



Article

Statistical Analysis of the Vocabulary of “Shajarayi tarokima” by Parts of Speech

Begmatova Marjona Jomgirovna

1. Tashkent State named after Alisher Navoi, University of Uzbek language and literature 2 Foundation doctoral Student
* Correspondence: -

Abstract: This study investigates the vocabulary of the work *Shajara-i Tarākima* through a statistical lens, focusing on parts of speech. The analysis is anchored in the importance of language as a communication tool that reflects the social, economic, and cultural status of its time. With the advancement of computational linguistics, this research employs statistical methods to analyze and compare the frequency of lexical categories, aiding in the understanding of language structure and style. Despite the relevance of historical works in shaping modern language, there is a lack of comprehensive statistical analyses of classic texts like *Shajara-i Tarākima*, particularly in terms of parts of speech and their semantic classifications. The lexicon of *Shajara-i Tarākima* was digitized and analyzed using Excel, categorizing 14,093 lexemes into various parts of speech, including nouns, verbs, adjectives, and other lexical categories. The frequency of each category was calculated, and a semantic classification was conducted to provide deeper insight into the vocabulary. The analysis revealed that nouns, particularly proper nouns such as anthroponyms, toponyms, and ethnonyms, formed the largest portion of the lexicon. Verbs and adjectives followed, with a smaller proportion of auxiliary words and conjunctions. Notably, the work exhibits historical and archaic lexical forms, which distinguish it from modern Uzbek. The study concludes that the vocabulary of *Shajara-i Tarākima* is rich in Turkic onomastic units, showing minimal grammatical deviation from the modern Uzbek language but containing archaic lexemes. This research contributes to understanding the evolution of the Uzbek language, offering a statistical framework for further linguistic studies and comparisons across historical texts. It also highlights the relevance of linguistic tools in studying classical works for cultural, historical, and literary analysis.

Keywords: Lexeme, parts of speech, words denoting a person and an object, words denoting an action, words denoting a quality, words denoting number and quantity, lexemes denoting reference.

1. Introduction

Language serves as a fundamental tool for communication, expressing human thoughts, cultural values, and societal changes. As an evolving medium, the study of language through literary texts offers critical insights into historical, cultural, and linguistic developments. This article focuses on the vocabulary of *Shajara-i Tarākima*, analyzing its lexicon through statistical methods. The primary objective is to examine the frequency and distribution of parts of speech in the text, offering a lens through which linguistic shifts and patterns can be understood. By focusing on lexical categories such as nouns, verbs, adjectives, and their semantic groups, this study bridges the gap between linguistics and historical texts, highlighting the role of language in reflecting socio-political and cultural dynamics of the 17th–18th centuries [1].

While numerous linguistic studies have explored historical works, few have applied statistical methods to examine vocabulary distribution in classic texts. Existing literature

Citation: Jomgirovna, B. M. Statistical Analysis of the Vocabulary of “Shajarayi tarokima” by Parts of Speech. Central Asian Journal of Literature, Philosophy, and Culture 2025, 6(4), 880-886.

Received: 10th Jul 2025
Revised: 16th Aug 2025
Accepted: 24th Sep 2025
Published: 01th Oct 2025



Copyright: © 2025 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>)

on the *Shajara-i Tarākima* often discusses its cultural, historical, and poetic significance, but lacks comprehensive linguistic analysis using modern techniques like linguostatics. Previous studies have predominantly focused on the historical and ethnographic aspects of the text, without investigating its lexical structure systematically. This study seeks to address this gap by applying statistical tools to analyze the frequency of lexemes, categorizing them into distinct semantic groups.

The method employed in this research involves digitizing the text and categorizing its 14,093 lexemes using Excel, followed by a detailed frequency analysis. The expectation is to uncover patterns in the use of different parts of speech, providing insights into the linguistic characteristics of the text and its relationship to modern Uzbek [2]. The findings are anticipated to reveal that the *Shajara-i Tarākima* features a significant proportion of proper nouns and verb forms, reflecting the socio-cultural context of its time. This analysis is expected to contribute to the study of historical Turkic languages and enrich our understanding of language evolution. The results have implications for both linguistics and literary studies, offering a framework for analyzing other classical texts using modern linguistic methods.

