



FREQUENCY COMPARISON OF SINGLE-COMPONENT RUSSIAN LEGAL TERMS IN THE NATIONAL CORPUS OF THE RUSSIAN LANGUAGE AND THE RUSSIAN-LANGUAGE CORPUS OF THE CONSTITUTIONS OF UZBEKISTAN

Anastasia A. Popova, Research Applicant
Urgench State University named after Abu Raykhan Beruni

Abstract. The article provides a comparative overview of the frequency indicators of Russian legal terms in the National Corpus of the Russian Language and in the Russian-language corpus of the Constitutions of Uzbekistan for the period 1927–2023. The methodological framework of the study includes the automated compilation of a list of legal terms, frequency analysis using the IPM metric in the National Corpus, and morphological processing of the constitutional corpus with the MyStem tool. The assembled corpus covers five editions of the Constitution of Uzbekistan, which enables diachronic analysis of terminological changes. The results demonstrate that legal terms with high frequency in the National Corpus maintain prominent in constitutional discourse, particularly in modern editions of the Basic Law. It is established that terms related to state structure, the judicial system, and the regulation of individual rights exhibit stable and increasing frequency. At the same time, differences between general-language and legally specialized usage are revealed, largely determined by genre and normative features of the texts. The diachronic analysis identifies the intensification of rights-related and state-structural terminology in the post-Soviet period. The data obtained in this study may be used for further research in legal linguistics, terminography, and the analysis of normative texts.

Keywords: legal terminology, frequency analysis, National Corpus of the Russian Language, constitutional discourse, legal discourse, diachronic analysis, legal terms, constitutional texts, Uzbekistan Constitution.

Introduction

The study of legal terminology in diachronic and cross-corpus perspectives is a significant task of contemporary legal linguistics, as it makes it possible to trace the dynamics of terms that reflect the development of statehood, law-making practices, and legal consciousness. With the increasing use of corpus-based methods, the need for comparative analysis of the functioning of legal terminology across various types of texts – from general-language corpora to specialized legal collections – becomes more evident.

The present study aims to identify correlations in the frequency of Russian legal terms between the National Corpus of the Russian Language (NCRL) and the Russian-language corpus of the Constitutions of Uzbekistan for the period 1927–2023. The objective is to determine the extent to which the frequency of legal terms in the national language corresponds to their distribution within specialized constitutional-legal discourse.

Scientific Novelty of the Study. The approach proposed in this article offers several significant advantages:

- It is based on cross-corpus comparison, a method that is rarely applied in legal linguistics;
- It combines traditional corpus data (the National Corpus of the Russian Language) with automated processing of a legal corpus using MyStem tool;
- It reveals the dynamics of the distribution of legal terms in both the national language and specialized legal discourse;
- It also makes it possible to classify legal terms according to the degree of their “legal specificity,” thereby identifying their functional status within the legal lexicon.

Literature Review

A substantial body of research in contemporary linguistics is devoted to the systemic, semantic, cross-cultural, and methodological analysis of legal terminology. Significant contributions to the development of the theory of the legal term have been made by both international and domestic scholars. For example, the work of A. Jopek-Bosiacka [1] examines the specificity of defining legal terms in a cross-cultural context, emphasizing discrepancies in conceptual models of law across different legal traditions. The author highlights the need to consider cognitive and cultural factors when defining legal terms, as this ensures adequate interpretation of legal units in multilingual communication.

Within the field of theoretical and terminological analysis, the studies of G. Gulyamova [2], V. Lazariiev [3], and A. Jumabayeva [4] are of particular importance. G. Gulyamova explores the relationship between the notions of “term” and “concept” in legal terminology, proposing a refined conceptual framework and identifying specific features of legal conceptualization. V. Lazariiev focuses on the functioning of legal terminology in the European Union, where the interaction of multilingual legal systems necessitates harmonization and reveals *highlights* the dynamic nature of legal terminology. A. Jumabayeva provides a lexico-semantic analysis of legal terms, identifying typical semantic relations and key tendencies in their derivational development.

