

# AI Development for Translating Indian Historical Devanagari Scripts into the Pali Language : A Technical Approach

1<sup>st</sup> Sangita R. Gudadhe  
dept. of Computer Science & Engineering  
Sipna College of Engineering &  
Technology, Amravati  
line 4: City, India

2<sup>nd</sup> Aashish A. Bardekar  
dept. of Computer Science & Engineering  
Sipna College of Engineering &  
Technology, Amravati  
Amravati, India

3<sup>rd</sup> Amitkumar B. Ranit  
dept. of Civil Engineering  
Prof. Ram Meghe College of Engineering  
& Management, Badanera  
line 4: City, Country

**Abstract**—Identification and categorization of ancient artifacts written in languages such as Pali are critical in the field of archaeology study. A vast amount of literature and knowledge can be found in Pali, an ancient Prakrit language. To comprehend and acquire the Pali language, you can depend on a Pali-to-English dictionary. For any script to Pali script, there isn't a dictionary or translation tool available, though. Translating the Pali language is done using the Devanagari script. An automatic method for translating text written in Devanagari script into Pali is suggested by this research study. To help people understand the contents written in the Devanagari script, the system has been developed.

**Keywords**—Devanagari script, mapping, text, Pali Language.

## I. INTRODUCTION (HEADING I)

Text translation between languages is made possible by machine translation technology. Rule-based machine translation creates translations that follow accepted language norms and structures by using linguistic dictionaries and rules. The way words and phrases from the source language should be translated into the target language is outlined in these rules. Numerous machine learning systems are available for multiple languages, including English, Sanskrit, and Bangla. The Devanagari script is very old and has a rich cultural legacy[1]. Most Indian languages are written using this script, either entirely or in conjunction with other scripts. More than 120 languages and dialects, including Sanskrit, Prakrit, Pali, Hindi, Marathi, Nepali, Konkani, Bodo, Sindhi, and Maithili, use this writing system, making it one of the most popular and widely accepted writing systems in the world. For improved comprehension, it may be useful to translate the Devanagari script into another language[2]. A translator is an instrument for translating between different languages. It is imperative to acknowledge that Devanagari is merely a script and not a language[3]. Consequently, it is essential to translate the Devanagari script into a language before beginning the translation process. Traditionally, the Devanagari script was read aloud in Pali. Translating the Devanagari script into Pali helps comprehend its meaning in the contemporary world[4]. Grammar is not a significant consideration in the translation process because Pali is also written in the Devanagari script. Using a one-to-one mapping, every character in the Devanagari script is independently translated into the Pali language to create the translation system. Information on previous studies conducted in this field is provided in Section II. Details regarding the Devanagari script are given in Section

III. In Section IV, details regarding the Pali language of the Devanagari script are covered. The methodology for the translation process, including the output steps, is presented in Section V. Section VII provides a closing statement, while Section VI displays the system results.

## II. LITERATURE REVIEW

Early research on artificial intelligence (AI) identified machine translation as a subfield of natural language processing[5]. Machine learning and deep neural network techniques have been widely applied to numerous classification-based problems and automated character and digit recognition in various languages [6]. Decoding the meaning of the source text and re-encoding it into the target language is the simplest way to sum up machine translation. However, the creation of machine translation systems has gained attention as a research topic because of the intricacy of natural language. Depending on the translation technique, machine translation systems can be broadly divided into two categories: direct translation systems and indirect translation systems. Word-to-word or phrase-to-phrase mappings are used by direct translation systems to translate from one language to another. An interlingua or transfer approach is used by indirect translation systems. Apart from the aforementioned common approaches, machine translation systems can be categorized into seven groups. rule-based, statistical, example-based, knowledge-based, hybrid, agent-based, and human-assisted [7]. The most fundamental method of machine translation is dictionary-based translation. This kind of machine translation software is simple to set up and makes use of powerful tools to provide translations for ambiguous search terms. Dictionary-based translation typically consists of translating text using a bilingual dictionary. Most systems include source and target language generators, a morphological analyzer, and bilingual dictionaries. Sindhu et al. [8] developed a dictionary-based machine translation from Kannada to Telugu, for example. Very little research has been done on translating Pali text into other languages, given the nature of the Pali language. A prototype rule-based machine translation system from Phoson and colleagues has been developed for Pali to Thai. With this system, a sentence's Pali language structure can be analyzed to produce a Thai language structure. The Pali character recognition system was created in 2012 by SR Suralkar, S. P. Ramteke, and Kiran S. Mantri[9]. Translations from script to script and from script to language have been the subject of

numerous studies. Because Pali is written in Devanagari script, which is grammatically independent, translating it is a special process. Therefore, a one-to-one mapping approach needs to be used to create a translation system for Pali. With this method, every character in the Devanagari script is translated one at a time into Pali. Therefore, mapping is the most effective way to translate the Devanagari script into Pali language.

