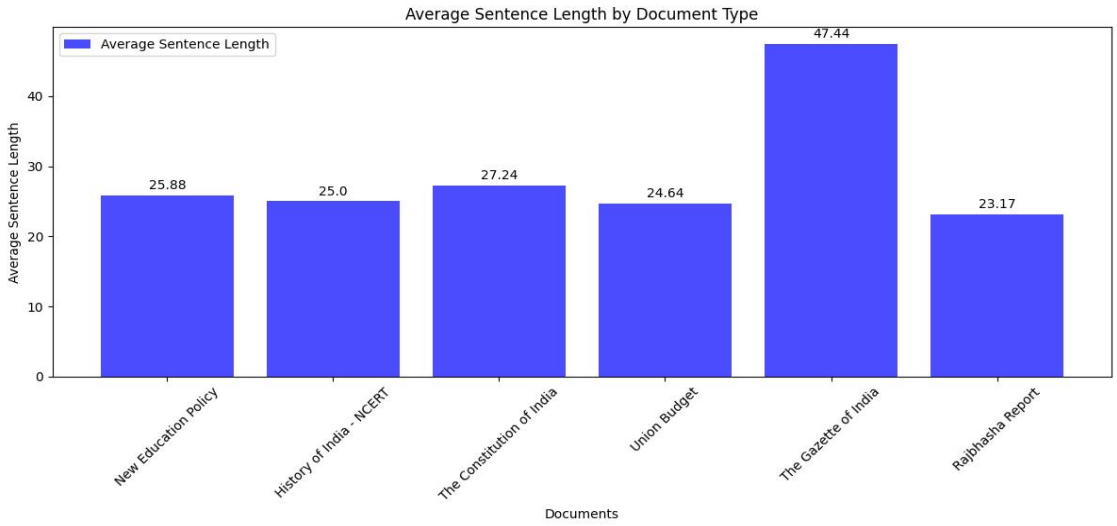
A brief analysis of the data is provided in the table below to convey the essence of the dataset and illustrate how it covers sentences for analysis from various government fields such as finance, education, advisory, and legislative.

Table 1: *Table showing Segregation of Source Texts*

Source Text	**Number of Sentences**	**Average Length of Sentence**
New Education Policy	18	25.88
History of India - NCERT	6	25.00
The Constitution of India	40	27.24
Union Budget	20	24.64

The Gazette of India	10	47.44
Rajbhasa Report	6	23.17
Total	**100**	

Image 2: *Graph showing avg. number of words in a sentence in different source text*



This analysis dives deep into the various types of errors observed in translations performed by Google Translate, specifically focusing on its preference for code-mixing, challenges in translating noun phrases from English to Hindi, structural ambiguity, literal translation and lexical anomalies.

Google Translate's preference for code-mixing

A significant feature was observed when sentences were translated using Google Translate from English to Hindi. It was found that the Hindi translation of English official texts incorporated content words from different linguistic registers like Persian, Arabic, and English. Using words from these languages makes the translation more readable for the general public, on the other hand it makes the translation less acceptable in formal correspondence. The use of such words in translation makes the text sound less formal. As a result, the propriety and effectiveness of formal texts are diminished. Below are some examples along with their analysis:

The word ‘दफ्तरों’ has been used for ‘offices’. The word ‘दफ्तर’ originates from Persian. A more suitable translation would be ‘कार्यालयों’ because ‘कार्यालय’ has its roots in Sanskrit, and Sanskrit-derived vocabulary is preferred for official purposes, making this translation more formal and effective. (APPENDIX 1, SENTENCE 1)

The phrase ‘Needless to say that’ has been translated as ‘कहने की जरूरत नहीं कि’. This translation is not very appropriate for formal texts and office documents. The word ‘जरूरत’ is borrowed from Persian and is generally used by the general public. A better translation of this phrase would be ‘कहने की आवश्यकता नहीं है कि’. This translation preserves the formality and propriety required in official texts. (SENTENCE 3)

