Mubin RAKHIMOV,
Teacher of Bukhara State University
E-mail: mubinrahimovbuba@gmail.com

Based on the review by Associate Professor of Bukhara State University, DSc Mehrinigor Ahmedova

PROBLEMS OF CREATING A LINGUISTIC CORPUS BASED ON JURISDICTIVE TEXTS
Annotation

The creation of linguistic corpora based on jurisdictive texts raises several issues. The research aimed to analyze a system of creating linguistic corpora based on the jurisdictive texts. For solving issues, the comparative analysis method has been distinguished. The study further discusses the role of national constitutions as foundational legal documents, with a comparative analysis of the legal frameworks of the United States, the United Kingdom, Australia, and Uzbekistan. Election laws, being integral to governance, are also analyzed, showcasing how different legislative systems shape electoral processes.

In conclusion, the study pointed out the complexity of developing a linguistic corpus for jurisdictive texts influenced by terminological, ethnolinguistic, and cultural phenomena. The research establishes a shape in creating a structured and functional corpus that can aid in the systematic study of legal language across different jurisdictions.

Key words: linguistic corpus, jurisdictive texts, constitution, electoral processes, MySQL database.

ПРОБЛЕМЫ СОЗДАНИЯ ЛИНГВИСТИЧЕСКОГО КОРПУСА НА ОСНОВЕ ЮРИДИЧЕСКИХ ТЕКСТОВ
Аннотация

Создание лингвистических корпусов на основе юридических текстов вызывает ряд проблем. Целью исследования было проанализировать систему создания лингвистических корпусов на основе юридических текстов. Для решения возникающих вопросов был выделен метод сравнительного анализа.

В исследовании также обсуждается роль национальных конституций как основополагающих правовых документов, а также проводится сравнительный анализ правовых систем США, Великобритании, Австралии и Узбекистана. Анализируются законы о выборах как неотъемлемая часть государственного управления, демонстрируя, каким образом различные законодательные системы формируют избирательные процессы.

В заключение исследование отмечает сложность разработки лингвистического корпуса для юридических текстов, обусловленную терминологическими, этнолингвистическими и культурными явлениями. Исследование формирует основу для создания структурированного и функционального корпуса, который может помочь в систематическом изучении правового языка в различных юрисдикциях.

Ключевые слова: лингвистический корпус, юридические тексты, конституция, избирательные процессы, база данных MySQL.

QONUN HUJJATLARI MATNI LINGVISTIK KORPUSINI YARATISH MUAMMOLARI
Annotatsiya

Huquqiy matnlarga asoslangan linguistik korpuslarni yaratish bir qator muammolarni keltirib chiqaradi. Tadqiqotning maqsadi huquqiy matnlarga asoslangan linguistik korpuslarni yaratish tizimini tahlil qilishdir. Yuzaga keladigan masalalarni hal qilish uchun solishtirma tahlil usuli tanlangan.

Tadqiqotda milliy konstitutsiyalar asosiy huquqiy hujjatlar sifatidagi roli ham muhokama qilinadi, shuningdek AQSh, Britaniya, Avstraliya va O'zbekiston huquq tizimlarini solishtirma tahlili o'tkaziladi. Davlat boshqaruvining ajralmas qismi sifatida saylov qonunlari tahlil qilinadi, turli huquqiy tizimlar saylov jarayonlarini qanday shakllantirishini namoyon etadi.

Xulosa qilib aytish mumkinki, huquqiy matnlar uchun linguistik korpusni ishlab chiqish terminologik, etnolingvistik va madaniy holatlar sababli murakkabdir. Tadqiqot turli yurisdiktsiyalar huquqiy tilini tizimli o'rganishda yordam berishi mumkin bo'lgan tuzilmali va funksional korpus yaratish uchun asos yaratadi.

Kalit so'zlar: Linguistik korpus, huquqiy matnlar, konstitutsiya, saylov jarayonlari, mysql ma'lumotlar bazasi

Introduction. The creation of linguistic corpora is directly linked to various spoken and written text styles. The text creation process requires specific knowledge, skills, and expertise. In any language, the jurisdictive terminology and texts based on legislation terms require not only jurisdictive knowledge and literacy related to national values but also linguistic competence.

Literature review. The selection of jurisdictive terminology is important to choose lexical units that are comprehensible to all social layers while maintaining a distinct style. The proper selection of grammatical tools such as word formation structures, moods, tenses, and syntactic structures, plays a crucial role in establishing the fundamental principles of jurisdictive texts. In addition to the correct use of grammatical phenomena, mastering the skill of accurately incorporating jurisdictive terminology in texts is a key responsibility of text creators. Commenting on this item, the scientist H. A. Kerimov highlights the following data: "The misuse or improper application of a word or phrase can create discrepancies between the intended meaning and its textual expression, sometimes leading to severe consequences - especially in the field of lawmaking" [8]. In the expression of jurisdictive texts, a new terminology known as "legal language" has emerged, and scholars have conducted several studies on

this subject. For instance, French scholar of jurisprudence F. Jenny suggested that there are specialized linguistic tools and methods used to shape legislative thought in legislative activity and legal practice [7]. These tools help distinguish legislative style from general literary speech by defining the specific objectives of the law, the unique method of presenting topics, and the specialized vocabulary employed to express the lawmaker's intent.

