

A Morphologically Annotated Database of Uzbek Verbal Phrasemes: Classification Methodology and Empirical Findings

FARANGIZ QOSIMJONOVA
Termez State University, Uzbekistan

ABSTRACT

Although the lexico-semantic description of phraseological units is well-developed in Uzbek linguistics, the systematic description of their morphological features and their representation in an electronic database have remained largely unstudied. The purpose of this study is to morphologically annotate Uzbek verbal phrasemes and to establish principles for adapting the system of grammatical forms, originally developed for single lexemes, to multi-component phraseological units. The research material, compiled from Sh. Rahmatullayev's 1978 and 2022 dictionaries, encompassed 5,439 phraseological units integrated into a unified UGPD database. Using descriptive, component analysis, comparative-typological, and statistical methods, the database was verified in four stages. As a result, 569 of the 4,259 verbal phrasemes were fully morphologically annotated. During the annotation process, grammatical homonymy between the causative voice suffix and the participial form was identified and resolved based on context and syntactic function. Furthermore, 12 units belonging to the historical layer and 43 units classified as "Proverbs" or "Compound words" were excluded from the scope of annotation. The results indicate that UGPD-Verb has created a reusable coding framework for the morphological description of Uzbek verbal phrasemes.

Keywords: Uzbek phraseology, morphological annotation, verbal phrasemes, corpus linguistics, phraseological database, grammatical homonymy

INTRODUCTION

A phraseological unit (idiom) is a stable combination of two or more independent words that expresses a holistic, figurative meaning and functions as a single constituent within a sentence (Rahmatullayev et al. 2022). Such units occupy an intermediate position in the language system between a word and a phrase: they are holistic like a word but, like a phrase, are composed of several components. It is this dual nature-holism and multi-component structure -that makes their grammatical description a more complex issue than that of ordinary words. This is because determining which component (or components) within the unit dictates its grammatical form requires a separate analysis.

In modern Uzbek linguistics, research on idioms has primarily focused on elucidating their meaning, imagery, and stylistic features. Well-developed lexicographical sources in this area have already been created. Specifically, B. Yuldashev (1998), A. Mamatov (2000), and M. Vafoyeva (2009) have researched the stylistic, formational, and synonymic aspects of idioms, respectively. However, alongside this, the grammatical structure of idioms, particularly their morphological features, remains largely unexamined as a distinct and systematic object of research. A similar gap has been noted in Russian linguistics, where it has been emphasized that a phraseological unit should be described not only as a lexical entity but also as a separate grammatical sphere of the language (Khusnutdinov 2015). This line of inquiry in the Russian school of phraseology began with the classical classification of V. V. Vinogradov (Cowie 1998) and continued in the lexico-grammatical research of Shansky (1996) and Kunin (1970), as well as in the conceptual approach of Baranov & Dobrovolsky (2013). Although the functional-discursive and linguacultural approaches also developed extensively after Vinogradov (Teliya 1996), our research is most closely aligned with the lexico-grammatical approach of Shansky and Kunin. Therefore, Rahmatullayev's principle of classification was not formed in isolation but rather within the broader theoretical context of the Russian school of phraseology.

Sh. Rahmatullayev's *Explanatory Phraseological Dictionary of the Uzbek Language* (Rahmatullayev 1978) and its 2022 expanded edition (Rahmatullayev et al. 2022) are currently the most comprehensive lexicographic sources for Uzbek phraseology. They cover a total of more than 5,500 phraseological units, each accompanied by a semantic definition and stylistic characterization (Rahmatullayev et al. 2022). Rahmatullayev (2006) divides these units into five categories—nominal, adjectival, verbal, adverbial, and predicative phrasemes (the principle of this classification is discussed in detail below in the 'Materials and Methods' section) but this classification operates at a categorical level and does not express the precise morphological form of each phrase (e.g., tense, mood, voice) through formal codes. It is precisely this gap between the two levels—categorical classification and precise morphological coding that constitutes the starting point of this study.

In parallel, the first attempts to create a grammatical dictionary have also emerged in Uzbek linguistics in recent years. Drawing on similar grammatical-lexicographic traditions in Russian (Zaliznyak 1987; Okuneva 2000; Tolmacheva 1995) and Turkic (Kenesbaev 1977; Japarov et al. 2015; Isanbet 1989–1990) languages, Hamroyeva, Xoliyorov & Abdualimova developed a system of grammatical forms using the example of a single verb (*kiymoq*) (Hamroyeva et al. 2021). However, this system has so far only been tested within the scope of a single lexeme and has not been applied to phrases, which constitute a more complex layer of the language compared to words.

Thus, the semantic description of phraseology and the system of morphological forms of the verb have developed separately from each other in Uzbek linguistics as two independent fields. Filling this methodological gap requires the annotation of phraseological units based on a system of morphological forms, which in turn necessitates a clear distinction between the base lexical form of a phraseological unit and its actual usage in speech (Naciscione 2010).

This gap is not unique to the Uzbek language. Internationally, the identification and annotation of multiword expressions (MWEs) has become a key area of computational

linguistics, as such units do not fully adhere to compositional analysis and consistently pose challenges for syntactic analysis, machine translation, and other natural language processing tasks. Indeed, they have been described as "a pain in the neck for NLP" (Sag et al. 2002: 1). For instance, the PARSEME initiative has developed general principles for identifying and annotating verbal MWEs in about twenty languages (Savary et al., 2017), but it did not aim for systematic morphological annotation. To our knowledge, a methodological framework based on the systematic morphological annotation of verbal phrasemes for Uzbek and related Turkic languages has not yet been described. This is particularly significant because the highly agglutinative nature of Turkic languages (Johanson & Csató 1998) complicates the direct application of standard annotation schemes (Öztürk et al. 2022; Bedir et al. 2021). In this sense, UGPD-Verb enriches the field of international MWE annotation with a morphological dimension based on Uzbek language material. In the future, it can serve as a linguistic resource for training parsers and machine translation systems, searching for MWEs in a corpus (concordancing), and testing how large language models process Uzbek phraseological units.

In this article, we present the results of the UGPD-Verb (Uzbek Grammatical Phraseological Database, Verb) project. The database was compiled based on Rahmatullayev's 1978 and 2022 dictionaries and currently encompasses 5,439 phraseological units. The portion pertaining to the verb category has been morphologically annotated using the system of grammatical forms developed by Hamroyeva et al. (2021). During this process, two key methodological issues required specific solutions: first, inconsistencies arising from adapting a system designed for a single lexeme to multi-component phraseological units (for example, the homonymy between the causative voice suffix and the participial form) were systematically resolved; second, units from the historical layer in the source dictionary were excluded from the scope of the current annotation to avoid confusing diachrony with synchrony (see the results section for details). The system of grammatical forms developed by Hamroyeva et al. (2021) had previously only been

tested on the example of a single lexeme, the verb *kiymoq* (to wear), and had not yet been applied to multi-component linguistic units. The present study addresses this gap. In this respect, the study establishes a methodological basis for the systematic morphological annotation of Uzbek verbal phrasemes. Its scientific contribution is demonstrated not only in the 569 annotated idioms but also in the set of methodological rules, tested during their creation, which can be replicated by other researchers.