The translation of ‘Provided that nothing in this clause’ is ‘बशर्ते कि इस खंड में कुछ भी’. The word ‘बशर्ते’ is of Persian origin and does not sound formal, making its usage less preferred for official texts. A more proper and effective translation would be ‘परन्तु इस खंड की किसी बात’. (SENTENCE 7)

The phrase ‘by law’ is better translated as ‘विधि द्वारा’ instead of ‘कानून द्वारा’. The word ‘कानून’ has been borrowed into Hindi from Persian.(SENTENCE 9)

The sentence ‘A Money Bill shall not be introduced in the Council of States’ is translated as ‘धन विधेयक राज्य सभा में पेश नहीं किया जाएगा।’. The word ‘पेश’ has Persian roots. It would be better to use ‘पुरस्थापित’ in its place as it makes the sentence sound more formal.

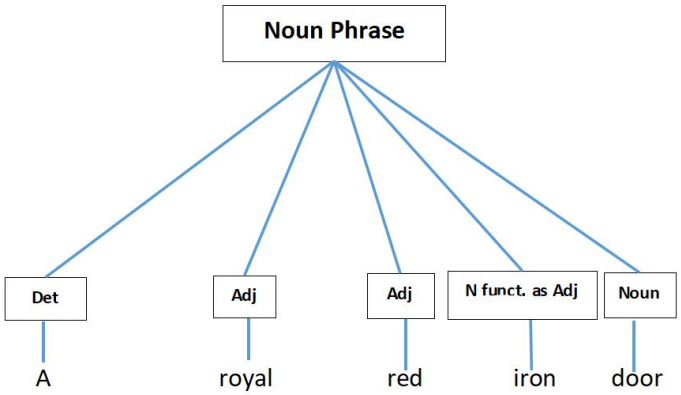
Some words have not been translated from English; instead, their transliterations have been used in Hindi. These include words from English like ‘quorum’ (कोरम) (SENTENCE 11) and ‘design’ (डिज़ाइन) (SENTENCE 37). ‘Quorum’ could be translated into Hindi as ‘गणपूर्ति’ and ‘design’ as ‘तैयार’.

From this analysis, we can conclude that Google Translate does not always consider the formality and propriety required in translations. When translating from English to Hindi, especially for official and government texts, it is crucial to maintain formality, appropriateness, propriety, and effectiveness. Using Sanskrit-derived vocabulary is considered more formal and suitable for such purposes.

Challenges in Translating Noun Phrase

The structure of noun phrases in Hindi is similar to that in English, but there are significant differences that pose challenges in translation. Understanding these differences is crucial for accurate and coherent translation, especially when using tools like Google Translate. This analysis explores the variations of noun phrase structures in both languages and highlights common translation errors, offering solutions to improve accuracy.

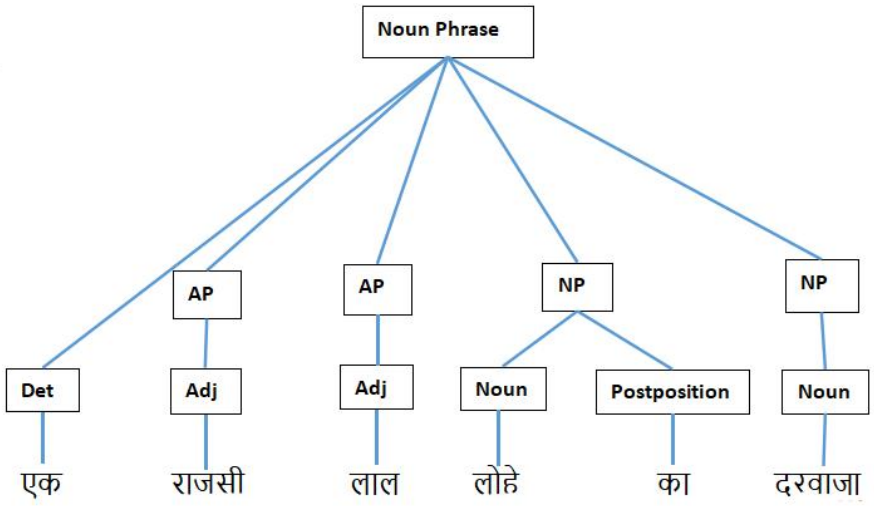
In English, a noun phrase typically consists of a noun head preceded by pre-nominal modifiers such as determiners, adjectives, quantifiers, and nouns functioning as adjectives. For instance, in the phrase "a red iron door," the noun head "door" is modified by the adjectives "red" and "iron," and the determiner "a." Native English speakers find it natural for a noun functioning as an adjective to be immediately followed by another noun. A representation of such structure of Noun phrase of English is given in the tree diagram below:



