BGPulseJournal
ISSN:1 http://www.bgpulsewa.com/index
BG PULSE JOURNAL

PRINCIPLES OF DEVELOPING AN ENGLISH–UZBEK LEGISLATIVE TEXT CORPUS: A CORPUS-BASED LEGAL LINGUISTIC APPROACH

Mubin M. Rakhimov
Bukhara State University, Uzbekistan

Abstract. The development of corpus linguistics has created new opportunities for studying legal discourse, translation processes, and language variation through digital resources. This article examines the theoretical and practical principles of creating an English–Uzbek legislative text corpus, focusing mainly on constitutional and electoral documents. The research discusses the main issues related to the formation of linguistic resources, digitization, text processing, and the organization of electronic legal materials.

Legal texts have a specific linguistic nature because they require accuracy, consistency, and stability of meaning. Therefore, the process of creating a legal corpus cannot be limited only to collecting texts; it also requires careful consideration of terminology, grammatical structures, and functional features of legal discourse.

The developed corpus includes constitutional and electoral legislation of Uzbekistan, the United States, the United Kingdom, and Australia. The corpus was created using PHP, JavaScript, and MySQL technologies, which provide possibilities for searching, frequency analysis, n-gram statistics, contextual analysis, and visualization of linguistic data. The findings show that the corpus can serve as a useful resource for legal linguistics, translation studies, and comparative analysis of legislative discourse.

Keywords: legal corpus, corpus linguistics, legislative discourse, English–Uzbek translation, legal terminology, computational linguistics.

Introduction. The creation of linguistic corpora has become one of the important directions in modern linguistics. With the development of digital technologies, researchers have gained the opportunity to examine language materials based on authentic texts and large-scale electronic databases. Corpus linguistics provides effective methods for identifying lexical, grammatical, and semantic patterns in different types of discourse (McEnery & Hardie, 2012).

Among different types of corpora, legal corpora occupy a special place because legal texts have their own structural and linguistic characteristics. Legislative documents regulate social relations and therefore require a high level of accuracy and clarity. In legal communication, even a minor change in the choice of a word or grammatical structure may influence the interpretation of a legal norm.

The creation of legal texts is a complex process that requires not only knowledge of law but also linguistic competence. Each legal system develops its own terminology and stylistic features based on historical, social, and cultural factors. Therefore, the formation of a legislative corpus requires careful selection and analysis of legal materials.

As Kerimov states, an incorrectly chosen word or expression in legal discourse may create differences between the intended meaning and its textual realization, which can lead to serious consequences, particularly in legislative activities (Kerimov, 2015).

Legal texts differ from other types of written discourse due to their functional stability and terminological consistency. The main feature of legal terminology is that terms are generally used within a specific context and are expected to maintain a fixed meaning.

The concept of "legal language" has developed as a separate linguistic phenomenon. According to Gény, legislative discourse has its own linguistic means and methods of expression that distinguish it from other forms of literary language. These features include a special vocabulary, formal structures, and specific ways of presenting legal ideas.

The accuracy of legal texts depends greatly on the appropriate use of grammatical structures. Morphological and syntactic elements such as verb forms, modality, sentence organization, and noun phrases contribute to maintaining the formal character of legislation.

A legislative text should preserve both linguistic stability and legal meaning. For this reason, the writer of legal documents must carefully select terminology and grammatical constructions. The misuse of linguistic elements may affect not only the style of the document but also its legal interpretation.

Materials and Methods. The development of the English–Uzbek legislative text corpus was carried out through several stages, including the selection of legal documents, electronic processing, text normalization, and database organization. Since legal texts represent a highly specialized type of discourse, the selection of corpus materials required both linguistic and legal considerations.

The main objects of the corpus were constitutional and electoral legislative documents of Uzbekistan, the United States, the United Kingdom, and Australia. These countries were selected because their legal systems represent different historical and structural models of legislation.

The Constitution of the Republic of Uzbekistan, adopted on 8 December 1992 and renewed through the national referendum held on 30 April 2023, was selected as one of the main sources for the corpus. As the fundamental legal document of the country, the Constitution serves as the basis for other legislative acts, including codes, laws, and subordinate legal documents.

The corpus also includes the Constitution of the United States, adopted in 1787, which is considered one of the oldest written constitutions still in force. In addition, constitutional documents of the United Kingdom and Australia were included in the corpus. The British constitutional system represents a unique model because it consists of both written and unwritten constitutional sources, including historical documents such as Magna Carta and later parliamentary acts. The Australian Constitution, adopted in 1901, represents another important example of a constitutional system within the English-speaking legal tradition.

