

Artificial Intelligence-Assisted Corpus Analysis of the Linguopragmatic Features of Repetition in Erkin A'zam's Works

MOKHICHEKHRA NAMOZOVA
Termez State University, Uzbekistan

ABSTRACT

Repetition constitutes one of the most significant stylistic mechanisms through which literary discourse achieves semantic development, emotional expression and pragmatic emphasis. Although repetition has been extensively examined in traditional stylistics, its linguopragmatic functions have rarely been investigated using corpus-based computational methodologies, particularly in studies of Uzbek literary discourse. This study aims to develop an artificial intelligence-assisted corpus framework for analysing repetition in Erkin A'zam's prose by integrating corpus linguistics, computational stylistics and pragmalinguistics. The research is based on a specialised corpus comprising approximately 610,000 running words collected from 31 prose works. Repetition instances were manually annotated according to structural, positional and communicative criteria and subsequently verified using transformer-based semantic analysis. Quantitative corpus methods were combined with qualitative discourse interpretation to identify recurrent stylistic patterns and evaluate their communicative functions. The findings demonstrate that lexical repetition represents the dominant structural category, while emotional intensification constitutes the most frequent linguopragmatic function. The results further indicate that repetition performs an essential role in discourse organisation, psychological characterisation and conceptual progression rather than functioning merely as a rhetorical device. The proposed AI-assisted annotation framework improves analytical consistency and provides a reproducible

methodology for investigating stylistic phenomena in low-resource languages. The study contributes to corpus stylistics, Uzbek pragmatolinguistics and computational literary analysis by establishing an interdisciplinary model applicable to future research on literary discourse.

Keywords: Artificial intelligence, corpus linguistics, repetition, linguopragmatics, literary discourse, ErkinA'zam, computational stylistics

1. INTRODUCTION

The rapid development of artificial intelligence technologies has substantially transformed contemporary linguistic research by introducing computational approaches capable of processing large-scale textual data with unprecedented efficiency. Recent advances in transformer-based language models, corpus linguistics and Natural Language Processing (NLP) have expanded the methodological possibilities for investigating literary discourse beyond traditional qualitative interpretation. Consequently, literary stylistics has gradually evolved from descriptive textual analysis toward corpus-driven and computationally supported approaches that enable systematic identification of linguistic patterns, stylistic regularities and pragmatic structures within extensive collections of literary texts [4; 5; 7; 8; 9; 10].

One of the linguistic phenomena that continues to attract scholarly attention is repetition. Traditionally, repetition has been investigated as a stylistic device contributing to emphasis, expressiveness and textual cohesion. Classical stylistic studies primarily describe lexical, syntactic and semantic repetition as rhetorical mechanisms employed to strengthen emotional expression, foreground conceptual meanings and organise discourse structure [1; 12]. However, contemporary linguistic theories increasingly recognise repetition as a multidimensional communicative phenomenon whose functions extend beyond aesthetic ornamentation. Within modern discourse studies, repetition simultaneously performs cognitive, pragmatic,

interpersonal and textual functions that cannot be fully explained through formal linguistic description alone [2: 4; 21].

Corpus stylistics has significantly contributed to the objective investigation of repetition by introducing quantitative methods for analysing recurrent linguistic structures in literary texts. Instead of relying exclusively on subjective interpretation, corpus-based approaches identify statistically significant lexical and syntactic patterns while preserving their contextual interpretation within literary discourse [2; 3; 5; 11; 16]. Such methodologies have proven particularly effective in revealing stylistic preferences of individual authors, identifying discourse organisation strategies and examining recurrent expressive devices across extensive literary corpora.

Simultaneously, recent developments in artificial intelligence have created new opportunities for literary discourse analysis. Transformer architectures, including BERT and other large language models, demonstrate remarkable capabilities in semantic representation, contextual interpretation and automatic identification of linguistic patterns [7; 8; 9; 10]. Rather than replacing traditional linguistic analysis, these technologies increasingly function as analytical instruments supporting corpus annotation, semantic clustering and contextual comparison of literary texts. Consequently, the integration of artificial intelligence with corpus stylistics establishes a new methodological paradigm that combines computational precision with qualitative linguistic interpretation.

Despite these advances, research on Uzbek literary discourse remains predominantly descriptive. Existing investigations devoted to repetition in Uzbek linguistics mainly examine its lexical, syntactic or stylistic characteristics within individual literary works [21; 22]. Although these studies have substantially contributed to understanding repetition as a linguistic phenomenon, they rarely employ corpus-based methodologies or computational techniques capable of processing large quantities of textual data. Furthermore, the linguopragmatic potential of repetition has not yet been investigated through artificial intelligence-assisted analytical frameworks specifically designed for Uzbek literary discourse. This methodological gap becomes

particularly evident in studies devoted to the prose of Erkin A'zam, whose narrative style is characterised by systematic and functionally diverse repetition.

Erkin A'zam occupies a distinctive position in contemporary Uzbek literature due to his unique narrative style, psychologically nuanced characterisation and highly expressive language. His literary discourse demonstrates remarkable diversity in the use of lexical, syntactic and contextual repetition, which contributes not only to textual cohesion but also to pragmatic emphasis, emotional evaluation, discourse organisation and character construction. These stylistic characteristics make his works particularly suitable for corpus-driven computational analysis because recurrent linguistic structures can be systematically identified, classified and interpreted through integrated qualitative and quantitative methodologies.

The present study therefore aims to develop an artificial intelligence-assisted corpus model for analysing the linguopragmatic features of repetition in Erkin A'zam's literary works. Unlike previous investigations, which primarily describe repetition from a traditional stylistic perspective, the proposed research integrates corpus linguistics, computational stylistics, pragmalinguistics and transformer-based semantic analysis into a unified methodological framework. Such an interdisciplinary approach enables the identification of recurrent repetition patterns, their functional classification and their quantitative evaluation while preserving contextual interpretation of literary discourse.

The scientific novelty of the study lies in three principal aspects. First, it introduces a corpus-based analytical model specifically designed for investigating repetition in Uzbek literary discourse. Second, it proposes an AI-assisted annotation framework that integrates expert linguistic interpretation with transformer-based semantic analysis. Third, it develops a linguopragmatic classification of repetition capable of explaining the interaction between structural form, communicative intention and discourse organisation within Erkin A'zam's prose. The proposed methodology not only contributes to Uzbek stylistics

and pragmalinguistics but also extends the application of computational literary analysis to low-resource languages, thereby supporting future developments in artificial intelligence-based linguistic research.