The above diagram of an English noun phrase structure shows that multiple adjectives can precede a noun without additional function words. In contrast, Hindi

requires careful structuring with appropriate postpositions to ensure clarity and grammatical correctness.

Hindi noun phrases follow a different structure. In Hindi, a pre-nominal modifier that is a noun functioning as an adjective cannot be directly followed by the noun head. Instead, a postposition is required. For example, the English phrase "a red iron shelf" cannot be directly translated to Hindi as "एक लाल लोहा दरवाजा." It must be



translated as "एक लाल लोहे का दरवाजा," using the genitive case marker 'का' to link the modifiers and the noun. A correct representation of noun phrase structure is given in the tree diagram below:

Hindi can also employ the inflections or Samas (समास) system to condense phrases. For instance, "लोहे का दरवाजा" can be simplified to "लौहद्वार," making the phrase more concise and maintaining grammatical correctness.

Google Translate often fails to adhere to these rules, resulting in errors. For example, translating "general school curriculum" as "सामान्य स्कूल पाठ्यक्रम" instead of the more appropriate "सामान्य स्कूली पाठ्यक्रम" or "सामान्य स्कूल का पाठ्यक्रम" shows a lack of understanding of Hindi phrase structure. Similarly, "grade 3 examination" should be "ग्रेड 3 की परीक्षा" rather than "ग्रेड 3 परीक्षा."

Google Translate often fails to adhere to these rules, resulting in errors. For example, translating "general school curriculum" as "सामान्य स्कूल पाठ्यक्रम" instead of the more appropriate "सामान्य स्कूली पाठ्यक्रम" or "सामान्य स्कूल का पाठ्यक्रम" shows a lack of understanding of Hindi phrase structure. Similarly, "grade 3 examination" should be "ग्रेड 3 की परीक्षा" rather than "ग्रेड 3 परीक्षा."

Common error occurs with phrases like "the school system." Google Translate might render this as "स्कूल शिक्षा प्रणाली," but a more accurate translation would be "स्कूली शिक्षा प्रणाली" or "स्कूल की शिक्षा प्रणाली." These examples illustrate how Google Translate often neglects the necessity of function words or postpositions in Hindi noun phrases, leading to awkward and ambiguous translations.

Google Translate's literal translation approach often overlooks the syntactic rules governing Hindi noun phrases, leading to errors and ambiguity. Accurate translation requires understanding the necessity of postpositions and function words in Hindi. By improving its handling of complex sentence structures and paying attention to linguistic nuances, Google Translate can provide more coherent and grammatically correct translations. For now, users should be aware of these limitations and manually adjust translations for accuracy and clarity.

Structural Ambiguity

Chomsky discusses ambiguity in terms of syntactic structures, defining it as “a phenomenon where a single sentence can be interpreted in multiple ways due to its syntactic structure” (p. 47).
Chomsky identifies structural ambiguity, resulting from different possible syntactic structures, and lexical ambiguity, where words have multiple meanings.

Structural ambiguity often occurs in translation, especially when using automated tools like Google Translate. This is due to differences in syntactic structures, grammatical rules, and contextual variations between languages.

