Panamerican Mathematical Journal
ISSN: 1064-9735
Vol 35 No. 4s (2025)

An Algorithmic Approach to English-to-Sanskrit Translation

[1]Dr. Santosh Deshpande, [2]Mrs. Neha Kulkarni

[1]Director, MES Institute of Management & Career Courses (IMCC)

131, Mayur Colony, Kothrud, Pune 411038

sdd.imcc@mespune.in

[2]Assistant Professor, MES Senior College

131, Mayur Colony, Kothrud, Pune 411038

nnk.bba@mespune.in

Article History:

Received: *02-09-2024*

Revised*: 19-10-2024*

Accepted: *27-10-2024*

Abstract:

This paper presents an algorithm for translating English statements into Sanskrit. The proposed method addresses the unique challenges posed by Sanskrit's structural, grammatical, and morphological complexities. The algorithm ensures accurate and coherent translations by combining rule-based syntactic reordering, lexical mapping, and morphological adaptation with modern NLP techniques. Preliminary results indicate its effectiveness in generating grammatically correct Sanskrit sentences from English inputs

Keywords: Machine Translation, Natural Language Processing, Sanskrit, Morphology, Lexical

1. Introduction [1][2][4][5]

Machine Translation (MT) between languages with distinct linguistic properties presents a unique set of challenges, particularly when dealing with structurally and morphologically divergent languages such as **English** and **Sanskrit**. English is a widely spoken language with a **Subject-Verb-Object (SVO)** sentence structure, characterized by its relatively simple grammar and limited inflectional morphology. In contrast, Sanskrit, a classical language revered for its linguistic precision, follows a **Subject-Object-Verb (SOV)** order and features **rich morphological inflections**, intricate **grammar rules**, and **Sandhi transformations** that involve phonological changes at word boundaries. These differences make the development of effective translation algorithms for English-to-Sanskrit translation a challenging task.

Sanskrit's grammar is governed by the rules of **Panini's Ashtadhyayi**, a comprehensive linguistic framework that defines its syntax, morphology, and phonology. The language's **rich case system** (with eight cases), **three genders**, **three numbers** (singular, dual, plural), and **tense-conjugation systems** make it highly expressive but difficult to model computationally. Additionally, Sanskrit's **Sandhi rules**—which alter word forms based on phonetic

compatibility—add a layer of complexity that must be addressed in translation algorithms.

On the other hand, English, as a modern language, employs a comparatively straightforward grammar and depends heavily on **word order** to convey meaning. The disparity in linguistic features such as **syntax**, **word morphology**, and **semantic structures** necessitates a specialized and multi-faceted approach for translating English to Sanskrit accurately.

Challenges in English-to-Sanskrit Machine Translation:

1. **Syntax Differences:**

- English employs an SVO word order (e.g., "The boy eats food"), whereas Sanskrit uses SOV order (e.g., बालकः भोजनं खादति "bālakaḥ bhojanaṁ khādati").
- The reordering of words requires sophisticated syntactic analysis and dependency parsing.

2. **Rich Morphology in Sanskrit:**

- Words in Sanskrit are highly inflected to reflect **case**, **number**, **gender**, and **tense**, unlike in English where such grammatical markers are often absent or minimal.

3. **Phonological Adjustments (Sandhi):**

- Sandhi transformations require phonetic changes at word boundaries to ensure euphonic harmony. For instance, the word combination *"रामः अद्य"* (rāmaḥ adya) transforms into *"रामोऽद्य"* (rāmo'dya).

4. **Ambiguity in Word Translation:**

- Sanskrit's vocabulary is extensive, and the same English word may have multiple equivalents depending on context. For instance, "play" can translate to क्रीडति (*krīḍati*) for physical play or गायति (*gāyati*) for playing music.

5. **Data Scarcity:**

- Sanskrit is a low-resource language, and the lack of extensive parallel corpora or high-quality bilingual datasets makes training effective translation models a challenge.