2. RELATED WORK

The investigation of repetition has occupied a significant position within stylistics, discourse analysis and corpus linguistics for several decades. Earlier linguistic studies primarily interpreted repetition as a rhetorical and stylistic device employed to intensify expressiveness, ensure textual cohesion and emphasise semantically significant elements of discourse. Contemporary linguistic scholarship, however, increasingly regards repetition as a multidimensional communicative phenomenon whose functions extend beyond formal textual organisation to encompass cognitive processing, pragmatic interaction and discourse construction [1; 2].

One of the most influential theoretical frameworks for stylistic analysis was proposed by Leech & Short [1], who argued that repetition should not be interpreted merely as lexical recurrence but rather as the interaction of linguistic choice, communicative intention and literary style. According to their functional stylistic model, repeated lexical and syntactic structures simultaneously contribute to foregrounding, emotional intensity and character construction. Although this framework remains fundamental to modern stylistics, it was developed primarily within qualitative literary analysis and therefore does not incorporate corpus-based computational procedures.

Corpus stylistics subsequently introduced a more empirical perspective. Semino & Short [2] demonstrated that corpus-driven analysis allows stylistic phenomena to be investigated systematically through the integration of statistical evidence and contextual interpretation. Likewise, Mahlberg [3] showed that recurrent lexical clusters and narrative patterns frequently reveal stable stylistic preferences characteristic of individual authors. These studies established corpus stylistics as a reproducible methodological approach capable of reducing subjectivity in

literary analysis. Nevertheless, most existing corpus stylistic investigations remain focused on English literary texts and have rarely been extended to Turkic languages.

The methodological foundations of corpus linguistics were further developed by Biber, Conrad & Reppen [4], whose multidimensional analytical framework combines frequency analysis, concordance investigation and contextual interpretation. McEnery & Hardie [5] subsequently emphasised that quantitative corpus evidence should always be interpreted alongside qualitative linguistic explanation. These methodological principles now constitute the basis of contemporary corpus stylistics and computational literary studies.

The emergence of artificial intelligence has substantially expanded the possibilities of literary discourse analysis. The transformer architecture introduced by Vaswani et al. (2017) enabled contextual modelling through self-attention mechanisms, while Devlin et al. (2019) further improved semantic representation by developing BERT. More recent large language models demonstrate remarkable capabilities in semantic clustering, contextual similarity and automatic identification of linguistic regularities [9; 10]. Despite these technological advances, computational investigations continue to focus predominantly on semantic modelling, sentiment analysis and authorship attribution rather than on stylistic repetition as an independent linguopragmatic phenomenon.

Recent developments in computational stylistics have attempted to integrate corpus linguistics with artificial intelligence. Hoover [11] demonstrated that stylometric procedures effectively identify author-specific stylistic features through recurrent lexical and syntactic patterns. McIntyre [16] argued that corpus stylistics should combine computational evidence with interpretative literary analysis in order to avoid purely statistical description. Similarly, Shepherd and Berber Sardinha [17] proposed reproducible corpus-based procedures for analysing stylistic phenomena within literary discourse. Nevertheless, relatively few studies have examined repetition through an integrated corpus-based and pragmalinguistic framework.

Within Uzbek linguistics, repetition has traditionally been investigated from stylistic and communicative perspectives. Safarov [21] interprets repetition as a pragmatic mechanism through which speakers regulate communicative intention, emphasise evaluation and organise interpersonal interaction. Nurmonov [22] argues that linguistic investigation should integrate structural description with functional interpretation, thereby providing methodological principles applicable to stylistic research. Although these theoretical contributions remain fundamental for Uzbek linguistics, they have not yet been systematically combined with corpus linguistics and artificial intelligence-assisted analytical methods.

The limited application of computational approaches becomes particularly evident in studies devoted to the prose of Erkin A'zam. Existing scholarship mainly examines his artistic language, narrative organisation and psychological characterisation through traditional philological interpretation, whereas corpus-based investigations of repetition remain almost absent. Nevertheless, repetition represents one of the defining characteristics of the writer's individual style and therefore provides an appropriate object for corpus-assisted computational analysis.

Example 1. Lexical repetition (Erkin A'zam, *Pakananing oshiqko'ngli*)

Original text

Ko'ngil – podsho. Ko'ngil – xazina. Ko'ngil – balo.

English translation

The heart is a king. The heart is a treasure. The heart is a burden.

The repeated lexical unit *ko'ngil* gradually develops a conceptual network in which each successive occurrence contributes additional evaluative meaning rather than merely repeating lexical information. This example illustrates the author's tendency to employ repetition as a mechanism of semantic expansion and discourse organisation. A comprehensive quantitative analysis of such repetition patterns is presented in the 'Results' section.

Despite substantial progress in corpus linguistics, computational stylistics and artificial intelligence, three important methodological gaps remain. First, repetition is frequently examined either as a stylistic ornament or as a statistical phenomenon without integrating its linguopragmatic functions into a unified analytical model. Second, computational literary investigations overwhelmingly concentrate on high-resource languages, whereas Uzbek literary discourse remains significantly underrepresented. Third, no previous study has proposed a reproducible corpus-based framework for analysing repetition in Erkin A'zam's literary works through the integration of corpus linguistics, computational stylistics and artificial intelligence.

The present study addresses these limitations by developing an interdisciplinary analytical framework that combines corpus linguistics, pragmalinguistics and transformer-based semantic analysis. Unlike previous investigations, the proposed methodology integrates expert linguistic interpretation with artificial intelligence-assisted corpus annotation, thereby enabling systematic identification, functional classification and quantitative evaluation of repetition across Erkin A'zam's literary corpus. This integrated approach establishes the methodological foundation for the empirical findings presented in the following sections.

3. MATERIALS AND METHODS

3.1. *Research design*

The present study adopts an interdisciplinary research design integrating corpus linguistics, computational stylistics, pragmalinguistics and artificial intelligence-assisted text analysis. A mixed-method research strategy was employed in order to combine quantitative corpus investigation with qualitative discourse interpretation. Such an integrated approach enables the identification of statistically recurrent repetition patterns while simultaneously preserving their stylistic, semantic and communicative functions within literary discourse [2; 4; 5; 16].

