

## CORPUS-BASED MODELLING OF SCRIPT VARIATION AND DIGITAL LEXICAL INNOVATION IN MODERN UZBEK: A FUNCTIONAL APPROACH TO NORMATIVE STABILIZATION

Rakhmatullayeva Dilafruzkhon Shukhratovna  
PhD, Associate Professor, Department of Uzbek  
Language and Literature Kokand State University  
E-mail: dilafruzxonrahmatullayeva53@gmail.com  
ORCID: <https://orcid.org/0009-0007-0568-3137>

### ABSTRACT

This article investigates one of the most urgent problems of contemporary Uzbek linguistics: the functional stabilization of the literary norm under conditions of digital communication, script coexistence, rapid lexical borrowing and corpus-based language description. The central argument is that modernization of Uzbek cannot be reduced to orthographic regulation alone, because real linguistic development is now formed at the intersection of education, official communication, digital journalism, social media, machine transliteration, speech technologies and academic terminology. The study proposes a functional model for analysing modern Uzbek through five interrelated dimensions: script variation, lexical innovation, morpho-syntactic regularity, digital-register differentiation and educational codification. The research is based on qualitative synthesis of corpus-linguistic, sociolinguistic and applied-linguistic approaches, with particular attention to the challenges faced by an agglutinative, historically multi-script and comparatively low-resource language in the digital environment. The findings indicate that Latin-Cyrillic coexistence, English-based digital borrowings and inconsistent orthographic habits should not be treated merely as signs of disorder. They constitute empirical material through which linguists can identify actual usage patterns and design more precise lexicographic, educational and technological solutions. The article concludes that the future of Uzbek literary language development depends on evidence-based codification, balanced terminology policy, high-quality annotated corpora and systematic integration of linguistic research with digital humanities and language technologies.

**Keywords:** Uzbek linguistics; corpus linguistics; script variation; digital communication; lexical borrowing; literary norm; language policy; normative stabilization.

### Аннотация

В статье исследуется одна из наиболее актуальных проблем современной узбекской лингвистики — функциональная стабилизация литературной нормы в условиях цифровой коммуникации, сосуществования графических систем, интенсивного лексического заимствования и корпусного описания языка. Основная идея исследования заключается в том, что модернизацию узбекского языка нельзя сводить только к орфографическому регулированию, поскольку реальные процессы языкового развития сегодня формируются на пересечении образования, официальной коммуникации, цифровой журналистики, социальных сетей, машинной транслитерации, речевых технологий и академической терминологии. В статье предлагается функциональная

модель анализа современного узбекского языка по пяти взаимосвязанным параметрам: графическая вариативность, лексические инновации, морфосинтаксическая регулярность, цифровая регистровая дифференциация и образовательная кодификация. Исследование основано на качественном синтезе корпусно-лингвистических, социолингвистических и прикладных подходов с учетом специфики узбекского языка как агглютинативного, исторически многописьменного и сравнительно малоресурсного языка. Сделан вывод о том, что будущее развития узбекской литературной нормы зависит от научно обоснованной кодификации, сбалансированной терминологической политики, качественных аннотированных корпусов и интеграции лингвистических исследований с цифровой гуманитаристикой.

**Ключевые слова:** узбекская лингвистика; корпусная лингвистика; графическая вариативность; цифровая коммуникация; лексическое заимствование; литературная норма; языковая политика; нормативная стабилизация.