Structural ambiguity may take place due to many reasons like word for word translation, unfamiliarity with the rules of either source language or target language and wrong use of punctuation.

The table given below lists such type of errors found in the data:

12	If for a period of sixty days a member of either House of Parliament is without permission of the House absent from all meetings thereof, the House may declare his seat vacant.	यदि संसद् के किसी सदन का कोई सदस्य साठ दिन की अवधि तक सदन की अनुज्ञा के बिना उसके सभी अधिवेशनों से अनुपस्थित रहता है तो सदन उसके स्थान को रिक्त घोषित कर सकेगा	यदि साठ दिनों की अवधि के लिए संसद के किसी भी सदन का कोई सदस्य सदन की अनुमति के बिना उसकी सभी बैठकों से अनुपस्थित रहता है, तो सदन उसकी सीट रिक्त घोषित कर सकता है।
29	The President may make such provision as he thinks fit for the discharge of the functions of the Governor of a State in any contingency not provided for in this Chapter.	राष्ट्रपति ऐसी किसी आकस्मिकता में, जो इस अध्याय में उपबंधित नहीं है, राज्य के राज्यपाल के कृत्यों के निर्वहन के लिए ऐसा उपबंध कर सकेगा जो वह ठीक समझता है।	राष्ट्रपति इस अध्याय में प्रदान नहीं की गई किसी भी आकस्मिक स्थिति में किसी राज्य के राज्यपाल के कार्यों के निर्वहन के लिए ऐसा प्रावधान कर सकता है जैसा वह उचित समझता है।

Google Translate has done a lot of development in its translation efficiency and accuracy in the recent years. However, presence of structural ambiguity is often observed when sentences are translated from English to Hindi by the use of Google Translate. From the data many sentences carrying such structural ambiguities were found.

In the sentence 12, the English to Hindi translation has error of structural ambiguity. The ambiguity occurs due to occurrence of time phrase ‘यदि साठ दिनों की

अवधि के लिए' before the phrase 'संसद के किसी भी सदन का कोई सदस्य'. Here 'साठ दिनों की अवधि के लिए' is a prepositional phrase and in Hindi prepositional(known as post positional) phrases occur after the Subject. The Hindi translation may not be considered completely wrong but still such structures are generally not used by native speakers. However, in English such structures are generally used and are considered correct.

In sentence 33, the meaning is not clear as what the translated text wants to say. It becomes hard for the native person to get the meaning of the translation. Propriety of sentence reduces due to the redundant use of 'और'. The redundancy might affect the clarity and smoothness of the sentence. The placement of parts of speech in the translated sentence is also less appropriate. The use of 'आजीवन सीखने के लिए' is less appropriate to use instead a more appropriate and formal 'जीवन-भर' could be used. The translation of 'is a necessary foundation and an indispensable prerequisite' is 'एक आवश्यक आधार और एक अनिवार्य शर्त है'. It would be more appropriate to use 'एक आवश्यक आधार और अनिवार्य शर्त है'. After conjunction 'और' there is no need of using a determiner 'एक ' as its use makes the determiner redundant.

In the sentence 35, 'शिक्षकों की सहायता के लिए' is a prepositional phrase and has been used before the Noun phrase 'तकनीकी हस्तक्षेपों को '. In a correct and more appropriate Hindi sentence structure, the noun phrase is used before any prepositional and postpositional phrases.

Literal Translation

Literal translation refers to the direct, word-for-word translation of text from one language to another, preserving the original order and structure as much as possible. This type of translation focuses on maintaining the exact wording and grammar of the source text rather than adapting it to the natural flow and idiomatic expressions of the target language.

Machine Translation tools often do the literal translation of the sentences. Many MT systems lack a sophisticated understanding of context. They translate words and phrases independently without considering the broader sentence or discourse context, leading to more literal translations.

