International Journal of Advanced Technology and Engineering Exploration, Vol 11(113)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2023.10102163

Research Article

Bloom's Taxonomy based automatic Marathi question generation

Pushpa M. Patil[1*], R. P. Bhavsar[2] and B. V. Pawar[3]

Assistant Professor, Department of Computer Science, Pratap College, Amalner, Maharashtra, India[1]
Professor, School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, Maharashtra, India[2]
Director, KCES's Institute of Management and Research, Jalgaon, Maharashtra, India[3]

Received: 20-August-2023; Revised: 23-April-2024; Accepted: 25-April-2024

Abstract
In the current era of automation, various fields, including education, are undergoing transformations to enhance their existing processes. One crucial aspect in the field of education is examination management. Automatic question generation (AQG) for creating evaluation systems and question papers represents a significant transformation that schools, colleges, and universities are experiencing. Although significant research has been conducted in AQG for foreign languages, there is a scarcity of such work in Indian regional languages. Considering this, a novel working model for AQG for Marathi language texts was presented. The proposed research generates a diverse set of questions automatically through various natural language processing (NLP) pipeline activities, including tokenization, parts of speech (POS) tagging, stemming, named entity recognition (NER), shallow parsing, and dependency parsing. The generated questions fall into the categories of context-based and grammar-based questions, each elaborated in detail with scientific interpretation. This process contributes to the validation and refinement of our question generation methodology. A benchmarking approach using Bloom's Taxonomy was employed to validate the accuracy of the generated questions, ensuring they were aligned with educational objectives and targeted the desired levels of cognitive complexity. The empirical evaluation of the proposed methodology is conducted using the bilingual evaluation understudy (BLEU) score and manual scoring. The accuracy achieved using the BLEU score is 90.37% for the 'wh' questions, based on the corpus created from the sixth standard science textbook published by the Maharashtra State Board, Maharashtra, India. A diverse set of high-quality Marathi language questions has been successfully curated, suitable for compiling question papers aligned with Bloom's taxonomy levels.

Keywords
Context based questions, Grammar based questions, NLP pipeline, POS, Parsing.

1.Introduction
In the contemporary era of automation, various fields, including education, are undergoing transformative changes to enhance their operational processes. Within the education sector, examination management stands out as a crucial aspect, and the integration of automatic question generation (AQG) has emerged as a significant transformation. This innovation plays a pivotal role in shaping evaluation systems and crafting question papers for educational institutions such as schools, colleges, and universities. Questions are essential components of the teaching and learning process [1], since effective questioning is crucial for learning [2].

Generating suitable questions is a cognitively demanding task [2]. The evaluator can assess the learner's knowledge by framing cognitive questions. Bloom's taxonomy offers detailed instructions on how to evaluate students' learning based on their responses to questions. The production of questions at higher Bloom's taxonomy levels, which necessitate a deeper level of cognitive processing and comprehension, is an area of ongoing research in the natural language processing (NLP) and question generation communities [3]. Traditionally, question creation is a time-consuming task [4] that requires domain expertise. However, without the aid of a domain expert, one can use AQG to generate a varied set of questions from a given text. AQG has gained importance due to its wide-ranging applications across various domains, including text summarization

*Author for correspondence

[5], intelligent tutoring systems [6], the education sector [7–9] to generate the question paper, dialogue management systems [10], and games [11].

In the current scenario, AQG is performed by employing various approaches like rule-base [3, 12, 13, 14], neural network-based [15] statistical approach using inverse document frequency [6], text summaries-based question generation [4, 16], and genetic algorithm-based question generation [11]. Neural network-based approaches employ bidirectional encoder representations from transformers (BERT) [10, 16], recurrent neural networks (RNN), graph neural networks (GNN) [17], bidirectional and auto-regressive transformers (BART) [18], and generative pre-trained transformer (GPT)-2 for AQG. Notable works have been reported for AQG in Indian regional languages such as Hindi [19], Bangla [20], Malayalam [21], and Marathi [14]. Few studies focusing on regional languages like Marathi have been reported by Gaikwad et al.; they presented a rule-based approach for generating 'wh' questions (who, where, when, what, and how).

Unlike data-driven or machine-learning approaches, the mentioned rule-based system provides transparency and enables fine-grained control over the generated questions, making it highly adaptable to different domains and languages. But these approaches depend on the input text, dataset, corpus, etc. Natural languages are known for their dynamic nature; hence, these approaches may fail to capture the linguistic aspects of a given text, which is crucial in the extraction of questions. Extracting a question from a given text requires a linguistic evaluation of the language. Existing techniques that rely on machine learning models fail to a certain extent when it comes to accurate AQG. Hence, a scientific study of the linguistic and grammatical aspects of Marathi has been conducted, and a set of rules and patterns has been crafted that facilitate AQG in a linguistically correct manner. While substantial research has been conducted in the area of AQG for foreign languages such as English [18, 22–25], Bahasa [26], Chinese [27], and Portuguese [28], there is a notable gap in AQG focusing on Indian regional languages. Addressing this gap, current research presents a novel working model for AQG crafted specifically for the Marathi language. This proposed framework utilizes NLP pipeline activities, including tokenization, parts of speech (POS) tagging, stemming, named entity recognition (NER), shallow parsing, and dependency parsing. This research paper presents a novel approach to AQG that traverses the NLP pipeline, covering multiple levels of linguistic analysis to extract patterns and rules for AQG. A novel rule-based technique is reported in the proposed work, which generates a diverse set of questions from various stages of the NLP pipeline. The outcome of the research resulted in the generation of nineteen different types of context-based and grammar-based questions. The generated questions are mapped against Bloom's taxonomy, which classifies the question at the cognitive level. The AQG approach described here attempts a systematic exploration of question generation across the NLP pipeline using a rule-based paradigm.

Following are the objectives of the proposed research work.

1. To perform grammatical and linguistic analysis of various NLP pipe line activities.
2. Identify the patterns and rules for framing a question based on the syntactic and semantic compositionality of
1. Marathi language.
2. Mapping of generated questions with Bloom's taxonomy.

The key contributions of this study are twofold: first, it provides a flexible and interpretable method for generating questions, allowing us to explicitly capture various linguistic phenomena at each level of the NLP pipeline. Second, the extensive range of question types generated showcases the robustness and depth of the approach. We explore the syntactic, semantic, and pragmatic aspects of a given text, generating interrogatives, 'wh' questions, true/false questions, and more.

The paper is organized into six sections. Section 2 provides a concise overview of related work in AQG. Section 3 delves into a comprehensive explanation of the detailed methodology of AQG. Section 4 engages in a thorough discussion of the results and their empirical evaluation. Section 5 is dedicated to addressing the limitations of the proposed model. Finally, Section 6 concludes the research work and outlines paths for future research.

2. Literature review

Multiple studies related to AQG are carried out by many researchers in various languages and contexts.

Das et al. [12] have generated 'wh' questions using a rule-based approach at the sentence level. They have transformed the declarative sentence into an interrogative by applying the rules for identifying the chunks in the sentence and replacing them with the wh word. Rules contain the specific sequence of the POS tags that define the verb phrase (VP) or a noun phrase (NP). They have generated who, what, whose, whom, when, where, how much, and 'how many' types of questions. Their system may generate erroneous questions if prepositions are excluded from the list provided in their system. Also, the AQG fails in cases of wrong POS tagging. However, they have reported an agreement rate of 80% amongst human evaluators for the quality of the generated questions.

Moron et al. [29] have utilized linguistic information such as POS, semantic role labeling (SRL), and NER to design the rules. After pre-processing the input text, rules were used to replace the part of the text having a particular semantic role and the named entity with a question word. They developed 'wh' questions like who, where, what, and when. Further, these questions can be used for teaching English as a second language. Their AQG has achieved highest F-score of 0.842 for 'who' type questions and the lowest F-score of 0.333 for 'where' type questions. They have offered a manual correction facility for inaccurate generated questions.

Chali and Hasan [13] proposed the topic-to-question method, which uses about 350 general-purpose rules to transform the declarative sentence into an interrogative sentence at the paragraph level. Here, the rules are based on linguistic information such as the named entity and the predicate argument structure of the sentence. They used latent dirichlet allocation (LDA) to extract the important subtopics from the given text. The integration of a tree kernel function to assess the syntactic correctness of the generated questions is one of the advantages of their proposed system. They have utilized the data set provided in Question Generation Shared Task and Evaluation Challenge 16 (2010) for question generation and achieved high scores in both 'topic relevance' and 'syntactic correctness', with values of 3.45 and 3.50, respectively.

Garimella et al. [30] explored the challenges of parsing natural language questions in the finance and weather domains. They have proposed a framework for automatic labeled domain question generation by utilizing domain knowledge and seed domain questions. Which improves parser accuracy by 49% ± 9% for domain-specific questions.

Wijanarko et al. [26] have generated the questions by combining the question template with the key phrase. The key phrases are the NPs that were extracted using chunk parsing from the text. The question templates are based on Bloom's taxonomy of question verbs. Context-free grammar rules are used for NP extraction. Their system generated 60,000 questions, having Bloom's levels 3, 4, and 5, which were derived from 1,432 sentences. The proposed system achieved a bilingual evaluation understudy (BLEU) score of 0.9566 and a Cohen-Kappa coefficient of 0.34589. However, their system is unable to generate multiple-choice questions.

Xu et al. [18], with the expertise of education professionals, have crafted an extensive dataset with 10,580 questions for children's narrative comprehension named FairytaleQA. The dataset includes seven relationship types. They employed the BART model for question generation on both FairytaleQA and NarrativeQA datasets. FairytaleQA's questions closely resembled ground-truth patterns, while NarrativeQA showed changes in question semantics during training. BART achieves a recall-oriented understudy for gisting evaluation (ROUGE)-L F1 score of 0.527, when it is fine-tuned on the FairytaleQA dataset, while on the NarrativeQA dataset, it achieves a lower ROUGE-L F1 score of 0.442. This comparison suggests that BART performs significantly better when fine-tuned on FairytaleQA compared to NarrativeQA.

Huber and Hagel [22] have utilized the GPT-2 text generation model for software engineering textual exercises for unified modeling language (UML) class diagrams. The model learns through grammar, sentence structure, vocabulary, and context. The generated text of the model needs manual correction. 19 software engineering students manually evaluated the textual exercises; 57.9% of them understood the grammar and sentence structure, and 84.2% would prefer the textual exercises for exam preparation.

Hou et al. [23] proposed a novel approach for sentence-level question generation by employing prefix-adjusted soft prompt learning and a syntactic information-based model. The proposed model aims to improve the fluency, relevance, and answerability of generated questions by effectively using syntactic information to guide question generation. The model's effectiveness is validated through experiments on benchmark datasets, the Stanford question answering dataset (SQuAD) and Microsoft machine reading comprehension (MS MARCO), demonstrating its ability to outperform previous question generation approaches in both automatic and human evaluation metrics. The limitations of the proposed question generation model are primarily related to the reliance on pre-trained language models and the use of benchmark datasets for performance evaluation. Also, there is a need to improve the diversity of question generation. The researchers manually evaluated the fluency and answerability of the generated questions on two datasets, SQuAD1.1 and MS MARCO. Dataset SQuAD1.1 received scores of 4.38 and 4.59, while dataset MS MARCO received higher scores of 4.51 and 4.7 for fluency and answerability, respectively.

Kumar et al. [31] proposed a data augmentation method introducing diverse question generation from a larger language model for the same context and answer pair. They utilized an over-generate-and-rank approach to select the optimal question. One notable advantage of the proposed model is its ability to generate challenging ‘implicit’ questions where answers are not directly present in the context of the text. The proposed model encounters a character coreference resolution error.

Maheshwari et al. [32] introduced pre-training with extracted gap-sentences for abstractive summarization (PEGASUS)-large and BART-large language models, designed for generating entity-level factual questions by delexicalizing uncommon words. For training, they employed the dataset, explain like I’m Five (ELI5). However, a notable drawback of their approach is the considerable time investment needed for both the training and inference phases. Among the three approaches, namely normal fine-tuned PEGASUS, rare word de-lexicalization plus multiple generation, and span copy without global relevance, the rare word de-lexicalization plus multiple generation approach consistently outperforms the other approaches across all datasets, achieving the highest F1 score.

Dugan et al. [4] conducted a feasibility study on answer-agnostic question generation for textbook passages. The findings indicated that, despite the use of answer-agnostic question generation, posing questions based on summarized text yielded better results. Answer-agnostic question generation sometimes generates questions that are out of context. They utilized the BART language model on the convolutional neural network (CNN)/DailyMail dataset for summarization. There is a significant increase in the acceptability of generated questions, from 33% to 83%.

Chen et al. [15] have introduced a novel bidirectional graph-to-sequence (Graph2Seq) model for realistic knowledge graph (KG) question generation, surpassing previous approaches focused on single KG triples. The model employs a unique node-level copying mechanism in the RNN decoder for direct attribute transfer. They have reported syntactic error patterns such as repeated words and the absence of important pieces of information in the generated questions.

Das et al. [33] proposed an approach that generates subjective questions from course curriculum key phrases, utilizing a multi-criteria decision-making method to assess student responses against model answers. The study showcases the efficiency of the automated system in minimizing manual assessment efforts, highlighting its significance in the fields of NLP and educational research. The proposed approach introduces automation for subjective question generation and single-sentence answer evaluation, employing a keyword-driven method and multi-criteria decision-making.

Fei et al. [24] proposed a controlled question generation (CQG) framework for multi-hop question generation, addressing the challenge of ensuring question complexity and reasoning over multiple pieces of information from input passages. The CQG framework employs a graph attention network-based key entity extractor and a controlled transformer-based decoder with flag tags to ensure question complexity and quality. Experimental results show that the CQG model outperforms existing models by 25% in terms of BLEU points on the hotpot question answering (HotpotQA) dataset, showcasing its efficiency in multi-hop question generation.

