

Date: 25th September-2026

ON THE PROCESS OF FORMING CHULPAN'S AUTHORIAL CORPUS

Rabbimova Gulchehra Toyir qizi

Doctoral Student, Samarkand State Institute of Foreign Languages

E-mail: gulchehraqoshmonova@gmail.com

Abstract. *This article examines the vocabulary of the prose works of Abdulhamid Sulaymon ugli Chulpan, a prominent representative of early twentieth-century Uzbek national renaissance literature, through linguostatistical and lexicographical approaches. The novel Night and Day (Kecha va kunduz) and the author's short stories were selected as the research material. The study identifies the word occurrences and certain rare words necessary for the formation of an electronic corpus of the author's texts. In addition, the socio-political, ethnographic, dialectal, and archaic lexical layers characteristic of Chulpan's prose are classified lexicographically, and their semantic and functional role in revealing the semantic core of the texts is substantiated. The findings may serve as an important practical source for the development of authorial lexicography and the compilation of a Frequency Dictionary of Chulpan's Prose.*

Keywords: *Chulpan's prose, linguostatistics, lexicography, Night and Day, frequency dictionary, authorial corpus, lexico-semantic layer, Jadid literature.*

Relevance and significance of the research topic. The prose, poetic, and dramatic heritage of Abdulhamid Sulaymon ugli Chulpan, a prominent representative of Uzbek national renaissance literature, constitutes an important stage in the formation of Uzbek artistic thought and literary language. The writer's novel Night and Day, as well as his early short stories such as The Victim of Ignorance and Doctor Muhammadyor, not only objectively reflect the socio-political, spiritual, and psychological landscape of Uzbek society at the beginning of the twentieth century, but also expand the lexico-semantic and stylistic potential of the Uzbek literary language.

In linguistics, the study of literary texts solely through traditional descriptive methods no longer fully meets contemporary requirements. Modern philology calls for the investigation of literary texts within the frameworks of corpus linguistics, computational linguistics, lexicography, and linguostatistics. The quantitative and qualitative analysis of the vocabulary of Chulpan's prose works, the evaluation of the author's mastery of word usage through mathematical and linguistic methods, and the compilation of frequency dictionaries of literary speech are among the most pressing tasks of contemporary Uzbek lexicography and linguostatistics.

It is well known that Chulpan's literary legacy has been extensively studied from the perspective of literary criticism by a number of scholars, including O. Sharafiddinov, N. Karimov, U. Normatov, and D. Quronov. From a linguistic perspective, the stylistic features and artistic devices employed in the writer's works have also been analyzed. However, the formation of a comprehensive corpus of the vocabulary of Chulpan's works, the determination of the linguostatistical characteristics of word forms and lexemes, and the



Date: 25th September-2026

creation of an authorial lexicographical database on this basis have not yet been studied in a comprehensive and systematic manner.

The aim of the research is to examine the lexico-semantic composition of Chulpan's prose works, primarily the novel *Night and Day* and his short stories; to calculate their linguostatistical indicators and the frequency of word usage; and to investigate the lexicographical parameters of the author's literary works.

The objectives of the research are:

- to form an electronic text database, or corpus, of Chulpan's prose works and take the initial steps toward their linguistic annotation and lemmatization;
- to determine the ratio between total word occurrences and rare word forms, or types, in the texts;
- to classify lexicographically the socio-political, ethnographic, dialectal, and archaic lexical layers characteristic of Chulpan's prose;
- to compile a dictionary of high- and medium-frequency words in the literary texts and substantiate their semantic and functional role in revealing the semantic core of the works.

This study employs a combination of modern corpus-linguistic, linguostatistical, and authorial-lexicographical methods to analyze the vocabulary of literary texts from both quantitative and qualitative perspectives. The research process is carried out in four main stages: digitizing the texts and making initial attempts to form the corpus, linguistic annotation and lemmatization, calculating statistical indicators, and conducting lexico-semantic classification.

The empirical basis of the research consists of Abdulhamid Sulaymon ugli Chulpan's prose works, including the novel *Night and Day* and such short stories as *The Victim of Ignorance* and *Doctor Muhammadyor*. Printed editions of the works are digitized, converted into textual form using the necessary software, and proofread. The files are prepared by removing spelling errors, punctuation inconsistencies, numbers, and technical elements unrelated to the text, such as page numbers and headings.

Computer-linguistic software is used to determine the quantitative characteristics of the texts. During the analysis, the total number of word occurrences and the number of unique word forms are identified. This is important for assessing the richness of the author's vocabulary and lexical diversity.

According to the frequency of occurrence of words in the author's vocabulary, it can be divided into three parts. The high-frequency zone includes function words, pronouns, and key conceptual nouns. The medium-frequency zone consists of the lexical layer that determines the subject matter and semantic core of the work. The low-frequency zone includes words that occur only once in the text, as well as rare stylistic and ethnographic units.

