



UDK: 811.512.133'35

Sevinch KENJAYEVA,
Magistrant, O'zbekiston Milliy universiteti, Toshkent, O'zbekiston
E-mail: kenjayevas@nuu.uz, ORCID: 0009-0007-0563-1302

PhD O.Yulbarsov taqrizi asosida

**LINGUISTIC ANALYSIS OF SPELLING ERRORS IN THE UZBEK LANGUAGE (PHONETIC,
MORPHOLOGICAL, KEYBOARD ERRORS, ETC.)**

Annotation

The article presents a linguistic analysis of spelling errors in Uzbek by their sources: phonetic, morphological, graphic (keyboard), orthographic, lexical-semantic, and Latin–Cyrillic conversion errors. For each group, typical examples, causes, and the requirements they impose on automatic correction algorithms are identified. The results provide a theoretical basis for developing automatic spelling correction systems for Uzbek.

Key words: Uzbek language, spelling errors, linguistic analysis, phonetic error, morphological error, keyboard error, transliteration, automatic correction, natural language processing.

**ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ ОРФОГРАФИЧЕСКИХ ОШИБОК В УЗБЕКСКОМ ЯЗЫКЕ
(ФОНЕТИЧЕСКИЕ, МОРФОЛОГИЧЕСКИЕ, КЛАВИАТУРНЫЕ ОШИБКИ И ДР.)**

Аннотация

В статье представлен лингвистический анализ орфографических ошибок в узбекском языке по источникам их возникновения: фонетические, морфологические, графические (клавиатурные), орфографические, лексико-семантические ошибки и ошибки латинско-кириллической конвертации. Для каждой группы определяются характерные примеры, причины появления и требования к алгоритмам автоматической коррекции. Результаты служат теоретической основой для разработки систем автоматической коррекции орфографии узбекского языка.

Ключевые слова: Узбекский язык, орфографические ошибки, лингвистический анализ, фонетическая ошибка, морфологическая ошибка, клавиатурная ошибка, транслитерация, автоматическая коррекция, обработка естественного языка.

**O'ZBEK TILIDAGI IMLO XATOLARINING LINGVISTIK TAHLILI (FONETIK, MORFOLOGIK, KLAVIATURA
XATOLARI VA BOSHQALAR)**

Annotatsiya

Maqolada o'zbek tilidagi imlo xatolari kelib chiqish manbalariga ko'ra lingvistik tahlil qilinadi: fonetik, morfologik, grafik (klaviatura), orfografik, leksik-semantik hamda lotin-kirill konvertatsiyasi bilan bog'liq xatolar ko'rib chiqiladi. Har bir guruh uchun xarakterli misollar, kelib chiqish sabablari va avtomatik tuzatish algoritmlariga qo'yiladigan talablar aniqlanadi. Natijalar imlo xatolarini avtomatik tuzatish tizimlarini ishlab chiqish uchun nazariy asos bo'lib xizmat qiladi.

Kalit so'zlar: O'zbek tili, imlo xatolari, lingvistik tahlil, fonetik xato, morfologik xato, klaviatura xatosi, transliteratsiya, avtomatik tuzatish, tabiiy tilni qayta ishlash.

Kirish. Hozirgi raqamli davrda yozma matnlar savodxonligi nafaqat ta'lim, balki axborotni qayta ishlash texnologiyalari uchun ham muhim [1]. PF-6084-son Farmon va o'zbek tilini rivojlantirish konsepsiyasida davlat tilini axborot texnologiyalari bilan integratsiya qilish, elektron lug'at va dasturiy mahsulotlar yaratish vazifalari belgilangan [17]. Bu imlo xatolarini avtomatik aniqlash va tuzatish texnologiyalarini dolzarb masalaga aylantiradi. Bunday tizimlarni yaratishda muhim bosqich — xatolarning kelib chiqish sabablarini lingvistik o'rganishdir, chunki har bir xato turi o'ziga xos yondashuvni talab qiladi: klaviatura xatosiga Levenshtein masofasi, fonetik xatoga esa fonetik kodlash (Soundex, Metaphone) zarur. O'zbek tilining agglutinativ tabiati, ikki yozuvning parallel qo'llanilishi va shevalar ta'siri imlo xatolarining o'ziga xos manzarasini hosil qiladi. Tadqiqotning ilmiy yangiligi — imlo xatolarini an'anaviy tilshunoslik va kompyuter lingvistikasi nuqtai nazaridan birgalikda tahlil qilishdir [8]. Maqsad — xatolarni kelib chiqish manbasi va lingvistik tabiatiga ko'ra tahlil qilib, har bir guruhning avtomatik tuzatish algoritmlariga qo'yiladigan talablarni aniqlashdir.