MATERIALS AND METHODS

The primary sources for the study were two editions by Sh. Rahmatullayev: the *Explanatory Phraseological Dictionary of the Uzbek Language* (1978) (Rahmatullayev 1978) and its revised and expanded 2022 edition, co-authored with a team (Mahmudov, Xolmanova, O'razova & Rixsiyeva; Rahmatullayev et al., 2022). Entries from both publications were consolidated into a single electronic database, the UGPD (Uzbek Grammatical Phraseological Database). The database currently contains 5,439 phraseological units. The main difference between the two editions is that the 2022 edition is not only enriched with new units, but some definitions and stylistic annotations from the 1978 edition have also been revised. For this reason, in cases of differing classifications, we have consistently prioritized the classification from the latest, revised edition.

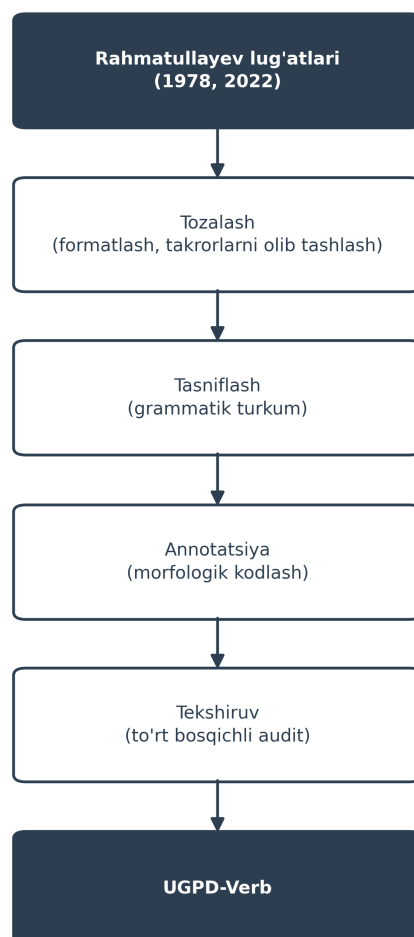


Figure 1. *The process of creating the UGPD-Verb database: from source dictionaries to a verified database*

RESEARCH METHODS

The research is based on the following set of methods:

1. The descriptive method was used to determine the grammatical category of phraseological units and to describe their morphological features;

2. The component analysis method was used to break down each phraseological unit into its constituent parts (a nominal component and a verbal component), and then to determine the morphological structure of each part separately;
3. The comparative-typological method was used to identify the differences between the 1978 and 2022 editions;
4. The statistical analysis method was utilized to quantitatively calculate and express in percentages the distribution of categories in the database, as well as the internal ratios of verb forms (finite verb, verbal noun, gerund, participle).

Each phrase was first classified according to its grammatical category and divided into verbal, adjectival, adverbial, nominal, interjective, and modal phrases. The semantic criterion was adopted as the primary measure for classification; that is, in determining the categorical belonging of a phraseological unit, the general categorical meaning of the entire unit served as the basis, rather than the grammatical form of the individual words within it. This approach is fully consistent with the principle advanced by Rahmatullayev in his work *Modern Literary Uzbek Language* (Rahmatullayev 2006).

Rahmatullayev (2006) divides phrasemes into two major groups, primarily based on the categorical seme inferred from their semantic content: lexical phraseological units (phrasemes) and grammatical phraseological units (Rahmatullayev 2006: 441). The first group is subdivided into the five aforementioned categories (noun, adjective, verb, adverb, predicative); the second group, the grammatical phraseological units, according to the author's own account (pp. 446-447), consists of modal phraseological units and interjective phraseological units, and is not included in the five-category list. Therefore, the use of the "modal word" and "interjection" categories as separate classes in this study is not a deviation from Rahmatullayev's theoretical system, but rather the result of unifying its two-level architecture the distinction between a phraseme and a grammatical phraseological unit-into a single practical coding system. Such a solution is consistent with the general practice for resources developed for corpus linguistics and natural language

processing systems: modern universal tagging standards like Universal Dependencies also define independent and special parts of speech as a single, flat list rather than a hierarchical one.

As for the "predicative" category, it was not used as a separate practical label in this study. The empirical basis for this is provided by the UGPD database itself: in the source, units that can serve as examples of constructions "formed on the basis of a predicative lexeme, equivalent to a sentence-form" (including Rahmatullayev's own examples - *maza (si) yo'q* [is tasteless], *yurag (i) da o't (i) bor* [has passion], *yurag (i) da o't (i) yo'q* [lacks passion]) are consistently classified under the adjective category during the practical classification process. This indicates that in the current database, the "predicative" does not practically occur as a separate unit distinct from the other four categories. Units with such semantics are fully covered within the adjective category. For this reason, "predicative" was considered not as an independent label, but as a structural manifestation of the adjective category.

A number of practical verification methods were developed to resolve ambiguous and borderline cases during the classification process. Specifically, if a phrase could combine with the verbs *qilmoq* (to do) or *bo'lmoq* (to be) and function as part of the predicate, it was classified as a noun. If a phrase could precede a noun and function as a modifier, it was classified as an adjective. For example, the phrase *ko'nglito'q* (content, satisfied) was classified as an adjective because it cannot combine with the verb *bo'lmoq* and can modify a noun (*ko'nglito'qodam* – a content person). In contrast, *ko'ngliog'rimoq* (to feel hurt/pity) was classified as a verb because it can function as an independent predicate. At the end of this stage, it was determined that 4,259 of the 5,439 phrases, or approximately 78 percent, belong to the verb category.

During our analysis, we observed that both editions of the Rahmatullayev dictionaries under review included proverbs and compound words that cannot be considered phraseological units. In total, 38 proverbs and 5 compound words were identified in the dictionaries. Interestingly, this issue is addressed in the source itself, specifically in the introduction to Rahmatullayev's

2022 dictionary. The author distinguishes between a broad and a narrow understanding of phraseology: in a broad sense, phraseology also includes proverbs, sayings, and aphorisms. However, he emphasizes that "only one type of these phenomena stable combinations of words can be regarded as a linguistic unit" (Rahmatullayev et al. 2022: 7). Thus, while the author acknowledges the narrow approach to phraseology as more valid, the practical content of the dictionary does not fully adhere to this position.

The grammatical basis for this inconsistency is as follows: according to Rahmatullayev's definition, a phraseological unit, within a sentence, "functions as a whole, answers a single question, and serves as a single part of the sentence" (Rahmatullayev et al. 2022: 7). A proverb, on the other hand, is an independent, complete didactic-generalizing thought that manifests as a separate sentence. For this reason, a proverb does not satisfy the main functional criterion established for a phraseological unit and, strictly speaking, should be considered an object of study for paremiology, not phraseology.

The same logic applies to the category of "compound words." According to Rahmatullayev's definition, any phraseological unit must consist of at least two independent words. These words must be syntactically connected, meaning they are used together while each retains its own grammatical properties (Rahmatullayev et al. 2022). In other words, although the components of a phraseological unit convey a new, figurative meaning as a whole, they can still be perceived as separate words. A compound word, however, does not meet this condition. It is the result of two roots merging into a single word (e.g., *temiryo'l* [railway], *qorako'l* [karakul]). Here, we no longer have two separate words, but a single, complete lexeme. Its internal parts cannot be separated as individual words (another word cannot be inserted between them, nor can the first part be modified independently). For this reason, a compound word does not satisfy the "at least two independent words" criterion required for a phraseological unit. It is a phenomenon of word formation (creating a new word). This criterion corresponds to the principle of "separability" (Russian: *razdel'noformlennost'*) used by

Kunin in Russian-English phraseology (Kunin 1970). Based on these considerations, when creating the UGPD-Verb database, we excluded 43 units classified as proverbs and compound words from the scope of morphological annotation, but we did not remove them from the database entirely. We have preserved them for future research as a separate category of "borderline cases."

