AMERICAN JOURNALS
Open Access, Peer Reviewed Journals

American Journal of Research in Humanities and Social Sciences
ISSN (E): 2832-8019 Volume 53, | October - 2026

CRITERIA FOR LINGUISTIC EVALUATION OF GENERATIVE ARTIFICIAL INTELLIGENCE TEXTS IN THE UZBEK LANGUAGE

Lola Abdurashitovna Uzaqova
Acting Associate Professor, Uzbekistan State World Languages University,
Doctor of Philosophy (PhD) in Philosophical Sciences
Tashkent, Uzbekistan
E-mail: lolauzakova00@gmail.com
ORCID: https://orcid.org/0009-0006-0060-2708

ABSTRACT

This article examines the theoretical and methodological foundations for evaluating Uzbek texts generated by artificial intelligence systems. The relevance of the study stems from the growing use of Uzbek in digital environments and the challenges of accurately representing its morphological, lexical-semantic, syntactic, and pragmatic features in automatically generated texts. The article argues that linguistic evaluation should extend beyond grammatical correctness to include semantic coherence, contextual appropriateness, stylistic consistency, and culturally grounded meaning. The agglutinative structure of Uzbek, suffix ordering, lexical relations, and sociolinguistic conventions are considered important dimensions of evaluation. A multidimensional framework for assessing AI-generated Uzbek texts is proposed. The discussion highlights the need to account for the structural, contextual, and cultural characteristics of Uzbek when developing and evaluating language technologies.

KEYWORDS

Artificial intelligence, generative text, Uzbek language, computational linguistics, large language models, morphology, semantics, pragmatics, linguistic evaluation, natural language processing.

Introduction

The development of digital technologies is significantly impacting processes related to the generation, processing, and dissemination of language. In particular, the advancement of generative artificial intelligence (AI) systems enables the automation of tasks such as text generation, translation, editing, summarization, and question-answering. Such technologies are increasingly widely applied in education, scientific research, mass media, and everyday communication.

As the text-handling capabilities of generative AI expand, the issue of scientifically evaluating the quality of the texts it produces is becoming increasingly urgent. This is because the outwardly fluent appearance of a text does not fully guarantee that it is grammatically correct, semantically coherent, or appropriate for a specific communicative situation. Especially in languages like Uzbek, where grammatical meanings are frequently expressed through affixes, examining the text generation process on the basis of linguistic criteria acquires special significance.

The agglutinative structure of the Uzbek language requires that word forms be generated through the specific sequential combination of roots and affixes. Grammatical meanings such as case, possession, plurality, tense, person-number, and others are expressed via corresponding morphological tools. These features necessitate the correct selection of grammatical forms, the normative application of affixes, and the assurance of mutual compatibility among units within a sentence.

Furthermore, the meanings of words in the Uzbek language can manifest differently depending on context, style, communicative purpose, and social relations. For instance, the difference between the pronouns "sen" (informal singular "you") and "Siz" (formal/respectful "you") is not limited solely to expressing grammatical person. They are also tied to social distance, respect, and the speech situation between communication participants. Therefore, in evaluating the linguistic correctness of a text, alongside grammatical forms, their pragmatic functions must also be taken into account.

The development of generative AI strengthens the connections between linguistics and computational linguistics. Bender et al. (2021) emphasize the importance of considering data quality, the social context of language, and technological constraints in creating and applying large language models. Meanwhile, Brown et al. (2020) investigated the capabilities of large language models in performing various tasks based on prompts. Although these works provide an important theoretical foundation for understanding the capabilities of generative systems, they do not eliminate the necessity of studying the specific grammatical and communicative characteristics of each individual language.

In computational linguistics research on the Uzbek language, issues concerning the formation of specialized language resources and the creation of language-adapted models are also being explored. The BERTbek model, introduced by Kuriyozov, Vilares, and Gómez-Rodríguez (2024), was tested across a series of natural language processing tasks as a pre-trained language model adapted for the Uzbek language. The authors demonstrate that the volume and quality of training data directly affect model outcomes. This study highlights the importance of specialized linguistic resources for the accurate representation of the Uzbek language in digital technologies.

In the study by Solidjonov and Najmiddinov, which directly examines Uzbek generative texts, cases related to spelling, morphology, syntax, lexical selection, and pragmatic compatibility in texts generated by large language models were analyzed (Solidjonov and Najmiddinov, 2026). The authors note that even outwardly fluent texts may contain inconsistencies concerning certain grammatical and socio-communicative features. These studies substantiate the necessity of linguistically evaluating Uzbek generative texts. At the same time, it is important to clearly define a system of criteria that serves to examine various language levels and communicative characteristics in an interconnected manner.