Mavzuga oid adabiyotlar tahlili. Imlo xatosi — yozma matndagi so'zning amaldagi orfografik me'yorlardan og'ishi. Tor ma'noda u orfografik qoidalardan chetga chiqishni, keng ma'noda esa tinish belgilari, qo'shib yoki ajratib yozish xatolari va so'z tartibidagi buzilishlarni ham qamrab oladi. Kompyuter lingvistikasida asosiy mezon — so'zning lug'atda (vocabulary) mavjudligi. Shunga ko'ra xatolar lug'atda bo'lmagan (non-word) va lug'atda mavjud, ammo kontekstga mos kelmaydigan (real-word) xatolarga bo'linadi [9]. Birinchisini aniqlash oson, ikkinchisi kontekstual tahlilni talab qiladi: "o'qibdi" o'mniga "o'qidi" yozilsa, har ikki shakl ham lug'atda bo'lgani uchun oddiy tekshiruvchi xatoni topa olmaydi. Shuning uchun zamonaviy tizimlar statistik til modellari va neyron tarmoqlardan foydalanadi. Tadqiqotlar shuni ko'rsatadiki, xatolarning 80 foizdan ko'prog'i so'zdan atigi bitta belgi bilan farq qiladi: belgi tushib qolgan, qo'shilgan, almashtirilgan yoki ikki qo'shni belgi o'rin almashgan bo'ladi. Bu F. Damerau (1964) taklif etgan to'rt asosiy operatsiya nazariyasiga asos bo'lgan [7]. So'nggi yillarda o'zbek tili uchun tabiiy tilni qayta ishlashda salmoqli natijalar qo'lga kiritildi: morfologik tahlil veb-tizimi [13],

MorphUz analizatori [5], shevadan adabiy me'yorga o'tkazish usullari [12] va kam resursli (low-resource) yondashuvlar [14]. Leksik-semantik tahlil uchun UzWordNet [6], SimRelUz [19], kontekstual tahlil uchun UzBERT [11] muhim resurslardir; mazkur tadqiqot an'anaviy tilshunoslik mezonlarini kompyuter lingvistikasi yondashuvi bilan birlashtiradi.

Tadqiqot metodologiyasi. Tadqiqot tavsifiy-tahliliy metodga asoslanadi. Manba sifatida amaldagi imlo lug'ati [16] va elektron korpus [20] ma'lumotlari tahlil qilinib, xatolarni guruhlashga empirik asos bo'ldi. Tahlilda ikki mezon birlashtirildi: xatoni keltirib chiqargan til sathi (fonetik, morfologik, leksik) va paydo bo'lish manbasi (talaffuz, bilim yetishmasligi, klaviatura omili, yozuvlararo konvertatsiya). Har bir guruh uchun misollar, kelib chiqish sabablari va algoritmik talablar aniqlanib, natijalar jadvalda umumlashtirildi.

Tahlil va natijalar. Imlo xatolari quyidagi asosiy guruhlarga ajratildi.

1. Fonetik xatolar

Fonetik xatolar so'z talaffuzga qarab yozilishidan yuzaga keladi va o'zbek tilida bir necha sabab bilan bog'liq.

Birinchi sabab — shevaga xos talaffuzning yozma matnga ko'chishi. Adabiy til qarluq lahjasiga tayansa-da, ayrim shevalarda “keladi” so'zi “kelodi” yoki “kelad” tarzida yoziladi; bunday shakllarni adabiy me'yorga avtomatik o'tkazish alohida tadqiq etilgan [12].

Ikkinchi sabab — yaqin yangraydigan tovushlarning aralashtirilishi. O'zbek tilida q va k, g' va g, h va x tovushlari ko'pincha aralashtiriladi: “ovqat” o'rniga “ovkat”, “qog'oz” o'rniga “kog'oz”.