The table given below lists some of the errors associated to Literal Translation:

28	Subject to the foregoing provisions of this article, a Governor shall hold office for a term of five years from the date on which he enters upon his office.	इस अनुच्छेद के पूर्वगामी उपबंधों के अधीन रहते हुए, राज्यपाल अपने पदग्रहण की तारीख से पांच वर्ष की अवधि तक पद धारण करेगा	इस अनुच्छेद के पूर्वगामी प्रावधानों के अधीन, एक राज्यपाल अपने कार्यालय में प्रवेश करने की तारीख से पांच वर्ष की अवधि के लिए पद धारण करेगा।
39	The mandated content will focus on key concepts, ideas, applications, and problem solving.	यह विषय-वस्तु अब मुख्य अवधारणाओं, विचारों, अनुप्रयोगों और समस्या-समाधान पर केंद्रित होगी।	अनिवार्य सामग्री प्रमुख अवधारणाओं, विचारों, अनुप्रयोगों और समस्या समाधान पर केंद्रित होगी।

Many sentences in the data reveal errors made by Google Translate when translating from English to Hindi.

For example, in Sentence 28, the Hindi translation of "a Governor shall hold office for a term of five years from the date on which he enters upon his office" results in "एक राज्यपाल अपने कार्यालय में प्रवेश करने की तारीख से पांच वर्ष की अवधि के लिए पद धारण करेगा।" It is evident from the translation of "date on which he enters upon his office" to "अपने कार्यालय में प्रवेश करने की तारीख" that this translation is literal. Official phrases often do not have a literal translation. A more accurate translation of this English phrase would be "अपने पदग्रहण की तारीख से।" The same error also appears in Sentence 31.

In Sentence 39, the Google translation of "the mandated content" as "अनिवार्य सामग्री" should instead be "यह विषय-वस्तु."

In Sentence 40, the English sentence has been completely literally translated to Hindi. The phrase "questions will be encouraged," referring to questions asked by students, is translated as "प्रश्नों को प्रोत्साहित किया जाएगा," which is ambiguous and unclear. A better and more appropriate translation would be "प्रश्न पूछने को प्रोत्साहित किया जाएगा।"

While machine translation tools like Google Translate are useful, they often produce literal translations that can result in inaccuracies, especially when dealing with idiomatic expressions and official phrases. They sometimes don't take into consideration the appropriateness, propriety and effectiveness during translation. Improving these tools to better understand context and idiomatic usage can lead to more accurate and meaningful translations. For critical texts, human oversight remains essential to ensure clarity and precision.

Lexical Anomalies

Lexical anomalies refer to irregularities or deviations in the use of words that can lead to confusion, misunderstanding, or errors in communication. These anomalies occur when words are used in ways that are unconventional, unexpected, or incorrect, often resulting in sentences that do not convey the intended meaning. Lexical anomalies can arise from various sources, including errors in word choice, ambiguous language, improper usage, and cultural differences.

Google translate committed errors associated with wrong use of lexis. In sentence 4, the translation of “no language can survive without” has been translated to “से जुड़े बिना जीवित नहीं रह सकती।”. But the translated word for “survive” is “जीवित”, hence in this case it is wrong. “जीवित” is used for living things so its usage

here referring to survival of language is wrong. A more appropriate and better word would be "पनप". Hence, the correct sentence would be "वर्तमान युग में कोई भी भाषा सूचना एवं संचार प्रौद्योगिकी से जुड़े बिना पनप नहीं रह सकती।"

In sentence 7, the google translation of "common to both Houses of Parliament" has turned out to be "संसद के दोनों सदनों के लिए सामान्य पदों". Here, in this context the term 'common' meant 'posts which are available in both houses of the parliament'. But google translate took meaning of "सामान्य" as 'general'. This error was caused by Google Translate's inefficiency of fully understanding the context.

In sentence number 10, the term 'shall not vote in the first instance' has been translated to 'पहली बार में मतदान नहीं करेगा'. But this translation is not correct in light of Formal correspondence, . Its correct translation would have been "प्रथमतः मत नहीं देगा". Here the correct translation of first instance would be 'प्रथमतः'.
Formality and propriety are not very well preserved at lexical level in translation by Google Translate.