Steuer et al. [25] investigated the transferability of non-educational answer selection models to the educational domain for AQG. For this, they utilized three data sets, among which SQuAD and natural questions (NQA) are non-educational datasets while textbook question answering with answer (TQA-A) is a novel educational dataset, developed for the study. Further answer selection is performed using six machine learning algorithms, viz., BERT, a robustly optimized BERT pretraining approach (RoBERTa), decoding-enhanced BERT with disentangled attention (DEBERTa), a distilled version of BERT (DistilBERT), a lite BERT (ALBERT), and SpanBERT, on the mentioned datasets. The research aims to answer two main questions: the extent to which answer selection models select the correct answers on dissimilar datasets, and the extent to which these models transfer between educational and non-educational datasets. The results show that SpanBERT gives the best result for phrase level answer selection among the six mentioned machine learning algorithms. The findings suggest that non-educational models do not necessarily transfer effectively to the educational domain, highlighting the need for larger educational training datasets or completely different answer selection methods for effective educational question generation.

Nguyen et al. [3] utilized materials from a graduate-level introductory data science course for generating questions. Data was extracted in extensible markup language (XML) format, organized by unit, module, and topic. They have employed MOOCCubeX for concept hierarchy extraction and Google's T5 model for question generation. Google's T5 model was initially fine-tuned for the SQuAD dataset. Out of the 203 generated questions, 151 (74.38%) questions by the GPT-3 module and 115 (56.7%) questions by human evaluators were classified as pedagogically sound. However, the study reported limited question diversity, primarily centered on 'what' questions.

Gopal [19] proposed a methodology that utilizes the GPT model, the text-to-text transfer transformer model, and syntactic post-processing for AQG. He utilized 'Pratham' course material to generate paragraph-level questions for Hindi and Marathi. The limitations of the research include the dissimilarity in sentence structure between English and Indian regional languages, as well as the need for robust evaluation systems for AQG tasks.

Rathi et al. [34] have applied NLP approaches to English AQG. By using the nouns in the parsed text, they have created objective and subjective questions using the templates. Following the manual evaluation, 73% of the generated questions were grammatically correct. Additionally, they have produced answers to the questions using the cosine similarity method. Only a few of the questions generated in their system have syntactic errors. The proposed system can generate the questions 'what,' 'define,' 'write a brief note on,' 'explain,' and other similar ones.

So it is clear from the literature survey that questions can be generated at the sentence [12, 23], paragraph [13, 10], and document [16, 6, 26] levels. Throughout the literature survey of AQG, a divergence in question generation for different natural languages, approaches, datasets used, and evaluation techniques was observed. The reported works employ neural network-based [18, 19, 22, 25], template-based [26, 34], and rule-based [3, 12, 13, 29] approaches.

Throughout the literature, it has been figured out that rule-based approaches are well suited when there is a requirement for the highest rate of accuracy and precision. But while dealing with scalable models and large datasets, machine learning approaches outperform rule-base approaches. Since the AQG engine is crafted here to provide accurate and refined question generation, a rule-based approach is employed. In addition to this, the proposed approach addresses context and grammar-based question generation at each level of the NLP pipeline. Furthermore, the generation of such a diverse set of questions contributes significantly to the field.

3. Methods

The overall framework of AQG is depicted in the following *Figure 1*. Initially, a valid Marathi sentence is provided as an input to the proposed model. The input sentence then undergoes several stages of pre-processing. At each stage, respective questions are generated, and the preprocessed sentence is transferred to the subsequent stage of the NLP pipeline. The questions generated by the AQG engine are later mapped to Bloom's taxonomy. The generated questions undergo post-processing, where the questions are modified. As "Marathi is a free word language" [35], the post-processing module changes the word order of the question and constructs a question with a different word order. The AQG engine can generate nineteen different types of questions. The pre-processing of the input text, along with linguistic study and the question generation methodology, by the AQG engine is detailed in Section 3.1. Revision of the generated questions is explained in Section 3.2, followed by Section 3.3, where a rule-based mapping between the question and Bloom's level is carried out.

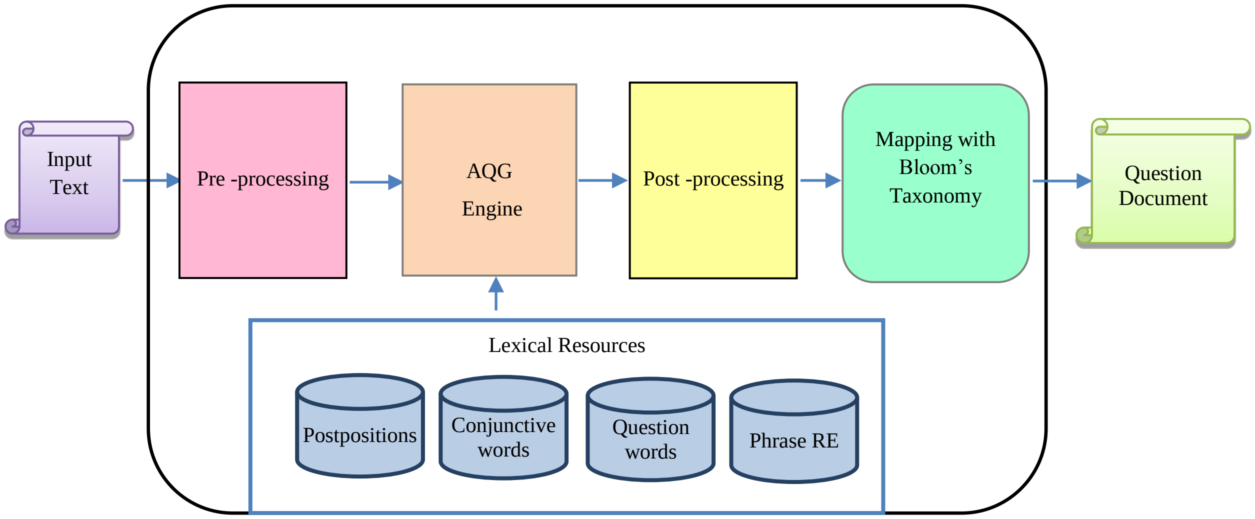

Figure1 Framework of AQG

3.1Preprocessing and question generation

Text preprocessing is an integral part of any NLP application. For AQG, the input text is preprocessed at various levels. The different steps for pre-processing the text are tokenization, POS tagging, stemming, NER, shallow parsing, and dependency parsing. The NLP pipeline consists of several steps that are used to analyze the natural language text at the syntactic and semantic levels. In syntactic analysis, one can analyze the grammatical structure of the sentence. Syntactic analysis is done with the help of tokenization, POS tagging, shallow parsing, dependency parsing, etc. In semantic analysis, one can understand the meaning of the sentence. Semantic analysis involves NER, word sense disambiguation (WSD), sentiment analysis, natural language inference (NLI), etc. The following sections give a brief discussion of each process in the NLP pipeline.

3.1.1Tokenization

Tokenization is the process of breaking down a text into smaller pieces, called tokens. These tokens are usually words from a given text. The proposed algorithm uses tokenization to generate questions of type 'Arrange the words in the correct order'. Such types of questions would check the student's knowledge of grammar, syntax, context, and meaning of the words. Here, the words of a sentence have been separated, and then these words are randomly shuffled. The student is asked to arrange these words in the same order as the original sentence. According to Bloom's taxonomy, such types of questions fall into the 'analysis level' (Level 4). It can check the student's ability to analyze the word's meaning and place, which involves grammatical study of the particular language.

The initial step involves tokenizing the input Marathi sentence. The natural language toolkit (NLTK) [36] tokenizer has been used for word tokenization. The Word_Shuffle_Algorithm is used to generate the question of type: 'Arrange words in the correct order and write a meaningful sentence'.

Word_Shuffle_Algorithm:

Input: Marathi sentence, punctuation symbol lexicon
Output: sentence with shuffled words
Step 1: Start
Step 2: Tokenize the sentence
Step 3: Remove the punctuation symbols
Step 4: Count the number of tokens as n.
Step 5: if n > 3 then randomly shuffle the tokens.
Step 6: Output the sentence with shuffled words
Step 7: End

3.1.2POS tagging

It gives us the grammatical category of each word in a sentence. There are eight types of POS categories in Marathi: noun, pronoun, adjective, verb, adverb, prepositions, conjunction, and interjection or exclamation. A noun is a word that is the name of something, like a person, location, thing, animal, time, date, etc. The noun is further subdivided into three categories: common noun, proper noun, and abstract noun. Proper nouns are the specific name for a particular person, thing, place, or animal. Proper nouns are used to generate the factoids. To identify the entity type, whether it is a person name, organization name, or anything else, the next level of the NLP pipeline is needed, i.e., NER. Trigrams'n' tags (TnT) POS tagger [37] is trained using the 'marathi.pos' corpus from NLTK and conducted POS tagging by leveraging the TnT POS tagger. Later, shallow POS tagging is performed in situations where TnT is unable to recognize the POS tag. The following section details the linguistic study of the POS tag with respect to question generation.

Adjective: An adjective is a word that describes the noun or pronoun in a sentence. Adjective words in a sentence have been used to generate questions for the tutorial dialogues [38] and vocabulary questions [39]. *Table 1* shows the types of adjectives and the types of questions that can be generated from the adjective words. The subtypes of adjectives and the questions that can be generated from the adjective words are discussed here.

Qualitative adjectives: A qualitative adjective is a word that describes different qualities of the noun. For example, *चांगला मुलगा* (good boy, Cāṅgalā mulagā). *गोड सफरचंद* (sweet apple, Gōḍa sapharacanda). *शूर सरदार* (brave soldier, Śūra saradāra). Here, the 'how' type of question is be formed by consuming the qualitative adjective. Questions types based on grammar, like 'find out the synonyms and antonyms of the given word,' can be formed by using the qualitative adjectives in the sentence. These questions have Bloom's level 2, i.e., 'comprehension level'.

Quantitative adjectives: A quantitative adjective is a word that counts the noun or gives some numerical information about the noun. The counting adjective shows the number of nouns, and from the counting adjectives, factoid questions like 'how many,' 'how,' and 'what is' can be formed. For example, the counting adjectives are *दहा मुली* (ten girls, Dahā mulī), *चौदा भाषा* (fourteen languages, Caudā bhāṣā), and *अर्धा तास* (half an hour, Ardhā tāsa). The sequential adjectives explains about the sequence of the nouns in the sentence. As an example, *प्रथम श्रेणी* (first class, Prathama śrēṇī), *पाचवा बंगला* (fifth bungalow, Pācavā baṅgalā). The question of the form *कितवा* (what is, Kitavā) can be constructed from the sequential adjective word. The repetition adjectives indicate the number of repetitions of the noun. For example, *चौपट* (four times, Caupaṭa), *दहापट* (ten times, Dahāpaṭa), *दुहेरी* (twice, Duhērī). 'How many times' type questions can be formed using the repetition adjective. The quantitative adjective words are also used to construct the 'true/false' type of questions, which have Blooms level 1, which is knowledge level. Thus, quantitative adjectives play an important role in question generation tasks where the questions are about numerical information. *Table1* provides a summary of the types of adjectives and the corresponding types of questions generated from them.

Verb: A verb is a word that completes the meaning of a sentence. *Table 2* shows the list of verbs that were used to generate different types of questions. The questions of type 'true/false' are formed after processing the auxiliary verbs in the sentence. Here, a set of specific verbs within the Marathi language that are employed to construct 'define' and 'who said to whom' type questions has been investigated. The study delves into the linguistic mechanisms underlying these verbs, aiming to uncover their syntactic and semantic properties that facilitate the generation of questions.

Table 1 Types of adjectives and the question words

Type of adjective	**Sub type of adjective**	**Questions**
Qualitative		How (कसा, कशी, कसे)
Quantitative	Counting	How many (किती), True/False questions
	Sequential	What is (कितवा, कितवी कितव्या, कितवे), True/False questions
	Repetition	How many (किती), True/False questions

Table 2 Verbs and the question words

Verbs	**Questions**	**Blooms level**
म्हणतात, म्हणतो(called as)	What is, define, fill in the blanks	Level 1
म्हणाले, म्हणालो, म्हणाली (said to)	Who said to whom Grammar-based questions	Level 1

Verbs	Questions	Blooms level
Auxiliary verbs	True/False	Level 1

Conjunctions: Two distinct categories of conjunctions exist: coordinating conjunctions and subordinating conjunctions. The ’true/false’ type of question can be formed by using subordinating conjunctions. Since each subordinate conjunction states a different type of relation between the main clause and the subordinate clause of the sentence, question generation varies with respect to the subordinate conjunction used in the sentence. The relationship between the main clause and the subordinate clause is applied in the question generation process. ‘Wh’ and ‘fill in the blanks’ questions can be formed using the relation expressed by the subordinate conjunctions in the sentence [40]. The discourse connectives have been used to generate the why, when, yes/no, ‘give an example’ type of questions [41]. In contrast to the system reported [14], proposed model excels in generating 'why' types of questions, showcasing an enhancement in question diversity and depth. *Table 3* presents a comprehensive overview of various subordinate conjunctions, question words, and their corresponding Bloom's cognitive levels. This categorization provides a structured understanding of the relationships between subordinate conjunctions, the associated question words, and the cognitive complexity levels, facilitating a nuanced analysis of question generation patterns.

Table 3 Types of subordinate conjunctions and the question words

Subordinate conjunctions	Context-based question word	Grammar-based questions	Blooms Level
म्हणजे (means, Mhaṇajē)	‘What is’, define, ‘Fill in the blanks’		Level 1
जर–तर (If then, Jara-tara) , जेव्हा–तेव्हा(when, jēvhā-tēvhā), जेव्हापासून–तेव्हापासून (since, jēvhāpāsūna-tēvhāpāsūna)	When, ‘What will happen if’	Join the following two sentences using proper conjunctive word	Level 1
कारण(because, Kāraṇa)	Why		Level 2
म्हणून (so, Mhaṇūna)	Why		Level 2
परिणामी (As a result, Pariṇāmī)	Why		Level 2

The tokenized sentence is then be passed to the POS tagger. At this stage, depending on the POS category of each word, it is negated to convert the sentence polarity from positive to negative and vice versa. The output of the ‘TF_Algorithm’ is a document containing multiple true or false questions.