### **Annotatsiya**

Ushbu maqolada zamonaviy o'zbek tilshunosligining eng dolzarb masalalaridan biri — raqamli kommunikatsiya, yozuv tizimlari qo'shaloqligi, tezlashgan leksik o'zlashuvlar va korpusga asoslangan til tavsifi sharoitida adabiy me'orning funksional barqarorlashuvi tadqiq etiladi. Tadqiqotning asosiy g'oyasi shundan iboratki, o'zbek tilining modernizatsiyasini faqat imlo yoki alifbo masalasiga bog'lab talqin qilish yetarli emas, chunki bugungi til taraqqiyoti ta'lim, rasmiy nutq, raqamli jurnalistika, ijtimoiy tarmoqlar, mashinaviy transliteratsiya, nutq texnologiyalari va akademik terminologiya kesishmasida shakllanmoqda. Maqolada zamonaviy o'zbek tilini tahlil qilish uchun besh tarkibiy yo'nalishga asoslangan funksional model taklif qilinadi: yozuv variativligi, leksik innovatsiyalar, morfosintaktik muntazamlik, raqamli registrlar farqlanishi va ta'limiy kodifikatsiya. Tadqiqot korpus lingvistikasi, sotsiolingvistika va amaliy tilshunoslikka oid yondashuvlarning sifatli tahliliy sinteziga tayanadi. Natijalar shuni ko'rsatadiki, lotin-kirill yozuvlari qo'shaloqligi, inglizcha raqamli o'zlashmalar va imloviy beqarorliklar faqat me'yor buzilishi sifatida emas, balki real til qo'llanishini aniqlash va ilmiy asoslangan yechimlar ishlab chiqish imkonini beruvchi empirik material sifatida baholanishi zarur.

**Kalit so'zlar:** o'zbek tilshunosligi; korpus lingvistikasi; yozuv variativligi; raqamli kommunikatsiya; leksik o'zlashuv; adabiy me'yor; til siyosati; me'oriy barqarorlashuv.

### **INTRODUCTION**

Modern Uzbek linguistics is situated at a methodological and sociolinguistic turning point. On the one hand, the Uzbek language has a rich literary heritage, a developed grammatical tradition, an established school of normative description and a strong symbolic function in national identity. On the other hand, the contemporary communicative environment has radically changed the mechanisms through which linguistic norms are produced, circulated, challenged and stabilized. In the pre-digital period, the main agents of normativity were school textbooks, dictionaries, official documents, literary publications, academic grammars and

print media. Today these agents coexist with Telegram channels, digital journalism, online education platforms, mobile keyboards, automatic spell-checkers, machine translation systems, speech recognition tools, search engines and informal social-media discourse. This shift does not merely create new technical conditions for language use; it changes the epistemology of linguistic research itself, because the linguist can no longer rely only on selected literary examples or manually observed materials but must also interpret large-scale digital usage, orthographic variation, code-mixing, transliteration practices, lexical innovation and register differentiation. Uzbek is especially sensitive to these processes because it combines three important features. First, it is an agglutinative Turkic language with complex morphological productivity, where a single lexical root can generate many surface forms through inflectional and derivational suffixation. Second, it has a historically layered script situation involving Arabic-based, Latin-based and Cyrillic traditions, and the current coexistence of Latin and Cyrillic continues to affect education, publishing, official communication and digital searchability. Third, it remains comparatively under-resourced in global language technologies when compared with English, Russian, Chinese or Turkish, although recent work on Uzbek educational corpora, machine transliteration, speech resources and morphological analysis demonstrates a clear transition toward computationally supported language description [1; 9; 10; 12; 13]. These developments are not peripheral technical projects. They directly affect the future of literary norm, education, lexicography, language policy and digital sovereignty. The present article therefore addresses the following research problem: how can Uzbek linguistics develop an evidence-based model of normative stabilization that accounts for script variation, digital lexical innovation and corpus-based description without reducing living language processes to mechanical prohibition or purely prescriptive correction? The relevance of this problem is heightened by the continued coexistence of Latin and Cyrillic in public, educational and digital spaces. This coexistence creates not only spelling uncertainty but also search fragmentation, archive fragmentation and stylistic uncertainty. A student may read twentieth-century sources in Cyrillic, submit academic assignments in Latin, communicate informally in a mixed transliterated mode and encounter English technological terms in their unadapted form. Such a situation cannot be resolved through simplified slogans about “correct” and “incorrect” language. It requires a more precise theory of functional normativity. At the same time, English-based digital borrowings, especially in technology, business, marketing, social media and youth discourse, are entering Uzbek at a speed that traditional dictionaries cannot immediately register. A purely defensive attitude toward such borrowings is scientifically weak, because every living language absorbs external elements. However, an unregulated acceptance of unstable forms is equally problematic, because it may weaken terminological coherence, educational clarity and official style. For this reason, the article argues for a functional approach. The main question is not whether a form is “new” or “foreign” in itself, but whether it is phonologically adaptable, morphologically integrable, semantically necessary, stylistically appropriate, educationally explainable and compatible with the long-term development of Uzbek literary language. Normative stabilization should therefore be understood as a dynamic scientific process based on observation, classification, evaluation and codification, not as a static administrative command. The aim of this article is to propose a conceptual model for studying