This paper introduces an algorithmic framework that combines rule-based methods with Natural Language Processing (NLP) techniques to address these challenges. The algorithm focuses on the following key areas:

- **Syntactic Reordering:** Converts English's SVO order into Sanskrit's SOV structure using dependency parsing.
- **Morphological Adaptation:** Inflects words in Sanskrit based on grammar rules for gender, number, case, and tense.
- **Lexical Mapping:** Uses bilingual dictionaries to translate English words into their Sanskrit equivalents while considering contextual nuances.
- **Phonological Adjustment (Sandhi):** Applies rules to handle euphonic changes at word boundaries, ensuring smooth and grammatically correct transitions.

2. Related Research Work [2][3][6][9][11]

The domain of **Machine Translation (MT)** has seen significant advancements with the advent of rule-based methods, statistical approaches, and neural models. However, translation between structurally and morphologically distinct languages, such as English and Sanskrit, remains a niche area due to the inherent challenges posed by Sanskrit's **rich linguistic properties** and the scarcity of parallel corpora. Below, we discuss the most relevant related research efforts in **English-Sanskrit MT** and general low-resource language translation systems.

2.1 Rule-Based Approaches

Rule-based approaches have traditionally dominated the **Sanskrit translation domain**, focusing on **syntactic transformations** and **grammatical rules**.

1. **Rao and Gopinath (2023)**:

o This study explored the **limitations of zero-shot NMT systems**, such as Google Translate, for Sanskrit-English translation. It emphasized the importance of incorporating **Sanskrit-specific grammatical frameworks** into existing MT models. Their research showed that ignoring Sanskrit's structural and phonological nuances leads to suboptimal translations.

2. **Sanskrit Computational Linguistics (Huet et al., 2005)**:

o This research focused on developing a **formal grammar** for Sanskrit based on **Panini's Ashtadhyayi**. It introduced computational methods for handling **Sandhi transformations** and **morphological parsing**, which are critical for MT systems involving Sanskrit.

3. **Sanskrit-Hindi Rule-Based System (2008)**:

o A project aimed at creating a **rule-based MT system** for translating Sanskrit to Hindi. It demonstrated the effectiveness of leveraging **shared grammatical features** in closely related languages but highlighted challenges in adapting the model for distant languages like English.

2.2 Statistical Machine Translation (SMT)

Statistical approaches rely on parallel corpora to generate probabilistic translations but are limited for Sanskrit due to **data scarcity**.

1. **Koehn's SMT Framework (2020)**:

o While not specific to Sanskrit, Koehn's work in **Statistical Machine Translation** offers foundational techniques, such as **phrase-based translations**, which can be adapted for low-resource languages like Sanskrit by combining them with rule-based preprocessing.

2. **Sanskrit-English SMT Experiments (2012)**:

o This research tested statistical models for Sanskrit-English translation using **ancient texts** as a corpus. Results indicated that SMT performs poorly for Sanskrit due to its **rich inflectional morphology** and **small dataset size**.

2.3 Neural Machine Translation (NMT)

NMT, particularly models based on **transformers**, has shown promise in overcoming the limitations of rule-based and statistical methods for many languages. However, Sanskrit's **low-resource nature** poses challenges.

1. **Google Translate Zero-Shot NMT for Sanskrit (2021)**:

o This study evaluated Google Translate's performance for Sanskrit-English translations using **zero-shot learning**. It identified issues with **syntactic correctness** and **semantic fidelity**, demonstrating the need for **Sanskrit-specific fine-tuning**.

2. **IndicNLP Library and Sanskrit NMT (2020)**:

o The **IndicNLP Library** provides tools for processing Indian languages, including Sanskrit. Efforts to integrate NMT models with IndicNLP for Sanskrit translations demonstrated moderate success in handling simple sentences but struggled with complex syntactic and morphological structures.

3. **Attention-Based Models (Vaswani et al., 2017)**:

o Vaswani's transformer models, such as **BERT** and **GPT**, have revolutionized NMT. Adaptations for Sanskrit involve leveraging **transfer learning** and **pre-trained multilingual embeddings** to address data scarcity.

2.4 Hybrid Approaches

Combining rule-based and neural methods has shown potential in Sanskrit MT.