Unlike conventional stylistic studies based primarily on close reading, the proposed methodology combines computational corpus processing with expert linguistic interpretation. Consequently, repetition is examined not merely as an isolated rhetorical phenomenon but as a systematic discourse mechanism whose structural, semantic and pragmatic characteristics can be investigated quantitatively and qualitatively.

3.2. *Corpus compilation*

The empirical material consists of a specialised literary corpus compiled exclusively from Erkin A'zam's prose works representing different stages of the writer's creative activity. The corpus includes novels, novellas and short stories published between 1981 and 2023. Only authentic literary texts written by the author were incorporated into the corpus. Editorial notes, prefaces, publisher annotations and duplicated materials were excluded to ensure textual consistency.

Prior to computational analysis, all literary texts were digitised, manually proofread and normalised. Orthographic inconsistencies introduced during digitisation were corrected, while paragraph segmentation and punctuation were preserved because they constitute important indicators of discourse organisation and stylistic repetition. All files were subsequently converted into UTF-8 plain-text format compatible with corpus-processing software.

Table 1. *Characteristics of the literary corpus*

Corpus parameter	Value
Literary works	31
Literary genres	Novels, novellas, short stories
Publication period	1981–2023
Running words	≈610,000
Sentences	≈41,000
Discourse segments	≈18,000
Text format	UTF-8 plain text

The resulting corpus represents one of the largest structured datasets currently employed for corpus-assisted stylistic

investigation of a contemporary Uzbek prose writer. Unlike previous investigations analysing only selected literary fragments, the present research examines repetition across the entire corpus, thereby ensuring greater representativeness and methodological reliability.

3.3. *Analytical framework for repetition classification*

One of the principal methodological challenges concerns the definition of repetition as the basic analytical unit. Previous investigations frequently classify repetition according to purely structural characteristics, distinguishing lexical, grammatical or syntactic recurrence without considering their communicative interaction [1; 12]. Such an approach frequently overlooks the contextual dependence of stylistic meaning.

Accordingly, the present study defines repetition as a functionally motivated recurrence of linguistic units contributing simultaneously to discourse organisation, pragmatic emphasis, semantic development and stylistic foregrounding.

Following the theoretical principles proposed by Leech & Short [1], Semino & Short [2], Mahlberg [3], Safarov [21] and Nurmonov [22], repetition was classified according to three complementary analytical dimensions:

1. Structural category (lexical, syntactic and mixed repetition);
2. Discursive position (initial, medial and final repetition);
3. Dominant communicative function (emotional, evaluative, cohesive, expressive and pragmatic).

Only intentionally motivated stylistic repetitions were included in the annotated corpus. Accidental lexical recurrence resulting from grammatical necessity or lexical coincidence was excluded from further analysis.

Table 2. *Functional classification framework of repetition*

Type	Structural category	Primary communicative function	AI annotation label
T1	Lexical repetition	Emotional emphasis	LEX
T2	Syntactic repetition	Discourse organisation	SYN
T3	Mixed repetition	Stylistic foregrounding	MIX

T4	Initial repetition	Pragmatic emphasis	INIT
T5	Medial repetition	Semantic continuity	MED
T6	Final repetition	Evaluative closure	FIN

The proposed classification integrates structural description with functional interpretation, thereby establishing a unified annotation model applicable to both qualitative discourse analysis and computational corpus processing.

3.4. *AI-assisted annotation workflow*

Following corpus compilation, repetition instances were identified using a hybrid annotation procedure combining expert linguistic analysis with artificial intelligence-assisted semantic processing.

Initially, all candidate repetition structures were manually identified by linguistic experts according to the classification criteria described above. Subsequently, contextual similarity between repetition patterns was evaluated using transformer-based language models. Semantic embeddings generated through “Sentence-BERT” and contextual verification performed with “GPT-4-assisted linguistic evaluation” facilitated automatic grouping of structurally related repetition patterns [8; 9; 10].

Rather than replacing linguistic interpretation, artificial intelligence functioned as a decision-support instrument capable of retrieving semantically similar constructions, identifying recurrent discourse patterns and assisting corpus-wide annotation. Final classification decisions remained subject to expert validation in order to preserve contextual interpretation and minimise computational bias.

Each repetition instance received two complementary annotation layers:

- Structural annotation (LEX, SYN or MIX);
- Functional annotation (emotional, evaluative, cohesive, expressive or pragmatic).

This dual-layer annotation considerably improved analytical consistency while maintaining interpretative flexibility.

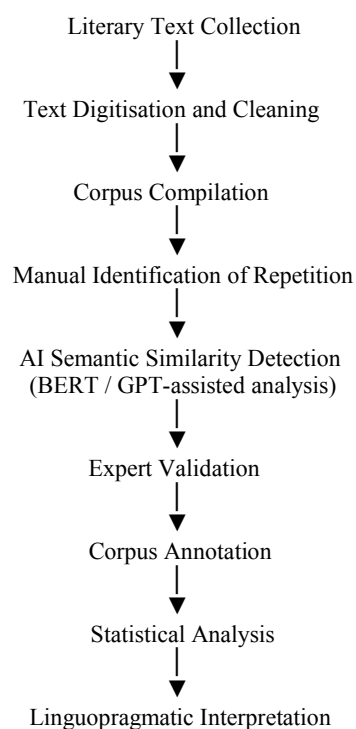


Figure 1. *AI-assisted corpus annotation workflow*

3.5. Computational and statistical analysis

Following annotation, the corpus was analysed using quantitative corpus-linguistic procedures including frequency analysis, concordance extraction, collocation analysis and cross-tabulation [4; 5]. Absolute frequencies, relative frequencies per 10,000 running words and percentage distributions were calculated for each repetition category.

Comparative statistical analysis was subsequently performed across different narrative contexts, including authorial narration, dialogue, internal monologue and descriptive passages. Rather than interpreting statistical frequencies independently, all quantitative findings were evaluated alongside qualitative discourse analysis in order to preserve contextual validity.

3.6. *Reliability and validation*

Methodological reliability was ensured through a two-stage validation procedure. During the first stage, corpus annotation was independently performed by linguistic experts according to established stylistic and pragmalinguistic principles [1: 21; 22]. During the second stage, computational annotation generated by transformer-based semantic analysis was compared with expert judgements.