modern Uzbek through the interaction of corpus linguistics, sociolinguistics and applied language planning. Its objectives are: to define the linguistic nature of script variation in Uzbek digital communication; to explain how digital borrowings and neologisms should be evaluated from a functional perspective; to demonstrate the importance of morphological and corpus-based tools for an agglutinative language; to identify the educational implications of corpus-informed normativity; and to formulate principles for integrating Uzbek linguistic research with digital humanities and language technologies. The scientific novelty of the article lies in its attempt to connect three domains that are often discussed separately: alphabet and orthography, lexical modernization and corpus-based computational description. The practical significance of the study is that the proposed model can be used in university-level Uzbek linguistics courses, academic writing instruction, terminology development, corpus annotation projects, digital dictionary design and language-policy discussions. The theoretical basis draws on corpus linguistics, language planning theory, sociolinguistics of writing systems, digital discourse analysis and Uzbek applied linguistics. In this perspective, Uzbek modernization is not a matter of replacing one form with another; it is the creation of a disciplined but flexible system through which the language can function in science, education, law, technology, literature and everyday communication.

### **MATERIALS AND METHODS**

The study is designed as a qualitative theoretical-analytical investigation rather than as a narrow experimental measurement of a single dataset. This methodological choice is deliberate, because the research problem concerns the structural relationship between several domains of Uzbek language development: orthography, script coexistence, digital borrowing, corpus description, morphological modelling and educational normativity. The materials of the study include contemporary scholarly discussions on Uzbek corpus development, machine transliteration, Uzbek speech resources, morphological analysis, language policy and digital lexical innovation, as well as general theoretical works in corpus linguistics, sociolinguistics and language planning. The research is methodologically based on five procedures. The first procedure is conceptual synthesis, through which different scholarly approaches are brought into a single explanatory frame. Uzbek digital language development cannot be adequately described by one discipline alone. Traditional grammar explains morphological and syntactic regularity, corpus linguistics explains frequency and distribution, sociolinguistics explains variation and identity, while applied language planning explains codification and institutional implementation. The second procedure is functional classification, which evaluates linguistic phenomena not only by origin but also by communicative role. For example, a borrowed English digital term may function as a necessary technical term, a prestige marker, a youth identity marker, a temporary slang item or an unnecessary duplication of an existing Uzbek equivalent. These functions must be separated before any normative conclusion is drawn. The third procedure is corpus-oriented reasoning, which treats frequency, distribution, genre, register and collocation as essential indicators of norm formation. Even where a full empirical corpus is not constructed in this article, the logic of corpus linguistics is applied as an analytical standard. A form becomes linguistically significant not because it appears once, but because it shows stable usage across speakers, genres and communicative situations. The