III. DEVANAGARI SCRIPT

This script has not been combined with any other scripts or symbols; it is in its original form. The script slightly alters when combined with other symbols and scripts. In Devanagari, the vowels appear first, with the long form of each short vowel coming after it. Consonants are arranged both by location and, within a location, by articulation style. Velar, palatal, retroflex, dental, and labial are the rows of consonants. They go nasal, unaspirated and aspirated voiced, and unaspirated and aspirated voiceless within each row.



Fig. 1. Examples of Devanagari manuscripts

An Indo-Aryan language called Prakrit (Middle Indic) was represented by Devanagari[10]. It was distributed widely but erratically throughout Central Asia, northwest India, eastern Afghanistan, and northern Pakistan. One of the oldest languages in the world, Samskrita, has been represented for several centuries by the ancient Indian script known as Devanagari[11]. This script is a member of the ancient and well-known Brahmi script family. Numerous other languages, including Prakrit, Hindi, Marathi, Nepali, and more, are also written in Devanagari.

Brahmi	𑀧	𑀘	𑀡	𑀢	𑀣	𑀤	𑀥	𑀦	𑀧	𑀨	𑀩	𑀪	𑀫	𑀬	𑀭	𑀮	𑀯	𑀰	𑀱	𑀲	𑀳	𑀴	𑀵	𑀶	𑀷	𑀸	𑀹	𑀺	𑀻	𑀼	𑀽	𑀾	𑀿
Gupta	𑀧	𑀘	𑀡	𑀢	𑀣	𑀤	𑀥	𑀦	𑀧	𑀨	𑀩	𑀪	𑀫	𑀬	𑀭	𑀮	𑀯	𑀰	𑀱	𑀲	𑀳	𑀴	𑀵	𑀶	𑀷	𑀸	𑀹	𑀺	𑀻	𑀼	𑀽	𑀾	𑀿
Devanagari	क	ख	ग	घ	ङ	च	छ	ज	झ	ञ	ट	ठ	ड	ढ	ण	त	थ	द	ध	न	प	फ	ब	भ	म	य	र	ल	व	श	ष	स	ह

Fig. 2. Evolution from Brahmi script to Devanagari script

It is extremely scientific in its application and closely adheres to the Lipi Shastra. Characters written in the Devanagari script are suspended from a horizontal line at the top of the character, known as the headstroke[12]. English letters are written up from a line below them, which is not the case here. Writing in Devanagari is done from left to right using a combination of syllabary and alphabet.

IV. PALI LANGUAGE

Pali is a highly literary and ancient Prakrit language with a vast body of knowledge. The ancient languages in India number in the thousands. Pali is an ancient language that possesses a sober and sweet nature among others[13]. since the language of Pali is Prakrit. The Pali language is also

referred to as the liturgical and classical language of Theravada Buddhism[14]. It has been used to write Pali since approximately 1000 BC, and its rich history aids in the understanding of the Buddha's ideas, culture, and teachings by archaeologists. With 32 consonants, 6 vowels, 2 diphthongs, and one accessory nasal sound known as Niggah'ta, Pali has 41 alphabets in total. The first Pali-English dictionary and the first Pali-translated English text were released in 1869 and 1872, respectively.

V. REPRESENTATION OF THE SCRIPT COMPONENTS

Vowels and consonants of the Devanagari script, along with their pronunciation in the Pali language, are listed in Table 1. Compound characters are used for Pali language consonants and vowels respectively (Table 1).

A. Vowels (Swar)

In the Devanagari script, vowels can be used alone or in conjunction with consonants. They can be found at the start of a word or attached to a consonant, and they are necessary for pronouncing words in Pali. In Pali, these vowels can be combined with consonants to create particular sounds and syllables, which help to form a variety of words and phrases.

B. Consonants (Vyanjan)

In the Devanagari script, consonants form the foundation of words and are joined with vowels to form syllables. Vowels can be added to consonants to change their inherent vowel sound, which is usually 'a' (अ). According to the Devanagari script's rules, these consonants are combined with vowels to form syllables and words in Pali.

C. Compound Characters

Combining a consonant with a vowel or another consonant creates compound characters in Pali. Conjuncts or "ligatures" are compound characters that are used to represent sounds that occur consecutively without a break. The production of more complex sounds and syllables is made possible by these compound characters, and this is crucial for understanding and pronouncing Pali words correctly.