Current Uzbek literary your language, brother Turkish languages with Uzbek language between connection, closeness wants to learn was man first ancient and old Turkish language sources if you learn to the goal appropriate would be [3]. "Family tree "tarokima" is also from the 17th-18th centuries Turkish of the language socio-political in life situation, word of the categories frequency in determining, not only language maybe history, ethnography, toponomics, cultural studies and political science sciences also important for source become service do takes. The work language current Uzbek literary to the language proximity and this with together appropriation from the floor not to use movement done reason that's it in the era Uzbek of the language status full showing give takes [4].

2. Materials and Methods

The methodology employed in this study involves a systematic statistical analysis of the lexicon of *Shajara-i Tarākima*, focusing on parts of speech. First, the text was digitized and processed into a dataset consisting of 14,093 lexemes, including repetitions. The dataset was organized using Microsoft Excel, allowing for detailed categorization and analysis. Each lexeme was classified into its respective part of speech, including nouns, verbs, adjectives, pronouns, and auxiliary words. The analysis aimed to quantify the frequency of each part of speech and to examine the semantic groups associated with them.

Once the text was organized into lexemes and categories, the frequency of each part of speech was calculated, providing an overview of the distribution of lexical items. Special attention was given to identifying the types of nouns, particularly proper nouns such as anthroponyms, toponyms, ethnonyms, and hydronyms, as well as other lexemes denoting objects, actions, qualities, and numbers. The analysis also extended to adjectives and verbs, with a particular focus on the formation of compound words and adjectives derived from nouns [5]. Lexical items were further analyzed for historical and morphological features, noting any deviations from contemporary usage in modern Uzbek. The results were compared to other historical and contemporary texts to identify patterns in vocabulary usage and linguistic development. This methodological approach, combining statistical analysis with semantic classification, aims to offer insights into the lexical structure of *Shajara-i Tarākima* and its reflection of historical language use. The findings will contribute to the broader understanding of the evolution of the Uzbek language and its relationship with historical Turkic linguistic forms [6], [7].

3. Results and Discussion

Person and of something name meaningful words as follows classification possible:

1. Work in the text used notable horses as follows:

- a. Notable nouns

Work in the composition notable horses the most part, that is, 444 organization as they own inside anthroponyms, toponyms, hydronyms, ethnonyms, book names such as onomastic units own inside takes.

b. Similar nouns.

1. Person nouns: āta, oʻgil, xātin, qiz, pādshāh, gadāy, payğambar, dōst, oʻgʻlān, elči, navkar, bek, yigit, ini, imām, bekāvul, vazir, baxshi, kadxudā, onbegi, mulla, šayx, chavuldur, kaniz, biy, kāhin
2. Item nouns: suv, āsh ovqat; qalam, dengiz, yigʻāc daraxt, yogʻoch; siyāh to write for used liquid, black color; luğat, turpaq, kiyik, kabāb, et goʻsht; kölik goods, cargo increasing transportation for cart; söngak, araba, hazina, uruğ, karvon such as many thing item names there is become current Uzbek literary from the language almost difference does not.
3. Place names : mamlakat oʻlka, viloyat, mintaqā, qitʻa; yurt oʻlka, mamlakat, viloyat, mintaqā, qitʻa; tōshak in the work, yuqori tōshak means the sky, and pastki toʻshak means the earth; saroy is the building where the kings and their families live; šahar is a larger and more centralized area than a village; tom is a house; kent is a larger and more centralized area than a village; šimāl is one of the four poles; davlat is an area consisting of several regions and cities; pāytaxt is the center of the state; qalʻa is an area surrounded by a wall [8];
4. Abstract nouns. Duā tilak, istak; to beg God for a good deed, a request, a goal, a long life; otunč to beg forgiveness; eʻtibār to pay attention to something; gunāh is a wrong deed or action; savāb is a good deed; fātiha is a request, a goal, a goal; aql is related to the human mind; zikr is to praise God or to say something;

2. Words that mean color

In Turkic languages, in particular in Uzbek, the adjective has historically not had a special morphological indicator. According to experts, the adjective did not initially exist as an independent word class, but later grew and developed from the noun class. Written monuments in the language morphological in a way adjectives horses and from verbs We are studying in the source general 140 qualities without repetition word to the category related lexeme there is of which 79 are basic adjectives, 61 are fake are adjectives [9].

