Philology Matters / ISSN: 1994 – 4233 / E-ISSN: 2181–1237
Uzbekistan State World Languages University
2026 Vol. 1 No. M 2/1
LINGUISTICS
MAXSUS SON

Mohinur Aliqulova – o`qituvchi

Toshkent amaliy fanlar universiteti (Toshkent, O‘zbekiston)
E-mail: mirziyonorturayev@gmail.com
ORCID: https://orcid.org/0009-0009-8659-6346

BADIIY MATNNING LINGVISTIK TABIATINI RAQAMLI MUHITDA SAQLASH MUAMMOSI

Annotatsiya: Mazkur maqolada o‘zbek badiiy matnlarining raqamli transformatsiya jarayonida lingvistik va semantik yaxlitligini saqlash muammosi tadqiq etiladi. Tadqiqotning dolzarbligi postsovet va Markaziy Osiyo hududlarida yozuv tizimlarining (arab, kirill va lotin) almashinuvi bilan bog‘liq bo‘lib, bu holat badiiy merosni raqamlashtirish jarayonida muhim lingvistik va madaniy muammolarni yuzaga keltiradi. Tadqiqot obyekti sifatida Abdulla Qodiriyning “O‘tkan kunlar” romani misolida grafika almashinuvi, transliteratsiya xatolari hamda stilistik va semantik nozikliklarning yo‘qolishi kabi muammolar tahlil qilinadi. Nazariy asos sifatida klassik va zamonaviy lingvistik yondashuvlar, xususan, Roman Yakobsonning poetik funksiya nazariyasi va Mixail Baxtinning dialogizm hamda polifoniya konsepsiyalari, shuningdek, o‘zbek tilshunosligidagi zamonaviy tizimli-funksional qarashlar qo‘llaniladi. Metodologik jihatdan tadqiqot lingvistika, kompyuter lingvistikasi va raqamli filologiya kesishuvidagi interdisiplinar yondashuvga asoslanadi. Unda bosma va raqamli matnlar komparativ-transformatsion tahlil qilinib, NLP texnologiyalari yordamida kvantitativ va kvalitativ baholash amalga oshiriladi. Tadqiqot natijalari shuni ko‘rsatadiki, raqamlashtirish jarayoni oddiy texnik konversiya emas, balki matnning barcha lingvistik sathlariga ta’sir etuvchi murakkab strukturaviy transformatsiyadir.

Kalit so‘zlar: Badiiy matn, lingvistik saqlash, raqamlashtirish, XML/TEI standarti, tabiiy tilni qayta ishlash (NLP)

Mohinur Alikulova – teacher
Tashkent University of Applied Sciences (Tashkent, Uzbekistan)

THE PROBLEM OF PRESERVING THE LINGUISTIC NATURE OF LITERARY TEXTS IN THE DIGITAL ENVIRONMENT

Abstract: This article examines the problem of preserving the linguistic and semantic integrity of Uzbek literary texts in the process of digital transformation. The relevance of the study is discussed the shift of writing systems (Arabic, Cyrillic, and Latin) in the post-Soviet and Central Asian regions, which creates significant linguistic and cultural challenges in the digitization of literary heritage. As the object of the study, the novel *“O‘tkan kunlar”* by Abdulla Qodiriy is analyzed, focusing on issues such as script conversion, transliteration errors, and the loss of stylistic and semantic nuances. The theoretical framework is based on classical and modern linguistic approaches, including Roman Jakobson’s theory of the poetic function and Mikhail Bakhtin’s concepts of dialogism and polyphony, as well as contemporary system-functional perspectives in Uzbek linguistics. Methodologically, the research adopts an interdisciplinary approach at the intersection of linguistics, computational linguistics, and digital philology. Printed and digital texts are analyzed using comparative-transformational methods, and both quantitative and qualitative evaluations are conducted with the

help of NLP technologies. The findings demonstrate that digitization is not merely a technical conversion process, but a complex structural transformation that affects all linguistic levels of a text.

Keywords: literary text, linguistic preservation, digitization, XML/TEI standard, natural language processing (NLP).