Research Methodology. Jurisdictive texts are defined as a collection of actions regulated and enforced by a given society, expressed by linguistic means. Each country's constitution serves as its primary legal framework, guiding the formulation of all other legal documents while maintaining consistency with its principles. Constitutional texts are structured based on the social contract, formation, and cultural aspects of a given society. When drafting constitutional texts, priority is given to using terminology that is comprehensible to all social groups, maintaining a formal style, ensuring proper punctuation, and correctly applying grammatical structures.

Legislation norms such as laws, codes, resolutions, decrees, orders, and other normative documents adhere to a specific, unchanging format and style. The primary guiding factor in structuring texts is the standards established in a

country's constitution. In the creation of linguistic corpora for jurisdictive texts, selecting which legal document to include poses a challenge, as the availability of national legal text translations plays a significant role. Therefore, it is advisable to compare the legislative documents of English-speaking countries with those of the national legal framework.

Uzbekistan's current legal framework consists of normative documents adopted after 1991. The Constitution of Uzbekistan was adopted on December 8, 1992, and was amended through a referendum on April 30, 2023. Research has been conducted on forming a corpus of legal texts, including the new constitution adopted in 2023.

As part of this study, the constitutions of other countries were also analyzed. The U.S. Constitution, adopted in 1787, includes the "Bill of Rights." The British Constitution, developed since the 13th century, comprises written and unwritten laws, starting with the Magna Carta Libertatum and including the Representation of the People Act (1928). The Australian Constitution, adopted in 1901, was last amended in 2023. Given the importance of constitutions as fundamental legal documents, they were selected as primary objects for linguistic corpus creation.

In addition to constitutions, the legal frameworks of these countries, including their election laws, were also analyzed:

Country	Constitution	Codes	Laws	Other Legal Norms
USA	Adopted in 1787	54 sections of the U.S. Code	State laws, common law system	-
UK	Collection of uncodified laws from the 12th to 18th centuries	Not applicable	Common law system, status laws, Scots law	The Magna Carta (Magna Charta Libertatum); The Petition of Right; The Habeas Corpus Act; The Bill of Rights "An Act Declaring the Rights and Liberties of the Subject and Settling the Succession of the Crown." [6]
Australia	Adopted in 1901	State codes, criminal code	19 sets of laws, Indigenous Australian laws	Statutes, regulations
Uzbekistan	Adopted in 1992, amended in 2023	19 codes	17 active laws	Resolutions, decrees, orders

Comparative Analysis of Election Laws

The election system, as a fundamental aspect of political life and one of the governing legal processes, was also analyzed.

USA: Election laws are outlined in Section 52 of the U.S. Code but not compiled into a single unified code. The legal texts specify the electoral procedures for Congress, the President, and state governors [9].

UK: Election laws have evolved over several centuries and include the Political Parties, Elections and Referendums Act (2000), parliamentary acts (1911, 1949), and other acts defining voting rights [10]. The election legislation includes texts regarding the elections of the UK Parliament, the Scottish Parliament, the Northern Ireland Assembly, the Welsh National Assembly, local self-governing bodies, county and district councils, mayors, and police commissioners.

Australia: Unlike other countries, Australia has consolidated its election laws into a single legislative framework. A notable difference is that voting is mandatory for all eligible voters [11].

The comparative analysis in the table illustrates the unique legislative frameworks of different governance systems. It also highlights that each country has undergone distinct developmental phases in shaping it's legal frameworks.

Both Uzbek- and English-speaking countries develop and implement constitutional and election laws. However, several challenges arise in creating a linguistic corpus of legal texts:

Legal texts evolve and improve over periods;

Linguistic corpora are not fixed entities but are continuously updated and expanded;

A speaker of one language cannot fully assimilate the cultural background of another, yet legal texts often reflect cultural and ethnic histories;

Governance structures vary across regions;

Technical limitations in corpus-building software.

The research based on creating linguistic corpora includes complex processes. Thus, choosing an appropriate legal document for analysis is also essential in the given process.
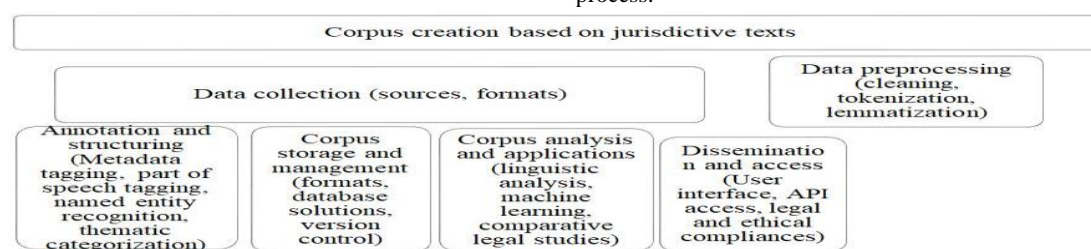
pic. 1. System of creating a linguistic corpus based on jurisdictive text

Analysis and results. According to this scheme, the creation of the linguistic corpus based on jurisdictive texts associate data-based analysis, connected with programming processes. Searching these phenomena, to ensure an effective corpus for jurisdictive texts, the following requirements had been proposed:

The corpus should include legal documents applicable to the languages studied, particularly constitutions and relevant legal materials.