- a. words that describe a person or thing. Pahlavon, āsiy, sādīq, yaxshi, yāmān, čebar, āqil, tirik, yālğiz, ābida, zāhida, sāliha, bahādir, yālğān, aqlli, davlatli, umrli, nāmusli, hurmatli and etc.
- b. Words that indicate the color of a person or thing: āq, qāra, āla, kök, ādmi, āč sariq;
- c. Words that describe the appearance, size, or shape of a person or thing: uluğča, kichikča, yārti, butun, balant, kičik;
- d. Words that indicate the state of a person or thing: ğalat, durust, qari, yetim, yālğuz, pinhoni, muškul, burunğı, quvish and etc.
- e. words expressing the taste of something: širin, tuzlaq, ači, qimizli.

Regardless of which side one approaches the issue of adjective formation, the facts, opinions, and general conclusions are close to each other, they may differ only in size. The lexemes related to the adjective word class used in the text are formed morphologically (affixation) and syntactically (composition):

Morphological method. Adjectives expressing the meanings of relationship, specificity, belonging, and state are formed through the forms-lik, (-liq). The -lig(k), which is found in the language of Navoi's works, -liγ(q), -lūg(k), -luγ(q) suffixes exist in the modern Uzbek literary language, but some forms have undergone some phonetic changes [10]. This suffix is It was considered very productive in that period and in the current Uzbek language, and was used to form nouns and adjectives. In "Sh.T" as well, a number of lexemes were formed from nouns to adjectives and adjectives to adjectives with the forms – laq, -lik. tuzlaq, mardānaliq, fazānaliq,

The -li formative affix is considered the most productive formative affix in both the Old Turkic language and the modern Uzbek literary language [11]. In "Sh.T" the main part of the formed adjectives is formed in the form [noun + -li] using this affix: aqlli, davlatli, umrli, hurmatli, qulachli, nāmusli, qimizli qānli, ko'rkli, tuğšili, qisraqli, o'čli, jinli, qāyali, aybli, o'ğlāqli, kāšinkārli,

There are also adjectives formed using Persian-Tajik affixes such as -zāda , -kār , -be , -gar: gunāhkār , begunāh , halālzāda , harāmzāda , bedādgar , mādarizād , beshumār. Made with the syntactic method: sāhibjamol, turkmānand, pāra-pāra,

3. Words expressing quantity.

The word "S " also expresses the idea of a sign, like the adjective and the adverb, and in this respect is close to those categories. The adjective expresses the sign of an object, the adverb expresses the sign of an action, and the number The p r e dmet expresses the quantity, number and order of the subject. In the work taken as the object of the study, there are 60 lexemes related to the number word category, 39 of which are compound, 16 are simple and 5 are numerative. In the text numbers into 3 groups we separated:

- a. ordinal number: učlanči, bešlanči, to'quzlanči, törtunči, bešinči
- b. count number: ön olti, sakson, yetmish bir, uç, qirq, ötüz, ön ikki, yigirma ön olti, sakiz yuz, olti yuz, uç yuz
- c. Numeratives: misqāl, qatra, bōlak, tōp, tōgram.