Postposition lexicon: PP
Quantifier lexicon: Q
Negation of copula word: $\bar{C}$
Negation of the adjective word: $\bar{A}$
Negation of the adverb word: $\overline{Adv}$
Negation of the postposition word: $\overline{PP}$
Negation of the quantifier word: $\bar{Q}$

TF_Algorithm(S):

Input: Valid Marathi sentence (S), Verb lexicon, quantifiers lexicon, conjunctive words lexicon, postposition lexicon
S_{pos} : POS tags of S
Output: True-False questions (TF)
If S is a simple sentence then
 If S is copular sentence then:
  Replace copula (C) with $\bar{C}$
  TF ← TF + S
 For each word of S and POS of S_{pos}
  If POS is an adjective/adverb (*A/Adv*) then
   Replace the word with $\bar{A}/\overline{Adv}$
   TF ← TF + S
  EndIf
  If POS is a noun then:
   If the noun ends with a postposition (*PP*)
    Replace the *PP* with $\overline{PP}$
    TF ← TF + S
  EndIf
  If the word is Quantifier (Q) then:
   Replace Q with $\bar{Q}$
   TF ← TF + S
  EndIf
  If word is Negation then remove it.
   TF ← S ₋ Negation
  EndIf
Else if S is a complex sentence then:
 Extract the main clause and subordinate clause
TF_Algorithm (main clause)

When presented with a complex sentence containing subordinate conjunctions, the AQG engine is poised

to generate a spectrum of diverse questions, reflecting the nuanced relationships between the main and subordinate clauses. Rules are defined to generate the questions by using the input sentence, the conjunctive word lexicon, and the question word lexicon. Here, the complex sentence is divided into two or more simple sentences. The grammar-based question, asking to join these two sentences using an appropriate conjunctive word, is generated at this stage.
Example.

Input sentence: मृदा प्रदूषणामुळे जलप्रदूषणाचा धोका वाढतो कारण विषारी द्रव् मृदेमधून जवळच्या पाणीसाठ्यात किंवा पाझरून भूर्भ्वजलात प्रवेश करतात.
Transliteration: Mr̥dā pradūṣaṇāmuḷē jalapradūṣaṇācā dhōkā vāḍhatō kāraṇa viṣārī drav mr̥dēmadhūna javaḷacy āpāṇīsāṭhyāta kinvā pājharūna bhūrbhvajalāta pravēśa karatāta.
Translation: Soil pollution increases the risk of water pollution as toxic substances seep through the soil into nearby water bodies or seep into groundwater.

Output question: प्रश्न. मृदा प्रदूषणामुळे जलप्रदूषणाचा धोका का वाढतो?
Transliteration: Praśna. Mr̥dā pradūṣaṇāmuḷē jalapradūṣaṇācā dhōkā kā vāḍhatō?
Translation: Question. Why does soil pollution increase the risk of water pollution?

The above question can test the learner's understanding level. The above question can be asked in another way, as follows:
Example.

Output question: प्रश्न. कारणे द्या.
–मृदा प्रदूषणामुळे जलप्रदूषणाचा धोका वाढतो.
Transliteration: Praśna. Kāraṇē dyā.
-Mr̥dā pradūṣaṇāmuḷē jala pradūṣaṇācā dhōkā vāḍhatō.
Translation: Question. Give reasons.
-Soil pollution increases the risk of water pollution.

The AQG engine can generate 'when,' 'what will happen if,' 'complete the sentence,' and 'what is the result of …' types of questions at the POS level, expending subordinate conjunctions.
The following example demonstrates the question generated at the POS level from the quantitative adjectives.
Example.

Input sentence: शामचा शाळेत पाचवा क्रमांक आहे.
Transliteration: Śāmacā śāḷēta pācavā kramāṅk aāhē.
Translation: Sham is fifth in school.
Output question**:** प्रश्न. शामचा शाळेत कितवा क्रमांक आहे?
Transliteration: Śāmacā śāḷēta kitavā kramāṅk aāhē?
Translation: Question. What is Sham's number in school?

3.1.3 Stemming

In this NLP task, the root word is extracted by removing the suffixes. After creating a suffix lexicon for the Marathi language, the lexicon-based suffix-stripping stemming technique is used. The suffixes are called 'vibhakti pratyaya' in Marathi; they can be used to generate the questions. The grammar-based questions like 'find the root word' and 'correct the sentence by using appropriate postpositions' can be formed at this level of NLP pipeline. Asking a student to use an appropriate suffix requires them to apply their knowledge of word structure and language rules to modify words correctly, which aligns with the application level of Bloom's Taxonomy. These types of questions are exclusively posed in the language subject.

Now the POS-tagged sentence is passed to the stemmer. The stemmer extracts the suffix from the words in an input sentence. In Marathi, when a suffix is added to nouns in a sentence, it sometimes modify the form of the root word. For example, शाळेच्या (school's), here the root word is 'शाळा' (school). Custom rules are designed to extract words with postpositions. Especially in language subjects, the learner is questioned to write the answers to the questions mentioned below. These types of questions can be asked only in the grammar study of the Marathi language. The words with a suffix have been used by the AQG engine for question generation.
Example.

प्रश्न. खालील वाक्यातील अधोरेखित शब्दाचे मुळरूप व सामान्यरूप लिहा.
Transliteration: Praśna. Khālīla vākyātīl aadhōrēkhita śabdācē muḷarūpa va sāmān'yarūpa lihā.
Translation: Question. Write down the root word of the underlined word in the following sentence.
Input: मोहन <u>शाळेच्या</u> आवारात खेळत आहे.
Transliteration: Mōhana **<u>śāḷēcyā</u>** āvārāta khēḷat aāhē.

Translation: Mohan is playing in <u>school’s</u> premises.

3.1.4NER

The extraction of entity types is facilitated through the use of NER. The entity can be a person's name, an organization's name, a date, or a time etc. Named entities [27, 28, 42] and SRL [43] have been utilized to generate factoid questions. *Figure 2* shows the types of proper nouns and possible factoid questions that can be generated from each type of proper noun. These types of questions check the learner’s ability to recall facts or concepts. They are at Bloom's level 1.

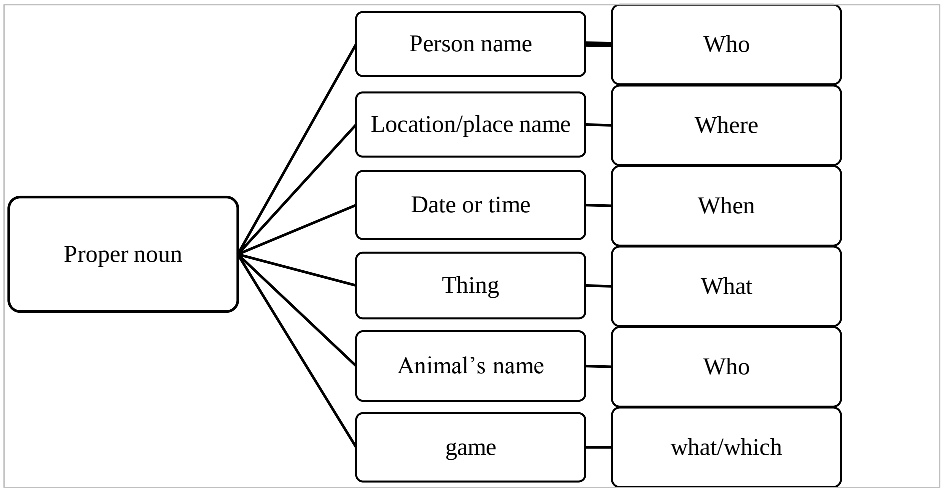

Figure 2 Nouns and factoid questions

After stemming, the preprocessed sentence is further passed on to the named entity recognizer. Here it identifies the type of entity, i.e., person name, organization name, date or time, etc. Depending on the type of entity, the question can be constructed using a set of custom rules. Here, the SpaCy NER tool [44] is used to extract the type of entity from the text. SpaCy is an open-source library providing NLP tools for the Python programming language (version 3.5.3).

Custom rule for entity translation: In the process of question generation, a custom rule has been implemented to enhance the precision of the generated output. This rule is designed to handle entities within the text and specifically addresses instances where the entity denotes a location.

Rule description: Identification of entity type: Text is subjected to an entity recognition process to identify the type of entity within the content.

Location entity detection**:** The rule focuses on detecting entities that represent locations within the text.

Entity replacement**:** If an entity is identified as denoting a location, it is replaced with the equivalent interrogative word that conveys the sense of location for example ‘कोठे’.

Example.

सीमा दिल्लीला जात आहे.

Transliteration: Sīmā dillīlā jāt aāhē.

Translation: Sima is going to Delhi.

प्रश्न. सीमा कोठे जात आहे?

Transliteration: Praśna. Sīmā kōṭhē jāta āhē?

Translation: Where Sima is going?

Person entity detection: The rule focuses on detecting entities that represent person entity within the text.

Entity replacement: If an entity is identified as denoting a person, it is replaced with the equivalent interrogative word that conveys the sense of person.

कोण दिल्लीला जात आहे?

Transliteration: Praśna. kōṇa Dillīlā jāta āhē?

Translation: Who is going to Delhi?

3.1.5Shallow parsing

Shallow parsing is the process of extracting phrases from a sentence, which means analyzing the sentence to identify the constituents like noun groups, adjective groups, verb groups, etc. However, it does not specify their internal structure or their role in the main sentence. It works on top of POS tagging. It uses POS tags as input and provides chunks (phrases) as output. Shallow parsing can break sentences into phrases that are more useful than individual words and yield meaningful results, which is important for question generation. A group of related words make up phrases, and there are three major categories.

1. NP

2. VP
3. Adjective phrase (ADJP)

For extracting phrases, NLTK's regular expression-based shallow parser is employed. The NP shallow parse tree is generated using the shallow parser, and the tree grafting approach is used to generate the questions from the tree. The AQG engine, employing distinct types of phrases, a postposition lexicon, and a question word lexicon, has the capability to generate questions types categorized as 'Fill in the blanks' and 'wh'. Rules play a crucial role in identifying question words based on the relationships expressed by postpositions in the provided sentence. This study delves into a linguistic analysis of postpositions and the resultant questions derived from them. During this stage of the NLP pipeline, questions categorized under Bloom’s levels 1 and 2 are generated. Level 1 questions are characterized as factoid questions, while Level 2 questions, falling under the comprehension level, includes inquiries such as ‘why’ or ‘give reasons.’ The extraction of specific VP is instrumental in identifying key terms within the sentence. These key terms subsequently serve as the foundation for constructing diverse question types, including ‘define,’ ‘what is,’ ‘who said to whom,’ ‘fill in the blanks,’ and more. In this study, a linguistic analysis of postpositions and the corresponding questions generated from them is thoroughly analyzed.

Example

Input sentence: पावसाळ्यात आकाशामध्ये जेव्हा वीज चमकते तेव्हा हवेतील नायट्रोजन आणि ऑक्सिजनचा संयोग होऊन नायट्रिक ऑक्साइड तयार होते०

Transliteration: Pāvasāḷyāt aākāśāmadhyē jēvhā vīja camaka tētēvhā havētīl a nāyaṭrōjan akinvāŏksij anacā sanyōga hō'ūna nāyaṭrika ŏksā'iḍa tayāra hōtē.

Translation: When lightning flashes in the sky during monsoons, nitrogen and oxygen in the air combines to form nitric oxide.

रिकाम्या जागी योग्य शब्द लिहा

पावसाळ्यात आकाशामध्ये जेव्हा वीज चमकते तेव्हा हवेतील ------------------ संयोग होऊन नायट्रिक ऑक्साइड तयार होते.

Transliteration: Pāvasāḷyāt aākāśāmadhyē jēvhā vīja camakatē tēvhā havētīla------------------ sanyōgahō'ū nanāyaṭrika ŏksā'iḍa tayāra hōtē.

Translation: Fill in the blanks

During monsoons, when lightning flashes in the sky, nitric oxide is formed by the combination of ------------------ in the air.

The NP ‘नायट्रोजन आणि ऑक्सिजनचा’ (nitrogen and oxygen) is utilized by the AQG engine for AQG.

3.1.6Dependency parsing

It involves examining a sentence's grammatical structure to determine the relationships between its words. Dependency focuses on the relationships between the words. The dependency relations are further used for AQG. Here, the study has been carried out to figure out the various types of question types from the dependency relations. For Marathi, the ‘vibhakti pratyay’ plays an important role along with the dependency relations to generate the precise questions. The questions formed at this stage have Bloom’s level 1. These types of questions test the learner’s ability to recall facts. Afzal et al. have generated multiple-choice-based questions using dependency-based patterns [45]. The dependency relations information has been used to improve the generated questions by making them more precise and relevant to the topic [17]. Here, the STANZA dependency parser is used to extract the dependency relationship between the words of a sentence. Rules are crafted to generate questions by leveraging dependency relations, morphological information from the dependent word, and the question word lexicon.

The output of the DepRel_Based_QG_Algorithm is a document containing various types of questions.

The questions generated from the dependency relations ‘oblique nominal (obl)’ and ‘nominal subject (nsubj)’ are shown in *Figure 3* and *Figure 4*.