The purpose of this article is to systematize the main criteria for the linguistic evaluation of Uzbek texts generated by generative artificial intelligence and to define the theoretical and methodological principles of their application. The primary research question of the article is formulated as follows: what criteria should be considered in evaluating the linguistic quality of Uzbek texts generated by generative AI, and how should they be interrelated? Within the framework of this question, morphological and syntactic correctness, lexical-semantic compatibility, text coherence, pragmatic appropriateness, and the expression of national-cultural content are established as the main dimensions to be studied.

METHODOLOGY

This article is theoretical-methodological in nature, utilizing methods of scientific literature analysis, comparative analysis, description, and the systematization of linguistic criteria. The primary task of the research is not to empirically measure the errors of a specific AI model, but rather to scientifically substantiate the criteria that can be employed for evaluating Uzbek generative texts.

In analyzing theoretical sources, reliance is placed on studies dedicated to large language models, natural language processing, the morphological features of the Uzbek language, and the linguistic quality of generative texts. Specifically, the work of Brown et al. (2020) serves as a theoretical source for understanding the capabilities of generative models, while the study by Bender et al. (2021) illuminates the limitations and socio-linguistic responsibilities of large language models. The research by Kuriyozov et al. (2024) addresses the issue of Uzbek-adapted language models and data quality. The work of Solidjonov and Najmiddinov (2026) serves as an important source for the direct linguistic analysis of Uzbek generative texts.

During the analysis process, the principle that text quality should not be evaluated through a single criterion alone is adhered to. A grammatically correct sentence may be semantically ambiguous; conversely, certain minor grammatical flaws may not seriously hinder the comprehension of a text's overall meaning. Therefore, it is advisable to evaluate different linguistic features separately and subsequently examine their interrelations.

The proposed evaluation model encompasses five main directions:

1. Morphological and grammatical compatibility.
2. Syntactic correctness.
3. Lexical-semantic precision.
4. Discursive and pragmatic appropriateness of the text.
5. Stylistic and national-cultural conformity.

These directions complement one another. To apply them in practical research, the verifiable markers, evaluation procedures, and error classification for each criterion must be predetermined.

If an empirical study is conducted in a subsequent phase, texts generated by various generative models should be collected based on identical or semantically equivalent tasks. The model name, version used, prompt text, data collection date, and the edited or unedited state of the text must be recorded. These requirements enhance reproducibility and the verification of results.

Furthermore, independent evaluation of texts by multiple experts and determining inter-rater reliability can strengthen the reliability of practical results. However, such empirical procedures are viewed not as research conducted within this article, but as methodological recommendations for future practical investigations.

RESULTS

Based on theoretical analysis, it was substantiated that five interrelated directions can be distinguished for evaluating the linguistic quality of Uzbek texts generated by generative artificial intelligence. These directions allow for the separate examination of a text's grammatical form, semantic precision, and communicative function.

1. Morphological and Grammatical Compatibility. The Uzbek language belongs to the category of agglutinative languages. In such languages, grammatical meanings are expressed by appending suffixes to roots in a specific order. Consequently, the verification of generative texts requires

accounting for the correct selection of affixes, the mutual compatibility of grammatical forms, and the syntactic roles of words.

For example, in the sentence *"Men kitobni do‘stimga berdim"* ("I gave the book to my friend"), the form *"kitobni"* expresses the accusative case, while *"do‘stimga"* expresses the dative case. These forms determine the semantic and grammatical relationships of participants in the sentence. Their incorrect application may alter the meaning of the sentence or reduce its naturalness.

Morphological evaluation examines the following aspects:

- Normative combination of roots and affixes;
- Correct application of case, possessive, and plural forms;
- Agreement of verb tense, mood, and person-number forms;
- Grammatical relationships among word forms;
- Adherence to orthographic norms.

Here, any unusual form should not automatically be considered an error. Dialectal units, authorial usages, or forms demanded by a specific genre must be examined separately.

2. Syntactic Correctness. Syntactic compatibility encompasses the correct combination of words in a sentence, relationships among sentence parts, and the coherent construction of complex sentences.

Although the Uzbek language possesses a standard word order for sentence components, word order may vary depending on communicative purpose, emphasis, and stylistic function. Therefore, syntactic analysis must consider not only word order itself, but also the grammatical structure and semantic coherence of the sentence.

For example, in the sentence *"Tadqiqotchi yangi usulni taklif qildi"* ("The researcher proposed a new method"), the subject, object, and predicate relationship is clearly expressed. The artificial complication of this sentence or the inappropriate placement of its constituent parts can hinder comprehension.

Syntactic evaluation applies the following criteria:

- Grammatical connection of sentence components;
- Agreement between subject and predicate;
- Normative structure of word combinations;
- Relationships among clauses of compound sentences;
- Correspondence of punctuation marks to syntactic functions.