The formation of the Russian scholarly tradition in the study of legal terminology has been significantly shaped by the works of B. V. Baranovskaya [5], D. I. Miloslavskaya [6], T. P. Nekrasova [7], and N. G. Sichinava [8]. B. V. Baranovskaya systematizes the notion of a legal term, classifies its types, and demonstrates the functional features of legal nomination within professional communication. D. I. Miloslavskaya proposes a comprehensive systemic description of modern Russian legal terminology, outlining its structural organization, derivational principles, and terminological criteria. T. P. Nekrasova examines the specifics of translating Russian legal terminology into English, emphasizing the importance of cross-linguistic equivalence, legal realia, and genre conventions of legal style. N. G. Sichinava studies lexico-semantic derivation in the formation of Russian legal terminology, identifying stages in the development of the terminological system and mechanisms of semantic change in key legal concepts.

The methodological dimension of research is represented by O. S. Tarasenko [9], who analyzes the specifics of teaching legal vocabulary in Russian as a foreign language. The author highlights the peculiarities of semantic interpretation of legal terms, describes strategies for

their introduction into the learning process, and emphasizes the need for a differentiated approach to developing a professionally relevant legal lexicon among foreign learners.

Research Methodology

1. Compilation of the Initial List of Legal Terms

At the first stage of the study, the identification and preliminary selection of key legal units functioning in contemporary Russian were carried out. To obtain a representative set of legal vocabulary, an automated linguistic extraction method was employed. Based on algorithmic data analysis, artificial intelligence generated a list of 50 legal terms characterized by high frequency and stable usage in modern Russian.

The selection process relied on several criteria:

- lexico-semantic affiliation with the legal domain;
- frequency of use in a wide range of legal texts (legislation, judicial acts, official-administrative documents);
- terminological stability established in professional usage;
- single-component structure, required for comparable statistical analysis.

The resulting list represents a model set of high-frequency legal nominations and can serve as an objective tool for subsequent cross-corpus comparison.

2. Determining the Frequency of Legal Terms in the National Corpus of the Russian Language.

At the second stage, quantitative indicators of the selected terms were determined using the National Corpus of the Russian Language (NCRL). The main frequency metric used was IPM (instances per million) — the number of occurrences of a lemma per one million word tokens in the corpus.

The choice of IPM is explained by the fact that this metric normalizes frequency relative to corpus size, ensuring comparability across corpora of different volumes.

The data extraction procedure included:

- querying each term in the NCRL interface in lemma-search mode;
- recording the absolute number of occurrences;
- retrieving the automatically calculated IPM value provided by the corpus system.

These results made it possible to determine the degree of frequency relevance of legal terms in the contemporary national linguistic landscape.

3. Formation of the Russian-Language Corpus of the Constitutions of Uzbekistan.

At the third stage, a separate specialized corpus of Russian-language constitutional texts of the Republic of Uzbekistan was created. This preparation ensured comparability with the NCRL data and allowed the reduction of system noise associated with structural differences between texts.

4. Determining the Frequency of Legal Terms in the Constitutional Corpus Using MyStem.

The next stage involved automatic morphological annotation of the constitutional corpus using MyStem, a tool developed by Yandex and widely used for morphological analysis of Russian texts

The use of MyStem made it possible to:

- perform lemmatization, ensuring compatibility with the queries used in the NCRL;
- automatically calculate the absolute frequency of each term under analysis.

The tool ensures a high level of precision in morphological analysis, which is particularly important for legal texts, as constitutional documents possess a highly formalized structure and a stable terminological system.

5. Cross-Corpus Comparison of Frequency Data.

The final stage of the research involved a comparative analysis of the frequency indicators of legal terms obtained from the National Corpus of the Russian Language and from the specialized corpus of the Constitutions (processed with MyStem).

The resulting model of legal vocabulary illustrates:

- which legal terms are most stable within the Russian-language legal discourse;
- which terms significantly increase their frequency specifically within constitutional-legal texts;
- how the status and prominence of legal terms shift when moving from general-language usage to specialized legal discourse.