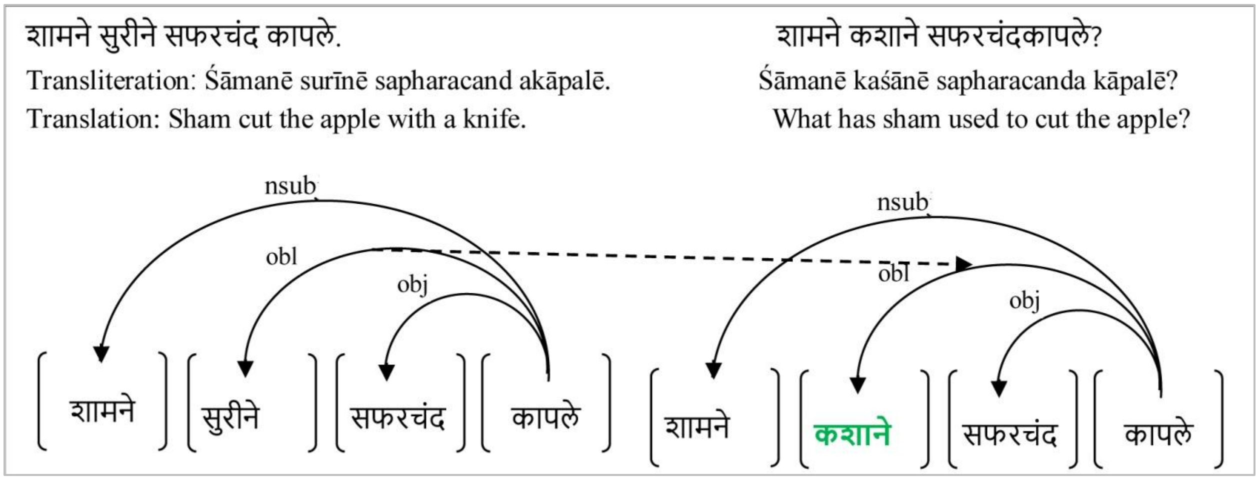

Figure 3 Question generation using the dependency relation 'obl'

DepRel_Based_QG_Algorithm:

Input: Valid Marathi sentence(S), Dependency Relations list, Suffix (Vibhakti) lexicon, Question word lexicon,

Output: set of Questions

1. Start
2. Parse the sentence (S) and create a dependency relation list (DpRlList)
3. Repeat step 3 to step 5 for each dependency relation(DpRl) from the DpRlList till DpRl ≠ NULL
4. Perform morphological analysis on dependent word to extract the morphologically inflected question word
5. Generate the question by utilizing the custom rules, morphological analysis, and the question word lexicon
6. End

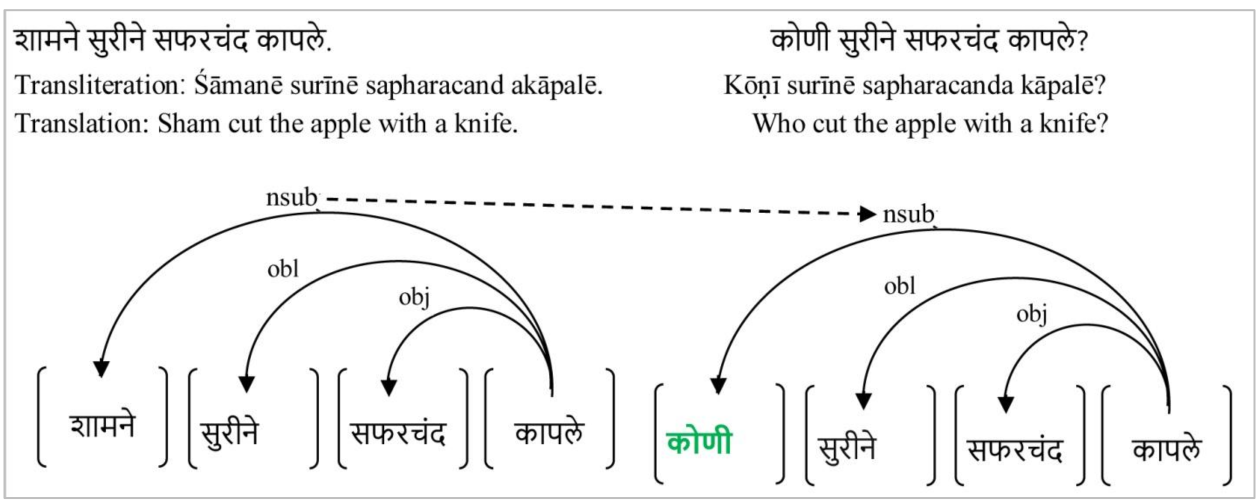

Figure 4 Question generation using the dependency relation 'nsubj'

The 'nsubj' dependency relation in dependency parsing represents the syntactic relationship between a verb and its subject. It identifies the noun or pronoun that performs the action or state described by the verb. A question asking for information about who has performed the action can be generated from the dependency relation 'nsubj'. The 'obl' dependency relation in dependency parsing identifies a noun or NP that provides additional information about the action or relationship expressed by the verb. The 'obl' can indicate how the action is performed and the time and duration of the action.

The possibility of multiple interpretations within a sentence is effectively addressed during the preliminary stages of sentence processing. These stages include POS tagging, NER, chunking, and parsing, which analyzes the syntactic structure of the sentence. Through these processes, the intended meaning of the sentence becomes clearer as ambiguous elements are disambiguated based on their grammatical roles and relationships within the sentence structure. In the context of question generation, the presence of linguistic ambiguity stemming from the same word having different meanings in different contexts may not significantly influence the question word. This is because question generation primarily relies on syntactic and semantic structures rather than the specific meanings of individual words.

For instance, consider the word 'हार, ' which can refer to a necklace or a loss. In the sentence, 'उर्मिलाने हार विकत घेतला' (Urmila bought a necklace). Here, the word 'हार' refers to a 'necklace.' In contrast, in the sentence 'उर्मिलाला खेळात हार पत्करावी लागली,' (Urmila had to lose the game), the word 'हार' refers to 'loss.' Despite the different meanings of 'हार' in each sentence, the syntactic structure and overall context provide enough information for question generation without requiring a deep understanding of the word's specific meaning. Therefore, during the pre-processing stages like POS tagging, chunking, and parsing, the system can effectively handle such ambiguity by focusing on the structural and contextual cues to generate questions accurately.

उर्मिलाने काय विकत घेतला?

Translation: What did Urmila purchase?

उर्मिलाला खेळात काय पत्करावी लागली?

Translation: What did Urmila have to face in the game?

3.2Post-processing

Here, the questions generated by the AQG engine exhibit variability. This variation can involve changing the position of the question word or, in certain instances, altering the question word itself. In Marathi, the question word 'कशामुळे' (why) can be replaced by the question word 'का' (why). Both questions have the same meaning, except for the question word.

3.3Mapping of questions with Blooms taxonomy

Bloom's taxonomy is an essential element in the teaching and learning process to assess the learning ability of the learner. The taxonomy was proposed in 1956 by Benjamin Bloom [46], an educational psychologist at the University of Chicago. Bloom's Taxonomy is a framework for classifying educational goals and objectives. The taxonomy consists of six hierarchical levels, which are knowledge, comprehension, application, analysis, synthesis, and evaluation. The main objective of asking a question to a learner (student) is to evaluate his or her learning ability, knowledge, and understanding. Assessing the learner (student) and self-assessment can be carried out by questioning. Hence questions play an important role for assessment and evaluation. If the questions are framed on the basis of Bloom's taxonomy, then the student is assessed at different levels. Like, does the learner (student) have knowledge of the basic facts related to the topic learned? Does he or she understand the concepts? Can he or she use knowledge in new circumstances? This approach can help the educator identify areas where the student or learner needs help for improvement. The six levels of questions defined by Bloom's taxonomy are summarized here.

Knowledge: Recalling is the primary educational objective in the curriculum. One can test whether the student can recall previously learned information. Factoid questions that require fact-based answers falls under this category.

Comprehension: This level refers to the learner's understanding of facts. A learner can answer these types of questions if he has understood the concepts he has learned. The teacher or an evaluator can judge the understanding level of the student by asking him level 2 questions, which are 'comprehension' questions. 'Give scientific reason', 'why', 'give reasons', 'why this happens'—these types of questions are of Bloom's level 2.

Application: Level 3 is the application of the learner's knowledge in a new situation. Whether a learner can apply or use his previous knowledge in a new situation, these types of questions not only require an understanding of the concepts but also their use in a new scenario.

Analysis: Break down objects or ideas into simpler parts and find evidence to support generalizations.

The proposed AQG model can generate ten distinct context-based question types and nine distinct grammar-based question types. Karamustafaoglu et al. have classified biology questions as per Bloom's taxonomy by Biology teachers [47]. Various approaches, including naïve bayes, laplace smoothing [48], artificial neural networks [49], CNN [50], pre-trained language model [51], and rule-base [52] have been employed for question classification.

The various types of context-based and grammar-based questions generated through the AQG model are depicted in *Figures 5* and *6*.

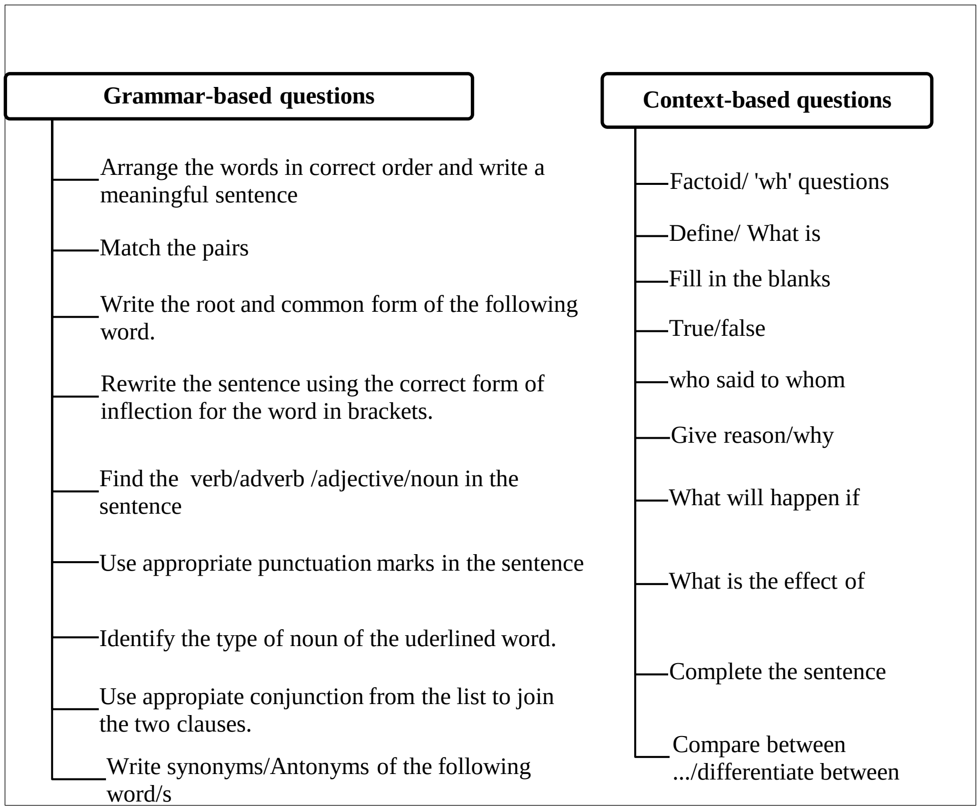

Figure 5 Types of questions generated by AQG model

The final model, where mapping of the generated questions with Bloom's taxonomy is carried out, receives the questions generated by the AQG engine and the questions processed through the post-processing model. Bloom's levels range from "simple to complex" [46]. All mapping is done using a rule-based approach. Custom rules are designed that checks which question falls under respective Bloom's level, depending on the question word. Omar et al. have proposed a rule-based automated analysis of the exam questions that utilizes keywords found in the question [52]. The factoid questions have fact-based answers; a learner has to recall the facts to answer the question. These questions are classified as Bloom's level 1 knowledge. Following *Figure 7* shows the mapping between NLP pipeline levels and the Bloom's taxonomy model.

Example

शामला शाळेत बक्षीस मिळाले.

Transliteration: Śāmalā śāḷēta bakṣīsa miḷālē.
Translation: Sham got a prize in school.

Multiple fact-based questions can be extracted from this sentence which is elaborated below. These questions can check the learner's recall of facts related to the given sentence.

Questions: कोणाला शाळेत बक्षीस मिळाले?
Transliteration: Kōṇālā śāḷēta bakṣīsa miḷālē?
Translation: Who won a prize at school?

शाळेत बक्षीस कोणाला मिळाले?
Transliteration: Śāḷēta bakṣīsa kōṇālā miḷālē?
Translation: Who won a prize at school?

शामला कोठे बक्षीस मिळाले?
Transliteration: Śāmalā kōṭhē bakṣīsa miḷālē?
Translation: Where did Sham get the prize?

Marathi language question words that are related to Bloom's level 1, i.e., knowledge level, are listed below in *Table 4*. Answers to the questions with these question words are specific facts only.