fourth procedure is sociolinguistic interpretation, which considers the role of age, education, institutional context, technological platform and script literacy in shaping language behaviour. This procedure is especially important for Uzbek because Latin and Cyrillic literacy are not merely technical writing choices; they are connected with educational generation, publication history, administrative practice, keyboard habits and cultural memory. The fifth procedure is normative evaluation, by which descriptive findings are translated into recommendations for literary-language stabilization. This final procedure avoids two extremes. It avoids extreme descriptivism, which accepts all usage without evaluation, and it avoids rigid prescriptivism, which ignores real speech and writing practices. The methodological position of the article may be defined as empirical normativity: the norm should be informed by documented usage, but documented usage should be interpreted through phonological, morphological, semantic, stylistic and pedagogical criteria. This approach is particularly suitable for Uzbek because its agglutinative structure creates high morphological productivity. Any lexical innovation must be examined not only as an isolated lexical unit but as a potential member of a derivational and inflectional network. A newly borrowed digital term may enter Uzbek in a nominative form, but its real integration begins when speakers attach Uzbek plural, possessive, case, derivational or verbalizing suffixes to it. Such processes cannot be captured adequately by simple word lists. They require annotated corpora, morphological parsers, lemmatization systems and genre-sensitive observation. The method also distinguishes between four levels of normativity. Graphic normativity concerns spelling and script. Lexical normativity concerns word choice, borrowing and terminology. Grammatical normativity concerns morphology, syntax and word formation. Pragmatic-register normativity concerns appropriateness in official, academic, educational, journalistic and informal digital contexts. These levels are interdependent but not identical. A spelling correction cannot solve terminological ambiguity, and a terminological recommendation cannot automatically determine stylistic appropriateness. The study also uses comparative interpretation, because Uzbek modernization is better understood when placed against general principles of language planning and corpus linguistics. Haugen's model of selection, codification, elaboration and acceptance remains relevant, but in the digital era each stage requires technological infrastructure: selection requires corpus evidence, codification requires digital dictionaries and style guides, elaboration requires terminological databases and academic writing resources, and acceptance requires educational integration and user-friendly tools [6; 15]. The article does not claim to provide a final regulatory dictionary; rather, it formulates a scientific model that can guide future empirical research, corpus design and educational codification. This limitation is methodologically important. Premature codification without corpus evidence can produce artificial rules that speakers do not follow, while pure observation without codification can leave schools, publishers, state institutions and technology developers without stable standards. The proposed method therefore treats Uzbek language modernization as a coordinated linguistic process requiring descriptive accuracy, technological infrastructure and pedagogical responsibility.

## RESULTS

The analysis produces a functional model of normative stabilization for modern Uzbek consisting of five interdependent dimensions: script stabilization, lexical-functional evaluation, morpho-corpus integration, digital-register differentiation and educational codification. The first dimension, script stabilization, shows that the coexistence of Latin and Cyrillic in Uzbek should not be treated as a simple binary opposition between “old” and “new” writing systems. It is a layered sociolinguistic reality in which institutional policy, generational literacy, archive accessibility, keyboard habits, publication history and digital searchability interact. A scientifically grounded approach should distinguish between three types of script variation: transitional variation, caused by the historical movement from Cyrillic to Latin; technological variation, caused by keyboard limitations, fonts, transliteration software and platform conventions; and identity-based variation, where users choose one script to signal education, nostalgia, modernity, officialness, informality or group belonging. This classification matters because not all variation requires the same response. Transitional variation requires educational support and accessible conversion tools. Technological variation requires improved input systems, spell-checkers and machine transliteration. Identity-based variation requires sociolinguistic understanding rather than immediate correction. The second dimension, lexical-functional evaluation, demonstrates that new digital borrowings must be assessed according to necessity, semantic precision, phonological adaptation, morphological compatibility and register. Some English-derived terms enter Uzbek because they name new realities for which no widely accepted Uzbek equivalent exists. Others enter because English has prestige in technology and business. Still others are short-lived slang items that will disappear without codification. Therefore, Uzbek linguistics should avoid two equally weak positions: rejecting all borrowings as linguistic contamination, or accepting all borrowings as inevitable modernization. A more precise model evaluates whether a borrowing fills a lexical gap, whether an Uzbek equivalent already exists, whether speakers can easily decline or derive forms from it, whether it is understandable to non-specialist audiences, and whether it belongs in official, academic, educational or informal discourse. The third dimension, morpho-corpus integration, confirms that Uzbek’s agglutinative structure requires more than surface-level lexical collection. A corpus of Uzbek must be able to handle suffix chains, morpho-phonetic alternations, derivational productivity, compound forms, auxiliary constructions and spelling variants. Without lemmatization and morphological annotation, corpus search results may seriously misrepresent frequency and meaning. A root may occur in dozens of inflected or derived forms, and a simple keyword search may fail to show its full distribution. Conversely, homographic forms may be counted together even when they represent different grammatical or semantic units. The development of Uzbek morphological analysis tools is therefore not a technical luxury but a central condition for modern linguistic research. Studies on Uzbek morphological analysis have emphasized the complexity of inflectional endings and the need to account for morpho-phonetic exceptions, data sparsity and word-level accuracy in computational modelling [12]. The fourth dimension, digital-register differentiation, shows that contemporary Uzbek exists in multiple communicative registers that require different normative expectations. Official documents, academic articles, school textbooks, literary prose, television speech, news portals, Telegram announcements, YouTube comments, advertising