Uchinchi sabab — tutuq belgisining tushib qolishi yoki noto'g'ri qo'yilishi: “ma'lumot” o'rniga “malumot”, “she'r” o'rniga “sher” (mutlaqo boshqa ma'noli so'z!). Bu hodisa real-word toifasidagi xatolarni keltirib chiqaradi. Bunday xatolarni Levenshtein masofasi bilan har doim tuzatib bo'lmaydi; ularni ajratish uchun maxsus fonetik kodlash, jumladan o'zbekchaga moslashtirilgan Soundex algoritmi zarur.

2. Morfologik xatolar

Morfologik xatolar affikslarning noto'g'ri qo'llanilishi yoki yozilishida namoyon bo'ladi. O'zbek tili agglutinativ bo'lib, bir o'zakka ketma-ket bir nechta qo'shimcha qo'shiladi; bu xatolarning xilma-xilligini izohlaydi.

Variantli qo'shimchalarning noto'g'ri tanlanishi. Jo'nalish kelishigining -ga / -ka / -qa variantlari o'zakning oxirgi tovushiga bog'liq: “tog'” + “-ga” = “toqqa”, “tilak” + “-ga” = “tilakka”. Foydalanuvchilar bu qoidani buzib, “oyoqqa” o'rniga “oyoqqa” yozadilar.

Fe'l zamonlari va shaxs-son qo'shimchalaridagi xatolar. O'tgan zamon fe'lining -gan/-kan/-qan variantlari va hozirgi zamon shakllari aralashtiriladi: “borgan” o'rniga “borkan”, “yozyapti” o'rniga “yozyabdi”.

Bo'lishsizlik shaklining buzilishi. Bo'lishsizlik qo'shimchasi -ma fe'l o'zagidan keyin keladi. Foydalanuvchilar ba'zan uni tushirib qoldiradi: “bormadim” o'rniga “bordim” yozilsa, ma'no teskarisiga aylanadi — real-word xatoga yorqin misol. Bularni tuzatish uchun morfologik analizator kerak: o'zbek tili uchun MorphUz [5] va morfologik veb-tizim [13] yaratilgan. Biroq ular to'g'ri yozilgan so'zlarga mo'ljallangani uchun imlo tuzatishda “toqatli” (fault-tolerant) rejim zarur.

3. Grafik (klaviatura) xatolari

Grafik xatolar klaviaturada yozish texnikasi bilan bog'liq bo'lib, ko'pincha barcha xatolarning yarmidan ortig'ini tashkil etadi; o'zbek lotin alifbosi uchun klaviatura va Unicode masalalari alohida tadqiq etilgan [4].

Belgi tushib qolishi (deletion). Eng keng tarqalgan klaviatura xatosi: “maktab” o'rniga “matab”; bu xatolar Levenshtein masofasi [10] asosida oson tuzatiladi.

Qo'shimcha belgi kiritilishi (insertion). Qo'shni tugmani yoki bir tugmani ikki marta bosish natijasi: “kitob” o'rniga “kittob”.

Belgilarning almashinishi (substitution). Kerakli harf o'rniga uning klaviaturadagi qo'shnisi bosiladi; bu xatolarni tuzatishda klaviatura tartibini hisobga olgan og'irlikli (weighted) Levenshtein masofasi yaxshi natija beradi.

Belgilar o'rnining almashinishi (transposition). Ikki qo'shni belgining o'rin almashishi: “maktab” o'rniga “matkab”; Damerau-Levenshtein algoritmi aynan shu turdagi xatolarni hisobga oladi.

Tilning noto'g'ri qatlamida yozish (layout switching). Foydalanuvchi tilni almashirishni unutib, ingliz qatlamida o'zbekcha so'z yozadi: “salom” o'rniga “cfkjv”; bunda klaviatura qatlamlari o'rtasidagi belgi xaritasi ishlatiladi. Klaviatura xatolari lug'atda bo'lmagani uchun oson aniqlanadi, biroq to'g'ri variantni topishda kontekst va til modeli muhim; P. Norvigning ehtimolikka asoslangan yondashuvi amaliy yechim beradi [15].