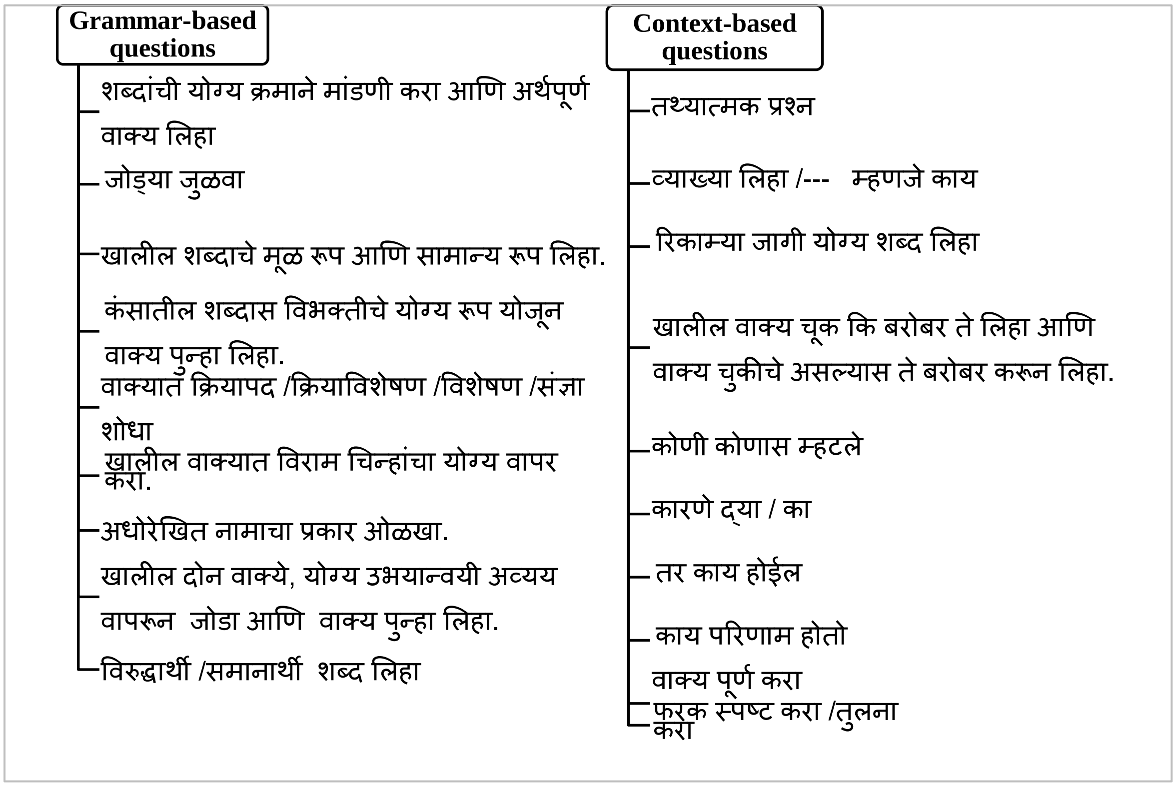

Figure 6 Types of questions generated by AQG model

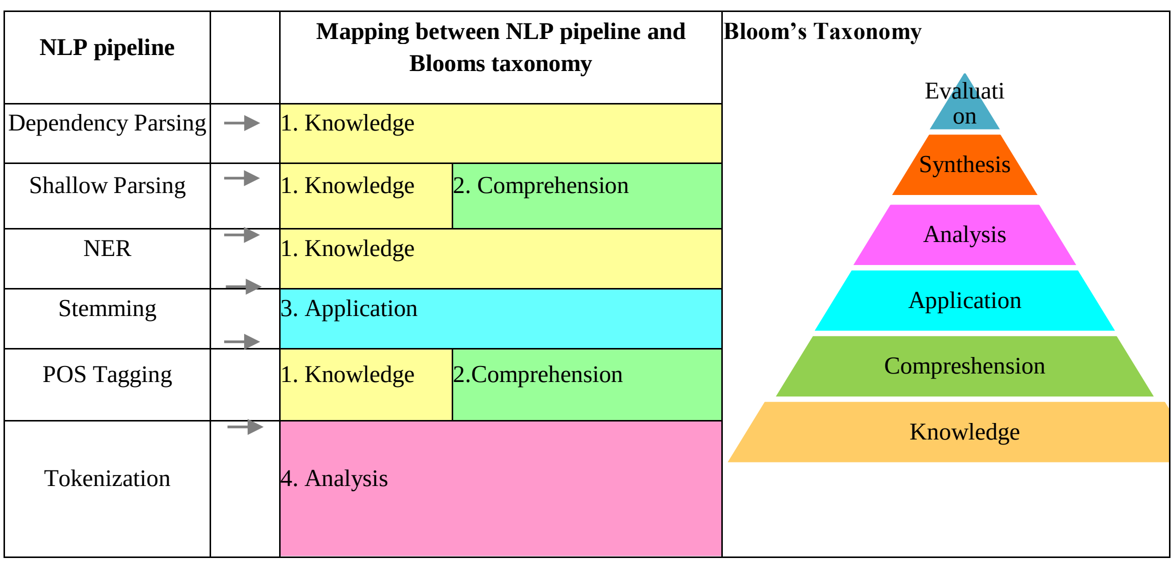

Figure 7 mapping of questions generated at each level of NLP pipeline and Bloom's taxonomy levels

Marathi language question words that are related to Bloom's level 1, i.e., knowledge level, are listed below in *Table 4*. Answers to the questions with these question words are specific facts only.

Table 4 Factoid question words

कोणाशी Kōṇāśī (with whom)	कोणास Kōṇāsa (to whom)	कशाने Kaśānē (with which)
कोणाला Kōṇālā (to whom)	कोणी Kōṇī (who)	कितवा Kitavā (how much)
कितव्या Kitavyā (how many)	किती Kitī (how many)	कशाशी Kaśāśī (with what)
कोणाशी Kōṇāśī (with whom)	कोणाभोवती	कोणासमोर (in front of whom)
कोठे Kaśāmuḷē (where)	कशासाठी Kaśāsāṭhī (for what)	कशाकरिता Kaśākaritā (for what)
कशानिमित्त kaśānimitta (for what reason)	कोठेKōṭhē (where)	कशामध्ये Kaśāmadhyē (in what)
कोणाजवळ kōṇājavaḷa (who)	कशापाशी Kaśāpāśī (why)	कोणासमक्ष Kōṇāsamakṣa (in front of whom)
कशाऐवजी kaśā'aivajī (instead of what)	कोणाविषयी Kōṇāviṣayī (about whom)	कोणाप्रमाणे Kōṇāpramāṇē (like who)
कोणाशेजारी kōṇāśējārī (who)	कोणाबद्दल Kōṇābaddal (about whom)	कोणामुळे Kōṇāmuḷē (due to whom)
कोणाकडून kōṇākaḍūna (from whom)	कोणाशिवाय Kōṇāśivāya (without whom)	कोणाखेरीज Kōṇākhērīja (apart from whom)
कोणाबरोबर kōṇābarōbara (with whom)	कोणासोबत Kōṇāsōbata (with whom)	कोणाव्यातिरिक्त Kōṇāvyātirikta (apart from whom)
कोणासह kōṇāsaha (with whom)	कोणासहित Kōṇāsahita (with whom)	कोणापेक्षा Kōṇāpēkṣā (than whom)
कशावरून Kaśyāvarūna (from what)	कशातून Kaśātūna(from what)	कशावर Kaśāvara(on what)
कोणाविरुद्ध Kōṇāvirud'dha (against whom)	कश्यामधून Kaśyāmadhūna (from what)	कोणाहून Kōṇāhūna (from whom)

The grammar-based questions to find synonyms and antonyms require the learners understanding of the words; therefore, they have Bloom’s level 2, i.e., comprehension.

A ‘why’ type of question corresponds to Bloom's level 4, which is analysis [46]. In Bloom's Taxonomy, ‘analysis’ is a higher-order thinking skill that involves breaking down composite information into smaller fragments and examining the relationships between them. ‘Why’ questions often require this type of analysis, as they typically require the student or learner to understand the underlying reasons or causes behind a particular phenomenon or situation. Hence, ‘Why’ questions are classified as analysis-level questions.

The learner or student needs to know the meaning of the given words in addition to understanding their use in a sentence and performing the analysis on the order of these words to form a meaningful sentence in order to answer the question, ‘Arrange the words in the correct order and create a meaningful sentence.’ This type of question also has Bloom’s level 4, i.e., analysis level. The proposed model's entire execution is shown in the example that follows.

Input Sentence: पावसाळ्यात आकाशामध्ये जेव्हा वीज चमकते तेव्हा हवेतील नायट्रोजन आणि ऑक्सिजनचा संयोग होऊन नायट्रिक ऑक्साइड तयार होते.

Transliteration: Pāvasāḷyāt aākāśāmadhyē jēvhā vīja camakatē tēvhā havētīla nāyaṭrōj anaāṇi ŏksijanacā sanyōga hō'ūna nāyaṭrika ŏksā'iḍa tayāra hōtē.

Translation: During the rainy season, when lightning flashes in the sky, nitrogen and oxygen in the air combine to form nitric oxide.

The above sentence is processed through a tokenizer, POS tagger, stemmer, NER, punctuation stripper, shallow parser, and dependency parser. The step-by-step output of each preprocessing module and the questions generated at the respective level are shown below in *Table 5*. This study aims to evaluate the AQG engine’s effectiveness in accurately generating questions from the text. A combination of automated and human-based metrics (adequacy, fluency, difficulty level, impact on overall understanding, answerability, etc.) is employed to assess the quality of the generated questions.

Table 5 AQG model's output demonstration

Output of tokenizer:	[**पावसाळ्यात**] [**आकाशामध्ये**] [**जेव्हा**] [**वीज**] [**चमकते**] [**तेव्हा**] [**हवेतील**] [**नायट्रोजन**] [**आणि**] [**ऑक्सिजनचा**] [**संयोग**] [**होऊन**] [**नायट्रिक**] [**ऑक्साइड**] [**तयार**] [**होते**] [.]
Question generated:	Arrange the words in correct order and write a meaningful sentence **होते आकाशा मध्ये जेव्हा वीज चमकते हवेतील तेव्हा नायट्रोजन आणि ऑक्सिजनचा संयोग होऊन नायट्रिक ऑक्साइड तयार पावसाळ्यात** Translation: Occurs in the sky when lightning flashes in the air, nitrogen and oxygen combine to form nitric oxide during rain
Output of POS tagger	**पावसाळ्यात**_NN **आकाशामध्ये**_NN **जेव्हा**_PRP **वीज**_NN **चमकते**_VM **तेव्हा**_CC **हवेतील**_NN **नायट्रोजन**_NN **आणि**_CC **ऑक्सिजनचा**_NNP **संयोग**_NN **होऊन**_VM **नायट्रिक**_NN **ऑक्साइड**_NNP **तयार**_JJ **होते**._VM ._SYM
Question generated	1)खालील वाक्य पूर्ण करा. **Translation**: Complete the sentence: पावसाळ्यात आकाशामध्ये जेव्हा वीज चमकते तेव्हा ----------------------------------- **Translation**: During the rainy season, when lightning flashes in the sky------------------------- 2)खालील वाक्य चूक कि बरोबर ते लिहा आणि वाक्य चुकीचे असल्यास ते बरोबर करून लिहा. **Translation**: Write the following sentences true or false and correct them if the sentences are false. पावसाळ्यात आकाशामध्ये जेव्हा वीज चमकते तेव्हा हवेतील नायट्रोजन किंवा ऑक्सिजनचा संयोग होऊन नायट्रिक ऑक्साइड तयार होते. **Translation**: When lightning flashes in the sky during monsoons, nitrogen or oxygen in the air combines to form nitric oxide. 3)प्रश्न. हवेतील नायट्रोजन आणि ऑक्सिजनचा संयोग होऊन नायट्रिकऑक्साइड केव्हा तयार होते? **Translation**: Question. When nitrogen and oxygen in the air combines to form nitric oxide? Questions based on grammar 4)प्रश्न. विरुद्धार्थी शब्द लिहा -तयार **Translation**: Write the antonym-ready प्रश्न. समानार्थी शब्द लिहा -तयार **Translation**: Write the synonym-ready 5) प्रश्न. जोड्या जुळवा **Translation**: Match the pairs नाम होते Noun was सर्वनाम तयार Pronoun ready विशेषण नायट्रिक Adjective Nitric क्रियापद तेव्हा Verb When 6) खालील दोन वाक्ये, योग्य उभयान्वयी अव्यय वापरून जोडा आणि वाक्य पुन्हा लिहा. **Translation**: Use appropriate conjunction from the list to join the two clauses. 1.ऑक्सिजनचा संयोग होऊन नायट्रिक ऑक्साइड तयारहोते 2. पावसाळ्यात आकाशामध्ये वीज चमकते हवेतील नायट्रोजन **Translation**: 1. Nitric oxide is formed by combining with oxygen 2. Lightning in the sky during rains Nitrogen in the air
Output of Punctuation stripper	This sentence has only one punctuation symbol. Therefore, question cannot be generated at this stage.
Output of Stemmer	पावसाळ्यात->पावसाळ्या ऑक्सिजनचा->ऑक्सिजन
Question generated	Questions based on grammar प्रश्न. खालील शब्दाचे मुळरूप व सामान्यरूप लिहा. पावसाळ्यात ऑक्सिजनचा **Translation**: Question. Write the root and common form of the following word. of oxygen during monsoon

	प्रश्न. कंसातील शब्दास विभक्तीचे योग्य रूप योजून वाक्य पुन्हा लिहा. **Translation**: Rewrite the sentence using the correct form of inflection for the word in brackets. (पावसाळ्या) आकाशामध्ये जेव्हा वीज चमकते तेव्हा हवेतील नायट्रोजन आणि (ऑक्सिजन) संयोग होऊन नायट्रिक ऑक्साइड तयार होते. **Translation**: When lightning flashes in a rainy sky, nitrogen and oxygen in the air combine to form nitric oxide.
Output of NER:	ऑक्सिजनचा MISC
Question generated	**पावसाळ्यात आकाशामध्ये जेव्हा वीज चमकते तेव्हा हवेतील नायट्रोजन आणि --------- संयोग होऊन नायट्रिक ऑक्साइड तयार होते**.
Output of Shallow parser:	NP : आकाशामध्ये NP : नायट्रोजन आणि ऑक्सिजनचा
Question generated	4) प्रश्न. पावसाळ्यात कोठे वीज चमकते तेव्हा हवेतील नायट्रोजन आणि ऑक्सिजनचा संयोग होऊन नायट्रिक ऑक्साइड तयार होते? **Translation**: Where during monsoons then lightning flashes, nitrogen and oxygen in the air combine to form nitric oxide? 5) प्रश्न. पावसाळ्यात कशामध्ये वीज चमकते तेव्हा हवेतील नायट्रोजन आणि ऑक्सिजनचा संयोग होऊन नायट्रिक ऑकसाईड तयार होते? **Translation**: Question. In which during rains, then lightning strikes, nitrogen and oxygen in the air combine to form nitric oxide? 6) रिकाम्या जागी योग्य शब्द लिहा. 1) पावसाळ्यात --------- चमकते तेव्हा हवेतील नायट्रोजनआणिऑक्सिजनचा संयोग होऊन नायट्रिक ऑक्साइड तयार होते. **Translation**: During monsoon, when ------- glows, nitrogen and oxygen in the air combine to form nitric oxide. 2) पावसाळ्यात आकाशामध्ये जेव्हा वीज चमकते तेव्हा हवेतील ------------------ संयोग होऊन नायट्रिक ऑक्साइड तयार होते. **Translation**: During monsoons, when lightning flashes in the sky, nitric oxide is formed by the combination of ------------------ in the air.
Output of Dependency parser:	Srno dependent dependency relation head 1 पावसाळ्यात obl 5 2 आकाशामध्ये advmod 5 3 जेव्हा advmod 5 4 वीज obl 5 5 चमकते root 0 6 तेव्हा mark 5 7 हवेतील advmod 5 8 नायट्रोजन nsubj 12 Srno dependent dependency relation head 9 आणि cc 10 10 ऑक्सिजनचा conj8 11 संयोग obj12 12 होऊन conj5 13 नायट्रिक amod14 14 ऑक्साइड nsubj16 15 तयार compound:lvc 16 16 होते conj5
Question generated	रिकाम्या जागी योग्य शब्द लिहा. **पावसाळ्यात आकाशामध्ये जेव्हा वीज चमकते तेव्हा हवेतील नायट्रोजन आणि ऑक्सिजनचा संयोग होऊन** ---------------- **ऑक्साइड तयार होते**. **Translation**: lightning flashes in the sky during monsoons, nitrogen and oxygen in the air combine to form ------- oxides.