slogans and private messaging do not follow the same norms, and they should not be evaluated by exactly the same criteria. A form that is acceptable in youth social-media slang may be inappropriate in a dissertation. A compressed transliterated form may be understandable in messaging but unacceptable in a school textbook. An English technical term may be necessary in a programming course but excessive in a general cultural essay. This does not mean that the literary norm becomes weak; it means that the literary norm must be supported by register awareness. The fifth dimension, educational codification, indicates that schools and universities remain decisive institutions in stabilizing Uzbek, but they can no longer rely only on memorized rules and isolated examples. Students must be taught how to understand language variation, how to distinguish literary form from informal digital usage, how to use dictionaries and corpora, how to evaluate borrowings, how to write consistently in Latin script and how to preserve stylistic adequacy in academic communication. The functional model therefore defines normative stabilization as a circular process. Real usage is collected through corpora. Usage is classified by linguistic analysis. Forms are evaluated by functional criteria. Appropriate norms are codified in dictionaries, textbooks and style guides. Codified norms are tested again in education, media and digital communication. This circular model is more effective than a linear command model, because it acknowledges that language users are not passive recipients of rules but active participants in the formation of contemporary Uzbek. A further result concerns the relation between national language and digital sovereignty. In the twenty-first century, language vitality depends not only on the number of speakers or the existence of official status, but also on the capacity of a language to function in search engines, e-government systems, educational software, mobile applications, automatic speech recognition, machine translation and electronic publishing. If Uzbek lacks reliable corpora, terminological databases and morphological tools, users will increasingly depend on foreign-language interfaces and conceptual systems. This dependence is not simply technological; it also affects knowledge production. The language that cannot efficiently express scientific, administrative and technological concepts in digital form risks being restricted to everyday and symbolic functions. For this reason, the study identifies corpus-based normalization as a strategic component of language development. The main result of the article is thus the claim that the future stability of Uzbek literary language depends not on suppressing variation but on scientifically managing it. Variation is not the enemy of normativity. Unanalysed variation is the problem. Once variation is collected, classified and functionally interpreted, it becomes the raw material for a stronger and more realistic literary norm.

## DISCUSSION

The proposed model has several theoretical and practical implications for Uzbek linguistics. The first implication is that script policy and linguistic normativity should be separated analytically even when they are connected socially. Orthographic reform may define the official graphic form of the language, but it does not automatically solve lexical, grammatical, stylistic or technological problems. A text can be written in the officially preferred script and still be stylistically weak, terminologically inconsistent or grammatically careless. Conversely, a Cyrillic Uzbek text may preserve high literary quality, rich vocabulary and correct grammar. Therefore, the real aim of modernization should not be a superficial replacement of one script

with another, but the formation of a stable, functional and teachable written standard. This requires reliable transliteration systems, digitized archives, parallel-script corpora, teacher training and consistent publishing norms. The second implication concerns lexical borrowing. Uzbek has historically developed through contact with Arabic, Persian, Russian and other languages, and the current wave of English borrowings should be understood within this broader contact-linguistic history. The difference is speed and medium. Digital platforms accelerate borrowing, shorten adaptation time and spread new forms before academic institutions can evaluate them. In this context, linguists must work faster but not more carelessly. A borrowing should be accepted into the literary language only when it demonstrates functional necessity, semantic stability and morphological adaptability. If an English-based technology term has no concise Uzbek equivalent and is already used in education, software documentation and professional communication, codification may be reasonable. However, if an English word merely duplicates a clear Uzbek word for prestige or fashion, its official use should be discouraged. This position is not linguistic purism; it is terminological hygiene. Languages do not become modern by importing every fashionable word. They become modern when they can name new realities through a balanced combination of internal resources, adapted borrowings and transparent terminology. The third implication concerns corpus development. A national corpus is not simply a large electronic archive. If it lacks metadata, genre classification, morphological annotation, spelling-variant mapping and search tools, it becomes a warehouse rather than a scientific instrument. For Uzbek, corpus design must include both Latin and Cyrillic materials, because excluding Cyrillic would cut researchers off from a major part of twentieth-century and contemporary usage, while excluding Latin would ignore current educational and official orientation. Parallel-script processing should therefore be a core principle of Uzbek corpus linguistics. The fourth implication concerns academic writing and higher education. Uzbek-language academic writing often faces a double pressure: it must remain faithful to national linguistic norms while also expressing international scientific concepts. This creates a terminological challenge. If every international concept is translated mechanically, the result may become artificial and obscure. If every concept is borrowed directly, the result may become dependent and stylistically fragmented. University departments of Uzbek language and literature should therefore become centres of terminological mediation. They should not only teach classical grammar and literature but also produce modern academic style guides, digital corpora for student writing, discipline-specific terminology lists and comparative analyses of Uzbek, Russian and English scientific discourse. The fifth implication concerns language technologies. Automatic translation, speech recognition, optical character recognition, spell-checking and educational applications will increasingly influence how users perceive correct Uzbek. If these tools are trained on inconsistent, unannotated or low-quality data, they may reproduce errors at scale. In the digital age, a bad algorithm can become a very energetic bad teacher; the joke is light, but the danger is real. This is why linguists must participate directly in language-technology development instead of leaving it entirely to programmers. Programmers can build systems, but linguists must define categories, annotation principles, norm-sensitive evaluation criteria and error typologies. The sixth implication concerns the philosophy of the literary norm. The literary norm should not be imagined as a museum object protected from