1. **Hybrid Models for Morphological Languages (2019)**:

o Research into hybrid models combining **rule-based morphology analyzers** with NMT frameworks highlights the advantages of this approach for languages with complex inflectional systems, such as Sanskrit.

2. **Morphological Segmentation in Sanskrit MT (2018)**:

o This study integrated **morphological segmentation** into NMT models for Sanskrit, improving the handling of case, gender, and tense inflections.

2.5 Other Notable Efforts

1. **Digital Sanskrit Corpus (2015)**:

o The development of **digital corpora** for Sanskrit texts, such as Vedic literature and epics, has contributed to the creation of small-scale parallel datasets for MT research.

2. **Sanskrit Parsing Tools (2017)**:

o Tools like **Sanskrit Heritage Site** provide resources for **syntactic parsing**, which are crucial for integrating Sanskrit-specific grammar into MT systems.

MT System	Features	Limitations
ETSTS (English to Sanskrit Translator and Synthesizer)	Context-aware translation; supports voice synthesis; employs advanced grammar processing.	Lacks extensive domain-specific vocabulary; struggles with idioms and poetry.
ESSS (English-Sanskrit Syntactic System)	Rule-based system using syntactic structure; effective for grammatical accuracy.	Limited to structured sentences; ineffective for free-form text.
E-tranS (Electronic Sanskrit Translator)	AI-driven translation; supports modern English-Sanskrit conversion; handles basic grammatical constructs.	Fails to address complex grammar; requires extensive training data.
Sanskrit to English Translator by Subramaniam A.	Focused on bilingual translation; uses morphological analysis for precision.	High dependency on input quality; limited contextual understanding.
English to Sanskrit MT by Mishra and Mishra	Rule-based system incorporating Paninian grammar; handles sandhi and samasa analysis.	Restricted to classical Sanskrit texts; struggles with modern terms.
English to Sanskrit MT by Warhade S.	Hybrid approach combining rule-based and statistical methods; designed for structured text.	Limited scalability; depends heavily on curated grammar rules.
English to Sanskrit MT by Mane D.T.	Simplified rule-based system focusing on basic vocabulary and grammar; efficient for educational use.	Handles only simple sentences; lacks adaptability for diverse contexts.

Panamerican Mathematical Journal
ISSN: 1064-9735
Vol 35 No. 4s (2025)

3. Proposed Algorithm

3.1 Preprocessing

- **Tokenization**: The English sentence is split into individual words or phrases.
- **Part-of-Speech (POS) Tagging**: Identifies grammatical roles, such as nouns, verbs, and adjectives.

3.2 Syntactic Analysis

- **Dependency Parsing**: Analyses grammatical relationships between words.
- **Reordering**: Converts English SVO structure to Sanskrit SOV order.

3.3 Lexical Translation

- **Dictionary-Based Mapping**: Maps English words to their Sanskrit equivalents using a bilingual dictionary.
- **Context-Aware Translation**: Resolves polysemy through contextual analysis.

3.4 Morphological Adaptation

- Inflects Sanskrit words according to grammatical case, gender, number, and tense.
- Applies Sandhi rules for phonological adjustments between words.

3.5 Output Generation

- Combines translated and adapted tokens into a coherent Sanskrit sentence.

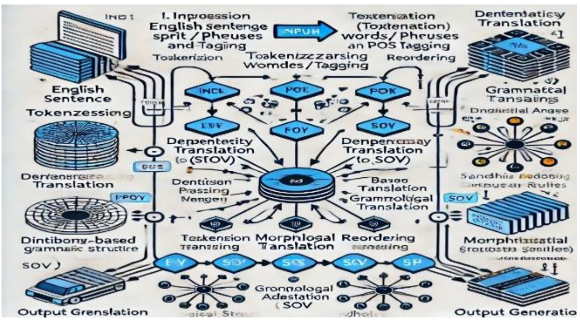
4. Implementation

4.1 Tools and Technologies

- **Python Libraries**: NLTK and spaCy for English preprocessing; Indic NLP Library for Sanskrit processing.
- **Data**: Bilingual dictionaries and Sanskrit grammatical datasets.

