International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Generative AI Framework for Marathi Grammar Learning

Dr. Gauri Deshpande[1], Prof. Vandana Sharma[2], Prof. Gayatri Dharap[3]

[1,2,3] *CSE-DS, Saraswati College of Engineering, Kharghar, India*

Abstract - Marathi is one of the popular oldest languages of India and possesses the greater syntactic complexity. Nouns, verbs and compound words of this language have very clear and simple rules that make the learning of the language an easy task for anybody. However, a more effective tool is required to consolidate the particularity of the higher-level grammar, especially the regional dialects. In the last few years there has been a lot of work done in the era of Artificial Intelligence (AI) and Natural Language Processing (NLP) where various languages related complicated tasks have been made easy. Pre-trained generative AI models like BERT (Bidirectional Encoder Representations from Transformers), GPT-3 (Generative Pre-trained Transformer), T5 (Text-to-Text Transfer Transformer) and many others hold much promise in different language uses like text generation, translation, and grammar checks. These models can formulate, interpreting and modifying any linguistic behaviour that might be useful in handling other challenges of Marathi language such as noun declensions and verb conjugations. AI model built from the transformer architecture can be adapted to handle several of linguistic problems in Marathi language that involves syntax analysis and error detection. This paper presents a generative AI model for learning the Marathi grammar which is further categorized into two parts. The first part of the study is devoted to the elementary grammar training, and the second part is dedicated to the intermediate and advanced levels. Using generative AI, this current model gives higher accuracy than rule-based system and presents effective ideas towards modern grammar. Also, this model became improving tools in computational linguistics for regional languages and for promoting language education.

Keywords: Marathi language, generative AI, BERT, GPT-3, T5, computational linguistics

1.INTRODUCTION

Marathi is one of the oldest and 23rd official languages in India today. The state government recognizes Marathi as its official language because it serves important roles in cultural activity, government work and political life. The language ranks fourth among Indian native speakers. Its distinct structure turns Marathi into a preferred language choice for scientists who examine language types and histories. The modern advancement of technology demands advanced Marathi language tools that help users to learn the language better while enabling faster translation and adapted experiences [2]. The Marathi language demonstrates a special formation that combines Dravidian and Indo-European language influences. The words maintain their correct pronunciation because Devanagari provides an accurate representation of sound in its written form. The sounds of Marathi make its language distinct because it contains a full set of vowels, consonants and combined letter combinations that create musical rhythms. The language builds words through adding prefixes and suffixes to base terms while showing the tense, gender and numbers [13]. Marathi sentences usually come in Subject-Object-Verb (SOV) order but switch patterns for emphasis purposes. The language of Marathi draws its words from Sanskrit, Persian, Arabic, Portuguese, and English because of its rich historical ties with these cultures. The complicated arrangement in Marathi serves as a foundation for diverse linguistic excellence in both poetry and official documents.

The real-time applications of Marathi grammar deliver essential functions that improve communication methods, educational processes and technological systems. Artificial Intelligence services support development of real-time applications that use Marathi grammar through various functional levels starting from simple to advanced technical capabilities. The basic functions of AI serve to create automatic speller systems along with grammar checkers and automated text predictors. By combining a set of algorithms and dictionary references, these applications make writing faster and more accurate. The main weaknesses of these systems stem from inadequate databases of words and their inability to understand context which results in erroneous suggestions. GPT models will analyse extensive datasets and get ability to identify context-based nuances and hence provide better context sensitive correction suggestions. [9]. Intermediate level of AI enables voice input-output systems to detect and produce Marathi verbalization. Hidden Markov Model and Recurrent Neural Network operate these systems while facing difficulties while managing various accents together with dialects. OpenAI's advanced Transformer architectures including Whisper develop their strength through multi-lingual speech training that enables dialect adaptation and effective speech synthesis accuracy [14]. At the advanced level, AI powers virtual assistants, chatbots, and machine translation systems. These programs need deep understanding of context while performing linguistic translations that preserve emotional content because Marathi contains intricate grammatical rules and cultural elements which present translation obstacles. The lengthy and multilingual large-scale corpora for generative AI models such as GPT-4 & T5 allow them to translate accurately and retain contextual dialogues through advanced syntactic and semantic patterns. These ones learn from new terminology and drifts to linguistic design through measures for instance few-shot learning and reinforcement coaching [16]. Generative AI models experienced high popularity because they demonstrate superior capability to understand complex syntactic structures together with contextual nuances and semantic relationships which enable them to process and generate Marathi language constructs efficiently.