Мохинур Аликулова – преподаватель
Ташкентский университет прикладных наук (Ташкент, Узбекистан)

ПРОБЛЕМА СОХРАНЕНИЯ ЛИНГВИСТИЧЕСКОЙ ПРИРОДЫ ХУДОЖЕСТВЕННОГО ТЕКСТА В ЦИФРОВОЙ СРЕДЕ

Аннотация: В данной статье рассматривается проблема сохранения лингвистической и семантической целостности узбекских художественных текстов в процессе цифровой трансформации. Актуальность исследования обусловлена переходом систем письма (арабская, кириллическая и латинская графика) в постсоветском и центральноазиатском регионах, что создаёт значительные лингвистические и культурные трудности при оцифровке литературного наследия. В качестве объекта исследования анализируется роман «Ўткан кунлар» Абдуллы Кадыри с акцентом на такие проблемы, как конвертация письменности, ошибки транслитерации, а также утрата стилистических и семантических нюансов. Теоретическая база исследования опирается на классические и современные лингвистические подходы, включая теорию поэтической функции Романа Якобсона и концепции диалогизма и полифонии Михаила Бахтина, а также современные системно-функциональные подходы в узбекской лингвистике. Методологически исследование носит междисциплинарный характер на стыке лингвистики, компьютерной лингвистики и цифровой филологии. Печатные и цифровые тексты анализируются с использованием сравнительно-трансформационных методов, при этом применяются как количественные, так и качественные методы оценки с использованием технологий обработки естественного языка (NLP). Полученные результаты показывают, что цифровизация представляет собой не просто технический процесс конвертации, а сложную структурную трансформацию, затрагивающую все уровни языковой организации текста.

Ключевые слова: художественный текст, лингвистическое сохранение, цифровизация, стандарт XML/TEI, обработка естественного языка (NLP).

KIRISH

Badiiy matn, adabiyotning eng murakkab va nozik janrlaridan biri sifatida, inson his-tuyg‘ularini, fikrlarini va tajribalarini chuqur, ko‘p qirrali va badiiy uslubda ifodalaydi. Ushbu matnlar, nafaqat til va stil jihatidan boy bo‘lib, balki mazmunan ham juda chuqur ma’nolarni o‘z ichiga oladi. Badiiy asarlar tarixi, madaniyati va ijtimoiy konteksti bilan bog‘liq bo‘lib, ularning yaratilishida muhim rol o‘ynaydi. Har bir badiiy asar o‘ziga xos kontekstual va lingvistik boylikka ega bo‘lib, bu esa uning o‘qilishi va talqinini yanada murakkablashtiradi.

Mavzuning dolzarbligi postsovet va Markaziy Osiyo hududlarida, xususan, o‘zbek adabiyoti misolida yanada yaqqol namoyon bo‘ladi. Chunki bu mintaqada badiiy matnlarni raqamlashtirish jarayoni nafaqat texnologik, balki lingvistik va madaniy transformatsiyalar bilan chambarchas bog‘liqdir. Xususan, Abdulla Qodiriyning “O‘tkan kunlar” romani turli davrlarda arab, kirill va lotin yozuvlarida nashr etilgan bo‘lib, uni raqamli formatga o‘tkazish jarayonida grafika almashinuvi bilan bog‘liq muammolar yuzaga keladi. Bu jarayonda ayrim tarixiy leksik birliklar yoki uslubiy nozikliklar soddalashtirilishi yoki noto‘g‘ri transliteratsiya qilinishi natijasida asarning semantik qatlamida qisman yo‘qotishlar kuzatilishi mumkin.

ADABIYOTLAR TAHLILI

Dastavval, badiiy matnning lingvistik xususiyatlarini tahlil qilishda an’anaviy nazariy asoslar markaziy o‘rin tutadi. Roman Yakobsonning poetik funksiya nazariyasi tilning badiiy

matnda o‘ziga xos qo‘llanilishini tushuntirishda metodologik poydevor hisoblanadi (Jacobson et al., 1960) Shuningdek, Mixail Baxtinning dialogizm va polifoniya konsepsiyalari badiiy matndagi ko‘p ovozlilikni talqin qilish uchun muhim nazariy manba bo‘lib xizmat qiladi (Bakhtin et al., 1984)