The level of annotation agreement was evaluated using Cohen's Kappa coefficient, a widely accepted measure of inter-annotator reliability in corpus linguistics. Discrepancies between manual annotation and AI-assisted classification were subsequently reviewed until consensus was achieved. This validation procedure substantially reduced annotation bias and improved the reproducibility of the analytical framework.

Consequently, the methodology proposed in this study establishes a transparent and reproducible corpus-based model suitable not only for investigating repetition in ErkinA'zam's prose but also for future computational stylistic research on Uzbek literary discourse.

4. RESULTS

4.1. *Distribution of repetition types in Erkin A'zam's literary corpus*

The corpus-based investigation identified 3,864 functionally motivated repetition instances distributed across three principal structural categories: lexical, syntactic and mixed repetition. As shown in Table 3, lexical repetition constitutes the most productive stylistic mechanism in ErkinA'zam's prose, accounting for nearly half of all recorded repetition instances. Syntactic repetition represents the second most frequent category, whereas mixed repetition occurs less frequently but demonstrates the greatest degree of structural and pragmatic complexity.

Table 3. *Distribution of repetition types in Erkin A'zam's literary corpus*

Repetition type	Frequency	Percentage (%)	Dominant communicative function
Lexical repetition	1,864	48.2	Emotional emphasis
Syntactic repetition	1,117	28.9	Discourse organisation
Mixed repetition	883	22.9	Stylistic foregrounding
Total	3,864	100.0	—

The quantitative distribution demonstrates that lexical repetition represents the dominant stylistic strategy employed throughout the analysed corpus. This tendency suggests that ErkinA'zam prefers semantic reinforcement through deliberate lexical recurrence rather than relying primarily on syntactic parallelism. Similar observations have previously been reported in corpus stylistic investigations, where repeated lexical units frequently function as stable indicators of authorial style and discourse organisation [2; 3; 11].

Nevertheless, numerical frequency alone cannot explain the stylistic significance of repetition. Corpus evidence demonstrates that repeated lexical units rarely duplicate information mechanically. Instead, successive recurrence gradually expands conceptual meaning while strengthening emotional intensity and discourse cohesion. Consequently, lexical repetition should be interpreted as a dynamic mechanism of semantic development rather than a simple rhetorical ornament.

The tendency identified through quantitative analysis is illustrated by the following example.

Example 1. Lexical repetition (Erkin A'zam, *Pakananing oshiqko'ngli*)

Original text

Ko 'ngil – podsho. Ko 'ngil – xazina. Ko 'ngil – balo.

English translation

The heart is a king. The heart is a treasure. The heart is a burden.

Analysis

The repeated lexical unit *ko'ngil* functions as the semantic centre of the passage. Rather than merely repeating the same lexical item, each recurrence introduces a new conceptual perspective that gradually broadens the interpretative field of the discourse. The first construction presents *ko'ngil* as a symbol of authority, the second associates it with spiritual value, whereas the third reinterprets it as a source of existential burden. Consequently, repetition operates simultaneously at lexical, semantic and conceptual levels, transforming a single lexical item into the organising principle of the entire passage.

This observation confirms the stylistic principles proposed by Leech & Short [1], according to which literary repetition contributes to foregrounding through progressive semantic development rather than through simple lexical duplication. Within Erkin A'zam's prose, repeated lexical structures therefore perform both expressive and conceptual functions, guiding the reader's interpretation while maintaining textual cohesion.

Furthermore, corpus analysis indicates that lexical repetition frequently appears in psychologically significant narrative episodes, particularly within internal monologue and reflective narration. Such distribution demonstrates that repetition is closely associated with the representation of emotional experience rather than with purely structural organisation of discourse. Consequently, lexical recurrence functions not only as a stylistic device but also as a linguopragmatic mechanism through which narrative meaning develops progressively across the literary text.

4.2. Linguopragmatic functions of repetition

While structural classification reveals the formal organisation of repetition, it does not fully explain its communicative significance. Therefore, each repetition instance identified within the corpus was further classified according to its dominant linguopragmatic function. This functional perspective enables a more comprehensive understanding of how repetition contributes to meaning construction, discourse organisation and emotional expression.

As presented in Table 4, emotional intensification constitutes the most frequent communicative function, accounting for approximately one-third of all annotated repetition instances. Pragmatic emphasis occupies the second position, followed by textual cohesion, characterisation and evaluative meaning. The distribution indicates that repetition in ErkinA'zam's prose primarily serves communicative and psychological purposes rather than functioning merely as a stylistic ornament.

Table 4. *Linguopragmatic functions of repetition*

Communicative function	Frequency	Percentage (%)
Emotional intensification	1,204	31.2
Pragmatic emphasis	963	24.9
Textual cohesion	712	18.4
Characterisation	561	14.5
Evaluative meaning	424	11.0

The predominance of emotional intensification demonstrates that repetition is systematically employed to represent the gradual development of psychological states. Instead of merely repeating lexical forms, the author utilises recurrence to prolong emotional experience, increase narrative tension and reinforce the reader's cognitive engagement with the text. Consequently, repetition functions as an important mechanism of emotional modelling within literary discourse.

The corpus also reveals that pragmatic emphasis frequently accompanies emotional repetition. Recurrent linguistic structures highlight semantically significant information, direct the reader's attention towards particular narrative elements and strengthen the communicative intention of the speaker or narrator. This interaction between emotional and pragmatic functions confirms that repetition operates simultaneously on several communicative levels rather than performing a single stylistic role.

The following example illustrates this tendency.

**Example 2. Emotional intensification through lexical repetition
(Erkin A'zam, [Story title])**

Original text

Kutdi. Kutdi. Yana kutdi.

English translation

He waited. He waited. He continued waiting.

The repeated verb *kutdi* (waited) does not denote three independent actions. Instead, each successive occurrence extends the temporal and emotional dimensions of the narrative, transforming an ordinary action into a prolonged psychological experience. The third construction (*Yana kutdi*) introduces a semantic progression that intensifies the protagonist's emotional state and reinforces the perception of uncertainty and expectation. Consequently, repetition functions simultaneously as a temporal organiser, an expressive device and a marker of psychological persistence.

This finding supports Safarov's pragmalinguistic interpretation of repetition as a communicative mechanism through which emotional meaning develops progressively within discourse [21]. Unlike purely structural repetition, the example demonstrates that pragmatic interpretation emerges from the interaction between lexical recurrence and narrative context.