4. Words that indicate a point.

Pronoun noun, adjective, number and sometimes other word categories in place used, subject and to him/her typical was of the sign existence shows. Pronoun objectivity or his/her sign permanent not, maybe certain in places expression does. In this respect pronoun independent word series although it is calculated, it is clear to the point has will not be [12]. The pronoun main meaning and which word categories in place usage text inside Pronouns are words with independent meaning that do not indicate a person, object, sign or quantity, such as nouns, adjectives, numerals, adverbs, but are used interchangeably, refer to them, and are considered their substitutes. In the source, pronouns are 31 without repetitions, and their meaning groups are as follows:

- a. personal pronouns. Men, biz, ānlar, ular.

Turkic languages have been formed since ancient times. The main part of the personal pronouns in the Old Uzbek language and modern Turkic languages also existed in ancient monuments. There are almost no differences between the Old Turkic language and the modern Uzbek literary language in the structure of pronouns. Some the pronoun I in sources instead of poor, fireplace words used in Navoi 's works [13]. The word "faqir" is used instead of the personal pronoun: Alarniñ avvalyi arba'inida bu faqir Hazrati Mavlānā Muhammad Tabādgāniy mulāzamatlariya barib erdim. In the current Uzbek literary language, the word kamina is used instead of faqir.

- a) Demonstrative pronouns. O'shal, ul, munğa, alqa, ulkim, anğa, bu, bundan, shundaq. According to M. Hamroyev, the semantics of demonstrative pronouns in Turkic languages, the peculiarities in their formation are examined on the basis of comparison. According to the author, demonstrative pronouns are used in these languages, as in other languages of the world, to indicate the spatial location of the object relative to the speaker.
- b) Interrogative pronouns. Although the interrogative pronouns of the Turkic languages are close to each other in terms of phonetic variants, they differ somewhat in terms of morphemic structure and meaning [14]. Interrogative pronouns: kim?, nimarsa?, qayusi?, nečasi?.
- c) personal pronoun. In the source identity pronoun own in the form of arrived.
- d) demonstrative pronouns. 4 markings in the source pronoun available: bir kun, har qayusi, har qayda, har qačān.

- e) also contains the Old Turkic pronoun *qamug'*, which means all, everyone, belonging to this type. Bizga böldi qamug' elin örne diqli.
- f) There are no pronouns of doubt or absence.

5. Action words.

Use of "Sh.T" to the level according to verbs horse word from the series next in place stands. Verbs total 223 without repetitions, 184 simple 39 compound. Based on the volume of lexemes expressing action and state in the work taken as the object of research, it was divided into the following groups.

- a. speech verbs: *aytmāq, temak//demak, sözla, sör, sözlayturg'an;*
- b. state verbs: *sargaymoq, qorqmāq, xušlamāq, aççıqlanmāq, ökindi;*
- c. action verbs: *qilmāq, yiglamāq, unutmāq, urušmāq, kelmāq, tögramāq, yātmāq, ötmāq, siğinmāq, yogmāq, yibarmāq;*
- d. verbs of action: *qurdurti, baqturdi, töldurtdi, pishirub, tikdi.*

The following compound verbs are present in the work: *in'am qilmāq, irim qilmāq, āt sālāmāq, tāb'e bo'lmāq, qarši bölmāq, ittifāq qilmāq, ğam yemāq, jar qildirmāq, jam qilmāq, šukr qilmāq, fosh qilmāq, böyin qilmāq va h.k.*

6. Words expressing state.

The word *ravish* is defined in textbooks as an invariable word that expresses the sign, quantity or state of an action, partly an object. If we look at the rules of the Old Turkic language, similar to other word groups, there are phenomena of assimilation and formation in *ravish*, as in the ancient and old Turkic languages. Historically created *ravishes* are not divided into bases and suffixes today. In the source we are studying, there are 30 lexemes related to the *ravish* word group, which belong to the following 3 semantic groups [15].

Quantitative-level adverbs: *köp, čāgliq, uluğča, azraq, kičikča, harčand, anča, munčaqlik, yarim, čandān, arqit-arqit, tulam;*

Adverbs of place: *saqli, sölli, ichimizda, yoqasida, yuqāriği, quyiği, taşqari, yāni, qatigā, ustina;*

Time adverbs: *hamisha, avvalqi, burun, ötgan, bu čaq.*

According to their structure, adverbs are divided into simple and complex types. Complex adverbs consist of two words, and "Sh.T." also contains *arqit-arqit* complex quantitative-degree adverbs.