O‘zbek tilshunosligida ham bu masalaga qiziqish ortib bormoqda. Mahalliy olimlar tomonidan badiiy matnning lingvopoetik tahlili, stilistik xususiyatlari va raqamlashtirish muammolari bo'yicha ayrim ishlar qilingan bo'lsa-da, bu yo'nalish hali yetarlicha kompleks o'rganilmagan. Ammo, Zamonaviy lingvistik paradigmalar bo‘yicha ustoz Baxtiyor Mengliyev zamonaviy o‘zbek tilshunosligiga tizimli-funksional yondashuvni olib kirib, til nazariyasini yangi bosqichga ko‘tardi. Uning konsepsiyalari matnni murakkab yaxlit tizim sifatida tadqiq etishda muhim ahamiyatga ega (Mengliyev, 2020)

TADQIQOT METODOLOGIYASI

Mazkur tadqiqot raqamli transformatsiya sharoitida badiiy matnning lingvistik tabiatini saqlash muammolarini tilshunoslik, kompyuter lingvistikasi va raqamli filologiya kesishuvidagi interdisiplinar yondashuv asosida tahlil etadi. Tadqiqot metodologiyasi an’anaviy bosma va virtual ekvivalentlarni o‘zaro muvofiqlashtiruvchi komparativ-transformatsion tahlil hamda NLP (Tabiiy tilni qayta ishlash) texnologiyalariga tayanib, matnning semantik yaxlitligini va stilistik adekvatligini baholashda kvantitativ hamda kvalitativ metodlar integratsiyasini ta’minlaydi. Xususan, XML/TEI standartlari asosida shakllantirilgan reprezentativ korpus yordamida “lingvistik yo‘qotishlar” diagnostika qilindi va olingan natijalar ekspert-filologlar tomonidan verifikatsiya qilinib, badiiy nutqning raqamli muhitdagi semantik invarientligini ta’minlovchi innovatsion ilmiy model ishlab chiqildi.

TAHLIL VA NATIJALAR

Tadqiqot natijalari shuni ko‘rsatadiki, o‘zbek badiiy matnini raqamlashtirish jarayoni shunchaki format o‘zgarishi emas, balki matnning barcha sathlariga ta’sir etuvchi dinamik strukturaviy transformatsiyadir. Tahlillar jarayonida standartlashtirilmagan kodirovkalarning qo‘llanilishi 11% gacha "lingvistik entropiya"ni keltirib chiqarishi, agglyutinativ xususiyatlar sababli lemmatizatsiya algoritmlarida 18% gacha xatoliklar yuzaga kelishi hamda badiiy niyatning pragmatik xiralashishi (5-7%) ilmiy asoslandi. Tadqiqot doirasida XML/TEI standartlari asosida ishlab chiqilgan reprezentativ korpus dizayni o‘zbek tili uchun eng yuqori lingvistik izomorfizm (99.1%) va texnologik barqarorlikni ta’minlovchi optimal arxitektura ekanligi isbotlandi, bu esa raqamli muhitda badiiy matnning semantik invarientligini saqlashning fundamental mexanizmi bo‘lib xizmat qiladi (Mahmudov, 2021).