It should be noted that classification by external structure (the number and category of components) is not new in Uzbek phraseology: in her research, Vafoyeva divided phraseological synonyms into groups of two, three, and four or more components, and within each group, she separately examined their relationship to parts of speech (Vafoyeva 2009). Our approach continues this tradition but differs in one important aspect: we encode not only the external structure but also the specific morphological form of each component (tense, mood, voice, etc.), rendering the result in a reusable format.

For the morphological modeling of idioms included in the verb category, the system of grammatical forms developed by Hamroyeva, Xoliyorov & Abdualimova was used (Hamroyeva et al. 2021). This system consists of standardized tags for tense, mood, voice, affirmation-negation, functional forms, person-number, possession, and case, and was initially developed using the example of a single lexeme (the verb *kiymoq* [to wear/put on]).

Applying this system to idioms and phrases, which are semantically and grammatically more complex than simple words, required a series of adaptive decisions. Specifically, the surface homonymy between the causative voice suffix -ar and the participial form -r/-ar led to the incorrect coding of some entries in the initial stage. In such cases, the syntactic function of each form in the sentence, as well as the grammatical and semantic relations within the phrase, were analyzed separately, and the coding was revised. Similar adaptive decisions were required at other points in the annotation process, as each specific case of applying a system designed for a single lexeme to a multi-component unit necessitates taking into account its unique grammatical context.

A special decision was also made regarding the limitation of the lexical selection. In the 2022 dictionary, a number of units were marked with *tar* (historical) and were quoted from the 14th-century work *Guliston bit-turkiy* by S. Saroyi. Since forcibly adapting such units to the synchronic morphological system would blur the boundaries between diachrony and synchrony, this layer was excluded from the scope of the current research.

A four-stage review was conducted to ensure the reliability of the database's content. In the first stage, the accuracy and internal consistency of the category assigned to each phrase were checked. In the second stage, the orthographic and technical uniformity of the entries such as the use of capitalization, spacing, and punctuation was standardized. In the third stage, phrases recorded in both source publications were examined separately. This review revealed that 99 such phrases had been entered as separate rows for each publication. After confirming they were the same phraseological unit, they were merged into a single row. In cases of discrepancy, the classification from the 2022 (revised) edition was prioritized. During this same stage, it was also discovered that within a single publication, the same phrase was recorded twice in the exact same category and form in eight instances. As these technical (duplicate) repetitions carry no independent linguistic information, unlike corresponding entries across different publications, they were individually reviewed and removed from the database.

In the fourth stage, the correctness of the morphological template codes was verified. This stage initially encompassed 570 idioms within the scope of the current study; later, during a more in-depth statistical analysis, one of them (*Borigabaraka*) was found to lack a verbal component, meaning it did not meet the criteria for a verbal phraseological unit, and was thus excluded from the analysis. Therefore, all indicators presented below are calculated based on the 569 verified entries.

RESULTS AND DISCUSSION

Currently, the UGPD database contains 5,439 phraseological units. Their distribution by grammatical category is shown in Table 1.

Table 1. *Distribution of the UGPD database by category*

Category	Number of entries	Share
Verb	4259	78.3%
Adjective	708	13.0%
Adverb	206	3.8%
Noun	120	2.2%
Interjection	63	1.2%
Modalword	40	0.7%
Borderline cases (Proverb, compound word)	43	0.8%
Total	5,439	

As is evident, verbal phrasemes constitute more than three-quarters of the database (78.3%). This specific ratio justifies prioritizing the verb category in the first stage of the research. On the one hand, this category provided the largest volume of material; on the other hand, it was considered the most convenient starting point from a methodological compatibility perspective, as the system of grammatical forms developed by the team of authors led by Hamroyeva was designed specifically for verbs. At the same time, the relatively small number of the remaining categories does not imply their insignificance. For instance, the share of adjectival phrasemes (13.0%) is significantly higher than that of adverbs (3.8%) and nouns (2.2%), which provides a useful guideline for determining which category to prioritize for annotation work in subsequent stages.

A review of the category column revealed a number of inconsistencies. During the verification process, it was discovered that two units were labeled as "Introductory word." Since introductory words indicate a syntactic function rather than an independent grammatical category, they were reclassified into their appropriate categories. Additionally, two other types of units not belonging to the six initially established main categories (verb, adjective, adverb, noun, interjection, modal word) were identified – "compound word" (5 units) and "proverb" (38 units) for which a separate "borderline cases" category was introduced.

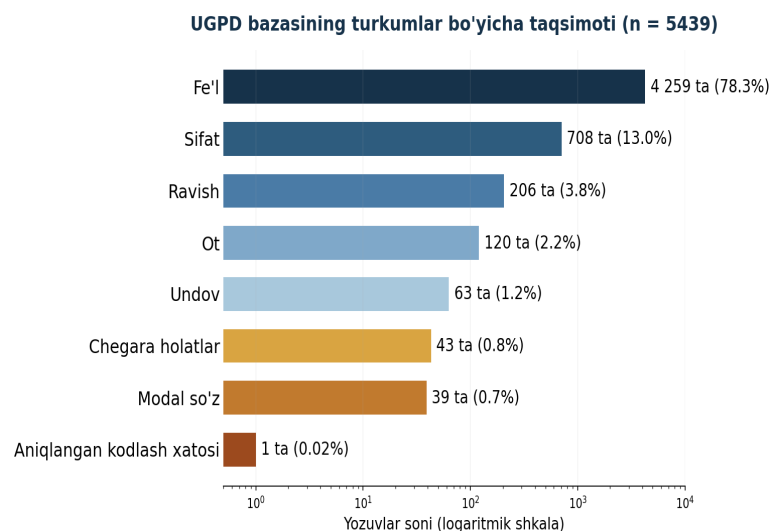


Figure 2. *Distribution of the UGPD database by grammatical categories*

During technical verification, a number of formatting inconsistencies in the phrases (inconsistent use of uppercase and lowercase letters, as well as extra spaces) were standardized. Additionally, eight actual duplicate phrases were removed, and 99 identical phrases entered separately from two publications were merged into a single table. These two types of cleanup technical (duplicates) and inter-publication (merging corresponding entries) were a decisive step in determining the final size of the database (5,439 entries), as without them, statistical indicators (e.g., percentages by category) would have been artificially distorted.

At the current stage, 569 out of 4,259 verbal phrasemes have been fully morphologically templated. In the coding system, the nominal component is defined by the case chain (k1-k6) and possessive-person indicators, while the verbal component is defined by indicators of voice, mood, tense, and functional form (e.g., n4 – causative voice, sh3bir – 3rd person singular, vhn – verbal noun); the boundary between the two components is

indicated by the "////" symbol. Table 2 provides several examples of annotated phrases.