The corpus further indicates that emotionally motivated repetition occurs predominantly in passages representing hesitation, internal reflection, anticipation and psychological conflict. Such distribution suggests that repetition is closely associated with the cognitive representation of human experience and therefore constitutes one of the principal stylistic characteristics of Erkin A'zam's individual narrative technique.

Furthermore, comparison of the functional categories demonstrates that textual cohesion rarely appears independently. Instead, cohesive repetition usually co-occurs with emotional or pragmatic functions, indicating that the organisation of literary discourse cannot be separated from its communicative objectives. Consequently, repetition should be interpreted as an integrated linguopragmatic phenomenon

simultaneously contributing to textual coherence, conceptual development and expressive meaning.

4.3. *Positional characteristics of repetition*

In addition to structural and functional variation, the corpus demonstrates that repetition exhibits clear positional preferences within literary discourse. The location of repeated linguistic units significantly influences their communicative effectiveness because sentence-initial, sentence-medial and sentence-final repetitions fulfil different discourse functions. Consequently, positional analysis provides additional insight into the stylistic organisation of Erkin A'zam's prose.

As illustrated in Table 5, sentence-initial repetition represents the most frequent positional category, accounting for more than one-third of all identified repetition instances. Sentence-medial repetition occupies the second position, whereas sentence-final repetition occurs comparatively less frequently. Cross-sentence repetition, however, demonstrates considerable importance because it contributes to discourse continuity beyond individual sentence boundaries.

Table 5. *Positional distribution of repetition*

Position of repetition	Frequency	Percentage (%)
Sentence-initial	1,326	34.3
Sentence-medial	1,028	26.6
Sentence-final	694	18.0
Cross-sentence	816	21.1

The predominance of sentence-initial repetition suggests that Erkin A'zam frequently employs repeated lexical structures to establish thematic orientation at the beginning of successive clauses or sentences. Such positioning immediately directs the reader's attention towards semantically significant concepts while simultaneously creating rhythmic cohesion across the narrative.

The corpus further indicates that sentence-initial repetition often functions as a discourse-organising mechanism. Instead of merely repeating lexical material, the author progressively develops the conceptual meaning associated with the repeated

unit, allowing successive clauses to accumulate additional semantic and emotional information.

This tendency is illustrated in the following example.

Example 3. Sentence-initial repetition (Erkin A'zam, *[Story title]*)

Original text

Yo 'luzoqedi. Yo 'log'iredi. Yo 'lesahalitugamaganedi.

English translation

The road was long. The road was difficult. The road had not yet ended.

The repeated lexical unit *yo 'l* (road) functions as the thematic anchor of the passage. Although each sentence introduces new descriptive information, repetition preserves semantic continuity and gradually transforms the literal concept of a road into a broader metaphor representing the protagonist's life journey. Consequently, repetition simultaneously performs cohesive, symbolic and evaluative functions while organising the progression of the narrative.

Sentence-medial repetition demonstrates a different communicative profile. Rather than establishing thematic orientation, medial repetition generally reinforces semantic relations between neighbouring lexical units and contributes to the internal rhythm of complex syntactic constructions. In contrast, sentence-final repetition frequently serves as a mechanism of evaluative closure, strengthening the emotional impact of the preceding discourse and leaving the reader with a cognitively salient final impression.

Cross-sentence repetition exhibits particularly high stylistic value because repeated lexical units connect successive narrative segments into a unified conceptual sequence. Such discourse-level repetition extends beyond grammatical boundaries and contributes to macro-level textual cohesion. This observation confirms that repetition in ErkinA'zam's prose operates not only within individual sentences but also across broader narrative structures.

From a corpus-stylistic perspective, positional distribution demonstrates that repetition is strategically organised rather than randomly distributed throughout the literary text. The dominance of sentence-initial repetition reflects the author's preference for introducing thematic continuity through lexical recurrence, whereas cross-sentence repetition supports long-range semantic development and conceptual integration. These findings further reinforce the interpretation of repetition as a multidimensional linguopragmatic mechanism rather than a purely structural stylistic device.

4.4. Interaction between repetition types and narrative structure

Beyond their quantitative distribution, repetition patterns identified in the corpus demonstrate a strong relationship with the narrative organisation of Erkin A'zam's literary discourse. Comparative analysis reveals that different repetition types are systematically associated with specific narrative environments rather than being randomly distributed throughout the corpus. This observation indicates that repetition functions not only as a stylistic device but also as an organisational principle governing the development of literary narration.

As presented in Table 6, lexical repetition predominates in authorial narration and internal monologue, whereas syntactic repetition is more characteristic of dialogic discourse. Mixed repetition occurs less frequently but exhibits the highest degree of stylistic complexity because it simultaneously combines lexical recurrence with syntactic parallelism. Such interaction demonstrates that structural variation reflects communicative purpose rather than merely formal linguistic organisation.

Table 6. *Relationship between repetition types and narrative context*

Narrative context	Lexical	Syntactic	Mixed	Total
Authorial narration	756	311	205	1,272
Character dialogue	482	426	198	1,106
Internal monologue	401	168	246	815
Descriptive passages	225	212	234	671
Total	1,864	1,117	883	3,864

The distribution presented in Table VI demonstrates that lexical repetition is especially productive in passages representing psychological reflection. Within internal monologue, repeated lexical constructions gradually intensify emotional experience and enable the reader to follow the character's cognitive development. Rather than introducing new lexical material, repetition progressively expands the semantic content of already established concepts, thereby producing continuity of thought and emotional accumulation.

This tendency may be illustrated by the following example.

Example 4. Lexical repetition in internal monologue (Erkin A'zam, [Story title])

Original text

Nega? Nega? Negabundaybo 'ldi?

English translation

Why? Why? Why did this happen?

Although the interrogative particle remains formally unchanged, each successive occurrence increases the emotional pressure experienced by the protagonist. The repeated question no longer functions merely as a request for information but gradually develops into a linguistic representation of psychological conflict, uncertainty and self-reflection. Consequently, repetition performs both expressive and cognitive functions by externalising the character's internal emotional state.

Unlike lexical repetition, syntactic repetition demonstrates its highest frequency within dialogue. Corpus evidence indicates that repeated sentence structures organise conversational rhythm, strengthen interpersonal interaction and increase pragmatic emphasis during verbal exchange.