TABLE 1. COMPOUND CHARACTERS OF PALI LANGUAGE FROM DEVANAGARI SCRIPT

Consonants		Vowels		Compound Characters
Pali Language	Devanagari Script	Pali Language	Devanagari Script	Pali Language + Devanagari Script
pa	प	a	अ	पाप
pa	प	ā	आ	पापा
pa	प	i	इ	पिपि
pa	प	ī	ई	पीपी
pa	प	u	उ	पुपु
pa	प	ū	ऊ	पूपू

VI. PROPOSED METHODOLOGY

The goal of the Devanagari script to Pali language translation system is to effectively translate Devanagari text into its equivalent Pali representation. The three primary components of this system architecture are input, mapping, and output. As shown in Figure 3, every stage is essential to guaranteeing the precision and consistency of the translation procedure.

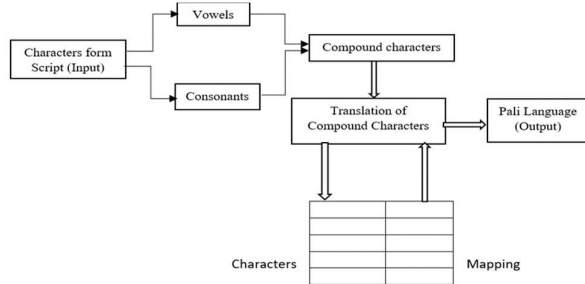


Fig. 3. Translation model for Devanagari script to Pali language

A. Input

When text written in the Devanagari script is fed into the system, the translation process starts with the input stage. In addition to being widely used for writing several languages, such as Hindi, Marathi, and Sanskrit, this script has historically also been used to write the Pali language.

1) Text Orientation:

Reading each line from top to bottom, the system analyzes the text line by line. Characters are processed sequentially from left to right within each line. This preserves the text's structure throughout the translation because it corresponds with the Devanagari script's natural reading order.

2) Character Recognition:

The Devanagari script allows for the individual recognition of each character. For the next mapping step to go smoothly, the system must correctly identify every character. Because any mistakes made here could spread throughout the system and result in inaccurate translations, the input phase is extremely important. Illustration of how the translation system reads and interprets the input text is shown in Figure 3.

B. Mapping for Translation

The main stage of the translation process is called mapping, during which every character in the Devanagari script is methodically swapped out for its equivalent in the Pali language.

1) Mapping:

Establishing a direct correspondence between two sets of characters is the process of mapping. The first set of characters in this system are those written in Devanagari script, and the second set is composed of characters written in Pali. Each Devanagari character is replaced with its corresponding Pali character as part of the mapping process.

2) Character Correspondence:

Table 2 provides a thorough mapping of Pali characters to Devanagari script characters. The system uses this table as a guide to make sure that every character is translated correctly. To preserve the meaning and phonetics of the original text, the

mapping takes into account the linguistic subtleties of Pali rather than just replacing words for words.

TABLE II. TRANSLATION OF PALI LANGUAGE FROM DEVANAGARI SCRIPT

Devanagari script	Pali language
प	P
पा	Pā
पि	Pi
पी	Pī
पु	Pu
पू	Pū
पे	Pe
पो	Po

3) Complexity of Mapping:

The mapping procedure takes into consideration Devanagari's common compound characters and ligatures. These are broken down and correspondingly mapped to their Pali counterparts. For example, when translating Devanagari to Pali, consonant clusters may need to be divided or treated differently.

C. Output

The translated text is created and displayed during the output phase.

1) Text Reconstruction:

Following the mapping phase, the translated Pali characters are arranged in the same order as the original Devanagari script by the system to reconstruct the text. This guarantees that the formatting and original structure of the source text are preserved in the translated version. Last-

2) Minute Presentation:

Using the matching characters from the Devanagari input, a text in the Pali language is the final product. Pali is a language with a rich phonetic and historical structure, so phonetic transcriptions into English are used to present the output. In doing so, it becomes more widely accessible while maintaining the original text's meaning and pronunciation. As an Example Illustration: Figure 4 shows an example of how the characters are translated and rearranged to help you better understand this process. This diagram shows how a Devanagari text's characters are methodically swapped out and rebuilt to create a Pali counterpart.

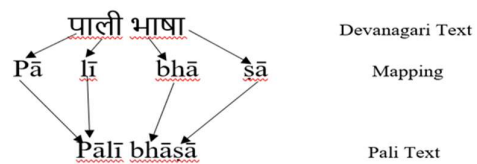


Fig. 4. Example of translated characters

VII. RESULT OF EXPERIMENT

The method for mapping Devanagari script characters into Pali language characters that correspond to them is very effective. This process ensures that nearly every Devanagari character is accurately identified and converted into its Pali

equivalent, with an astounding accuracy rate of 99.99%. With an accuracy rate this high, the system is extremely reliable for real-world applications, it errors out of 10,000 characters only once. Also, this process is likely to use sophisticated algorithms and language models capable of handling the nuances of both the Devanagari script and the Pali language. This includes accurately processing similar-looking characters, comprehending contextual differences, and addressing any script-specific complexities that could otherwise complicate translation. In conclusion, the system's 99.99% accuracy rate reflects its high performance and dependability in translating the Devanagari script to Pali, making it a valuable tool for linguists, scholars, and anyone working with these languages.