7. Auxiliary words.

Auxiliary words, when derived from their own name, do not have the same properties as independent words, cannot be formed, and do not serve as a basis for formation. They only make sense when combined with independent word groups, and serve to connect independent words and give them additional meaning.

Auxiliaries: The first complete definition of an auxiliary in the Uzbek language was given by A.N. Kononov: "Auxiliaries are a group of words that serve to create relationships between a noun or an object and a predicate, such as means, purpose, reason, time, distance, direction, analogy" [16]. In both the Old Turkic and Old Turkic languages, auxiliaries are divided into pure and co-functional auxiliaries and serve to connect words in a dominant-subordinate relationship in a sentence, grammatically close to conjunctions, and if auxiliaries are divided into semantic groups according to their properties, they will be close to the situation. The total number of auxiliaries in "Sh.T" is 6, and they are as follows:

Pure assistants: *birla// bilan, učun, sari;*

Colleague helpers: *ustigā, söngidin, söng.*

The work contains lexemes such as *tashqari, ichkari, qatig'a, ustina*, but we have included them in the group of verbs. This is because these lexemes are not functional helpers in sentences, but rather are semantically independent lexemes.

Conjunctions: Conjunctions differ from auxiliaries in that they serve to connect words in equal and subordinate relations. Auxiliaries, on the other hand, have only a subordinate

relation. As written monuments testify, most of the conjunctions used in modern Turkic languages were borrowed from Iranian or Arabic, and they began to be actively used from the 12th-14th centuries. In the introduction to the work, Abulgazi Bahodirkhan emphasized that he aimed not to use borrowed words, but to write in the simplest Turkic language, and perhaps for this reason there are 5 conjunctions in the source.

Equal conjunction: *va*;

Conjunctions: *ammā*, *agar*, *uĉun* (also present in the auxiliary), *taqi* (this conjunction dates back to the Old Turkic period and was used instead of the conjunction) [17].

Particles: In addition to connecting words, suffixes also perform functions such as strengthening the meaning, emphasizing, distinguishing, and simulating. In the Old Turkic language, the main function of suffixes is to add and strengthen additional meaning to the content of the sentence. The “connecting” function inherent in auxiliary words is weakened in suffixes. In “Sh.T,” suffixes are 2:

taqi is an auxiliary word of the ancient Turkic language that can act as both a preposition and a conjunction. It is a conjunction and is synonymous with its equal conjunction: *Qāraxān elga savun sāldi taqi uluġ tōy qildi, tōy kuni öġlānni ma’raka içiga kelturub.....* It is this conjunction that is the most frequently used auxiliary word in the source.

Even when it comes as a preposition, it is synonymous with its preposition: *Anuġ taqi taqi ātasi va uluġ ātasining šariatiga amal qilib ul taqi ātasitek to’qquz yuz ön ikki yil bu manzilda ölturub, ul manzilga ketdi.*

ham belongs to the group of emphatic prepositions and emphasizes and strengthens the meaning: *Xurāsān va Irāqda ham köp bölür, ānlar čigatāy elina qōšula türürlar.*

Separate word classes. Words belonging to this class have the characteristics of being able to stand alone as sentences, not being able to enter into syntactic relations with any part of the sentence in the structure of the sentence, not being able to combine with conjunctions, and therefore not having forms of inclination, tense, person-number. These word classes include the following:

Modal words: As is known, modal words serve to express the speaker's attitude to the thought being expressed, to express its certainty, truth, doubt or conditionality, and mainly act as introductory words [18]. They refer to the whole thought or to some part of it, are characterized by not changing in form, do not act as a part of a sentence, and do not enter into syntactic relations with other parts of a sentence. In the source The modal words are used: *bas*, *bor*, *yōq*, *šart*, *kerak*

Exclamations: It is known that one of the main characteristics of exclamations is that, in addition to not being able to express lexical meaning, they cannot act as conjunctions, possessives, number suffixes, or as part of a sentence. The number of exclamations in the source is 3: *Ey*, *balli*, *ofarin*.