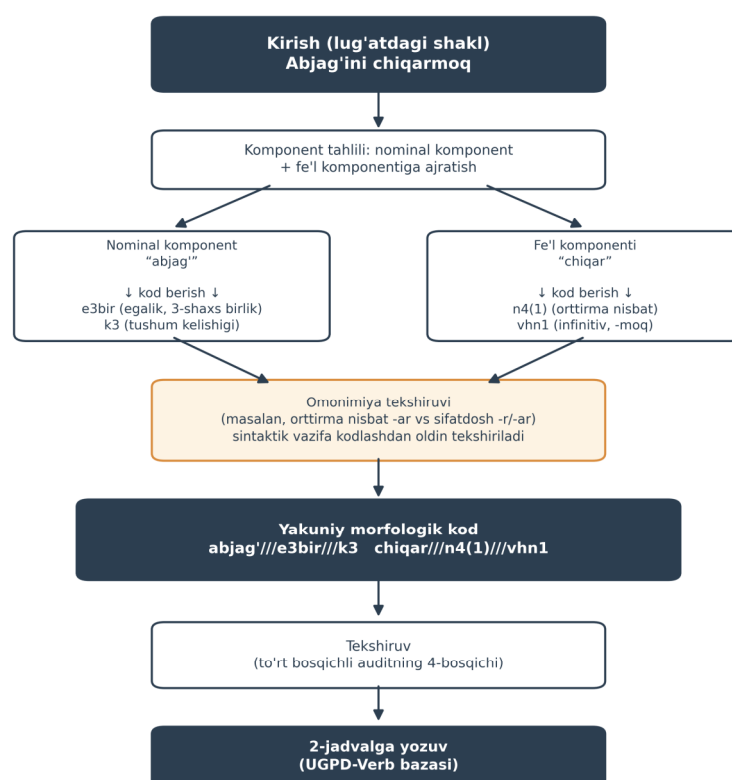


Figure 3. The morphological annotation process for a single verbal phrase (example: Abjag'inichiarmoq)

Table 2. Examples of morphological templating

Phrase	Morphological template	Comment
Abjag'inichiarmoq	abjag'///e3bir///k3 chiqar///n4 (1) ///vhn1	Nominal component + verb in causative voice
Achchig' (i) keldi	achchig'///e3bir///k1 kel///Z1 (1) ///sh3bir	Perfect (past tense) form
Toloseone'stemper	nerve///e3bir///k1 break///n3 (2) ///vhn1	Passive voice verb component
To be able to comprehend/understand	mind///e3bir///k1 reach///Z3 (1) ///sh3bir	Present-future tense form

In the annotation process, special attention was paid to two methodological aspects. Firstly, since the system we used for morphological coding was developed at the single-lexeme level, we encountered the phenomenon of grammatical homonymy when applying it to multi-component phraseological units. Specifically, the causative voice suffix *-ar* (n4 (1)) and another type of participle (*-r/-ar*, vsif (4)) are identical in appearance, which caused incorrect coding several times in the initial stage. The source itself notes that this form is also homonymous with another indicator the future-dubitative tense form Z3 (4) (Hamroyeva et al. 2021). Thus, this grammatical multi-layeredness is recognized as a point requiring caution not only in our annotation but also within the source's own system. In practice, this means that the annotator cannot automatically link every verb component ending in *-ar* to a single code. Its syntactic function within the phraseme (whether it is linked to the subject, or serves as the predicate) must be examined separately. All such identified cases were re-examined and corrected according to their grammatical function.

Secondly, the annotation pertains only to the specific form of the phraseme recorded in the dictionary (not to all possible paradigmatic forms). This resulted in some entries being encoded in the verbal noun form (*-moq*), while others were encoded in a pure, conjugated verb form (past or present-future tense) (see the difference in Table 2).

During the annotation process, two distinct categories were identified and deliberately excluded from the scope of the current study. The first category consists of 12 verbal phrasemes labeled with *tar* (historical) in the 2022 dictionary, quoted from S. Saroyi's 14th-century work *Guliston bit-turkiy*. Some lexical components within them (for example, the Old Turkic *elig/alin*, meaning "hand") do not occur as independent lexemes in modern Uzbek. The second category, totaling 43 units classified as "proverbs" (38) and "compound words" (5), does not fully meet the definition of a phraseological unit according to the theoretical criteria established in the materials and methods section.

At this point, it is necessary to distinguish between two levels of exclusion. The first is that 43 borderline cases (Proverb,

Compound word) were set aside as a separate "borderline cases" category, outside the verb class, during the classification stage (see Table 1). Therefore, they are not included in the 4,259 verb phrasemes at all. The second is that 12 historical units are part of the 4,259 verb phrasemes, but they were excluded from the scope of the current research precisely at the morphological annotation stage, in order not to blur the boundaries between diachrony and synchrony. Thus, the portion open for synchronic annotation within the verb class consists of $4259 - 12 = 4,247$ units. Of these, 569 have been fully coded in the current stage, and the remaining 3,678 are planned for annotation in future studies. Units from both categories (12 and 43) were not deleted from the database but were preserved and noted separately for future research.

In addition to the general coverage indicators mentioned above, the 569 annotated phrases also demonstrate a clear internal distribution according to the finite and non-finite verb forms used in the coding system. This distribution is presented in Table 3.

Table 3. *Distribution of verb form types among 569 morphologically annotated phrases*

Verb Form Type	Number of Entries	Share
Verbal Noun (vhn)	448	78.7%
Gerund (vrav)	36	6.3%
FiniteVerb (conjugated form)	77	13.5%
Participle (vsif)	8	1.4%
Total:	569	100%

In Rahmatullayev's dictionaries, verbal nouns (*vhn*), which are the typical citation form for idioms, constitute the largest portion of the annotated phrases (78.7%). Finite verb forms, recorded in a complete, conjugated form, hold the second position (13.5%). Gerunds (*vrav*) account for 6.3%, and participles for 1.4%. The Wilson 95% confidence interval, calculated based on the sample size ($n=569$), is as follows: 75.2-81.9% for verbal nouns, 11.0-16.6% for finite verbs, 4.6-8.6% for gerunds, and 0.7-2.7% for participles. This distribution is not the result of a random citation

method in the dictionary, but rather reflects the grammatical nature of Uzbek verbal phrasemes. That is, the majority of them are presented in the dictionary in a non-conjugated, lexicalized form, while fully conjugated forms are characteristic of only a relatively small number of idioms that have become fixed with a specific tense or person-number meaning. For example, the idiom *Achchig'ikeldi* (he/she got angry) in Table 2 is not a verbal noun but is encoded as a finite verb in the past tense. Morphological annotation made it possible to identify this regularity based on quantitative indicators. In traditional explanatory dictionaries, this phenomenon is not singled out as a distinct grammatical pattern. However, the annotated database presents it in a systematic and quantifiable manner.