living speech. It is better understood as a disciplined cultural mechanism that selects, stabilizes and transmits the most functional forms of the language. Such a norm must be conservative enough to preserve continuity and flexible enough to serve new domains. In Uzbek, this balance is especially important because the language must simultaneously preserve its classical literary heritage, support national identity, function as a language of science and higher education, and operate effectively in digital technologies. The proposed model also shows that Uzbek linguistics needs interdisciplinary cooperation. Traditional grammarians, corpus linguists, lexicographers, sociolinguists, educators, software developers and policymakers should not work in isolation. A spelling reform without corpus data is blind. A corpus without normative interpretation is mute. A dictionary without digital integration is slow. Language technology without linguistic expertise is risky. The most defensible direction is therefore an integrated research infrastructure in which Uzbek is studied as a living literary, social and digital system. The discussion also reveals that the concept of “correctness” requires a more precise hierarchy. At the lowest level, correctness may mean mechanical spelling accuracy. At a higher level, it includes morphological and syntactic acceptability. At an even higher level, it means stylistic and register appropriateness. Finally, at the academic level, correctness includes conceptual precision, terminological coherence and intertextual consistency. Uzbek language education often gives significant attention to the first two levels but less systematic attention to register, academic phraseology and digital genre competence. This imbalance creates a paradox: students may know formal rules but still produce weak academic texts, because they lack models of scholarly argumentation, citation style, terminological discipline and genre-specific cohesion. Corpus-based academic writing resources could solve part of this problem by showing students authentic patterns of scientific Uzbek, including collocations, sentence frames, citation constructions and terminology networks. Even when an article is written in English, as in the present case, the same logic applies to Uzbek linguistics as a discipline: researchers need to explain Uzbek phenomena through internationally understandable concepts while preserving the specific structural and cultural reality of the language. Finally, the model has implications for OAK-oriented publication practice. Scientific articles in linguistics should move beyond general declarations about the importance of language and provide clearer methodological grounding, explicit research problems, defined analytical categories, and references to contemporary empirical resources. This does not mean that every article must include programming or statistical modelling. It means that claims about language change, lexical borrowing, orthographic variation or normative instability should be supported by observable material, systematic classification and theoretically defensible interpretation. The stronger the evidence base, the stronger the norm. Normativity built only on authority is fragile; normativity built on evidence is more durable.