The software should provide a search function to facilitate the retrieval of words or phrases from texts.

The corpus should allow unlimited lexical entries.

Given the comparative nature of the study, legal documents should be provided in at least two languages.

The corpus should support N-gram analysis to generate statistical data.

Statistical analyses should be visually represented using diagrams and frequency charts.

The corpus should be structured based on specific legal genres.

By addressing these considerations, the development of a comprehensive and functional linguistic corpus based on jurisdictive texts can be achieved. The creation of a synchronic linguistic corpus of jurisdictive texts given in English and Uzbek needs an appropriate computer program. It is convenient to use the MySQL programming system. MySQL is an open-source database that serves as an efficient tool for storing and managing large volumes of data. Using MySQL, legislative document texts are stored in an organized and systematic manner. The database provides capabilities for searching, analyzing texts, and generating statistical data. The corpus offers the following key features:

N-gram Statistics: It allows extracting n-gram statistics for each word, showing how words are connected and distributed within the text.

Most Frequently Used Words and Phrases: It enables the identification and statistical analysis of the most frequently used words or phrases.

Paragraphs Containing Words or Phrases: It provides the ability to extract paragraphs that contain specific words or phrases, which helps in understanding the overall context of the text and analyzing word usage.

Visualization of Statistical Data: The ability to display statistical data in diagrams allows for comparative text analysis and visual representation of analytical results.

Conclusion. Creating a linguistic corpus for legal documents constitutes an important advancement in the areas of legal research, linguistics, and computational analysis. By methodically gathering legal documents from different jurisdictions, this study improves cross-border legal comprehension and language accuracy. Due to the rigidity and formal structure of legal language, employing corpus-based methods guarantees that legal terms are retained and precisely examined [1].

Additionally, the combination of MySQL with advanced computational techniques enables effective data storage, access, and linguistic analysis. Instruments like n-gram analysis and frequency statistics offer important insights into the syntactic and semantic structures of legal discourse, rendering the corpus an essential resource for both researchers and practitioners [2].

Despite issues concerning multilingual compatibility and the continuous changes in legal terminology, this research highlights the significance of creating standardized approaches for corpus development. Legal translation, comparative law, and legislative drafting can all gain advantages from a well-organized and thorough corpus, emphasizing the importance of linguistic analysis in legal research [3].

In the end, this study establishes a foundation for forthcoming advancements in legal text handling. As legal systems develop, a strong body of juridical texts will be vital for maintaining clarity, accessibility, and uniformity in legal communication among various languages and legal frameworks [4]. By combining linguistic and legal knowledge with technological innovations, this corpus acts as a fundamental resource for promoting worldwide legal clarity and knowledge sharing [5].

In summary, the complexity of creating a linguistic corpus for legal texts stems from terminological, linguistic-cultural, geographical, and ethnolinguistic factors. Additionally, due to constraints in automated grammatical analysis tools, human involvement remains essential in the research process.

REFERENCES

1. Bhatia, V. K. (2010). Legal Discourse: Opportunities and Challenges for Corpus Linguistics. International Journal of Legal Discourse, 1(1), 9-26.
2. Tiersma, P. M. (1999). Legal Language. University of Chicago Press.
3. Šarčević, S. (2016). Legal Translation and Legal Terminology: Challenges and Solutions. Routledge.
4. Mattila, H. E. S. (2006). Comparative Legal Linguistics: Language of Law, Latin and Modern Lingua Francas. Ashgate Publishing.
5. Gémar, J. C. (2018). Translation and Legal Discourse: The Dynamics of Legal Languages and Legal Systems. Springer.
6. Дайси А. В. Основы государственного права Англии. М.: Издание Тов-ва И. Д. Сытина, 1905.
7. Карташев В.Н. Юридическая техника, тактика, стратегия и технология. Н. Новгород, 2000. ст. 167.
8. Керимов. Д. А. Законодательная техника: науч.-метод. и учеб. Пособие. М.: НОРМАИНФРА-М 1998. ст.123
9. https://www.usa.gov/state-local-elections
10. https://cyberleninka.ru/article/n/pravovoe-regulirovanie-vyborov-v-soedinennom-korolevstve-velikobritanii-i-severnoy-irlandii/viewer
11. https://www.democracy.ru/library/foreign/countries/australia/rus_1991-2.html