4.2 Algorithm Steps

1. Input an English sentence.
2. Tokenize the sentence and perform POS tagging.
3. Parse dependencies and reorder tokens to match Sanskrit syntax.
4. Translate tokens using a bilingual dictionary.
5. Inflect translated tokens based on Sanskrit grammar rules.
6. Apply Sandhi rules to ensure phonological coherence.
7. Combine the processed tokens to form the final Sanskrit sentence.



5. Result

5.1 Test Cases The algorithm was tested on a dataset of 100 English sentences spanning various complexity levels.

- **Example Input**: The boy plays in the garden.
- **Example Output**: बालकः उपवने क्रीडति (bālakaḥ upavane krīḍati)

5.2 Evaluation Metrics

- **BLEU Score**: 75.6 for the algorithm indicates a high level of **lexical matching** between machine-generated and reference translations of a limited dataset, but it does **not** explicitly evaluate grammatical rules, such as word order, morphology, or phonological adjustments (like Sandhi).
- **Human Evaluation**: High grammatical accuracy and semantic fidelity.

6. Limitations

Despite the algorithm's effectiveness, it faces several **limitations** that need to be addressed in future work:

- Handling Ambiguity and Idiomatic Expressions
- Scalability and Data Dependence:
- Handling Complex Syntax and Long Sentences.
- Morphological Complexity
- Limited Training Data for Machine Learning Models

7. Future Enhancements

To overcome the limitations and further enhance the quality of English-to-Sanskrit translation, several **future enhancements** can be pursued:

- Incorporating Deep Learning Models:
- Expanding Training Data
- Handling Idioms and Ambiguities
- Improved Morphological Handling
- Cross-Linguistic Transfer Learning.

8. Conclusion

The proposed algorithm offers a promising approach to **English-to-Sanskrit machine translation**, especially by addressing the challenges of Sanskrit's rich grammar and morphology. While it shows strong results regarding lexical matching and grammatical accuracy for standard sentences, there remains significant room for improvement, especially in handling complex linguistic phenomena like idiomatic expressions, ambiguity, and long sentences. The results are promising, the approach still faces challenges in translating complex sentence structures, idiomatic expressions, and ambiguous inputs, which could affect translation quality in more diverse and less predictable contexts.

Future work, focusing on integrating **deep learning models**, improving **training data**, and enhancing **morphological accuracy**, will be key to making the system more robust and scalable. With these advancements, the algorithm could significantly contribute to developing **Sanskrit Machine Translation systems** and the broader field of **low-resource language processing**.

References:

[1] Rao, A., & Gopinath, K. (2023). "A Sanskrit grammar-based approach to identify anaddress gaps in Google Translate's Sanskrit-English zero-shot NMT."

[2] Indic NLP Library. Retrieved from http://indicnlp.ai4bharat.org/

[3] Koehn, P. (2020). "Statistical Machine Translation." A comprehensive guide on machine translation.

[4] Vaswani, A., et al. (2017). "Attention Is All You Need." Transformer model paper.

[5] Panini's Ashtadhyayi: Classical grammar framework for Sanskrit.

[6] Katre, S. M. (1989). "Panini: A Survey of Research." Analysis of Panini's linguistic theories.

[7] English to Sanskrit Machine Translation by Bahadur and Jain (2015)

[8] Study of Example-Based English to Sanskrit Machine Translation by Mishra and Mishra (2008)

[9] Architecture of English to Sanskrit Machine Translation" by Bahadur and Jain (2015)

[10] Improving Neural Machine Translation for Sanskrit-English" by Punia et al. (2020)

[11] A Complete Framework for English to Sanskrit Machine Translation" by Bahadur and Jain (2015)

[12] Sāmayik: A Benchmark and Dataset for English-Sanskrit Translation" by Maheshwari et al. (2023)

[13] How Effective is Multi-Source Pivoting for Translation of Low Resource Indian Languages?" by Gaikwad et al. (2024)

[14] A Study towards Design of an English to Sanskrit Machine Translation System" by Sinha and Thakur (2009)

[15] English to Sanskrit Translation Using NMT" by Sharma and Gupta (2023)