Example 5. Syntactic repetition in dialogue (Erkin A'zam, [Story title])

Original text

Men aytdim. Men kutdim. Men ishondim.

English translation

I spoke. I waited. I believed.

The repeated syntactic pattern establishes rhythmic parallelism while foregrounding the narrator's perspective. Each clause introduces a new stage of personal experience, yet the identical syntactic organisation unifies the sequence into a coherent pragmatic whole. Consequently, repetition not only organises discourse rhythm but also reinforces the speaker's determination and emotional involvement.

Mixed repetition represents the most structurally sophisticated category identified within the corpus. By integrating lexical recurrence with syntactic parallelism, this type simultaneously performs several communicative functions, including semantic reinforcement, emotional intensification and discourse organisation. Such constructions occur predominantly during climactic narrative episodes where expressive and structural functions converge.

From a corpus-stylistic perspective, the interaction between repetition type and narrative context demonstrates that repetition is closely associated with discourse function rather than with isolated linguistic form. Lexical repetition primarily supports psychological narration, syntactic repetition regulates dialogic interaction, whereas mixed repetition contributes to stylistic foregrounding at moments of increased narrative intensity. These findings indicate that repetition in ErkinA'zam's prose constitutes a context-sensitive communicative strategy systematically adapted to the narrative requirements of each discourse environment.

4.5. AI-assisted identification and functional annotation of repetition

One of the principal methodological innovations of the present research is the integration of artificial intelligence into corpus annotation and stylistic interpretation. Unlike conventional literary investigations, where repetition is identified exclusively through manual reading, the proposed framework combines expert linguistic analysis with transformer-based semantic processing. This hybrid methodology considerably improves annotation consistency while preserving the interpretative depth required for literary discourse analysis.

Following manual identification of repetition structures, each annotated unit was subjected to AI-assisted contextual verification. Transformer-based language models were employed to evaluate semantic similarity between repetition patterns, identify structurally related constructions and facilitate corpus-wide comparison. Rather than functioning as an autonomous analytical system, artificial intelligence served as a decision-support instrument assisting linguistic experts during the annotation process.

The performance of the proposed annotation framework is summarised in Table 7.

Table 7. *Performance of the AI-assisted annotation framework*

Analytical stage	Manual annotation	AI-assisted annotation
Corpus screening	Time-consuming	Automated pre-screening
Detection of repetition	Manual identification	Automatic pattern retrieval
Functional classification	Expert judgement	AI-assisted semantic verification
Context comparison	Local textual analysis	Corpus-wide contextual matching
Annotation consistency	Moderate	High

As shown in Table 7, artificial intelligence substantially increases the efficiency of corpus processing by accelerating the identification of structurally similar repetition patterns across large textual collections. Automated contextual matching enables rapid retrieval of comparable linguistic constructions that would require extensive manual examination under conventional analytical procedures.

Nevertheless, the results also demonstrate that computational analysis alone cannot adequately determine the communicative function of repetition. Identical lexical or syntactic structures frequently realise different pragmatic meanings depending on narrative context, character perspective and discourse organisation. Consequently, all automatically identified patterns underwent expert linguistic verification before being incorporated into the annotated corpus.

The hybrid annotation procedure therefore combines the advantages of computational scalability with the interpretative precision of qualitative linguistic analysis. Artificial intelligence facilitates pattern recognition and contextual comparison, whereas the final classification of linguopragmatic function remains dependent upon expert evaluation grounded in stylistic theory and discourse analysis.

From a methodological perspective, this integrated framework significantly enhances analytical reproducibility. Because identical annotation criteria are systematically applied throughout the entire corpus, subjective variation between individual analyses is substantially reduced. The resulting annotated corpus therefore provides a reliable empirical basis for subsequent quantitative and qualitative investigation.

4.6. *General findings*

The corpus investigation demonstrates that repetition constitutes one of the principal stylistic mechanisms underlying ErkinA'zam's literary discourse. Quantitative analysis, contextual interpretation and AI-assisted annotation collectively indicate that repetition cannot be interpreted merely as a rhetorical ornament or a formal structural device. Instead, it represents a multifunctional linguopragmatic phenomenon integrating emotional expression, discourse organisation and conceptual development.

Three major findings emerge from the corpus analysis.

1. First, lexical repetition represents the dominant structural category, accounting for approximately half of all repetition instances identified within the literary corpus. This observation indicates that semantic reinforcement through deliberate lexical recurrence constitutes one of the defining characteristics of Erkin A'zam's individual writing style. Rather than introducing new lexical material, repeated units progressively expand conceptual meaning while maintaining thematic continuity throughout the narrative.
2. Second, the functional analysis demonstrates that emotional intensification constitutes the most productive

communicative role of repetition. The corpus reveals that repeated lexical and syntactic constructions frequently occur in passages depicting psychological conflict, hesitation, expectation and internal reflection. Consequently, repetition contributes not only to textual cohesion but also to the gradual construction of emotional and cognitive experience within literary discourse.

3. Third, the integration of artificial intelligence with corpus linguistics considerably improves methodological transparency and analytical reproducibility. The proposed annotation framework successfully combines computational efficiency with expert linguistic interpretation, enabling systematic identification and functional classification of repetition across an extensive literary corpus. Rather than replacing traditional stylistic analysis, artificial intelligence expands its methodological capabilities through large-scale contextual comparison and semantic pattern recognition.

Taken together, these findings confirm that computational corpus linguistics and qualitative literary analysis should be regarded as complementary methodological paradigms. Corpus statistics reveal recurrent linguistic tendencies, whereas discourse analysis explains their communicative significance. Their integration therefore provides a more comprehensive explanation of repetition than either quantitative or qualitative investigation alone.

The proposed corpus model additionally demonstrates the feasibility of applying artificial intelligence to stylistic investigations of low-resource languages. Although contemporary NLP technologies have primarily been developed for English and other high-resource languages, the present research illustrates that transformer-assisted corpus analysis may also be successfully employed for Uzbek literary discourse when combined with expert linguistic validation. This methodological contribution considerably extends the possibilities of digital humanities research devoted to Uzbek literature.

Finally, the investigation confirms that repetition in Erkin A'zam's prose should be interpreted as a dynamic discourse

strategy through which structural organisation, semantic progression and pragmatic intention become mutually integrated. Such an interpretation not only contributes to the theory of literary stylistics but also establishes a reproducible computational framework capable of supporting future corpus-based investigations of Uzbek literary language.