#### VIII. CONCLUSION

The translation system presented in this research is incredibly accurate and is meant to translate text from the Devanagari script into Pali. When reading and comprehending Devanagari-written Pali texts, the system is especially helpful for individuals or researchers. Using an accurate mapping process, the system effectively converts text written in the Devanagari script into text written in Pali language. About every character is translated with remarkable accuracy—99.99% of the translations are successful. This helps to preserve and study Pali texts with minimal errors, making the system an invaluable resource for academics, linguists, and anyone working with these languages.

#### REFERENCES

- [1] S. R. Narang, M. K. Jindal, S. Ahuja, and M. Kumar, "On the recognition of Devanagari ancient handwritten characters using SIFT and Gabor features," *Soft Comput.*, vol. 24, no. 22, pp. 17279–17289, Nov. 2020, doi: 10.1007/s00500-020-05018-z.
- [2] R. Ahmad, "Scripting a new identity: The battle for Devanagari in nineteenth century India," *Journal of Pragmatics*, vol. 40, no. 7, pp. 1163–1183, Jul. 2008, doi: 10.1016/j.pragma.2007.06.005.
- [3] J. Vaid and A. Gupta, "Exploring Word Recognition in a Semi-Alphabetic Script: The Case of Devanagari," *Brain and Language*, vol. 81, no. 1–3, pp. 679–690, Apr. 2002, doi: 10.1006/brln.2001.2556.
- [4] N. Gautam, S. S. Chai, and M. Gautam, "Translation into Pali Language from Brahmi Script," in *Micro-Electronics and Telecommunication Engineering*, vol. 106, D. K. Sharma, V. E. Balas, L. H. Son, R. Sharma, and K. Cengiz, Eds., in Lecture Notes in Networks and Systems, vol. 106., Singapore: Springer Singapore, 2020, pp. 117–124. doi: 10.1007/978-981-15-2329-8\_12.
- [5] Y. A. Mohamed, A. Khanan, M. Bashir, A. H. H. M. Mohamed, M. A. E. Adiel, and M. A. Elsadig, "The Impact of Artificial Intelligence on Language Translation: A Review," *IEEE Access*, vol. 12, pp. 25553–25579, Aug. 2024, doi: 10.1109/ACCESS.2024.3366802.
- [6] N. Alrobah and S. Albahli, "Arabic Handwritten Recognition Using Deep Learning: A Survey," *Arab J Sci Eng*, vol. 47, no. 8, pp. 9943–9963, Aug. 2022, doi: 10.1007/s13369-021-06363-3.
- [7] B. Hettige, A. Karunananda, and G. Rzevski, "Thinking Like Humans: A New Approach to Machine Translation," in *Artificial Intelligence*, vol. 890, J. Hemanth, T. Silva, and A. Karunananda, Eds., in Communications in Computer and Information Science, vol. 890., Singapore: Springer Singapore, 2019, pp. 256–268. doi: 10.1007/978-981-13-9129-3\_18.
- [8] D. V. Sindhu and B. M. Sagar, "Dictionary Based Machine Translation from Kannada to Telugu," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 225, p. 012182, Aug. 2017, doi: 10.1088/1757-899X/225/1/012182.
- [9] N. Gautam, R. S. Sharma, and H. Garima, "Ariyaka: A PALI Alphabet Recognition Script," in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, India: IEEE, Dec. 2015, pp. 293–295. doi: 10.1109/CICN.2015.65.
- [10] O. V. Hinüber, R. Salomon, and O. V. Hinuber, "Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and Other Indo-Aryan Languages," *Journal of the American Oriental Society*, vol. 121, no. 3, p. 517, Jul. 2001, doi: 10.2307/606699.
- [11] B. Kataria and H. B. Jethva, "Review of Advances in Digital Recognition of Indian Language Manuscripts," *IJSRSET*, pp. 1302–1318, Feb. 2018, doi: 10.32628/IJSRSET1841215.
- [12] K. Mehrotra, S. Jetley, A. Deshmukh, and S. Belhe, "Unconstrained handwritten Devanagari character recognition using convolutional neural networks," in *Proceedings of the 4th International Workshop on Multilingual OCR*, Washington D.C. USA: ACM, Aug. 2013, pp. 1–5. doi: 10.1145/2505377.2505386.
- [13] S. Collins, "Madness and Possession in P'li Texts," *BSRV*, vol. 31, no. 2, pp. 195–214, Jan. 2015, doi: 10.1558/bsrv.v31i2.195.
- [14] B. Levman, "The language of early Buddhism," *Journal of South Asian Languages and Linguistics*, vol. 3, no. 1, pp. 1–41, Mar. 2016, doi: 10.1515/jsall-2016-0001.