In doing so, deliberate stylistic shifts in literary or journalistic texts should not be confused with grammatical errors.

3. Lexical-Semantic Precision. Lexical-semantic compatibility refers to word meanings, collocability, and appropriateness to the text context. Even if a word's dictionary definition is correct, its application in a specific collocation may not be natural.

For example, in collocations such as *"muammoni hal qilmoq"* ("to solve a problem"), *"qaror qabul qilmoq"* ("to make a decision"), and *"xulosa chiqarmoq"* ("to draw a conclusion"), the words conform to the combinatory norms established in the Uzbek language. Substituting the words in these collocations arbitrarily with synonyms or near-synonyms does not always yield a natural result.

Lexical-semantic evaluation takes into account:

- Usage of words in a context-appropriate sense;
- Word collocability (combinatory potential);
- Correct selection of synonymous units;
- Appropriate interpretation of polysemantic words in context;
- Precise and consistent use of terminology;
- Presence or absence of semantic contradictions within the text.

This criterion is especially important for scientific and official texts. Because the incorrect selection of a term or the use of a single concept in varying meanings throughout the text reduces the precision of scientific thought.

4. Discursive and Pragmatic Compatibility. Discourse is associated with the organization of a text in a specific social and communicative situation. Pragmatics, meanwhile, studies the relationship of linguistic units to speech goals, situations, and communication participants.

The same content can be expressed differently in various situations. For instance, an expression used in informal communication among friends may not be appropriate in an official address. Therefore, evaluating the quality of a generative text requires considering to whom it is addressed, what function it performs, and what attitude it expresses.

The Uzbek forms *"sen"* and *"Siz"* can serve as a clear example of this issue. Their selection depends not only on grammatical person, but also on respect, social distance, and communication conditions. Consequently, the inappropriate use of an informal form in an official letter or academic address may not be a grammatical error, but it can be evaluated as a pragmatic mismatch.

Pragmatic evaluation encompasses the following aspects:

- Appropriateness of the address form to the situation;
- Correct selection of the degree of formality and informality;
- Clear expression of the speech intention;
- Logical connection between parts of the text;
- Consideration of the addressee's level of knowledge and communicative needs.

5. Stylistic and National-Cultural Conformity. The stylistic compatibility of a text entails the use of linguistic tools appropriate to its purpose, genre, and addressee. Scientific articles, official documents, literary stories, and daily communication are not structured under identical stylistic requirements.

National-cultural conformity helps evaluate how closely the concepts and relations expressed in a text align with the cultural experience formed in the Uzbek language. This aspect is particularly important in proverbs, phraseological units, forms of address, and devices expressing social etiquette.

For instance, the idiom *"ko‘z qorachig‘idek asramoq"* ("to guard like the pupil of one's eye") conveys the meaning of protecting and cherishing someone or something with extreme care. To use it correctly in a text, one must take into account not only the literal meanings of the words, but also the holistic figurative meaning of the idiom.

However, national-cultural conformity cannot be measured solely by the presence of traditional idioms. Modern Uzbek language also employs new words, borrowings, and novel communicative forms. Therefore, evaluation must also account for language variability, genre diversity, and speech situations.

6. Multi-Criteria Evaluation Model. Based on the preceding analyses, the following model is proposed for evaluating Uzbek texts generated by generative artificial intelligence:

Evaluation Dimension	Main Criteria	Potentially Identifiable Inconsistency
Morphological-Grammatical	Affixes, grammatical forms, word form compatibility	Incorrect grammatical form
Syntactic	Sentence structure, word linkage, inter-clause relations	Syntactic mismatch
Lexical-Semantic	Word choice, semantic precision, collocational norms	Incorrect or ambiguous meaning
Discursive-Pragmatic	Address form, communicative purpose, situational appropriateness	Mismatch with speech situation
Stylistic-Cultural	Genre, style, phraseology, and national-cultural content	Stylistic or cultural mismatch

The proposed table reflects evaluation criteria developed on the basis of theoretical analysis rather than ready empirical results. To apply them in practical research, an operational definition of each criterion, an evaluation scale, and a coding guideline reinforced with examples must be developed.

DISCUSSION

In evaluating Uzbek texts generated by generative artificial intelligence, although it is important to examine different language levels separately, their interrelationship must also be taken into account. A morphologically correct word combination may be semantically ambiguous. Similarly, a grammatically correct sentence may contain an inappropriate expression for a specific speech situation. Therefore, the overall quality of a text is not determined by a single grammatical check alone.