Electoral legislation was also selected as an important part of the corpus because elections represent a fundamental area of public law. The electoral texts demonstrate how legal systems regulate political participation, representation, and institutional procedures.

The corpus construction process involved collecting authentic legal documents from official sources. After collection, the texts were converted into electronic format and carefully reviewed. The preparation stage included removing unnecessary symbols, correcting formatting inconsistencies, and standardizing text presentation.

Such preparation was necessary because corpus quality directly influences the reliability of linguistic analysis. According to Sinclair (2004), a corpus should represent authentic language use and maintain consistency in order to provide meaningful research results.

The creation of the legislative corpus required the combination of linguistic analysis and technological solutions. The developed corpus was designed as an interactive electronic resource that allows users to search, compare, and analyze legislative texts.

The technical structure of the corpus was created using PHP and JavaScript programming languages together with a MySQL database management system. PHP was applied for server-side operations, while JavaScript was used to improve user interaction with the system.

The MySQL database provided an organized environment for storing large amounts of legislative data. Each document was structured according to its country, legal category, and textual characteristics. This organization made it possible to perform various types of linguistic analysis.

The developed corpus provides several analytical functions. First, the system enables frequency analysis of individual words and expressions. This function helps identify the most commonly used lexical units in legislative discourse.

Second, the corpus allows the extraction of n-gram structures. One-word, two-word, and three-word combinations can be analyzed to determine recurring patterns and phraseological structures in legal texts.

Third, users can retrieve the context in which a particular word or phrase occurs. This function is especially important in legal research because the meaning of a term often depends on its textual environment.

Finally, the corpus provides statistical visualization tools that make comparative analysis between different legal systems easier. The implementation of these functions demonstrates that digital corpus technologies can significantly support research in legal linguistics and translation studies.

Results and Discussion. The analysis of the developed corpus shows that legislative texts share several common linguistic characteristics despite differences among legal systems.

One of the most important findings is the stability of legal terminology. Unlike general language, legal discourse requires terms to maintain relatively fixed meanings. This feature allows researchers to identify repeated patterns and compare the use of legal concepts across different languages.

The corpus-based analysis also confirms the importance of grammatical structures in legislative communication. Legal documents frequently rely on standardized sentence patterns, formal constructions, and specific grammatical choices to avoid ambiguity.

The comparison of constitutional and electoral texts from Uzbekistan, the United States, the United Kingdom, and Australia demonstrates that each legal system has developed its own linguistic traditions. However, all of them share the common principle of accuracy and institutional stability.

The English–Uzbek comparison is particularly significant for translation studies. Legal translation requires not only replacing words from one language into another but also transferring legal meaning between different legal cultures. Therefore, corpus-based analysis provides an empirical basis for identifying similarities and differences between source and translated texts.

As Baker (1995) emphasizes, corpora enable researchers to study translated language through real textual evidence rather than relying only on theoretical assumptions. In this regard, the developed corpus can contribute to improving the quality of legal translation and terminology management.

Another important result is that the created corpus can support further studies in computational linguistics. The integration of database technologies with linguistic analysis provides opportunities for automatic retrieval, statistical processing, and comparative investigation of legal discourse.

Conclusion. The development of an English–Uzbek legislative text corpus demonstrates that corpus creation is a multidisciplinary process combining linguistic theory, legal knowledge, and information technologies.

Legal texts require special attention because their meaning depends on terminological accuracy, grammatical stability, and formal organization. Therefore, the creation of a legal corpus requires carefully selected authentic materials and reliable methods of electronic processing.

The developed corpus includes constitutional and electoral legislation from Uzbekistan, the United States, the United Kingdom, and Australia. Its structure allows researchers to conduct lexical, grammatical, contextual, and statistical analyses. The use of PHP, JavaScript, and MySQL technologies has transformed traditional legal texts into an interactive research resource. The corpus provides possibilities for studying legal terminology, translation equivalence, discourse patterns, and comparative features of legislative language.

The created English–Uzbek legislative corpus can serve as a foundation for future research in corpus linguistics, legal linguistics, translation studies, and digital humanities.

References

1. Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. Target, 7(2), 223–243.
2. Gény, F. (1927). Science et technique en droit privé positif. Paris.
3. Kerimov, H. A. (2015). Legal Language and Problems of Legislative Style. Tashkent.
4. McEnery, T., & Hardie, A. (2012). Corpus Linguistics: Method, Theory and Practice. Cambridge University Press.
5. Sinclair, J. (2004). Trust the Text: Language, Corpus and Discourse. London: Routledge.
6. Tiersma, P. M. (1999). Legal Language. Chicago: University of Chicago Press.