4. Conclusion

In conclusion, the statistical analysis of the *Shajara-i Tarākima* lexicon reveals significant insights into the historical and morphological aspects of the Uzbek language. The study highlights that the text predominantly consists of nouns, with a substantial proportion of proper nouns such as anthroponyms, toponyms, and ethnonyms, reflecting the socio-political and cultural context of the 17th–18th centuries. Additionally, verbs and adjectives form a considerable part of the lexicon, with many derived from noun forms, demonstrating linguistic evolution through affixation. The study's findings suggest minimal grammatical deviation from modern Uzbek, though the presence of archaic lexemes indicates the historical distance between the text's language and contemporary forms. These results not only contribute to understanding the evolution of the Uzbek language but also offer a framework for linguistic studies of other historical Turkic works. Further research could expand this methodology to analyze other classical texts,

deepening our understanding of language development and its reflection of cultural and historical dynamics. Additionally, exploring the semantic nuances and diachronic shifts in lexical items could provide a more comprehensive view of linguistic change over time.

REFERENCES

- [1] H. Dadaboyev, *O'zbek terminologiyasi*, Toshkent: Nodirabegim, 2020.
- [2] *Explanatory Dictionary of the Uzbek Language*, vols. I–V, Tashkent: Uzbekistan National Encyclopedia, 2006–2008.
- [3] *Old Turkic Dictionary (DTS)*, Leningrad, 1969, p. 516.
- [4] M. Xolmurodova, *Qutadghu Bilig leksikasi*, Ph.D. diss., Tashkent, 2018.
- [5] B. Abdushukurov, *Turkiy manbalar leksikasi*, Tashkent: BOOKMANY PRINT, 2022, pp. 141-142.
- [6] B. Abdushukurov, *Qissasi Rabg'uziy leksikasi*, Ph.D. diss., Tashkent, 2017.
- [7] Alisher Navoi, *Mahbub ul-qulub*, Complete Works, vol. 14, Tashkent: Fan, 1998.
- [8] A. Khojiyev, *O'zbek tili sinonimlarining izohli lug'ati*, Tashkent: O'qituvchi, 1974, p. 307.
- [9] A. Khojiyev, *Tilshunoslik terminlarining izohli lug'ati*, Tashkent: Uzbekistan National Encyclopedia, 2002.
- [10] Sh. Rahmatullaev, *O'zbek tilining etimologik lug'ati*, vol. II (Arabic words and their derivatives), Tashkent: Universitet, 2003.
- [11] I. Ismoilov, *Turkiy tillarda qavm-qarindoshlik terminlari*, Tashkent: FAN, 1966, p. 150.
- [12] A. Sodiqov, A. Abduazizov, and M. Irsqulov, *Tilshunoslikka kirish*, Tashkent: O'qituvchi, 1981, p. 95.
- [13] Sh. Shoabdurahmonov, *O'zbek adabiy tili va o'zbek xalq shevalari*, Tashkent, 1962, p. 10.
- [14] B. Abulgazi, *Shajarayi Tarokima*, Tashkent, 1995.
- [15] S. T. Durbin, "The Application of Linguostatistics in Turkic Language Studies," *Journal of Turkic Linguistics*, vol. 22, no. 1, pp. 45-58, 2021.
- [16] J. M. Stone, *Statistical Approaches to Linguistic Analysis of Classical Texts*, Cambridge: Cambridge University Press, 2019.
- [17] Z. M. Akhmedov, "Lexical Analysis of Old Turkic Texts Using Modern Computational Tools," *Asian Journal of Linguistics*, vol. 15, no. 3, pp. 120-133, 2020.
- [18] M. R. Jafarov, "The Role of Lexical Categories in Historical Texts: A Case Study of Shajara-i Tarākima," *Linguistic Research Review*, vol. 5, no. 2, pp. 88-102, 2022.