In sentence 15, term 'lapse' has been translated as 'समाप्त' by the google translate. It would be more appropriate to use 'व्यपगत' because it refers to something that has ended naturally or has ceased to be effective after a certain period. It is typically used in contexts like expiry dates, tenures, validity periods, or when referring to something that has come to an end due to the passage of time. 'समाप्त' is more general and can be used to describe the completion of events. Both terms are formal, but "व्यपगत" has a slightly more formal and legalistic tone compared to "समाप्त" .

Discussion

Google Translate has significantly improved in the last few years. In many sentences, the formality and propriety were preserved. In some cases, Google Translate delivered translations that were even better than those done by government translators. However, despite this development, Google Translate sometimes commits

errors related to lexis, sentence structure, and collocation. Such errors make the translations less formal and appropriate, making them unfit for official use.

The key findings of this research include issues with code-mixing, noun phrase structures, structural ambiguity, literal translation, and lexical anomalies.

One significant issue identified is Google Translate's preference for code-mixing, incorporating words from Persian, Arabic, and English into Hindi translations. While this approach may enhance readability for the general public, it detracts from the formality and propriety expected in official documents. The use of such mixed linguistic registers makes the translated text sound less formal and thus diminishes its effectiveness for formal correspondence. The translation of noun phrases presents another area of difficulty. The structural differences between English and Hindi noun phrases require careful handling to maintain grammatical correctness and coherence. The direct translation of pre-nominal modifiers from English to Hindi often leads to awkward or incorrect phrases. This points to a broader issue of syntactic mismatches between languages that automated translation tools must address to improve accuracy. Structural ambiguity is another frequent problem which takes place because of difference between word order of English and Hindi. The translation of complex sentences often results in ambiguous and ill formed sentences. Such errors and misalignment generally happen due to Google Translate’s lack of context understanding in complex sentences. Literal translation remains a significant challenge, as it often fails to convey the intended meaning, especially with idiomatic or formal expressions. Literal translation can miss the idiomatic usage of language, highlighting the need for translation tools to go beyond word-for-word translation and consider idiomatic and contextual appropriateness. Lexical anomalies, or the improper use of words, are another common issue. This error illustrates how Google Translate struggles with selecting the right lexical items based on context. Ensuring that translations maintain the correct level of formality and specificity is crucial for producing accurate and meaningful translations in official texts.

These findings highlight the limitations of current machine translation systems like Google Translate, particularly in handling the nuanced demands of formal and official texts. The prevalence of errors in formality, syntax, and lexical choices

suggests that relying solely on automated translations for official documents can lead to miscommunications and a lack of professionalism.

Google Translate is a free tool that is used by millions of people on a daily basis. Often times it delivers correct translation but still there is scope for improvement. Improving machine translation tools not only enhances communication but also ensures that official documents maintain their intended formality and propriety, ultimately supporting more effective governance and international relations.

Conclusion

The aim of this research is to evaluate the propriety and effectiveness of Google Translate's English to Hindi translations in official texts with a view to analyzing the errors found in translation of official texts . The study addresses three key questions: the most common types of errors, the extent to which formality and cultural propriety are maintained, and the contextual appropriateness of the translations for official use. The study yields the following findings:

Google Translate’s preference for code-mixing : While translating the official texts from English to Hindi, Google translate incorporates the use of lexis from various languages like English, Persian and Arabic. Use of words from such languages makes the translation easy to understand for general public but it makes the translation less formal and appropriate for official use. Such error was prevalent in the translation done by google translate.The use of such mixed linguistic registers makes the translated text sound less formal and thus diminishes its effectiveness for formal correspondence.

Challenges in translating the noun phrase from English to Hindi : The structural differences between English and Hindi noun phrases require careful handling to maintain grammatical correctness and coherence. The direct translation of pre-nominal modifiers from English to Hindi often leads to ill-formed or inappropriate phrases.