This research study presents generative AI framework for Marathi language rule education employing transformer-based AI models such as BERT, mBERT, T5 and GPT-3. This framework performed basic and advanced tasks like sentence

classification, compound word identification, avyay recognition, transitivity classification, grammar correction, and negation form detection, Sandhi Vigrah, tense identification, dialect understanding, idiomatic phrase identification, and polarity detection. By using this model, this study effectively addresses Marathi language learning problems.

2. RELATED WORK

Marathi is known to be one of the major spoken languages amongst the people of India and it plays a crucial role in the communication domain as well as in the academic sector. However, the more conventional approaches that have previously been used in grammar learning and teaching. This leads to learning challenges and low motivation among the learners. The language learning system is still changing along with technology as interactive, personalized, and adaptive learning, with the help of ML. These technologies offer immediate footprints as well as the surrounding environment to increase grammar understanding. Nevertheless, some problems in adopting AI models to Marathi language remained uncovered.

Early computational models of Marathi grammar learning based on rule-based systems that took advantage of pre-defined grammatical rules for handling language processing. These systems used hand crafted rules, lexicons, prescribed grammatical structures, syntactic patterns for analysing and generating Marathi text but facing difficulties such as contextual nuances and exceptions. Machine translation is one notable application of the rule-based systems. Many rule-based systems were developed for Marathi language processing. One of the systems was developed for translation which is based on sentence structure rearrangement and word sense disambiguation. The syntactic transformations are dealt with effectively, but the compound words and idiomatic expressions are difficult to handle [7]. Similarly, another rule-based system was presented for translating English sentences into Marathi in the realm of machine translation. This system parsed Marathi and English sentences through Stanford parser and reordered the sentence structure based on Marathi syntax and hand coded rules that were added for Marathi inflections. The system provided a balance between preserving grammatical integrity and faces challenges like handling complexity of structures and the contextual nuances [8] Rule-based system also used for Part of Speech tagging. One of the rule-based systems was developed for Marathi language text POS tagger which was based on grammatical rules to resolve word and morphology ambiguity. The performance of this system was high for well-formed sentences but struggled with ambiguous words [1]. A rule-based POS tagger was designed for the Marathi language pertaining to tokenization, morphological analysis and ambiguity resolution by the means of grammatical rules. However, their system simply managed to give POS tags to the words on any sentence but not dealing with unknown words and dialectal variations [6]. Syntactic rule-based systems also evolved in early studies for sentence type categorization and compound word identification in Marathi. One of the systems achieved moderate accuracy with rule-based parser which follows predefined grammatical rules and morphological analysis. Unfortunately, rules were rigid and sentences were not complex enough to be handled by their approach [12]. This work was extended by another study that used context free grammar (CFG) rules for compound word segmentation but could not handle ambiguous compound word and idiomatic expressions [4]. Some rule-based systems were developed for transitive (Sakarmak) as well as intransitive (Akarmak) sentence classification. One study made the distinction between transitive and intransitive verbs according to the pattern of verb conjugation and case markers. This study showed high precision, but it was susceptible to syntactic variants and exception [10]. Rule based approaches use negative particle patterns to identify negations. These models were accurate for simple sentences, however not so much for double negations or context dependent negations. Singular and plural form of Marathi is indicated by suffix based morphological changes. These have been mostly rule-based methods for singular and plural identification by suffix stripping and inflectional patterns. Many morphological analyzers were designed with suffix rules and lexical lookups using rule-based approach, but they were restricted by other forms of verbs in plurals and dialectal variations. Many rule-based systems can accurately recognize honorifics, but they struggle with context-aware usage [15]. The syntactic relations are represented in Marathi with postpositional case markers. However, one rule-based system was able to discover case markers by matching predefined postpositions (e.g. "la", "ne", "cha") on nouns and pronouns but facing problems due to homophones and ambiguous markers [11]. In the case of Sandhi vigrah (compound word separation), phonetic rule-based approach applied phonetic rules for vowel and consonant combination. Although accurate for most compounds but they failed on complex sandhi formations and lexical variations. A rule-based tense identifier was created that matched the standard suffixes of the verb to predefined tense rules with good accuracy but failed on compound tenses and aspectual variation [5]. The rule-based systems have expanded abbreviated sentences with the help of phrase structure rules but fail to provide semantic context handling or idiomatic expansion.

Rule-based methods are an important resource to Marathi grammar processing, but they failed with the linguistic intricate and variable contextual. The emergence of generative AI models BERT, GPT-3, and T5 gives more powerful alternatives to learn contextual relations and generate contextual sounding language. BERT's two-way context comprehension facilitates sentence segmentation and part-of-speech tagging, GPT-3 performs superior in generation and paraphrasing and T5 encoder-decoder based model provides extensive applicability. These models deal with complicated syntax and code-mixed data more successfully than workable techniques. However, the problems such as insufficient annotated data, dialectical situations and high computational cost are still the biggest problems. In order to overcome these challenges multimodal learning, low-resource injection and cross generative models are required for contextual understanding.