The overall quantitative indicators of the corpus are determined through the automatic processing and lemmatization of the text during the linguostatistical analysis. The preliminary results indicate that the low level of repetition in Chulpan's prose demonstrates



Date: 25th September-2026

the breadth of the author's lexico-semantic range. There is also no doubt that the volume of expressive vocabulary used by the author for individual stylistic purposes is considerable.

Some words and their functional-semantic characteristics are discussed below. Dedi ("said") is the principal verb used in the conversations of the characters; qiz ("girl") is a noun referring to Zebi and her tragic fate; ko'z ("eye") functions as a means of portraying emotional and psychological states; kun ("day") represents the concepts of time and destiny; mingboshi is a historicism denoting a social and administrative position; and xotin ("woman" or "wife") is a social unit associated with the status of women in society.

The lexicographical analysis identifies four principal lexico-semantic layers that determine the distinctive character of Chulpan's prose. An examination of the characteristic lexical layers found in Chulpan's prose reveals the following components. The category of ethnographic and everyday vocabulary includes such examples as chimmat, paranja, so'ri, chimildiq, sarpo, and chillashir. These lexical units serve to depict the period's cultural identity and the national way of life within the works.

Administrative and historical vocabulary, or historicisms, including mingboshi, strazhnik, ellikboshi, volost, poyloqchi, and ellikboshilik, serves to convey criticism of the administrative systems of Tsarist Russia and the khanates.

The category of Jadid and socio-political vocabulary includes such units as ma'rifat ("enlightenment"), taraqqiy ("progress"), millat ("nation"), islohot ("reform"), gazeta ("newspaper"), and madrasa. These words express ideas of enlightenment and national awakening.

Dialectal and colloquial units serve to individualize the speech of the characters stylistically. The analysis shows that in the novel *Night and Day* and his short stories, Chulpan avoided traditional bookish language patterns and made effective use of elements of the living vernacular. The use of dialectal and colloquial units ensured naturalness and dynamism in the dialogues.

The frequent use of administrative vocabulary such as mingboshi, strazhnik, and volost—for example, the word mingboshi occurs in nearly every second prose work—determined the socio-political pathos of the works. Ethnographic terms served to realistically portray the everyday life of the Uzbeks of the Fergana Valley and the social status of women.

The study of the vocabulary of Abdulhamid Sulaymon ugli Chulpan's works has led to the following scholarly conclusions:

1. Lexical richness. The standardized word-usage indicator calculated on the basis of the electronic corpus formed from Chulpan's prose works demonstrates the writer's high level of mastery in word usage and the breadth of his lexico-semantic potential.

2. Linguostatistical distribution. The fact that certain types in the text belong to the low-frequency zone, including instances of single occurrence, demonstrates that the author made effective use of unique lexical units.

3. Lexico-semantic character. The vocabulary of Chulpan's works is based on such layers as dialectal and colloquial units, ethnographic terms, administrative and historical vocabulary, and Jadid terminology.



Date: 25th September-2026

REFERENCES:

1. Chulpan, A. Night and Day: A Novel and Short Stories. Tashkent: Sharq, 2012. 352 p.
2. Chulpan, A. Collected Works in Four Volumes: Prose Works, Plays, and Articles. Tashkent: Akademnashr, 2016. 384 p.
3. Quronov, D. Chulpan's Poetics: Based on His Prose Works. Tashkent: Fan, 2004. 288 p.
4. Karimov, N. Chulpan: His Life and Literary Legacy. Tashkent: Uzbekistan, 2017. 312 p.
5. Abduzoirov, H. B. "Uzbek Literary Language and Literature." *Badiiy o'zbek tili va adabiyoti*. Tashkent, 2019, No. 4, pp. 45–51.
6. Polatova, M. S. "Methodology of Corpus Linguistics and the Automatic Processing of Literary Texts." *O'zbekistonda xorijiy tillar*. Tashkent, 2021, No. 2 (37), pp. 82–94.
7. Mengliyev, B. R. "The National Corpus of the Uzbek Language: Problems and Objectives." *Filologiya masalalari*. Tashkent, 2020, No. 1, pp. 12–19.
8. Rahmatullayev, Sh. *Lexicology of the Uzbek Language*. Tashkent: Universitet, 2010. 260 p.
9. Mamatov, A. E. *Uzbek Lexicography and Phraseography*. Tashkent: Navro'z, 2015. 210 p.
10. Axtamova, M. "Frequency Distribution of Lexemes in Literary Texts and Zipf's Law." *Ilmiy axborotnoma. Filologiya tadqiqotlari*. Samarkand, 2022, No. 3, pp. 67–73.