The agglutinative structure of the Uzbek language plays a crucial role in the linguistic evaluation of generative models. The order of affixes, the expression of grammatical meanings, and the contextual compatibility of word forms are tied to the systemic features of the language. The research by Kuriyozov et al. (2024) regarding the **BERTbek** model demonstrates the necessity of specialized models and quality data for the Uzbek language. However, the results of this model cannot be directly applied to all generative text-producing systems, as their tasks and architectures differ. Solidjonov and Najmiddinov (2026) studied morphological, syntactic, lexical, and pragmatic inconsistencies in Uzbek generative texts. The relationship between their work and the present article must be clearly defined: whereas the previous study empirically analyzed specific model results, this article focuses on systematizing linguistic evaluation dimensions and explaining their methodological functions. Nevertheless, the practical superiority of this criteria system can only be assessed after being tested against real data.

The issue of national-cultural conformity also requires special attention. The cultural features of language are not expressed solely through national words or traditional idioms. They also manifest in address forms, the category of respect, the degree of speech formality, irony, and other pragmatic devices. Therefore, evaluating national-cultural conformity requires relying on the actual usage norms of a given linguistic community. Here, an important methodological limitation must be kept in mind: the low frequency of a specific idiom does not mean it is incorrect. To determine whether a unit has been used correctly or incorrectly, its context, stylistic function, recording in dictionaries, and usage in real texts must be examined. Similarly, classifying all stylistic variations as errors reduces the reliability of scientific conclusions.

Bender et al. (2021) demonstrate the necessity of considering data quality and social context when evaluating the capabilities of large language models. This approach is equally important with respect to the Uzbek language: examining a text solely from a formal perspective fails to fully capture the social and cultural functions of the language.

The practical significance of the proposed evaluation model lies in its ability to predetermine criteria to be tested in future experimental studies. For example, texts can be obtained from various models based on identical prompts and evaluated individually against each criterion. In this process, the number of texts, prompt contents, and number of evaluators should be determined in accordance with the research question. If results are expressed in percentages or numbers, their calculation methods and data sources must be transparently indicated. Additionally, employing a unified guideline and set of criteria during expert evaluation is essential; otherwise, personal biases among different evaluators may affect the results. Utilizing methods to measure inter-rater reliability further enhances the credibility of conclusions.

Overall, the linguistic evaluation of Uzbek texts generated by generative artificial intelligence requires a multi-layered approach. Within this framework, grammatical correctness, semantic precision, text coherence, pragmatic situation, and national-cultural features must be studied in an interconnected manner. The ultimate efficacy of the proposed model must be confirmed in future empirical studies.

CONCLUSION

The linguistic evaluation of Uzbek texts generated by generative artificial intelligence tools is one of the urgent directions of modern linguistics and computational linguistics. In determining the quality of such texts, grammatical correctness alone is insufficient. The semantic precision, syntactic coherence, situational appropriateness, and national-cultural content of the text must also be taken into account.

In this article, five main dimensions for evaluating Uzbek generative texts—morphological-grammatical, syntactic, lexical-semantic, discursive-pragmatic, and stylistic-cultural compatibility—were systematized. These criteria enable the examination of texts across distinct language levels while analyzing their interrelations. The agglutinative structure of the Uzbek language, affix ordering, grammatical form compatibility, and the social functions of address devices demand special attention in evaluating generative texts. At the same time, examining national-cultural content relies importantly on phraseological units, stylistic features, and real speech usage.

As a theoretical-methodological result of the study, a multi-criteria evaluation model was proposed. This model can be applied in the future for the comparative study of generative models, the classification of linguistic errors, and the development of evaluation tools adapted to the Uzbek language. However, to confirm the practical efficacy of the proposed criteria, it is necessary to form

corpora of texts generated by various generative models, standardize the evaluation procedure, and analyze results based on expert review. Subjecting these criteria to empirical testing in future studies will help more accurately evaluate the quality of generative texts in the Uzbek language.

Thus, evaluating Uzbek generative texts requires the joint consideration of the grammatical, semantic, pragmatic, and cultural features of the language. This approach serves to ensure the full-fledged representation of the Uzbek language in the digital environment and to enhance the linguistic quality of language technologies.

REFERENCES

1. Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. – 2021. – P. 610–623. – DOI: 10.1145/3442188.3445922.
2. Brown T. B. et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 1877–1901.
3. Kuriyozov E., Vilares D., Gómez-Rodríguez C. BERTbek: A Pretrained Language Model for Uzbek // Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages @ LREC-COLING 2024. – Torino: ELRA and ICCL, 2024. – P. 33–44.
4. Solidjonov D., Najmiddinov M. Investigating Linguistic Errors in Large Language Model Generation of Uzbek Text // Cogent Arts & Humanities. – 2026. – Vol. 13, No. 1. – Article 2600519. – DOI: 10.1080/23311983.2025.2600519.
5. UNESCO. Guidance for Generative AI in Education and Research. – Paris: UNESCO, 2023. – DOI: 10.54675/EWZM9535.