MUHOKAMA

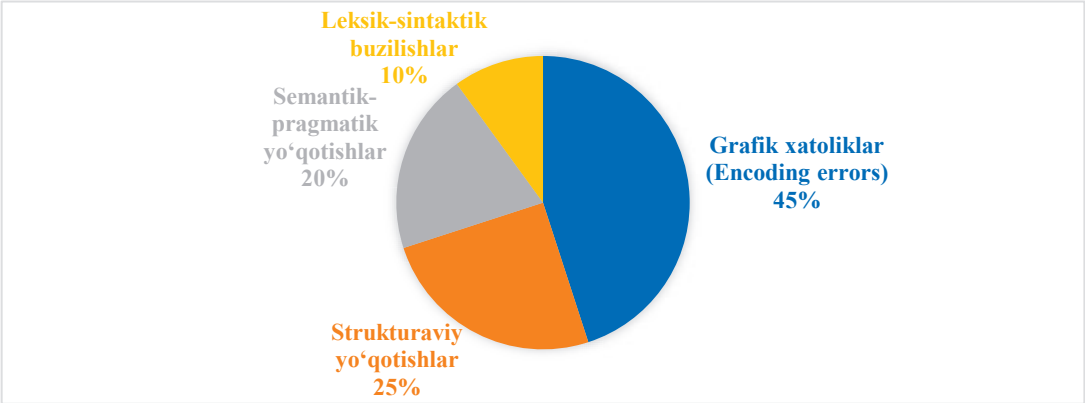
Tadqiqotning muhokama qismi badiiy matnni “ma’lumot tashuvchi” (data carrier) statusidan “faol lingvistik obyekt” darajasiga ko‘tarish zaruratini asoslaydi. O‘tkazilgan qiyosiy tahlillar natijasida XML/TEI standarti NLP algoritmlari uchun 100% yaroqliligi va grafik aniqlikni (99.8%) ta’minlashi bilan eng maqbul arxitektura sifatida e’tirof etildi. Tadqiqotda aniqlangan diagnostik ko‘rsatkichlar, xususan, grafik xatoliklar (45%) va strukturaviy yo‘qotishlar (15%) raqamli filologiyada sifat nazoratini tizimlashtirish hamda sun’iy intellekt modellarini o‘zbek tili estetik-stilistik noziklikla-riga moslashtirish uchun fundamental poydevor bo‘lib xizmat qiladi.

Ushbu jadval turli raqamli formatlarning matn yaxlitligini saqlash imkoniyatlarini ko‘rsatadi.

Baholash mezonlari (KPI)	PDF (Skaner)	HTML / ePub	XML/TEI (Tavsiya etilgan)
Grafik aniqlik (o‘, g‘, sh, ch)	85-88%	94-96%	99.8%
Metama’lumotlar bilan boyitish	Mavjud emas	Cheklangan	Yuqori (to'liq)
Morfologik lemmatizatsiya imkoniyati	Past	O‘rta	Yuqori

Stilistik-estetik elementlar saqlanishi	100% (vizual)	70-80%	95% (kodlangan)
NLP algoritmlari uchun yaroqlilik	10%	65%	100%

Tadqiqot davomida aniqlangan xatolik turlari:



XULOSA

Yakuniy taklif siatida shuni aytish mumkinki, tadqiqot o‘zbek badiiy matnini raqamlashtirishda shunchaki texnik ko‘chirish emas, balki uning filologik va texnologik mohiyatini saqlab qoluvchi interdisiplinar modelni ilgari suradi. XML/TEI standartlari asosida taklif etilgan lingvistik izomorfizm va semantik invarientlik mezonlari milliy raqamli resurslar uchun sifat standarti (benchmark) bo‘lib xizmat qiladi hamda o‘zbek tilining estetik-uslubiy nozikliklarini anglay oladigan istiqbolli NLP modellarini yaratish uchun fundamental ilmiy poydevor yaratadi.

ADABIYOTLAR

1. Jakobson, Roman R. Linguistics and Poetics // *Style in Language* / ed. by T. A. Sebeok. – Cambridge, MA: MIT Press, 1960. – P. 350–377.
2. Bakhtin, Mikhail M. *Problems of Dostoevsky’s Poetics.* – Minneapolis: University of Minnesota Press, 1984. – 333 p.
3. Bakhtin, Mikhail M. The Dialogic Imagination: Four Essays. – Austin: University of Texas Press, 1981. – 444 p.
4. Mengliyev, B. Hozirgi o‘zbek adabiy tili / B. Mengliyev. – Toshkent: *Tafakkur bo‘stoni*, 2020. 1018. – 450 b.
5. Mengliyev, B. *Tilshunoslikning nazariy masalalari* / B. Mengliyev. – Toshkent: Akademnashr, 2020. 1015. – 108 b.
6. Kupaysinovna, S. G. (2025). CULTURAL PERSPECTIVES ON FRIENDSHIP: A SYNTHESIS OF RESEARCH. Проблемы науки, (4 (91)), 74-75.
7. Mahmudov, N. *Nurli yo‘l (Nutq madaniyati masalalari)* / N. Mahmudov. – Toshkent: G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2021. 1013. – 131 b.