4. Results

The AQG model is tested for 213 randomly selected Marathi sentences from the sixth standard text book of 'SamanyaVidnyan' (Science,) published by the Maharashtra State Board, Maharashtra, India. The selected science textbook, being a specialized academic publication, adheres to a standardized and formal linguistic framework. The language in scientific texts is characterized by clarity, precision, and a formal tone. These attributes are central to the analysis. The 213 sentences selected for analysis were randomly sampled from the entire spectrum of content within the science textbook. This ensures a diverse representation of syntactic structures relevant to scientific discourse.

From these 213 sentences, a total of 2154 questions were generated, out of which 800 are context-based and 1354 are grammar-based questions. The following table shows the number of questions generated by the proposed AQG model. To evaluate the performance of the system as a whole and the accuracy of the generated questions, two different strategies have been employed. In the first strategy, context-based questions are evaluated with BLEU; in the second strategy, manual evaluation is performed by three evaluators. *Table 6* offers a concise summary detailing the distribution and types of questions generated. It provides a quantitative breakdown, specifying the number of questions corresponding to each distinct type. This tabulated presentation allows for a quick and insightful understanding of the variety and abundance of questions generated by the AQG engine.

Table 6 Summary of the generated questions

Context-based Questions	
Type of questions	**No. of questions**
Fill in the blanks	269
Factoid Questions	266
Define	29
Complete the sentence	52
What will happen if	24
Why (give reason)	57
True or False	93
Total context-based questions	800
Grammar-based questions	
Punctuation questions	42
Arrange the words in correct order and write a meaningful sentence	213
Join the following two sentences using proper conjunctive word	53
Match the pairs	106
Write synonyms and antonyms	248
Rewrite the sentence using the correct form of inflection for the word in brackets	274
Write the root form and common form of the following word	274
Identify the verb/adverb/adjective in the following sentence	144
Total grammar-based questions	1354
Total questions	**2154**

It's notable that the AQG engine excels at managing complex sentences. The AQG engine consistently generates questions accurately from these structures, showcasing its proficiency in capturing nuanced linguistic elements. The successful rendering of complex sentence structures underscores the strength and effectiveness of the AQG engine in handling diverse linguistic challenges.

4.1Evaluation with BLEU

For evaluation and benchmarking, questions have been compiled through evaluators. Here, the evaluators were provided with the sentences, and they were expected to derive questions from those sentences. The questions derived by the manual evaluators were compared with the output of the proposed AQG model to evaluate its accuracy and efficiency. BLEU is a metric for comparing a human translation of the text with machine translations [53]. Generally, the BLEU score is used to rate machine translations; however, it can be used for the AQG model to rate the syntactical structure of the generated factoids. *Table 7* illustrates two sample questions generated by the AQG model, two reference questions formulated by the evaluator, and their BLEU score. The evaluation of the AQG model's performance was conducted using a dataset consisting of 574 'wh' questions. For each question, the AQG model-generated question was compared against the human-formulated question extracted from the same set of sentences. The BLEU score was employed as the evaluation metric, measuring the degree of overlap between the AQG model-generated and human-created questions. Notably, the obtained BLEU score from this comparative analysis was recorded as 90.37, indicating a high level of linguistic similarity between the two question sets. The *Table 8* presents the BLEU scores for 1-gram, 2-gram, 3-gram, and 4-gram evaluations from three evaluators. The graph in *Figure 8* shows a comparison of the BLEU scores among the evaluators' assessments of 'wh' questions.

Table 7 Sample sentence BLEU score evaluation

Reference question	Generated question	BLEU score
प्रश्न. द्विनाम पद्धतीचा उपयोग का केला जातो ? **Translation**: Question. Why is binomial method used? **Transliteration**: Praśna. Dvināmapad' dhatīcāupayōgakā kēlājātō?	प्रश्न. द्विनाम पद्धतीचा अवलंब का केला जातो ? **Translation**: Question. Why is binomial method adopted? **Transliteration**: Praśna. Dvināmapad'dhatīcāavalambakākēlājātō?	59.694918
'खंडग्रास' सूर्यग्रहण होते तेव्हा सूर्यबिंब कोणामुळे पूर्णपणे झाकले जात नाही? **Translation**: When 'continental' solar eclipse occurs, the Sun is not completely covered by whom? **Transliteration**: 'Khaṇḍagrāsa' sūryagrahaṇa hōtē tēvhā sūrya bimba kōṇāmuḷē pūrṇapaṇē jhākalē jāta nāhī?	जेव्हा सूर्य बिंब कोणामुळे पूर्णपणे झाकले जात नाही तेव्हा 'खंडग्रास' सूर्यग्रहण होते? **Translation**: A 'continental' solar eclipse occurs when the Sun is not completely covered by whom? **Transliteration**:Jēvhā sūryabimba kōṇāmuḷē pūrṇapaṇē jhākalē jāta nāhī tēvhā 'khaṇḍagrāsa' sūryagrahaṇa hōtē?	62.014240

Table 8 BLEU evaluation score of AQG model for 'wh' questions

Evaluator	1-gram	2-gram	3-gram	4-gram
Evaluator 1	94.04	92.84	91.85	90.87
Evaluator2	93.86	92.61	91.40	90.31
Evaluator3	93.72	92.39	91.10	89.94
Average BLEU Score	93.87	92.61	91.45	90.37

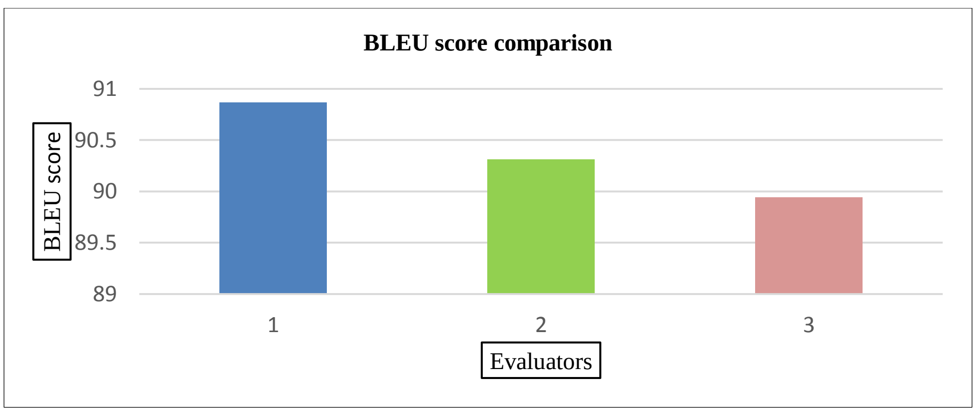

Figure 8 BLEU score comparison

The scores are relatively close to each other, indicating a degree of agreement among the evaluators in terms of their assessment of the generated question's quality. Variance of the BLEU Score for the generated set of questions is expressed in Equation 1.

$$Variance\ \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \qquad (1)$$

Where,
N : number of questions in the dataset
x_i : BLEU Score for the i^{th} question in the dataset.
$\bar{x}$: mean BLEU Score for all generated questions in the dataset.
The calculated variance of 0.04 in combination with the upper confidence interval of 0.94 and the lower

confidence interval of 0.91 provides valuable insights into the consistency and precision of the AQG process. The relatively small variance indicates a moderate degree of variability in the BLEU scores, while the narrow confidence interval suggests a high level of confidence in the accuracy of the AQG. This precision is crucial in assessing the reliability of the AQG model, and the confidence intervals provide a robust range within which the true question generation quality is likely to fall. The analysis implies a stable and dependable AQG performance, encouraging confidence in the overall quality of the generated questions.

4.2Manual evaluation

The evaluation of the proposed methodology in this research work is performed by human evaluators. The evaluators are primary school teachers with preliminary knowledge of different types of questions. The questions of type 'wh' questions, 'define', 'complete the sentence', 'what will happen if', and 'give reasons' are rated using below mentioned scoring scale with parameters viz. fluency, adequacy and answerability.

Fluency: Read the generated question and assign 0 to 1 points to the grammatical correctness of the question. If the question is correct, then assign 1; if it is partially correct, then assign 0.5; and assign 0 if it is incorrect.
Adequacy: Here also, the evaluator has to assign 0 to 1 points depending on the meaning of the question. Whether the generated question is meaningful (1), partially meaningful (0.5), or meaningless (1).
Answerability: The assessor must determine whether or not the question's answer fits within the context of the input sentence. He or she must assign a score of 0 for the answer being out of context, 0.5 for the answer being partially in context, and 1 for the answer being entirely in the context of the input statement.

The 'fill in the blanks' questions have the same sentence structure as the input sentence. Therefore, to evaluate the 'fill in the blanks' questions, the following three metrics are used: impact on overall understanding, completeness, and specificity.
Impact on overall understanding*:* Some questions may have a greater impact on overall comprehension or understanding of a subject. These questions may focus on critical concepts or relationships that significantly contribute to the overall learning outcomes. Such questions can be seen as more important. The evaluator has to assign 0 for the question having no impact on overall understanding of the topic, 0.5 for the question having a partial impact on overall understanding of the topic, and 1 for the question having a greater impact on overall understanding of the topic.
Completeness: The question should provide all the necessary information required to fill the blank accurately. It should not suppress important details that are essential to arriving at the right answer. If the information given in question is incomplete, then the evaluator has to assign 0. If it is partial, then assign 0.5, and if the information is complete to detect the correct answer, then assign 1.
Specificity: The question should be precise and focused, directing the learner towards the desired answer. It should avoid broad or general statements that could lead to multiple possible answers. If the generated question is precise and focused, then the evaluator has to assign 1; if the question is not precise and results in multiple answers, then assign 0, and assign 0.5 for the questions that are partially focused. Accuracy of the generated questions is expressed in Equation 2.

$$\text{Accuracy}=\frac{\sum_{i=1}^{N} w_i . q_i}{N+Q_0} \qquad (2)$$

$$W_i = [W_{1\,,} W_2, W_3]$$

$$W_i = [0, 0.5, 1]$$

$$Q_i = [Q_{0\,,}\ Q_1,\ Q_2]$$

Q_0 : Number of questions having score 0
$Q_{1,}$: Number of questions having score 0.5
Q_2: Number of questions having score 1
N: Total number of questions generated by the AQG engine

Here Q_0 is the number of questions having score 0 are used to penalize the AQG engine for incorrect accuracy.
The accuracy of the generated questions with respect to fluency is presented as follows:

$$=\frac{[(0.5\times Q_1)+(1\times Q_2)}{(N+Q_0)}$$

The results of the AQG engine for the factoid questions for fluency, adequacy, and answerability by each evaluator are shown in *Table 9*, and the evaluation of the 'fill in the blanks' questions on the basis of 'impact on overall understanding, completeness, and specificity is shown in *Table 10*. The evaluation of grammar-based questions based on accuracy and difficulty level is presented in *Table 11*.

Table 9 Results of AQG model for 'wh' questions

'Wh' Questions	Fluency (%)	Adequacy (%)	Answerability (%)
Evaluator 1	74	74	83
Evaluator 2	71	72	80
Evaluator 3	74	75	79
Accuracy	73	73.66	80.66

Table 10 Results of AQG model for 'fill in the blanks' questions

Fill in the blanks Questions	Impact on overall understanding (%)	Completeness (%)	Specificity (%)
Evaluator 1	99	98	92
Evaluator 2	98	97	92
Evaluator 3	97	97	91
Accuracy	98	97.33	91.66

Table 11 Results of AQG model for true/false questions

True/false Questions	***Impact on overall understanding*** **(%)**	Completeness (%)	Specificity (%)
Evaluator 1	67	72	67
Evaluator 2	68	71	66
Evaluator 3	67	72	66
Accuracy	67.33	71.66	66.33

The grammar-based questions are evaluated on the basis of accuracy and the difficulty level metric.
Accuracy of question: If the provided options for the grammar-based question are completely correct, then the evaluator has to assign 1, partially correct, assign 0.5, or incorrect, then assign 0. ***Difficulty level of the question***: In order to evaluate the grammar-based questions for difficulty level, there is a need to assign the class level first. Means for which grade students the question's difficulty level is checked. So here, the evaluators are going to evaluate the difficulty level of the questions for the sixth-grade students. It is 1 for very difficult questions, 0.5 for medium-difficult questions, and 0 for easy questions. *Table 12* shows the results of the evaluations of the grammar-based questions.

Table 12 Results of AQG model for grammar-based questions

Grammar based questions	Accuracy of the questions (%)	Difficulty level (%)
Evaluator 1	91	90
Evaluator 2	92	93
Evaluator 3	91.25	91.40
Accuracy	91.41	91.46

Grammar-based questions were generated with an impressive accuracy of 91 percent. This notable level of accuracy proves to the effectiveness of the approach used in formulating questions that strictly adhere to grammatical rules and structures. The 90 percent accuracy rate reflects the high quality and precision of the generated questions, rendering them suitable for a wide range of educational and evaluative purposes. The following *Figure 9* shows the comparative results of the context-based descriptive questions. The average accuracy of all three evaluators in assessing the 'wh' questions is 76%. This indicates a reasonably high level of accuracy in their evaluations, suggesting that they largely agree on the fluency, adequacy, and answerability of the questions under consideration.