A summary of the research findings is presented in the table below.

№	Legal Term	IPM Indicator	Constitution of 1927	Constitution of 1937	Constitution of 1978	Constitution of 1992	Constitution of 2023
1.	Территория	87	31	10	29	26	37
2.	Гражданин	108	11	43	18	57	81
3.	Свобода	171	8	7	26	26	51
4.	Партия	214	-	2	6	14	12
5.	Суд	222	4	25	34	75	78
6.	Государство	234	7	9	48	36	79
7.	Закон	295	9	8	63	120	157
8.	Власть	357	29	13	23	39	45
9.	Право	370	35	46	150	89	142
10.	Народ	532	1	14	22	24	29

The table represents the final section of a comparative analysis of the frequency of key Russian legal terms across two corpora:

(1) the National Corpus of the Russian Language, where the IPM indicator reflects the general-language significance and frequency of terms in contemporary Russian discourse;

(2) the Russian-language texts of the Constitutions of Uzbekistan from five historical periods (1927, 1937, 1978, 1992, and 2023), where their normative and functional roles within constitutional-legal style are recorded.

The comparison of these datasets made it possible to identify the diachronic dynamics of legal terminology and to determine the degree to which it corresponds to general Russian frequency trends. The analysis shows that legal terms with high frequency in the National Corpus (such as law, right, court, state, people) display similarly high levels of distribution in constitutional texts, particularly in the editions from the modern period (1992 and 2023). This indicates a strong correlation between general-language and normative-legal frequency profiles of the examined lexicon.

At the same time, the early constitutions (1927 and 1937) are characterized by lower terminological density, reflecting the general features of the Soviet legal model and the stage of institutional formation of Uzbekistan's legal system. Below, we examine the specifics of each term individually.

The term “территория”. With an IPM of 87, this term belongs to the group of stable state-legal nominations in the National Corpus of the Russian Language. In the constitutions, it demonstrates noticeable dynamics: from 31 occurrences in 1927 to 37 in 2023. The highest value is recorded in 1927 (31), which is explained by the early Soviet constitutional focus on legally defining the territorial and administrative status of the republic.

The term “гражданин”. Its IPM value (108) confirms its firmly established position in general-language discourse. In the constitutions, a steady increase in usage is observed: from 11 occurrences in 1927 to 81 in 2023, which correlates with the expansion of the legal status of the individual in the modern state.

The term “свобода”. With a high IPM (171), the term shows moderate frequency in early constitutions (7–8 occurrences) but exhibits a significant increase starting from 1978, reaching 51 occurrences in the 2023 edition. This reflects the strengthening of the rights-protective component of constitutional discourse.

The term “партия”. Despite having a high IPM (214), the term demonstrates considerable fluctuation. A noticeable rise is recorded in 1978 and 1992, corresponding to political transformations and changes within the party system.

The term “суд”. With a correspondingly high IPM (222), the term shows gradual but significant growth in frequency: from 4 occurrences in 1927 to 78 in 2023. The increase reflects the institutional strengthening of the judiciary and the expansion of its constitutional functions.

The term “государство”. The growth in frequency from 7 to 79 indicates increased constitutional attention to issues of state structure and principles of statehood — especially evident in the 1978, 1992, and 2023 editions.

The term “закон”. As one of the most frequent terms in the National Corpus (IPM = 295), it shows a sharp rise in constitutional usage: from isolated occurrences in earlier editions to 157 in 2023. This growth reflects the transition from the declarative model of the Soviet constitution to the normative model of a rule-of-law state.

The term “власть”. With a high IPM (357), the term appears consistently across all editions. Its frequency increases from 29 (1927) to 45 (2023), which is associated with the expanding constitutional regulation of the system of public authority.

The term “право”. With an IPM of 370, it is one of the most frequent terms in the National Corpus. In the constitutions, it demonstrates strong dynamics: from 35 occurrences in 1927 to 142 in 2023. The maximum value occurs in 1978 (150), due to the development of the Soviet concept of socialist legality.