## CONCLUSION

The article has examined the problem of normative stabilization in modern Uzbek through the interaction of script variation, digital lexical innovation and corpus-based linguistic description. The main conclusion is that Uzbek linguistics should move from episodic correction toward systematic evidence-based codification. Script coexistence, English-based

digital borrowings, social-media spelling habits and terminological instability should not be ignored, but they should also not be interpreted through panic or moral judgment. They are empirical signs of a language expanding into new communicative domains. The task of linguistics is to describe these signs accurately, classify them scientifically, evaluate them functionally and convert the results into educational and normative resources. The proposed functional model includes five dimensions: script stabilization, lexical-functional evaluation, morpho-corpus integration, digital-register differentiation and educational codification. Each dimension addresses a specific aspect of the contemporary linguistic situation, but their effectiveness depends on their integration. Script stabilization requires technological tools and educational consistency. Lexical modernization requires semantic and morphological evaluation. Corpus development requires annotation and metadata. Digital-register analysis requires sociolinguistic sensitivity. Educational codification requires practical materials that teach students not only rules but also linguistic judgment. The study also argues that Uzbek, as an agglutinative and historically multi-script language, needs corpus and language-technology tools designed according to its own structural features rather than copied mechanically from models developed for English or Russian. Future research should create balanced Uzbek corpora representing official, academic, educational, literary, journalistic and digital genres; develop reliable morphological annotation standards; compare Latin and Cyrillic usage across age groups and institutions; analyse the life cycle of digital borrowings; and design corpus-based academic writing resources for universities. The most important strategic conclusion is that the modernization of Uzbek should be neither passive borrowing nor rigid preservation. It should be a scientifically governed process in which the language remains historically continuous, socially inclusive, technologically functional and intellectually productive. Such an approach will strengthen the status of Uzbek not only as a national language but also as a language of science, education and digital civilization. For departments of Uzbek language and literature, this direction creates a concrete research agenda: the study of classical heritage must be connected with digital corpora, modern terminology, academic discourse and language technologies. Only then can Uzbek linguistics respond to contemporary challenges with both scholarly depth and practical effectiveness.

### REFERENCES

1. Abjalova, M., & colleagues. Educational Corpus of the Uzbek Language and Its Opportunities. Proceedings of the International Conference on Computer Science and Engineering, 2023.
2. Biber, D., Conrad, S., & Reppen, R. Corpus Linguistics: Investigating Language Structure and Use. Cambridge: Cambridge University Press, 1998.
3. Crystal, D. Internet Linguistics: A Student Guide. London: Routledge, 2011.
4. Fishman, J. A. Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages. Clevedon: Multilingual Matters, 1991.
5. Gries, S. T. Statistics for Linguistics with R: A Practical Introduction. Berlin: De Gruyter Mouton, 2021.
6. Haugen, E. Dialect, Language, Nation. *American Anthropologist*, 68(4), 1966, 922–935.

7. Herring, S. C. A Faceted Classification Scheme for Computer-Mediated Discourse. *Language@Internet*, 4, 2007.
8. McEnery, T., & Hardie, A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2012.
9. Mansurov, B., & Mansurov, A. *Uzbek Cyrillic-Latin-Cyrillic Machine Transliteration*. arXiv preprint arXiv:2101.05162, 2021.
10. Povey, A., & Povey, K. *FeruzSpeech: A 60 Hour Uzbek Read Speech Corpus with Punctuation, Casing, and Context*. Proceedings of the 7th International Conference on Natural Language and Speech Processing, 2024, 360–364.
11. Salaev, U., Kuriyozov, E., & Gómez-Rodríguez, C. *A Machine Transliteration Tool Between Uzbek Alphabets*. CEUR Workshop Proceedings, Vol. 3315, 2022.
12. Salaev, U. *UzMorphAnalyser: A Morphological Analysis Model for the Uzbek Language Using Inflectional Endings*. arXiv preprint arXiv:2405.14179, 2024.
13. Spolsky, B. *Language Policy*. Cambridge: Cambridge University Press, 2004.
14. Wardhaugh, R., & Fuller, J. M. *An Introduction to Sociolinguistics*. Oxford: Wiley-Blackwell, 2015.
15. Ўзбек тилининг изоҳли луғати. 5 жилдлик. Тошкент: Ўзбекистон миллий энциклопедияси, 2006–2008.
16. Ҳожиёв, А. Тилшунослик терминларининг изоҳли луғати. Тошкент: Ўзбекистон миллий энциклопедияси, 2002.
17. Қўнғуров, Р., Бегматов, Э., & Тожиёв, Ё. *Нутқ маданияти ва услубият асослари*. Тошкент: Ўқитувчи, 1992.
18. Jamolxonov, N. *Hozirgi o'zbek adabiy tili*. Toshkent: Talqin, 2005.