3. METHODOLOGY

In this research, a generative AI model is proposed for Marathi grammar tasks that employ transformer-based models like BERT, GPT-3, T5, and mBART, along with custom algorithms to improve accuracy. This model addresses challenges such as management of large, annotated datasets, shortening computation time, and address dialectal variations.

This methodology comprises three distinct phases:

1. Data Collection and Pre-processing
2. Designing the Generative AI Model for Basic and Advanced Grammar Tasks
3. Evaluation of the Model

1. Data set Collection and Pre-processing:

The first step includes collecting a rich set of Marathi data from books, articles and everyday conversations in various domains. By tokenizing the data set, lemmatization and tagging part-of speech, it is possible to pre-process most of this data to standardize the language. Special treatment is accorded to Marathi specific attributes such as honorifics, sandhi (compound word) splitting, and regional dialect. Specific tokens are also used to recognize the word boundaries, between different languages to handle complicated languages and so forth.

2. Developing the Generative AI Model for Basic and Advanced Grammar:

In the second phase, a combination of transformer-based architecture and custom algorithms will be used to build generative AI models. This model is designed for use as basic and complex Marathi grammar. This study provides a way for the utilization of Marathi linguistic tasks by fine-tuning BERT and GPT-3/4 models for specific tasks.

2.1 Basic and Advanced Tasks Implementation:

For Sandhi Vigrah (Word Combination & Split), mBART was employed along with a pattern recognition system to recognize the compound word structures and split them to the base forms adopting the morphological rules. Tense and Aspect Recognition utilized T5 and XLNet to identify tense markers on verbs and correlate them to temporal information for the correct annotation. Sentence Expansion (Dilation) leveraged a seq2seq model based on GPT-3 to expand simple sentences with clauses, adjectives and prepositional phrases adding context and semantics of the content. Comparison with Other Languages was carried out with mBART to examine sentence structures in Marathi, Hindi and English and asserted the syntactic and morphological changes for facilitating the linguistic translation and translation adaptation. Complex Sentence Generation & Transformation combined GPT-3 with a rule-based system for merging clauses and reshuffling sentences and always retaining syntactic accuracy and valid punctuation. Dialect and Regional Variation Detection employed BERT trained on Marathi-specific dataset to pick-up vocabulary change and phonetic variation. In Pragmatic and Contextual Interpretation, idioms are translated into literal or contextual sense using a syntactic device in combination with GPT-3 and T5. Anaphora Resolution employed BERT to identify and connect pronouns to their antecedents utilizing syntactic and semantic clues, thereby enhancing coherence and narrative fluency. Modality Handling and Polarity Detection merged T5 and BERT with rule-based system to detect negation and polarity keywords, improve sentiment analyse. Sentence Type Identification employed BERT and other transformer models with hand-coded rules based on punctuation and auxiliary verbs, professionally stuck with deep learning adjustments to give higher classification precision. Compound Word (Samas) recognition combined spaCy and BERT along with a rule-based system to detect probable morphemes and handily divide compound words. Transitivity Classification employed transformer models to keep track of verb-object relationships, consistently distinguishing Sakarmak (transitive) and Akarmak (intransitive) sentences. Avyay (Indeclinable Word) Identification used a sequence classifier based on BERT for contextual pattern and word position analysis, increasing the recognition rate. Singular/Plural Identification integrated spaCy with their own implementation of a singular and plural morphological marker-based algorithm based on Marathi grammatical rules. Grammar Correction was executed by GPT-3 and T5 together with a proprietary model to spot and figure out grammatical mistakes, such as subject-verb agreement mismatch. Honorifics Identification and Generation used spaCy and rule-based methods to detect sentences' tone and formality to correctly utilize honorifics in all social contexts. BERT was enabled to detect frequent negation patterns by sentence structure and type to enhance Negation Form Identification. These integrated techniques improve Marathi NLP jobs significantly by implying that grammatical accuracy, syntactic coherence, semantic comprehension, and dialectal adaptability are attained.