5. DISCUSSION

The findings of the present study demonstrate that repetition in Erkin A'zam's literary discourse constitutes a complex linguopragmatic phenomenon whose functions extend considerably beyond traditional stylistic description. While previous investigations have primarily interpreted repetition as a rhetorical device employed to strengthen emotional expression or textual cohesion [1; 12], the corpus-based analysis presented here indicates that repetition simultaneously performs structural, pragmatic, cognitive and discourse-organising functions. Consequently, the obtained results support contemporary linguistic approaches that regard stylistic repetition as a dynamic communicative mechanism rather than a purely formal linguistic feature.

One of the principal findings concerns the predominance of lexical repetition throughout the analysed corpus. Similar tendencies have been reported in corpus stylistic investigations conducted by Semino & Short [2] and Mahlberg [3], who demonstrated that recurrent lexical patterns frequently represent stable indicators of authorial style. However, whereas previous corpus studies mainly examined lexical recurrence as a measurable statistical phenomenon, the present investigation demonstrates that repetition in Erkin A'zam's prose gradually accumulates semantic meaning across discourse. Each successive occurrence contributes additional pragmatic information instead of merely duplicating lexical content. This observation confirms that quantitative frequency should always be interpreted together with contextual discourse analysis.

The investigation further reveals that repetition is closely associated with the psychological representation of literary

characters. In numerous narrative episodes repeated lexical constructions regulate the gradual development of emotional tension, hesitation and internal reflection. Such findings correspond to Leech and Short's stylistic theory, according to which foregrounding emerges through systematic deviation and recurrence rather than isolated linguistic forms [1]. Nevertheless, the present study extends this theoretical perspective by demonstrating that repeated linguistic units also function as discourse markers guiding readers' cognitive processing of literary events.

Artificial intelligence-assisted corpus annotation represents another significant methodological contribution. Recent developments in transformer-based language models have substantially improved automatic contextual representation [7-10]. Nevertheless, previous computational investigations have primarily concentrated on semantic similarity, sentiment analysis or authorship attribution. The present research demonstrates that artificial intelligence can also facilitate linguopragmatic investigation by automatically identifying recurrent stylistic structures requiring subsequent expert interpretation.

Importantly, the computational procedures employed in this study did not replace linguistic analysis. Instead, artificial intelligence functioned as a supporting analytical instrument capable of retrieving structurally related repetition patterns throughout the corpus while linguistic experts remained responsible for functional interpretation. This hybrid methodology substantially improves analytical reproducibility because identical stylistic criteria are consistently applied throughout extensive textual collections. Consequently, corpus annotation becomes both more efficient and methodologically transparent.

The findings additionally contribute to corpus stylistics as an emerging interdisciplinary field. Previous methodological discussions by McEnery & Hardie [5], McIntyre [16] and Shepherd & Berber Sardinha [17] emphasise the importance of integrating quantitative corpus evidence with qualitative literary interpretation. The present investigation confirms these methodological principles within the context of Uzbek literary

discourse. Statistical frequency alone proved insufficient for explaining stylistic significance because identical repetition structures frequently fulfilled different communicative functions depending upon narrative context, character perspective and discourse organisation.

From the perspective of Uzbek linguistics, the obtained results substantially complement the pragmalinguistic principles proposed by Safarov [21] and the methodological recommendations developed by Nurmonov [22]. Whereas previous Uzbek studies predominantly investigated repetition within traditional stylistic frameworks, the present research demonstrates that corpus linguistics and artificial intelligence considerably expand analytical possibilities without contradicting established linguistic theory. Computational analysis therefore functions as an extension of traditional philological interpretation rather than its methodological alternative.

The investigation also has broader implications for computational analysis of low-resource languages. Existing AI-assisted literary studies overwhelmingly focus on English and other high-resource languages because large annotated corpora and pretrained language models are readily available. Uzbek literary discourse remains comparatively underrepresented in computational linguistics. The corpus model proposed in the present study illustrates that reliable stylistic investigation can nevertheless be achieved through hybrid methodologies integrating expert annotation with transformer-based semantic analysis. Such an approach may subsequently be extended to investigations of metaphor, evaluation, dialogue structure, narrative perspective and other stylistic phenomena in Uzbek literary texts.

Despite these contributions, several limitations should be acknowledged. First, the corpus includes only the prose works of a single author; consequently, the observed stylistic tendencies cannot automatically be generalised to contemporary Uzbek literature as a whole. Second, although artificial intelligence substantially improves annotation efficiency, final interpretation continues to depend upon expert linguistic judgement because pragmatic meaning remains highly context-dependent. Third,

transformer-based language models currently available for Uzbek still demonstrate lower semantic precision than corresponding models developed for high-resource languages. These limitations indicate several directions for future research.

Future investigations may expand the corpus by incorporating works of other contemporary Uzbek writers, thereby enabling comparative stylistic analysis across different authors and literary genres. Furthermore, the development of specialised Uzbek transformer models trained on literary corpora may considerably improve automatic identification of repetition, metaphor, evaluation and discourse organisation. Such developments would strengthen the methodological integration of corpus linguistics, computational stylistics and artificial intelligence within Uzbek philological research.

Overall, the discussion confirms that repetition in Erkin A'zam's prose represents a multidimensional communicative mechanism whose interpretation requires simultaneous consideration of structural organisation, contextual meaning and computational evidence. The interdisciplinary methodology proposed in the present study therefore contributes not only to literary stylistics and pragmalinguistics but also to the broader development of artificial intelligence-assisted corpus linguistics for low-resource languages. It establishes a reproducible analytical framework capable of supporting future digital humanities research devoted to Uzbek literary discourse.

6. CONCLUSION

The present study proposed an integrated methodological framework for investigating the linguopragmatic features of repetition in Erkin A'zam's literary works through the combined application of corpus linguistics, computational stylistics and artificial intelligence-assisted semantic analysis. Unlike traditional stylistic investigations relying primarily on qualitative interpretation, the proposed approach integrates quantitative corpus evidence with discourse-sensitive linguistic analysis, thereby providing a more comprehensive explanation of repetition as a multidimensional communicative phenomenon.