When comparing our work, a notable distinction emerges. Das et. al. [33] have primarily focused on generating 'wh' questions using NP and VP. In contrast, the proposed approach extends beyond 'wh' questions; the proposed model successfully generated 'wh' questions, 'true or false' questions, and grammar-based questions. Wijanarko et al. [26] constructed questions by combining Bloom's verbs with key phrases while, in the current work, a different approach is followed by initially generating questions, and subsequently, through the utilization of predefined rules, assigned a Bloom's level to each question based on the question word. This method provides a unique perspective on question generation and Bloom's taxonomy integration.

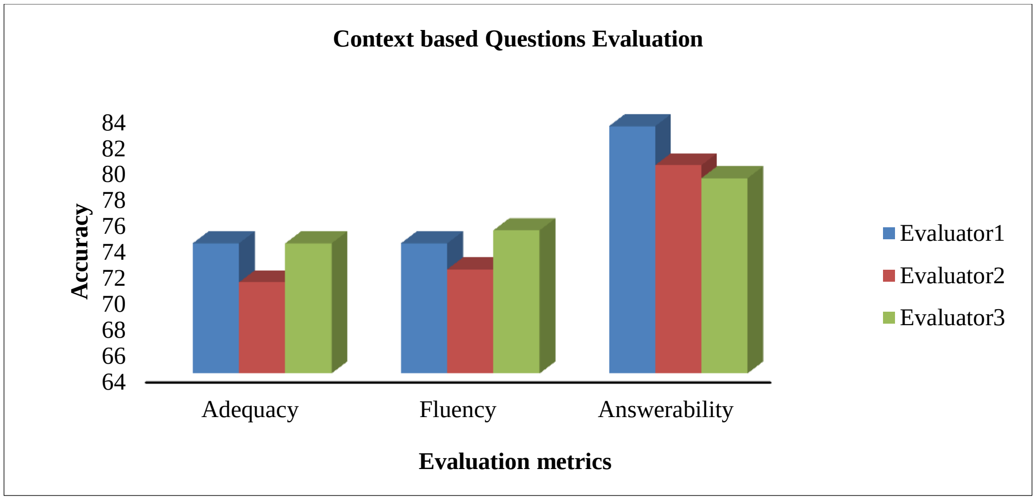

Figure 9 Comparative evaluation results of the context-based questions

5. Discussion

Moving on to the interpretation of our findings, the proposed novel question generation approach demonstrates a distinctive feature that sets it apart from conventional methods. In this study, a pioneering approach to AQG that traverses the entire NLP pipeline is presented, covering multiple levels of linguistic analysis. Unlike traditional approaches that merely generate a question into a single equivalent, the proposed methodology goes beyond, generating a diverse set of interrogations for the same sentence. The proposed methodology generates a diverse set of context-based and grammar-based questions including 9 types of context-based questions. Among these, the 'wh' questions were generated utilizing 42 different question words.

In exploring the outcomes of the AQG approach, a standout feature emerges: the ability to generate more than one 'true-false' questions from a single input sentence. The generation of multiple 'true-false' questions from a single source sentence holds particular significance in scenarios where the teacher or the instructor want to test the learners surface-level understanding during teaching. This nuanced approach enriches the generated questions, capturing a range of linguistic possibilities and offering a more nuanced representation of the input text.

Overview of key findings:

Mentioned results showcase the effectiveness of this novel approach, both through human evaluation and automatic metrics, specifically the BLEU score. The dual evaluation methodologies consistently yielded favorable outcomes, indicating the robustness and quality of the generated questions. During human evaluation, participants consistently recognized and appreciated the diversity of questions generated by the AQG engine. The use of BLEU scores in automatic evaluation further validates the efficacy of the proposed approach. The high scores obtained suggest a close alignment between the generated 'wh' question and human references, reinforcing the accuracy and fluency achieved by the AQG engine.

In conclusion, proposed novel question generation approach, offering a diverse set of questions, represents a significant advancement in the field. The positive outcomes from both human and automatic evaluations underscore the robustness and effectiveness of the proposed approach. This approach not only contributes to the richness of generated questions but also holds promise for applications requiring a more nuanced understanding of language shades.

5.1Comparative analysis with existing systems

Reported AQG engine excels at generating questions for a diverse set of linguistic elements, encompassing 19 different types of context-based and grammar-based questions. This extensive coverage ensures a refined and varied output. In contrast, Das et al. have employed a similar rule-based approach, but it is limited to generating only 'who', 'what', 'whose', 'how much', 'how many', 'whom', 'where', and 'when' questions [12]. While both systems share a common approach, the difference in linguistic coverage becomes evident when considering the broader scope of reported AQG engine.

However, the AQG engine can generate more types of questions compared to the system of Das et al., which handles only simple and complex sentences.

However, the reported AQG engine can tackle compound sentences too. This makes the AQG engine more versatile and effective, especially when dealing with a variety of sentence structures. It's important to note that these comparisons are based on different datasets. *Table 13* provides a comparative analysis between the AQG engine and various other related works in terms of the types of questions generated [12]. It showcases the diversity and characteristics of questions produced by both entities, offering insights into their respective question-generation capabilities.

Table 13 Comparison of Question Types Generated by AQG engine and Das et al. system

Reported work	**Types of questions generated**	**Approach used**
Das et al. system [12]	Who, what, whose, how much, how many, whom, where, when.	Rule-based
Gašpar, Grubišić, and Šarić-Grgić [43]	who, what, where, when, why and how	Rule-based
Ours	Context-based questions: How many, how, what, who, whom, define, when, why, what will happen if, what is the result of, where, complete the sentence etc. True or false Fill in the blanks Grammar-based question.	Rule-based

5.2Limitations

The proposed model generates some incorrect questions; the reasons behind these incorrect questions are listed below.

Shallow parser challenges in detecting the phrasal border

While proposed question generation approach exhibits notable strengths in generating diverse questions, it is essential to acknowledge a limitation tied to the performance of the shallow parser. In some instances, the shallow parser faces challenges in accurately detecting the correct entire phrase from the input text. This limitation can impact the fluency of the generated question, particularly when dealing with incomplete phrases.

When the shallow parser fails to capture the entirety of a phrase in the input text, it can result in incomplete questions, affecting the overall fluency of the question. The AQG engine's reliance on the parser for syntactic analysis means that inaccuracies in phrase identification may propagate into the generated question.

Example Sentence: काजूसारख्या काही फळांमध्ये बी थोडेसे बाहेरच्या बाजूस आलेले असते

Phrase detected: काही फळांमध्ये

Translation: In some fruits

Intended Phrase: काजूसारख्या काही फळांमध्ये

Translation: In some fruits like cashew nuts

Actual Question generated by the AQG engine (Impacted by Shallow Parser Limitation):

प्रश्न. काजूसारख्या कश्यामध्ये बी थोडेसे बाहेरच्या बाजूस आलेले असते?

Translation: Question. What kind of nut like cashew has the seed slightly outwards?

Intended question: प्रश्न. कश्यामध्ये बी थोडेसे बाहेरच्या बाजूस आलेले असते?

Translation: Question. In which case the seed is slightly protruding?

Linguistic Issue: Inaccurate negation of auxiliary verbs in 'True/False' questions.

Auxiliary verbs play a crucial role in indicating various grammatical aspects, including tense and mood. The negation of these verbs requires a nuanced understanding of the semantic context. In the example of 'होता' (was) the intended negation, such as 'नव्हता'(was not) and 'होत नाही' (does not happen) convey specific negations that may not always be accurately captured by the current model. The following *Table 14* shows an example sentence from which the AQG engine can generate different types of questions and also shows Bloom's level for each question.

A complete list of abbreviations is summarized in *Appendix I.*

Table 14 Example sentence

Input sentence and the questions generated by the AQG engine and mapping with Bloom's taxonomy			
Input Sentence	**Translation**	**Transliteration**	
सीमाने तिच्या मैत्रिणीसोबत सुंदर चित्र पटकन रंगविले.	Sīmānē ticyā maitriṇī sōbata sundara citra paṭakana raṅgavilē.	Seema quickly painted a beautiful picture with her friend.	
Questions	**Transliteration**	**Translation**	**Bloom's Taxonomy Levels**
प्रश्न. खालील वाक्यातील शब्दांचा योग्य क्रम लावा. रंगविले तिच्या सुंदर मैत्रिणीसोबत चित्र पटकन सीमाने	Praśna. Khālīla vākyātīla śabdāñcā yōgya karma lāvā. Raṅgavilēticyāsundaramaitri ṇīsōbatacitrapaṭakanasīmānē	Question. Arrange the words in the following sentence in the correct order. painted her beautiful with friend Seema picture quickly a	Level 4 Analysis
प्रश्न. विरुद्धार्थी शब्द लिहा -सुंदर प्रश्न. समानार्थी शब्द लिहा -सुंदर	praśna. Virud'dhārthīśabdalihā -sundara praśna. Samānārthīśabdalihā -sundara	Question. Write the opposite words - Beautiful Question. Write the synonyms - Beautiful	Level 2 Comprehension
प्रश्न. जोड्या जुळवा नाम रंगविले सर्वनाम सुंदर विशेषण चित्र क्रियाविशेषण तिच्या क्रियापद पटकन	Praśna. Jōḍyā juḷavā nāma raṅgavilē sarvanāma sundara viśēṣaṇa citra kriyāviśēṣaṇa ticyā kriyāpada paṭakana	Question. Match the pairs Name painted Pronoun beautiful Adjective picture Adverb her verb quickly	Level 4 Analysis
प्रश्न. कोणी मैत्रिणीसोबत सुंदर चित्र पटकन रंगविले? प्रश्न. मैत्रिणीसोबत सुंदर चित्र पटकन कोणी रंगविले?	Praśna. Kōṇīmaitriṇīsōbatasundaracitrapaṭakanaraṅgavilē? Praśna. Maitriṇīsōbata sundara citra paṭakana kōṇī raṅgavilē?	Question. Who quickly painted a beautiful picture with a girlfriend?	Level 1 Remembering
प्रश्न. सीमाने कोणासोबत सुंदर चित्र रंगविले?	Praśna. Sīmānē kōṇā sōbata sundara citra raṅgavilē?	Question. With whom Seema painted a beautiful picture?	Level 1 Remembering
Input Sentence	**Transliteration**	**Translation**	
पृथ्वीवर वातावरण जर नसते, तर आकाश दिवसा काळे दिसले असते.	Pṛthvīvara vātāvaraṇa jara nasatē, tara ākāśa divasā kāḷē disalē asatē	If there was no atmosphere on Earth, the sky would appear black during the day.	
Questions	**Transliteration**	**Translation**	**Bloom's Taxonomy Levels**
प्रश्न. आकाश दिवसा काळे केव्हा दिसले असते?	Praśna. Ākāśa divasā kāḷē kēvhā disalē asatē?	When was the sky black during the day?	Level 1Knowledge
प्रश्न. कशावर वातावरण जर नसते, तर आकाश दिवसा काळे दिसले नसते?	Praśna. Kaśyāvara vātāvaraṇa jara nasatē, tara ākāśa divasā kāḷē disalē nasatē?	Q. Where if there was no atmosphere, the sky would not appear black during the day?	Level 1Knowledge

Input sentence and the questions generated by the AQG engine and mapping with Bloom's taxonomy

Input Sentence	Translation	Transliteration	
प्रश्न. खालील वाक्य चूक कि बरोबर ते लिहा. वाक्य चुकीचे असल्यास ते बरोबर करून लिहा. १) पृथ्वीवर वातावरण जर असते, तर आकाश दिवसा काळे दिसले असते. २) पृथ्वीवर वातावरण जर नसते, तर आकाश दिवसा काळे दिसले नसते.	Praśna. Khālīla vākya cūka ki barōbara tē lihā. Vākya cukīcē asalyāsa tē barōbara karūna lihā. 1)Pr̥thvīvara vātāvaraṇa jara asatē, tara ākāśa divasā kāḷē disalē asatē. 2) Pr̥thvīvara vātāvaraṇa jara nasatē, tara ākāśa divasā kāḷē disalē nasatē.	Question. Write the following sentences true or false. Correct the sentence if it is wrong. 1) If the earth had an atmosphere, then the sky would have appeared black during the day. 2) If there was no atmosphere on Earth, the sky would not appear black during the day.	Level 1 Knowledge

6. Conclusion and future work

The AQG methodology proposed in this paper demonstrates a comprehensive approach for AQG through various stages of the NLP pipeline. This paper introduced an end-to-end framework that effectively processes a sentence and generates both context- and grammar-based questions as outputs. Using this approach, the successful generation of ten different types of context-based questions and nine types of grammar-based questions, spanning Bloom's cognitive levels 1 through 4, has been achieved. The AQG model generated a total of 2154 questions from 213 sentences, showcasing its capability to produce a diverse range of meaningful inquiries. The performance of the proposed AQG methodology has been evaluated using the BLEU score and manual evaluation on a corpus selected from the sixth standard science textbook prescribed by the Maharashtra State Board, India. The BLEU score for the 'wh' questions is 90.37. Manual evaluations through three levels of metrics—adequacy, fluency, and answerability—yielded accuracies of 73% for fluency, 74% for adequacy, and 81% for answerability. The 'fill in the blanks', 'true or false', and other grammar-based questions were evaluated using metrics such as 'impact on overall understanding', 'completeness', and 'specificity'. The 'fill in the blanks' questions achieved an accuracy of 98% for 'impact on overall understanding', 97.33% for 'completeness', and 91.66% for 'specificity'. Grammar-based questions were generated with an impressive accuracy of 91%.

Future research will focus on refining context-based question generation. The integration of NER with shallow parsing is seen as a promising development, signaling a significant advance in precision. By leveraging the strengths of both techniques, our goal is to enhance accuracy in NLP applications further. We also plan to expand the AQG model to generate questions at Bloom's levels 5 and 6 for paragraph-level texts.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Data availability

The data considered in this study is gathered from sixth standard science text book ('SamanyaVidnyan') prescribed by the Maharashtra State Board, Maharashtra, India. The data is publicly available.