The term “народ”. Being the most frequent among the listed terms in the National Corpus (IPM = 532), it has relatively low but stable representation in the constitutions. Its

frequency rises from 1 (1927) to 29 (2023), corresponding to the constitutional concept of popular sovereignty and the reinforcement of its rhetorical expression.

Discussion

The comparative analysis demonstrates that legal terms exhibit varying levels of frequency depending on the genre of the text, the historical period, and the stages of development of the political system and state-building processes.

The incorporation of legal concepts becomes particularly pronounced during periods of institutional reform (1978, 1992, and 2023). Terms related to statehood and regulatory mechanisms consistently display high frequency regardless of the period under consideration. At the same time, terms referring to individual rights and freedoms significantly increase their frequency after 1991, which correlates with the formation of a new political and legal model of the state.

Ideologically marked terms (for example, *party*) gradually lose their dominant position in the post-Soviet editions of the Constitution. The consistency between the frequency profiles observed in the National Corpus of the Russian Language and those in the constitutional texts confirms the systemic character of legal terminology.

At the same time, certain discrepancies have been identified: some lexemes that are highly frequent in general-language discourse (for instance, *people*) occur considerably less often in constitutional texts, which can be explained by the genre-specific characteristics of normative legal documents.

Conclusion

The conducted study made it possible to identify stable patterns in the functioning of legal terminology within both general-language discourse and specialized constitutional discourse. The cross-corpus comparison revealed a high degree of correlation between the general-language and normative frequency profiles of legal terms, clear diachronic trends reflecting transformations in state structure and legal ideology, and the growing prominence of terms related to state institutions, the judiciary, and individual rights in the contemporary editions of the Constitution of Uzbekistan.

The methodology based on the combined use of the National Corpus of the Russian Language, the MyStem morphological analyzer, and a historically stratified corpus of constitutional texts has demonstrated its effectiveness. This approach can be applied in further research in the fields of legal linguistics, the analysis of normative legal texts, and the study of the evolution of legal terminology.

References:

1. Jopek-Bosiacka, A. (2011). Defining law terms: A cross-cultural perspective. *Research in Language*, 9(1), 9–29.
2. Gulyamova, G. (2022). Term and concept in legal terminology. *Academic Research in Educational Sciences*, 1, 181–190.
3. Lazariiev, V. (2021). Features of the use of legal terminology in the countries of the European Union. *ScienceRise: Juridical Science*, 3(17), 4–8.
4. Jumabayeva, A. (2025). Lexico-semantic features of legal terminology. *Journal of Science-Innovative Research in Uzbekistan*, 3(4), 178–185.



5. Барановская, Б.В. (2023). Юридические термины в лингвистике: понятие и виды. Молодой учёный, 40(487), 139–142.
6. Милославская, Д.И. (2000). Системное описание юридической терминологии в современном русском языке (Дисс. ... канд. филол. наук). Москва.
7. Некрасова, Т.П. (2013). Особенности перевода юридической терминологии с русского языка на английский язык (Дисс. ... канд. филол. наук). Москва.
8. Сичинава, Н.Г. (2003). Лексико-семантическая деривация при становлении русской юридической терминологии. Тбилиси.
9. Тарасенко, О.С. (2015). Особенности обучения юридической лексике на занятиях по русскому языку как иностранному. Филологические науки. Вопросы теории и практики, 12(54), 180–182.
10. Конституция Узбекской Советской Социалистической Республики. (1927). Самарканд: Юридическое издательство СНК УзССР.
11. Конституция Узбекской Советской Социалистической Республики. (1937). Ташкент: Партиздат.
12. Конституция Узбекской Советской Социалистической Республики. (1978). Ташкент: Узбекистан.
13. Конституция Республики Узбекистан. (1992). Ташкент: Узбекистан.
14. Конституция Республики Узбекистан. (2023). Ташкент: Узбекистан.