Figure 6. The process of step-by-step filtering of entries in the UGPD database: from a total of 5,439 units down to the full annotation of 569

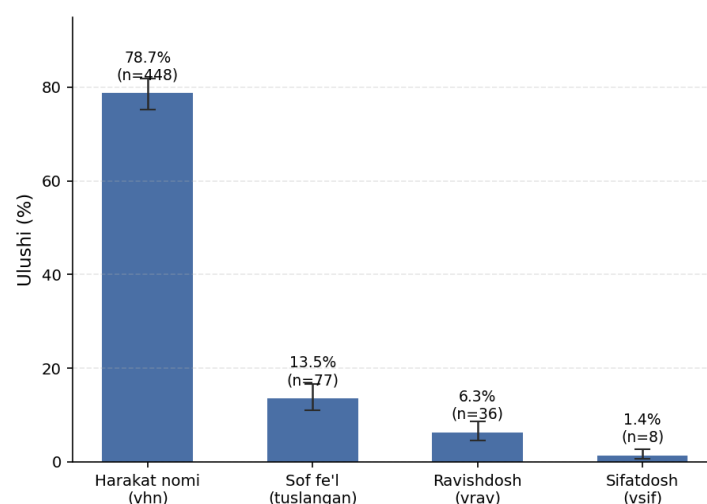


Figure 4. *Distribution of verb forms among annotated phrases (n=569, error bars – Wilson 95% confidence interval).*

In addition to the distribution of verb forms above, the 569 annotated entries were also analyzed by the voice indicator. Table 4 presents this distribution.

Table 4. *The total exceeds 100% because 12 entries simultaneously use two voice indicators (e.g., a combined form of the causative and passive voices)*

Voicetype	Number of entries	Share
n1 – activevoice	489	85.9%
n4 – causativevoice	55	9.7%
n3 – passivevoice	21	3.7%
n5 –reciprocalvoice	15	2.6%
n2 –reflexive voice	1	0.2%
Total:	581*	Figure 5. Distribution of voice indicators (n=569, error bars – Wilson 95% confidence interval).

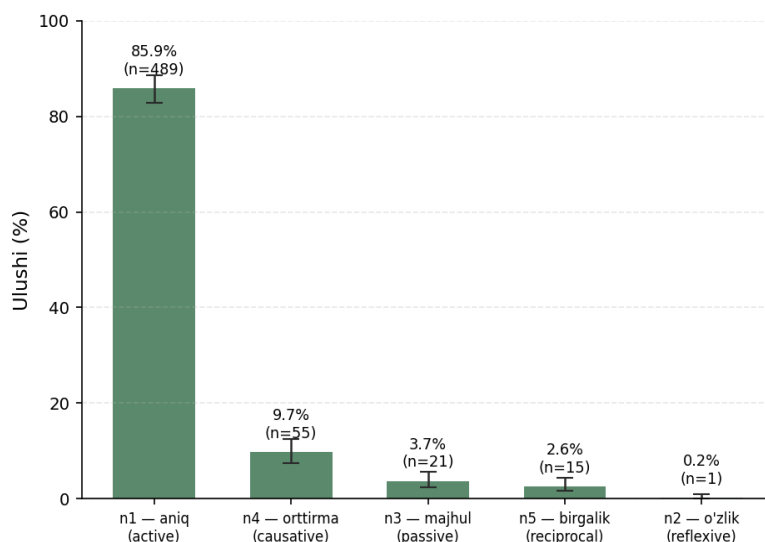


Figure 5. *Nisbat (voice) ko'rsatkichlariningtaqsimoti* ($n=569$, xatolikchiziqlari Wilson 95% ishonchoralig'i).

Jami ko'rsatkich 100% dan oshadi, chunki 12 ta yozuvda ikkinisbat ko'rsatkichibirvaqtda qo'llangan (masalan, orttirmavamajhulnisbatningqatlashganholati).

As is evident, the majority of the annotated phrasal verbs (85.9%) are used in the active voice, meaning they denote an action or state with a clearly defined agent. The causative (9.7%) and passive (3.7%) voices are significantly less common. The reflexive voice (n2) was noted in practically a single instance (0.2%). Furthermore, almost half of the 569 entries (284, or 49.9%) include a nominal component with a possessive marker (primarily third-person singular, e3bir). This indicates that most phrasal verbs are based on an inalienable possession construction of the "somatism + verb" type (typical for words referring to body parts), such as *abjag'inichiqarmoq* (to beat someone to a pulp; lit. "to bring out his/her mess"). The most frequent model for the combination of case, voice, and verb form is: nominative case + active voice + verbal noun (k1+n1+vh), which was recorded in 148 entries, or 26.0% (95% CI: 22.6-29.8%). This

shows that the typical grammatical "core" of Uzbek phrasal verbs consists of a noun component in the function of a subject and a verb component in its lexical base form (-moq).

DISCUSSION

The fact that verbal phrasemes constitute 78.3% of the database may seem like an expected result, given that the verb is the word class with the richest morphological paradigm in the Uzbek language. However, the sheer magnitude of this proportion warrants a specific explanation: it indicates that the central core of the Uzbek phraseological system is formed by units expressing action-state, while nouns, adjectives, and other word classes do not hold such a dominant position. This observation aligns with a qualitative assessment previously made by Rahmatullayev, who noted that verbal phrasemes make up "nearly four-fifths" of all phrasemes (Rahmatullayev 2006); our results, for the first time, substantiate this approximate estimate with a precise figure calculated from a complete dataset of 5,439 units. This finding is also consistent with the proportion reported in an independent study on phraseological synonyms (82% for verbal phraseological synonyms) (Vafoyeva 2009). This demonstrates that the predominance of verbal phrasemes in the Uzbek phraseological system is not incidental but rather a systemic characteristic, repeatedly confirmed across various cross-sections of the database. This characteristic can also be explained typologically: as Shansky noted, a phraseological unit's affiliation with a word class is determined not by its structural composition, but by its general categorical meaning (categorical seme) (Shansky 1996). Since the verb is the word class with the broadest semantic field expressing action-state, the need for figurative-phraseological expression is also most concentrated in this domain. In other words, the predominance of verbal phrasemes is not a random statistical observation but a natural consequence of the semantic relationship between a phraseological unit and a grammatical word class.

The identification of the problem of grammatical homonymy is also of particular importance: the causative voice suffix *n4* (1) and the *-ar/-r* form of the participle are identical in appearance, which led to incorrect coding in the initial stage. After the syntactic function of each instance was re-examined, the coding was corrected. This situation demonstrates that applying a morphological system developed at the single-lexeme level to a multi-component linguistic unit is not a mechanical transfer, but rather a process that requires a separate grammatical check in each case. A similar problem has been noted in the example of the typologically similar Turkish language: the negative impact of external homonymy, rooted in agglutinative morphology, on the quality of automatic and manual annotation of verbal MWEs in the PARSEME corpus has been shown (Öztürk et al. 2022). Thus, the case identified in our study is not a methodological flaw, but a manifestation of a systemic difficulty arising from the general morphological nature of agglutinative languages. This phenomenon of external homonymy can, in fact, also be explained from the perspective of the theory of the correlation of a phraseological unit to a word, as proposed by Kunin (Kunin 1970). The verb form within a phraseological unit does not fully retain its individual, word-level grammatical properties; instead, it is subordinated to the grammatical behavior of the phraseme as a whole. For this reason, morphological boundaries that are distinct at the word level (for example, the difference between *n4* (1) and *vsif* (4)) can become blurred at the phraseme level, creating the need for additional interpretation and verification by the annotators.