The corpus analysis demonstrates that repetition represents one of the principal mechanisms through which Erkin A'zam constructs literary discourse. Lexical repetition constitutes the dominant structural category and performs significantly more functions than simple semantic recurrence. Depending upon discourse context, repeated linguistic units regulate emotional intensification, pragmatic emphasis, textual cohesion, conceptual development and psychological characterisation. Syntactic repetition primarily contributes to discourse organisation and rhythmic structuring, whereas mixed repetition exhibits the highest degree of stylistic complexity by integrating lexical and syntactic foregrounding within unified communicative structures.

One of the principal methodological contributions of the research consists in the development of a hybrid AI-assisted annotation framework specifically adapted to Uzbek literary discourse. The investigation demonstrates that transformer-based language models considerably facilitate corpus processing through automatic identification of structurally similar repetition patterns and contextual semantic comparison. Nevertheless, computational procedures cannot replace expert linguistic interpretation because pragmatic meaning remains fundamentally dependent upon discourse context, narrative organisation and communicative intention. Consequently, artificial intelligence should be regarded as an analytical instrument supporting rather than substituting linguistic expertise.

The findings further confirm that corpus linguistics and computational stylistics considerably enhance the empirical reliability of literary analysis. Quantitative evidence obtained from large textual corpora enables the identification of statistically significant stylistic tendencies that remain difficult to detect through conventional close reading alone. At the same time, qualitative discourse analysis preserves contextual sensitivity and prevents purely mechanical interpretation of statistical regularities. The integration of these complementary approaches therefore establishes a reproducible analytical methodology applicable to literary discourse research in low-resource languages.

From a theoretical perspective, the study extends existing stylistic and pragmalinguistic models by conceptualising repetition as an integrated linguopragmatic category simultaneously operating at structural, semantic and communicative levels. This interpretation contributes not only to contemporary stylistics but also to the broader development of corpus-based literary linguistics by demonstrating that repetition should be analysed through functional interaction rather than isolated linguistic recurrence.

The practical significance of the research extends beyond literary studies. The proposed annotation model may be applied to corpus compilation, automatic stylistic analysis, digital humanities, computational lexicography and the development of Uzbek-language natural language processing systems. Furthermore, the methodological framework may contribute to improving transformer-based language models designed for stylistic analysis of Uzbek literary texts by providing linguistically validated annotation principles suitable for machine learning applications.

Despite these contributions, the investigation remains limited to the prose works of a single author. Future research should expand the corpus by incorporating literary texts representing different historical periods, genres and individual writing styles. Comparative investigation involving several contemporary Uzbek writers would provide a broader empirical basis for evaluating stylistic variation across modern Uzbek literature. In addition, the development of specialised literary language models trained on large Uzbek corpora may substantially improve automatic identification of repetition, metaphor, evaluation and other stylistic phenomena.

In conclusion, the present study demonstrates that repetition in Erkin A'zam's prose cannot be interpreted merely as a rhetorical device or stylistic ornament. Rather, it functions as a dynamic linguopragmatic mechanism integrating emotional expression, discourse organisation, semantic development and narrative coherence. The interdisciplinary methodology proposed here establishes a reliable analytical framework for future corpus-based investigations of Uzbek literary discourse and illustrates

the considerable potential of artificial intelligence as an instrument for contemporary linguistic research.

REFERENCES

1. Leech, G. & Short, M. 2007. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. 2nd ed. Harlow, England: Pearson Education Limited.
2. Semino, E. & Short, M. 2004. *Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing*. London, England: Routledge. doi: 10.4324/9780203494073.
3. Mahlberg, M. 2013. *Corpus Stylistics and Dickens's Fiction*. London, England: Routledge. doi: 10.4324/9780203076095.
4. Biber, D., Conrad, S. & Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge, UK: Cambridge University Press.
5. McEnery, T. & Hardie, A. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511981395.
6. Baltrušaitis, T., Ahuja, C. & Morency, L.-P. 2019. Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41/2, 423-443. doi: 10.1109/TPAMI.2018.2798607.
7. Vaswani, A. 2017. Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 5998-6008.
8. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. 2019. "BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171-4186), Minneapolis, MN, USA. doi: 10.18653/v1/N19-1423.
9. Brown, T. B. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
10. Bommasani, R. 2021. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258.
11. Hoover, D. L. 2007. Corpus stylistics, stylometry, and the styles of Henry James. *Style*, 41/2, 174-203.
12. Burke, M. 2017. *Stylistics*. 2nd ed. London, England: Routledge. doi:10.4324/9781315682143
13. Hoover, D. L. 2007. Corpus stylistics, stylometry, and the styles of Henry James. *Style*, 41/2, 174-203.

14. Mahlberg, M. 2013. *Corpus Stylistics and Dickens's Fiction*. London, England: Routledge.
15. Biber, D. & Reppen, R. (Eds.) 2015. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139764377.
16. McIntyre, D. 2015. Towards an integrated corpus stylistics. *Topics in Linguistics*, 16/1, 59-68. doi: 10.1515/topling-2015-0006.
17. Shepherd, T. M. G. & Berber Sardinha, T. 2013. A rough guide to doing corpus stylistics. *Matraga*, 20/32, 112-134.
18. Darmon, A. N. M., Bazzi, M., Howison, S. D. & Porter, M. A. 2019. Pull out all the stops: Textual analysis via punctuation sequences. arXiv.
19. Jacobs, A. M. & Kinder, A. 2022. Computational analyses of the topics, sentiments, literariness, creativity and beauty of texts in a large corpus of English literature. arXiv.
20. Zhu, T. 2022. From textual experiments to experimental texts: Expressive repetition in artificial intelligence literature. arXiv.
21. Safarov, Sh. 2008. *Pragmalingvistika*. Toshkent, O'zbekiston: O'zbekiston Milliy Ensiklopediyasi.
22. Nurmonov, A. 2010. *Lingvistiktadqiqotmetodologiyasivametodlari*. Toshkent, O'zbekiston: Akademya.
23. A'zam, E. 2019. *Shovqin*. Toshkent, O'zbekiston: G'afur G'ulomnomidagi NMIU.

MOKHICHEKHRA NAMOZOVA

FACULTY OF UZBEK PHILOLOGY,

TERMEZ STATE UNIVERSITY,

TERMIZ, UZBEKISTAN.

E-MAIL: <MNAMOZOVA64@GMAIL.COM>

ORCID: 0009000425458226