Author's contribution statement

Pushpa M. Patil: Conceptualization, investigation, methodology, data collection, writing – original draft. **R. P. Bhavsar:** Conceptualization, methodology, study investigation and supervision. **B. V. Pawar:** Study investigation, conceptualization and supervision.

References

[1] Craig J, Cairo IIIL. Assessing the relationship between questioning and understanding to improve learning and thinking (QUILT) and student achievement in mathematics: a pilot study. Appalachia Educational Laboratory at Edvantia (NJ1); 2005.

[2] No TC. 12: deeper learning through questioning. Teaching Excellence in Adult Literacy. 2013.

[3] Nguyen HA, Bhat S, Moore S, Bier N, Stamper J. Towards generalized methods for automatic question generation in educational domains. In European conference on technology enhanced learning 2022 (pp. 272-84). Cham: Springer International Publishing.

[4] Dugan L, Miltsakaki E, Upadhyay S, Ginsberg E, Gonzalez H, Choi D, et al. A feasibility study of answer-agnostic question generation for education. In

findings of the association for computational linguistics: ACL 2022 (pp. 1919–26).
[5] Gaikwad DK, Kadam V, Mahender CN. Question based text summarization under rule based framework. International conference on cognitive knowledge engineering 2016 (pp. 332-5).
[6] Shah R, Shah D, Kurup L. Automatic question generation for intelligent tutoring systems. In 2nd international conference on communication systems, computing and IT applications 2017 (pp. 127-32). IEEE.
[7] Heilman M, Smith NA. Good question! statistical ranking for question generation. In human language technologies: the annual conference of the North American chapter of the association for computational linguistics 2010 (pp. 609-17). ACL.
[8] Wang Z, Valdez J, Basu MD, Baraniuk RG. Towards human-like educational question generation with large language models. In international conference on artificial intelligence in education 2022 (pp. 153-66). Cham: Springer International Publishing.
[9] Sheu FR, Grissett JO, Hsu YL. Examining students' perceptions and preferences for traditional paper vs. multiple-choice question generation assignments. In the annual convention of
a. the association for educational communications and technology 2022 (pp. 298-306).
[10] Wang Y, Rong W, Zhang J, Zhou S, Xiong Z. Multi-turn dialogue-oriented pretrained question generation model. Complex & Intelligent Systems. 2020; 6:493-505.
[11] Langi AZ, Rosmansyah Y. Generation quiz with genetic algorithm based on bloom's taxonomy classification in serious game based virtual environments. In international conference on ICT for smart society 2016 (pp. 42-8). IEEE.
[12] Das R, Ray A, Mondal S, Das D. A rule based question generation framework to deal with simple and complex sentences. In international conference on advances in computing, communications and informatics 2016 (pp. 542-8). IEEE.
[13] Chali Y, Hasan SA. Towards topic-to-question generation. Computational Linguistics. 2015; 41(1):1-20.
[14] Gaikwad DK, Mahender CN, Namrata C. Question generation system for Marathi text. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2018; 3(3): 246-9.
[15] Chen Y, Wu L, Zaki MJ. Toward subgraph-guided knowledge graph question generation with graph neural networks. IEEE Transactions on Neural Networks and Learning Systems. 2023; 1-12.
[16] Yin X, Zhou L, Small K, May J. Summary-oriented question generation for informational queries. In proceedings of the 1st workshop on document-grounded dialogue and conversational question answering 2020 (pp. 81-97). Association for Computational Linguistics.
[17] Fei Z, Zhang Q, Zhou Y. Iterative GNN-based decoder for question generation. In proceedings of the conference on empirical methods in natural language processing 2021 (pp. 2573-82).
[18] Xu Y, Wang D, Yu M, Ritchie D, Yao B, Wu T, et al. Fantastic questions and where to find them:fairytaleqa-an authentic dataset for narrative comprehension. In proceedings of ACL 2022 (pp. 447–60). Association for Computational Linguistics.
[19] Gopal A. Automatic question generation for Hindi and Marathi. In international conference on advanced learning technologies 2022 (pp. 19-21). IEEE.
[20] Fahad AR, Al NN, Islam MA, Rahman RM. Answer agnostic question generation in Bangla language. International Journal of Networked and Distributed Computing. 2024: 1-26.
[21] Jayaseelan KA. Questions and question-word incorporating quantifiers in Malayalam. Syntax. 2001; 4(2):63-93.
[22] Huber F, Hagel G. Semi-automatic generation of textual exercises for software engineering education. In global engineering education conference 2022 (pp. 51-6). IEEE.
[23] Hou Z, Bi S, Qi G, Zheng Y, Ren Z, Li Y. Syntax-guided question generation using prompt learning. Neural Computing and Applications. 2024: 1-12.
[24] Fei Z, Zhang Q, Gui T, Liang D, Wang S, Wu W, et al. CQG: a simple and effective controlled generation framework for multi-hop question generation. In proceedings of the 60th annual meeting of the association for computational linguistics 2022 (pp. 6896-906). Association for Computational Linguistics.
[25] Steuer T, Filighera A, Tregel T. Investigating educational and noneducational answer selection for educational question generation. IEEE Access. 2022; 10:63522-31.
[26] Wijanarko BD, Heryadi Y, Toba H, Budiharto W. Automated question generating method based on derived keyphrase structures from bloom's taxonomy. ICIC Express Letters. 2020; 14(11):1059-67.
[27] Liu M, Rus V, Liu L. Automatic Chinese factual question generation. IEEE Transactions on Learning Technologies. 2016; 10(2):194-204.
[28] Leite B, Cardoso HL, Reis LP, Soares C. Factual question generation for the Portuguese language. In international conference on innovations in intelligent systems and applications 2020 (pp. 1-7). IEEE.
[29] Morón M, Scocozza J, Chiruzzo L, Rosá A. A tool for automatic question generation for teaching English to beginner students. In 40th international conference of the Chilean computer science society 2021 (pp. 1-5). IEEE.
[30] Garimella A, Chiticariu L, Li Y. Domain-aware dependency parsing for questions. In findings of the association for computational linguistics: ACL-IJCNLP 2021 (pp. 4562-8).
[31] Kumar NA, Fernandez N, Wang Z, Lan A. Improving reading comprehension question generation with data augmentation and overgenerate-and-rank. In proceedings of the 18th workshop on innovative use of

NLP for building educational applications (BEA 2023) (pp. 247–59). Association for Computational Linguistics.
[32] Maheshwari H, Shekhar S, Saxena A, Chhaya N. Open-world factually consistent question generation. In findings of the association for computational linguistics: ACL 2023 (pp. 2390-404).
[33] Das B, Majumder M, Sekh AA, Phadikar S. Automatic question generation and answer assessment for subjective examination. Cognitive Systems Research. 2022; 72:14-22.
[34] Rathi S, Wawage P, Kulkarni A. Automatic question generation from textual data using NLP techniques. In international conference on emerging smart computing and informatics 2023 (pp. 1-7). IEEE.
[35] Patil N, Patil AS, Pawar BV. Issues and challenges in Marathi named entity recognition. International Journal on Natural Language Computing. 2016; 5(1):15-30.
[36] https://www.nltk.org/. Accessed 23 March 2024.
[37] Carlberger J, Kann V. Implementing an efficient part-of-speech tagger. Software: Practice and Experience. 1999; 29(9):815-32.
[38] Olney AM, Graesser AC, Person NK. Question generation from concept maps. Dialogue & Discourse. 2012; 3(2):75-99.
[39] Brown J, Frishkoff G, Eskenazi M. Automatic question generation for vocabulary assessment. In proceedings of human language technology conference and conference on empirical methods in natural language processing 2005 (pp. 819-26). Association for Computational Linguistics.
[40] Patil PM, Bhavsar RP, Pawar BV. Automatic question generation for subordinate conjunctions of Marathi. In 8th international conference on advanced computing and communication systems 2022 (pp. 169-173). IEEE.
[41] Agarwal M, Shah R, Mannem P. Automatic question generation using discourse cues. In proceedings of the sixth workshop on innovative use of NLP for building educational applications 2011 (pp. 1-9). Association for Computational Linguistics.
[42] Kalady S, Elikkottil A, Das R. Natural language question generation using syntax and keywords. In proceedings of QG2010: the third workshop on question generation 2010 (pp. 1-10).
[43] Gašpar A, Grubišić A, Šarić-grgić I. Evaluation of a rule-based approach to automatic factual question generation using syntactic and semantic analysis. Language Resources and Evaluation. 2023; 57(4):1431-61.
[44] https://spacy.io/. Accessed 30 June 2024.
[45] Afzal N, Mitkov R. Automatic generation of multiple choice questions using dependency-based semantic relations. Soft Computing. 2014; 18:1269-81.
[46] Bloom BS. Taxonomy of educational objectives: the classification of educational goals: by a committee of college and university examiners. David McKay; 1971.
[47] Karamustafaoglu S, Karamustafaoglu O, Bacanak A, Degirmenci S. Classification of biology exam questions as to bloom. Energy Education Science and Technology Part B: Social and Educational Studies. 2011; 3(4):579-88.
[48] Setyaningsih ER, Listiowarni I. Categorization of exam questions based on bloom taxonomy using naïve bayes and laplace smoothing. In 3rd east Indonesia conference on computer and information technology 2021 (pp. 330-3). IEEE.
[49] Ifham M, Banujan K, Kumara BS, Wijeratne PM. Automatic classification of questions based on bloom's taxonomy using artificial neural network. In international conference on decision aid sciences and applications 2022 (pp. 311-5). IEEE.
[50] Gani MO, Ayyasamy RK, Sangodiah A, Fui YT. Bloom's taxonomy-based exam question classification: the outcome of CNN and optimal pre-trained word embedding technique. Education and Information Technologies. 2023; 28(12):15893-914.
[51] Romadhony A, Abdurohman R. Primary and high school question classification based on bloom's taxonomy. In 10th international conference on information and communication technology 2022 (pp. 234-9). IEEE.
[52] Omar N, Haris SS, Hassan R, Arshad H, Rahmat M, Zainal NF, et al. Automated analysis of exam questions according to Bloom's taxonomy. Procedia-Social and Behavioral Sciences. 2012; 59:297-303.
[53] Papineni K, Roukos S, Ward T, Zhu WJ. Bleu: a method for automatic evaluation of machine translation. In proceedings of the 40th annual meeting of the association for computational linguistics 2002 (pp. 311-8). ACL.

Pushpa M. Patil received Bachelor of Computer Science in 2003. She completed Masters in Computer Science in 2005. She has 17 years of experience in field of teaching. Currently she is pursuing Ph.D. (Computer Science) from School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. Currently, she is working as Assistance Professor in the department of Computer Science, Pratap College, Amalner, Jalgaon. Maharashta. Her research areas of interest include NLP.
Email: pushpapatil@pca.ac.in

R. P. Bhavsar obtained his B.C.S. degree from Pune University (1992) followed by MCA in 1995 from Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. He was awarded Ph.D. in Computer Science in 2016 from the same university. He has varied experience of 24 years which includes 03 years as Scientist at C-DAC (Pune), 07 years as a System Analyst, 15 years of teaching. Currently, he is working as Professor in School of

Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. His research area includes Machine Translation, Applied NLP, Capacity Building, Machine Learning, IT & Networking infrastructure setup and has handled software development projects. He co-guided 2 research scholars for their doctoral research in the area of language technology. Currently 4 research scholars are pursuing their Ph.D. under his supervision.
Email: rpbhavsar@nmu.ac.in

B. V. Pawar finished his B. E. (Production Engineering) from reputed VJTI Mumbai (1986), followed by M.Sc. (Computer Science) from Mumbai University (1988). He was awarded a Ph.D. degree in Computer Science by Kavayitri Bahinabai Chaudhari North Maharashtra University (formerly North Maharashtra University), Jalgaon, India in 2000. He has huge experience of 35 years teaching and research in the field of Computer Science. Formerly, he is working as Professor and Director of School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. Currently he is working as a Director of KCES's Institute of Management and Research, Jalgaon. His research interests include NLP; Web computing, Information Retrieval and Machine Translation, Machine Learning. He has guided 17 research Scholars for their doctoral research. He has published over 93 research/journal papers and 129 conference papers presentations and 05 invited talks.
Email: bvpawar@imr.ac.in

Appendix I

S. No.	Abbreviation	Description
1	ADJP	Adjective Phrase
2	ALBERT	A lite BERT
3	AQG	Automatic Question Generation
4	BART	Bidirectional and Auto-Regressive Transformers
5	BERT	Bidirectional Encoder Representations From Transformers
6	BLEU	Bilingual Evaluation Understudy
7	CNN	Convolutional Neural Network
8	CQG	Controlled Question Generation
9	DEBERTa	Decoding-enhanced BERT with Disentangled Attention
10	DistilBERT	A Distilled Version of BERT
11	ELI5	Explain like I'm Five
12	GNN	Graph Neural Network
13	GPT	Generative Pre-trained Transformer
14	Graph2Seq	Graph-to-Sequence
15	HotpotQA	Hotpot Question Answering
16	KG	Knowledge Graph
17	LDA	Latent Dirichlet Allocation
18	MS MARCO	Microsoft Machine Reading Comprehension
19	NER	Named Entity Recognition
20	NLI	Natural Language Inference
21	NLP	Natural Language Processing
22	NLTK	Natural Language Toolkit
23	NP	Noun Phrase
24	NQA	Natural Questions
25	nsubj	Nominal Subject
26	obl	Oblique Nominal
27	PEGASUS	Pre-Training With Extracted Gap-Sentences for Abstractive Summarization
28	POS	Parts of Speech
29	RNN	Recurrent Neural Network
30	RoBERTa	Robustly optimized BERT pretraining approach
31	ROUGE	A Recall-Oriented Understudy for Gisting Evaluation
32	SQuAD	Stanford Question Answering Dataset
33	SRL	Semantic Role Labeling
34	TnT	Trigrams'n'Tags
35	TQA-A	While Textbook Question Answering With Answer
36	UML	Unified Modeling Language
37	VP	Verb Phrase
38	XML	Extensible Markup Language
39	WSD	Word Sense Disambiguation