3. Evaluation of the Model:

Evaluation of the model involved testing its performance by measuring accuracy levels with efficiency while handling irregular forms and complex structures and regional dialects. The system was evaluated using limited annotated datasets alongside present methods for comparison. Precision, Recall combined with F1 Score became the performance metrics for both basic and complex tasks regarding identification of linguistic features that include sentence type and compound words alongside grammatical forms. The accuracy evaluation method was used for general classification tasks but BLEU (Bilingual Evaluation Understudy) alongside WER (Word Error Rate) and TER (Translation Error Rate) served as evaluation metrics for text generation activities including grammar correction, generation and translation functions.

4. RESULT AND DISCUSSION

A proposed generative artificial intelligence model evaluated Marathi sentences through basic and advanced linguistic assessment for this research. The analysis was conducted on 2,000 sentences specifically selected for retrieving adequate testing data. The evaluation of the model assessed its competence to generate proper syntactic constructions in reasonable time durations and resource allocation for different linguistic patterns. Performance evaluation of the proposed AI model and its accuracy level was conducted by directly

comparing it against traditional rule-based methodologies. An illustration of the proposed generative model's outputs appears below.

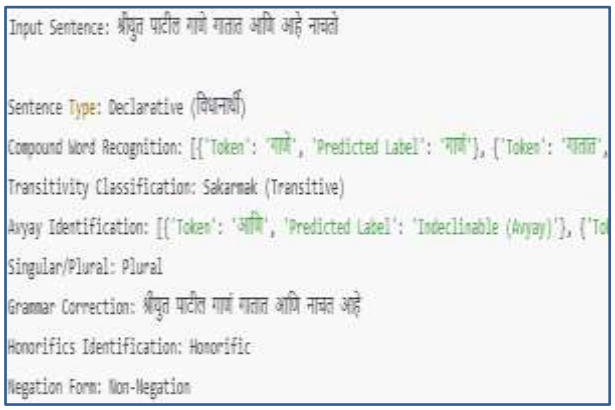

Figure 4.1: Basic Tasks of Marathi Grammar

Figure 4.2: Advanced Tasks of Marathi Grammar

The following table presents a comparative evaluation of AI models and Rule-Based approaches for linguistic tasks.

Table 4. 1: Performance Comparison: Rule-Based vs. AI Model

Tasks	Rule-Based				Proposed Model			
	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
Sentence Type Identification	76%	76%	76%	76%	92%	92%	92%	92%
Identifying Compound Words	73%	76%	76%	76%	92%	92%	92%	92%
Sakarmak vs. Akarmak Sentences	67%	67%	68%	66%	93%	93%	93%	93%
Singular or Plural Identification	65%	65%	65%	65%	90%	90%	90%	90%
Honorific Identification	75%	75%	75%	75%	95%	95%	95%	95%
Identifying Negation Forms	75%	75%	75%	75%	95%	95%	95%	95%
Case Markers Identification	75%	75%	75%	75%	95%	95%	95%	95%
Sandhi Vigrah	85%	80%	82%	83%	92%	91%	91%	92%
Tense Detection	87%	86%	87%	86%	92%	91%	92%	91%
Expanded Sentence	85%	83%	84%	85%	91%	90%	90%	91%
Tense and Aspect Handling	85%	84%	84%	85%	92%	91%	92%	91%
Context Interpretation	90%	85%	87%	88%	95%	93%	94%	93%
Idiomatic Expression Recognition	96%	80%	89%	90%	96%	96%	96%	96%
Modality Handling	96%	96%	96%	96%	96%	96%	96%	96%
Anaphora Resolution	96%	96%	96%	96%	96%	96%	96%	96%
Polarity Detection	80%	80%	80%	80%	90%	90%	90%	90%
Subject-Object Reversal	75%	75%	75%	75%	92%	91%	92%	91%

The proposed model outperforms the rule-based approach in the entire language processing task. In Sentence Type Identification, the proposed model obtained 92% for precision, recall, F1-score, and accuracy over the 76% for the rule-based method. For Identifying Compound Words, it had 92% of precision and recall, which is significantly better than the case of the rule-based approach's 73% and 76%, respectively. Most significant progress was made in Sakarmak vs. Akarmak Sentences where proposed model had 93% of precision and recall outperforming rule-based method 67%. Besides, Honorfic Inference, Negation Forms Inference, and Case Markers Inference got obvious improvement. The suggested algorithm always reached 95% precision and recall. The proposed model also out fraud the rule-based method in Sandhi Vigrah, Tense Detection, Sentence Extension Handling, Context Interpretation, this model could do its precision and recall at the range of 90-95%. Additionally, the tasks such as Idiomatic Expression Recognition, Modality Handling, Anaphora Resolution, and Polarity Detection achieved increased improvements with the proposed model achieving precision and recall of 90-96%. On average, the proposed model's accuracy over all tasks was around 92%, showing that it has high credibility and effectiveness. The obtained results convincingly demonstrate that the proposed model is more efficient than the traditional rule-based method, show how reliable and effective it is in various language processing tasks.