Our morphological annotation formalizes a phenomenon that Rahmatullayev himself described qualitatively. Analyzing the alternation between *ko'ngl (i) og'ridi* (a sentence-equivalent construction) and *ko'ngl (i) niog'ritmoq* (a phrase-equivalent construction), Rahmatullayev (2022) shows that this difference is linked to the formation of the causative voice, which concurrently changes the case of the noun component from the nominative to the accusative. In our database, this very phenomenon is expressed in a coded form: the first variant is denoted by *Z1* (1) + *sh3bir* (a finite verb form), and the second

by n4 (1) + vhn1 (causative voice and verbal noun). In other words, we have converted the phenomenon Rahmatullayev described qualitatively into a computable, replicable coding system. This, in turn, indicates the existence of a bridge between traditional phraseography and grammatical lexicography. Rahmatullayev's descriptive explanation provides the patterns of meaning and usage for the phraseological unit, while the UGPD-Verb brings these patterns into a formal representation that can be verified within the context of each individual idiom and generalized statistically. The two approaches are not mutually exclusive but rather complement each other.

Another pattern observed in the annotated sample can also be explained theoretically: some phrasemes are recorded in the dictionary as a verbal noun (*-moq*), while others are in a conjugated form. Nacişcionè (2010) had discussed the discrepancy between the base form of a phraseological unit in a dictionary and its use in speech. Our results empirically confirm this distinction using the material of Uzbek verbal phrasemes. This pattern is also significant in that it is not typologically universal. For example, in phraseological dictionaries of the Kazakh language, verbal phrasemes are typically recorded not as verbal nouns, but in a conjugated, indicative mood form (Kenesbaev 1977). Therefore, the base form in which a phraseme is presented in a dictionary depends not on a universal rule, but on the independent choice of each lexicographical tradition. This once again confirms that the coding system in UGPD-Verb is itself the result of a conscious methodological decision.

The concept of valency was first defined in international linguistics by Tesnière as the ability of a verb to attach grammatically necessary "actants" (e.g., subject, object) to itself (Tesnière 1959). Furthermore, in a separate study dedicated to the theory of syntactic-semantic connection (valency) of verb phrases, verb phrases are classified into one-, two-, and three-place types based on the number of connections, where connections with the subject, possessor, and object are considered individually (Rahmatullayev 1992). The case chain (k1-k6) that we use for the nominal component is, in fact, a practical manifestation of this valency theory: for instance, the k3

indicator (accusative case) signifies a connection with an object, while k1 (nominative case) signifies a connection with the subject. In this sense, the UGPD-Verb database can serve as an initial empirical foundation for implementing valency theory at the corpus level. Such a relationship is not accidental: Rahmatullayev himself did not separate valency and morphological form, instead viewing them as two sides of the holistic grammatical behavior of a phraseological unit (Rahmatullayev 1992). For this reason, valency indicators can be added as a separate column to the UGPD-Verb database in the future. This step would be a natural progression that would significantly increase the descriptive power of the database.

The team of authors led by Hamroyeva defines the purpose of their dictionaries as "preparing a linguistic base for educational corpora, a national corpus, automatic morphological analyzers, translation, and automatic spell-checking programs" (Hamroyeva et al. 2021). The UGPD-Verb database continues these same objectives at the level of phraseological units. This can be directly applied, in particular, to the development of systems for the automatic identification (phraseme-tagging) of phraseological units within the national corpus of the Uzbek language.

From a broader perspective, the UGPD-Verb project continues the tradition of historical interaction between Russian and Western phraseological lexicography (Cowie 1998) using Uzbek language material. Specifically, while the grammatical dictionaries of Zaliznyak, Okuneva, and Tolmacheva focus on systematizing morphological information at the word level (Zaliznyak 1987; Okuneva 2000; Tolmacheva 1995), UGPD-Verb transfers this same principle to the level of phraseological units. This is a significant step not only from a methodological standpoint but also in terms of the mutual enrichment of lexicographical traditions, as it complements traditional semantic-stylistic phraseography (perfected by Rahmatullayev) with a new type of resource adapted to the requirements of grammatical and computational linguistics.

The practical significance of the UGPD-Verb database can also be considered in the context of international natural

language processing research. While Sag et al. (2002) noted that the non-compositional nature of multi-word expressions poses an ongoing problem for computational systems, the PARSEME initiative, led by Savary et al. (2017), sought to address this issue based on common annotation standards for about twenty languages. However, as mentioned above, the direct adoption of these standards for Turkic language as seen in the case of Turkish (Öztürk et al. 2022; Bedir et al., 2021) requires further adaptation due to their agglutinative morphology. In this sense, UGPD-Verb provides a practical example of the adapted annotation approach needed for the Turkic language family, using Uzbek as a case study, and can serve as a foundational linguistic resource for developing automatic MWE identification systems for the Uzbek language in the future. In particular, the issue of external homonymy of the n4 (1) /vsif (4) type, which was resolved in this study, is likely to cause similar errors in any automatic Uzbek morphological analyzer or POS-tagger. Therefore, the 569 manually verified solutions in UGPD-Verb can serve as a ready-made reference set for such systems. Furthermore, the non-compositional nature of multi-word expressions has been identified as a regular source of literal interpretation errors in large language models and machine translation systems (Sag et al., 2002); the morphologically and semantically linked pairs in UGPD-Verb can serve as material for testing and preemptively identifying such errors. Furthermore, the non-compositional nature of polysemous units has been noted as a regular source of literal interpretation errors in large language models and machine translation systems (Sag et al. 2002);

LIMITATIONS OF THE STUDY

The scope and capabilities of this research are defined by certain limitations. The morphological annotation is currently in an initial testing phase: out of 4,259 phrases belonging to the verb category, 569 (or 13.4%) have been covered. The remaining five word classes (adjective, adverb, noun, interjection, modal word), borderline cases, and a total of 1,180 units related to identified coding errors have not been annotated at all. Therefore, at this

stage, the UGPD should be considered not as a fully annotated resource but as a pilot database designed to test the developed coding system in practice. The selection of 569 entries is not arbitrary: this volume was sufficient to cover all the main grammatical phenomena encountered during the annotation process (including homonymy between the causative voice and the participial form, borderline cases, and units belonging to a historical layer), thereby allowing for the verification of the coding system's stability against various complex situations. Furthermore, the four-stage manual procedure (see Figures 2 and 3) which requires component analysis, code assignment, and homonymy checks for each entry also dictates the speed of annotation. The number 569 reflects the volume that could be completed manually within the time allocated for the current stage of the research; in other words, this figure is not a theoretically justified limit but a natural outcome of the practical-methodological pilot phase. Nevertheless, the coding principles developed and the methodological difficulties identified during this very pilot phase (for example, cases of grammatical homonymy) will serve to ensure the quality of subsequent, more comprehensive annotation efforts.

Secondly, the current annotation was performed by a single researcher, meaning the level of inter-annotator agreement was not calculated. Determining this metric has been deferred to the next stage of the research. However, since all decisions made during the annotation process, the principles for resolving cases of grammatical homonymy, and the criteria for identifying borderline cases are thoroughly documented in this article, the necessary transparency is provided for future verification by a second annotator.

Thirdly, a total of 55 phrases, representing historical layers and inter-categorical borderline cases, were methodologically excluded from the current system. While this decision limits the absolute completeness of the database, it ensures the consistency of the synchronic grammatical description. In the future, if the historical layer is examined separately within the scope of diachronic research, it will be possible to conduct a comparative analysis of how it grammatically differs from modern verbal phrasemes.