Structural Ambiguity: The translation of complex sentences often results in ambiguous or ill-formed constructions. Google Translate was found ignoring the rule where postpositional phrases follow the subject in Hindi syntax. While the translation may not be entirely incorrect but it does not align with the conventions of a native speaker. Thus, highlighting the need for more context-aware translation algorithms that can handle syntactic differences effectively.

Literal Translation: Google Translate's reliance on literal translation often fails to convey the intended meaning, especially with idiomatic or formal expressions. This approach leads to inaccurate and contextually inappropriate translations, emphasizing the need for translation tools to go beyond word-for-word translation to ensure accuracy and effectiveness.

Lexical Anomalies: Lexical anomalies are prevalent in Google Translate's output, with improper usage of words leading to confusion and errors. These anomalies highlight the tool's struggle to select contextually appropriate terms, emphasizing the need for improved algorithms to ensure precise and meaningful translations in official texts.

Google Translate provides a valuable service, but its translations often fall short of the accuracy and formality required for official documents. Although the study was conducted on a sample of 100 sentences, the errors identified are likely to extend to various other domains.

Google Translate has made significant progress in recent years, especially with the introduction of the Transformer architecture. However, further development is needed to enhance the tool's efficiency and accuracy, and Google Translate is actively working towards these improvements.

References

Aiken, M., & Balan, S. (2011). *An Analysis of Google Translate Accuracy*. In Translation Journal 2000, Volume 16(2).

Aiken, M., Ghosh, K., Wee, J., & Vanjani, M. (2009). *An evaluation of the accuracy of online translation systems*. Communications of the IIMA, 9(4), 67-84.

Catford, J. C. (1965). *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. London: Oxford University Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton.

Even-Zohar, I. (1978). *The position of translated literature within the literary polysystem*. Poetics Today, 2(4), 21-27.

Karami, O. (2014). *The brief view on Google Translate machine.* Artificial Intelligence on Natural Language.

Levin, M., (2020). *Challenges in Legal Machine Translation*. Journal of Legal Technology.

Malik, P., & Baghel, A. S. (2021). *Performance Enhancement of Machine Translation Evaluation Systems for English – Hindi Language Pair.* International Journal of Modern Education and Computer Science (IJMECS), 13(4), 1-12.

Martínez-Gómez, P., & Sánchez-Martínez, F. (2017). *Formality and Register in Machine Translation*. Translation Studies Journal.

Munday, J. (2008). *Introducing Translation Studies*. New York: Routledge.

Muzaffar, S., & Behera, P. (2021). *A Qualitative Evaluation of Google's Translate: A Comparative Analysis of English-Urdu Phrase-based Statistical Machine Translation (PBSMT) and Neural Machine Translation (NMT) Systems*. Journal of Translation Studies, 29(2), 123-140.

Newmark, P. (1988). *A Textbook of Translation*. Prentice Hall, New York, USA.

Nida, E. A. (1964). *Toward a Science of Translating: With Special Reference to P rinciples and Procedures Involved in Bible Translating*. E.J. Brill.

Nida, E. A. (1991). *Theories of Translation. TTR: Traduction, Terminologie, Rédaction*, 4(1), 19-32.

Pym, A. (2013). *Translation Solutions for Many Languages.* Translation and Interpreting Studies.

Turovsky, B. (2016). *The Challenges of Context in Machine Translation*. Google AI Blog.

Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.

Vermeer, H. J. (1989). *Skopos and commission in translational action.* Readings in Translation Theory (pp. 173-187). Oy Finn Lectura Ab.

Sources for the Data

Excerpt from the Legislative Department:
https://legislative.gov.in/constitution-of-india/

Excerpt from the NEP document of Ministry of Education:
https://www.education.gov.in/nep/

Excerpt from the Rajbhasha Report of Department of Official Langauge:

https://rajbhasha.gov.in/hi/official-journals-articles

Excerpt from the The Gazette of India:

https://www.ugc.gov.in/