5. CONCLUSION

The proposed study uses different transformer models including BERT, GPT3 and T5 as a generative AI architecture to learn the Marathi grammar. In the context of various tasks, these models perform much better than rule-based approaches by effectively analyzing sentences, compound terms and contextual meanings. It is an attempt to build a strong AI driven framework that demonstrates strong results over both basic and advanced linguistic tasks and can be used as a promising tool for Marathi language processing and grammar education. The proposed framework also benefits computational linguistics beyond Marathi where it is found to be an effective methodology for learning regional languages and other Indian language with complex grammar structure. Future research could focus on adding dialectal modifications to make linguistic diversity coverage more dialect-centric or to use multi modal AI by combining speech and text-based learning to improve pronunciation, prosody and real time grammar correction. Finally, the framework could be extended to use more AI based grammar instruction such as semantic role labelling, discourse analysis, and syntactic tree generation to further refine a given use of AI based grammar instruction. Further, this system can also be used in educational platforms, AI driven tutoring systems, and auto generated language assessment tools as a means to determine its role in the real-world usage, and utility in Marathi language education.

ACKNOWLEDGEMENT

The authors express their sincere thanks to Saraswati College of Engineering, Kharghar, for their support and resources. We are grateful to our mentors and colleagues for their valuable guidance and encouragement throughout this research. We also

acknowledge the contributions of the research community in the field of AI and NLP, which inspired this work.

REFERENCES

1 Bagul, D. D., Patil, P. S., Kulkarni, P. A., & Tarte, S. V. (2014). Rule-based part of speech tagger for Marathi language. International Journal of Computer Science and Information Technologies, 5(2), pp.-2215-2218

2 Basutiya, V. N., & Hankare, V. K. (2023). Compendious study of Gujarati migrated people Marathi dialect in Vidarbha region with standard Marathi. International Journal of Research and Analytical Reviews, 10(2) ,p- 310.

3 Dabre, R., Amberkar, A., & Bhattacharyya, P. (2012). Morphological analyzer for affix stacking languages: A case study of Marathi. In M. Kay & C. Boitet (Eds.), Proceedings of COLING 2012: pp. 225–234.

4 Desai, P., & Kulkarni, A. (2020). Context-free grammar approach for compound word segmentation in Marathi. International Journal of Advanced Science and Technology, 29(4), 4866-4875.

5 Deshmukh, S., & Patil, R. (2021). Rule-based tense detection in Marathi using verb conjugation patterns. Journal of Linguistic Studies, 14(3), pp-321-334.

6 Gaikwad, P. S., Naik, D. S., & Mahender, C. (2018). Rule-based part-of-speech tagging for Marathi text. International Journal of Scientific Research in Science and Technology, 4(5), pp-1687-1693.

7 Garje, G. V., Kharate, G. K., & Kulkarni, S. (2014). Transmuter: An approach to rule-based English to Marathi machine translation. International Journal of Computer Applications, 92(15), pp-25-29.

8 Godase, A., & Govilkar, S. (2015). A novel approach for rule-based translation of English to Marathi. Advanced Computational Intelligence: An International Journal (ACII, 2(4).

9 Joshi, A., Deshpande, S., & Patil, M. (2022). Generative AI for Indian Languages: A Marathi Perspective. International Journal of Artificial Intelligence and Applications, 14(3), pp-89-98.

10 Joshi, M., & Patil, A. (2018). Rule-based approach for Sakarmak and Akarmak sentence classification in Marathi. Linguistic Journal of South Asian Languages, 12(2), pp-102-113.

11 Kulkarni, R., & Patil, V. (2019). Rule-based identification of case markers in Marathi. International Journal of Computational Linguistics, 7(4), pp-243-252.

12 Patil, M., & Deshmukh, S. (2019). Rule-based parser for sentence type classification in Marathi. Journal of Language and Linguistic Studies, 15(2), pp-53-65.

13 Patil, R., Vishwakarma, A., & Chindaliya, P. (2023). Linguistic structural study of Agri Marathi language: Special reference to lexis, sound, and structure. Urban India, 43(2)

14 Radford, A., Kim, J. W., & Xu, T. (2023). Robust Multilingual Speech Models: Whisper Architecture. Proceedings of the Neural Information Processing Systems Conference.

15 Shinde, P., & Joshi, R. (2021). Honorific identification in Marathi using rule-based systems. Indian Journal of Computational Linguistics, 9(1), pp- 87-96.

16 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30, pp- 5998-6008.