In the subsequent stages, the following directions are planned: first, to fully apply the coding system developed in the current stage to the remaining 3,678 verb phrasemes, which will transform the database from a pilot version into a comprehensively annotated one; second, to develop a separate annotation convention for nouns and adjectives that corresponds to their specific grammatical nature, as they have a different morphological paradigm than verbs; third, to conduct a separate diachronic study of the historical layer excluded from the scope of the current research, in particular, to determine its grammatical relationship with the modern language; fourth, to involve a second annotator in the future and statistically determine the level of inter-annotator agreement, which will allow for an objective assessment of the annotation system's reliability. Fifth, it is planned to explore the possibility of partially automating the manual coding process: for example, rule-based automatic tagging can be introduced for affixes that do not cause homonymy (such as case and possessive markers), while instances of grammatical homonymy (e.g., *n4* (1) /*vsif* (4)) will still require human verification. Sixth, it is envisioned that the practical effectiveness of the developed coding system will be tested using computational linguistics tools, particularly through integration with a morphological analyzer or parser. This is planned as a separate research direction requiring collaboration with computational linguistics specialists. Seventh, the possibilities of applying the developed methodology to other Turkic languages will be explored: although typologically common Turkic categories such as the case chain (*k1-k6*) and possessive-person markers can be transferred almost unchanged, the unique phraseological inventory and lexicographical tradition of each language (for instance, in Kazakh, verb phrasemes are typically recorded in inflected form, Kenesbaev 1977) will require specific adaptation.

CONCLUSION

In this article, we have outlined the initial stage of creating UGPD-Verb, a morphologically annotated database of Uzbek verbal phraseological units. The work began with a systematic

review of a database comprising 5,439 idioms, compiled based on Sh. Rahmatullayev's 1978 and 2022 phraseological dictionaries (Rahmatullayev 1978; Rahmatullayev et al. 2022). During this process, several classification inconsistencies were identified and corrected, including incorrect category labels, formatting errors, eight true duplicates, and 99 identical idioms that were entered as separate entries in the two editions.

As the main result of the research, 569 out of 4,259 phrasemes belonging to the verbal category were morphologically described based on the system of grammatical forms developed by Hamroyeva, Xoliyrov & Abdualimova (Hamroyeva et al. 2021). A number of methodological challenges arose during the annotation process; in particular, applying a grammatical system designed for single words to multi-component phraseological units, and instances of grammatical homonymy, required each idiom to be analyzed individually.

Furthermore, two categories of units were excluded from the scope of annotation: phrasemes belonging to a historical layer and entries incorrectly classified as "proverb" or "compound word." These units were not deleted from the database but were retained under a separate tag. This decision leaves open the possibility of their use in future research.

The primary significance of this study lies not only in the number of annotated idioms but also in the reproducibility of the developed coding rules and their applicability by other researchers. Specifically, the main scientific contribution of the study can be summarized in three aspects: first, the system of grammatical forms, initially developed at the single-lexeme level, was applied to multi-component phraseological units, forming a replicable methodology on this basis; second, clear verification protocols were developed to resolve methodological challenges that arose during this application process, such as grammatical homonymy (e.g., *n4* (1) /*vsif* (4)); and third, a complete morphological annotation of 569 verbal phrasemes was performed, creating a tested foundation for future research. It is now clearly defined which codes to use in which cases, how to resolve grammatical homonymy, and how to identify and

document boundary cases. In this sense, this article can be regarded less as a description of a single database and more as a foundational resource that establishes the theoretical and practical principles for the grammatical formalization of Uzbek phraseology. This, in turn, further strengthens the computational linguistics infrastructure for the Uzbek language.

Although the research is not yet complete, the annotation principles and coding rules developed at this stage provide the necessary methodological foundation for future work. In the future, they can be applied to 3,678 unannotated verbal phrasemes, as well as to phraseologisms belonging to nouns, adjectives, and other grammatical categories. In this sense, this study proposes a systematic methodology for the morphological annotation of Uzbek verbal phrasemes and creates a scientific basis for developing the UGPD database as a linguistic resource suitable for the needs of corpus linguistics and natural language processing. Against the background of international research on the annotation of multiword expressions (Savary et al. 2017; Öztürk et al. 2022), UGPD-Verb can serve as an initial model for the creation of similar resources for Turkic languages, using the Uzbek language as an example. This, in turn, empirically reinforces the general theoretical thesis that a phraseological unit manifests itself in a distinct, unique way within the unified grammatical structure of the language (Khusnutdinov 2015).

REFERENCES

- Baranov, A. N. & Dobrovolskiy, D. O. 2013. *Osnovy frazeologii* [Fundamentals of phraseology]. Flinta.
- Bedir, T., Şahin, K., Güngör, O., Üsküdarlı, S., Özgür, A., Güngör, T. & Öztürk Başaran, B. 2021. Overcoming the challenges in morphological annotation of Turkish in universal dependencies framework. In C. Bonial & N. Xue (Eds.), *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop* (pp. 112-122). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.law-1.12>
- Cowie, A. P. (Ed.) 1998. *Phraseology: Theory, Analysis, and Applications*. Clarendon Press. <https://doi.org/10.1093/oso/9780198294252.001.0001>

- Hamroyeva, Sh., Xoliyorov, O'. & Abdualimova, G. 2021. *O'zbektilining fe'lishakllarilug'ati ("kiymoq" fe'limisolida)* [Dictionary of Uzbek Verb Forms: Based on the Verb "Kiymoq"]. Vol. 1. Ter SU Publishing and Printing Center.
- Isanbet, N. S. 1989-1990. *Tatar telenefrazeologiksüzlege* [Phraseological Dictionary of the Tatar Language]. Vols. 1-2. Tatarstan Kitap Nāşriyatı.
- Japarov, Sh., Konkobaev, K. & Osmonova, Zh. 2015. *Kyrgyz tilinifrazeologiyalyksözdüğü* [Phraseological Dictionary of the Kyrgyz Language]. Vols. 1-2. Imak Offset.
- Johanson, L. & Csató, É. Á. (Eds.) 1998. *The Turkic Languages*. Routledge.
- Kenesbaev, I. K. 1977. *Qazaqtilinifrazeologijalyqsözdigi* [Phraseological dictionary of the Kazakh language]. Gylım.
- Kunin, A. V. 1970. *Anglijskayafrazeologiya (teoreticheskijkurs)* [English phraseology: A theoretical course]. Vysshayashkola.
- Mamatov, A. E. 2000. O'zbek tilifrazeologi zmlarining shakllanishimasalalari. [Issues in the Formation of Phraseological Units in Uzbek]. Abstract of doctoral dissertation, Alisher Navoiy Institute of Language and Literature, Uzbekistan.
- Naciscione, A. 2010. *Stylistic use of Phraseological Units in Discourse*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.159>
- Okuneva, A. P. 2000. *Russkiyglagol. Slovar'-spravochnik* [The Russian Verb: A Reference Dictionary]. Russkiyyazyk.
- Öztürk, Y., Hadj Mohamed, N., Lion-Bouton, A. & Savary, A. 2022. Enhancing the PARSEME Turkish corpus of verbal multiword expressions. In A. Bhatia, P. Cook, S. Taslimipoor & M. Garcia (Eds.), *Proceedings of the 18th Workshop on Multiword Expressions @LREC2022* (pp. 100-104). European Language Resources Association. <https://aclanthology.org/2022.mwe-1.14/>
- Rahmatullayev, Sh. 1978. *O'zbektiliningizohlifrazeologiklug'ati* [Explanatory phraseological dictionary of the Uzbek language]. O'qituvchi Publishing House.
- . 1992. *Fe'lfrazemalarningbog'lashuvi* [Valency of Verbal Phraseological Units]. Universitet Publishing House.
- . 2006. *Hozirgiadabiyo'zbektili (2-qism)* [Modern literary Uzbek: Part 2]. Universitet Publishing House.
- , Mahmudov, N., Xolmanova, Z., O'razova, I. & Rixsiyeva, K. 2022. *O'zbektilifrazeologiklug'ati* [Phraseological Dictionary of the Uzbek Language]. G'afur G'ulom Publishing and Printing Creative House.

- Sag, I. A., Baldwin, T., Bond, F., Copestake, A. & Flickinger, D. 2002. Multiword expressions: A pain in the neck for NLP. *Lecture Notes in Computer Science*, 2276, 1-15. Springer.
- Savary, A., Ramisch, C., Cordeiro, S., Sangati, F., Vincze, V., Qasemi Zadeh, B., Candito, M., Cap, F., Giouli, V., Stoyanova, I. & Doucet, A. 2017. The PARSEME shared task on automatic identification of verbal multiword expressions. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)* (pp. 31-47). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-1704>
- Shanskiy, N. M. 1996. *Frazeologiyasovremennogorusskogoyazyka* [Phraseology of the modern Russian language]. 4th ed. Spetsial'naya Literatura.
- Teliya, V. N. 1996. *Russkayafrazeologiya: Semanticheskij, pragmaticheskijilingvokul'turologicheskijaspekty*. [Russian phraseology: Semantic, pragmatic, and linguocultural aspects]. Shkola Yazykirusskoykul'tury.
- Tesnière, L. 1959. *Éléments de syntaxestructurale* [Elements of structural syntax]. Klincksieck.
- Tolmacheva, V. D. 1995. *Uchebnyyslovar' glagol'nykh form russkogoyazyka* [Study dictionary of Russian verb forms]. Rus. yaz. Kursy.
- Vafoyeva, M. Y. 2009. *O'zbektilidafrazeolog iksinonimlarv aularningstruktural-semantiktahlili* [Phraseological synonyms in Uzbek and their structural-semantic analysis]. Abstract of candidate dissertation, Alisher Navoiy Institute of Language and Literature, Uzbekistan.
- Khusnutdinov, A. A. 2015. Phraseological unit in the grammatical structure of language. *Vestnik of KSU named after N. A. Nekrasov*, 1, 158-161.
- Yo'ldoshev, B. 1998. *Frazeologikuslubiyatasoslari* [Foundations of Phraseological Stylistics]. Samarkand State University.
- Zaliznyak, A. A. 1987. *Grammaticheskislovar' russkogoyazyka. Slovoizmeneniye* [Grammatical dictionary of the Russian language: Inflection]. 3rd ed. Russkiyyazyk.

APPENDIX

Morphological coding system (codebook). The table below presents the main categories of the coding system used for the morphological annotation of the UGPD-Verb database. The

coding system is based on the system for marking grammatical forms developed by Hamroyeva, Kholiyorov, and Abdualimova (2021: 10=19). Due to its volume, all sub-variants of the "Tense" (Z) category are not included in the table. For a complete list, please refer to the original source.

Code group	Sub-code	Affix	Meaning (brief)
Polarity (bb)	bb1	-Ø	Affirmative
Polarity (bb)	bb2 (1-6)	-ma, emas, yo'q, -mas, -may, -maslik	Negative (variants)
Voice (n)	n1	-Ø	Active voice
Voice (n)	n2 (1-2)	-n/in, -l/il	Reflexive voice
Voice (n)	n3 (1-2)	-n/in, -l/il	Passive voice
Voice (n)	n4 (1-6)	-ar, -ir, -giz/kiz..., -iz, -dir/tir, -t/it	Causativevoice
Voice (n)	n5	-sh/ish	Cooperativevoice
Functional form – converb (vrav)	vrav (1)	-b/ib	Converb of manner/state
Functional form – converb (vrav)	vrav (2-3)	-gancha, -gudek	Manner/similitude
Functional form – converb (vrav)	vrav (4)	-gach/kach	Time (completion of action)
Functional form – converb (vrav)	vrav (5)	-guncha	Time (limit of duration)
Functional form – converb (vrav)	vrav (6)	-gani/gali	Purpose
Functional form – participle (vsif)	vsif (1)	-gan/kan/qan	Past participle
Functional form – participle (vsif)	vsif (2)	-yotgan	Presentparticiple
Functional form – participle (vsif)	vsif (3)	-ydigan	Future participle
Functional form – participle (vsif)	vsif (4)	-r/ar	Other participle type
Functional form – participle (vsif)	vsif (5-9)	-guvchi/uvchi, -gulik, -gur etc.	Additional participle types
Functional form – verbal noun (vhn)	vhn (1)	-moq	Verbalnoun
Functional form – verbal noun (vhn)	vhn (2)	-sh/ish	Verbalnoun
Functional form – verbal noun (vhn)	vhn (3)	-v/uv	Meaning close to a noun

Tense (Z)	Z1 (1-16)	-di, -gan, -ganedi etc.	Pasttense (16 sub-forms)
Tense (Z)	Z2 (1-18)	-yap, -moqda etc.	Present tense (18 sub-forms)
Tense (Z)	Z3 (1-15)	-y/a, -r/ar, -moqchi etc.	Future tense (15 sub-forms)
Mood (m)	m1	-Ø	Indicative mood
Mood (m)	m2.*sg/pl (1-4)	- (a) y (in), -gin, -sin etc.	Imperative-optative mood (by person-number)
Mood (m)	m3 (1-7)	-sa, -saedi etc.	Conditional mood
Person-number (sh)	sh1sg / sh2sg / sh3sg	-man, -san, -Ø/di	Singular (I/II/III person)
Person-number (sh)	sh1pl / sh2pl / sh3pl	-miz, -siz (lar), -Ø (lar)	Plural (1st/2nd/3rd person)
Case (k)	k1	-Ø	Nominativecase
Case (k)	k2	-ning	Genitivecase
Case (k)	k3	-ni	Accusativecase
Case (k)	k4	-ga/ka/qa	Dativecase
Case (k)	k5	-da	Locativecase
Case (k)	k6	-dan	Ablativecase
Possessive (e)	e1bir / e2bir / e3bir	-m/im, -ng/ing, -i/si	Singular (1st/2nd/3rd person)
Possessive (e)	e1ko'p / e2ko'p / e3ko'p	-miz/imiz, -ngiz/ingiz, -i (lari) /si (lari)	Plural (1st/2nd/3rd person)

Source: Hamroyeva, Xoliyorov, Abdualimova (2021: 10-19)

FARANGIZ QOSIMJONOVA

BASIC DOCTORAL STUDENT,

DEPARTMENT OF UZBEK LINGUISTICS,

TERMEZ STATE UNIVERSITY,

TERMEZ, UZBEKISTAN.

E-MAIL: <FARANGIZQOSIMJONOVA03@GMAIL.COM>