4. Orfografik (qoidaviy) xatolar

Orfografik xatolar imlo qoidalariga oid bilim yetishmasligidan, ko'pincha istisnolardan xabarsizlikdan sodir bo'ladi; bu qoidalar 1995-yilgi O'zbek tilining asosiy imlo qoidalari bilan belgilangan [18].

Qo'shma so'zlarning yozilishi. Qo'shma so'zlar qo'shilib, ajratib yoki chiziqcha bilan yoziladi; istisnolar ko'pligi atoga sabab bo'ladi: “har kim” o'rniga “harkim”, “ko'p qavatli” o'rniga “ko'pqavatli”.

Bosh harf bilan yozish qoidalari. Atoqli otlar bosh harf bilan yoziladi, ammo amalda “toshkent”, “o'zbekiston respublikasi” kabi shakllar uchraydi; ularni aniqlash uchun atoqli ot ro'yxatlari (named entity lists) kerak.

Olingan so'zlar imlosi. O'zlashgan atamalar imlosida ko'p xato qilinadi: “kompyuter” o'rniga “komputer”, “professor” o'rniga “profesor”; me'yoriy manba — amaldagi imlo lug'ati [16]. Bunday xatolarni tuzatishda qoidaviy (rule-based) tizimlar yoki korpusda o'qitilgan til modellari samarali natija beradi.

5. Lotin va kirill yozuvlari aralashmasidan kelib chiqadigan xatolar

O'zbek tili ikki yozuv tizimi parallel qo'llaniladigan kam sonli tillardan biri; 1993-yildan boshlangan lotin yozuviga o'tish to'liq yakunlanmagani o'ziga xos xatolarni keltirib chiqaradi.

Yozuvlarning aralash ishlatilishi. Bir matnda lotin va kirill harflarining, ayniqsa vizual o'xshash harflarning (lotin a va kirill a, lotin o va kirill o) aralash ishlatilishi muammo tug'diradi: ko'z bilan to'g'ri ko'ringan so'z texnik jihatdan xatoli bo'lib, lug'atda topilmaydi.

Sh, ch, ng kabi harf birikmalarining noto'g'ri yozilishi. Lotin yozuvida bir tovush ikki harf bilan ifodalanadi; foydalanuvchilar ba'zan ularni xato yozadi: “choyxona” o'rniga “coyxona”. Bu xatolarni tuzatish uchun Unicode normallashtirish (NFKC, NFKD), yozuvlararo konvertatsiya xaritalari va kontekstga sezgir qoidalar tizimi talab etiladi.

6. Leksik-semantik xatolar

Leksik-semantik xatolar eng murakkablari: so'zning o'zi to'g'ri bo'lsa-da, kontekstda noto'g'ri ishlatilgan (real-word). Ularni aniqlash chuqur kontekst tahlilini talab qiladi; transformer asosidagi UzBERT modeli bunda yangi imkoniyatlar ochdi [11].

Paronimlarni aralashtirish. Paronimlar — talaffuzi yaqin, ma'nosi har xil so'zlar; foydalanuvchilar ba'zan kerakli so'z o'rniga paronimini ishlatadi.

Omonimlarni aralashtirish. Omonimlar — yozilishi va talaffuzi bir xil, ammo ma'nosi turlicha so'zlar (masalan, “yuz” — chehra yoki yuz soni; “ot” — hayvon yoki ism). Imlo

jihattan to'g'ri bo'lsa-da, tuzatish tizimi ko'zda tutilgan ma'noni faqat kontekst orqali aniqlaydi.

Sinonimlar orasidagi uslubiy farqlar. “Yuz”, “chehra”, “bashara” so'zlari bir narsani anglatganda, qo'llanilish doirasi farq qiladi. Bunday nozik farqlarni aniqlash uchun semantik o'xshashlikni baholovchi SimRelUz to'plami [19] va UzWordNet bazasi [6] zarur.

1-jadval. Xato turlari va tavsiya etilgan algoritmik yondashuvlar

Xato turi	Asosiy aniqlash mexanizmi	Tavsiya etilgan algoritmlar
Fonetik xatolar	So'zning fonetik kodi bilan solishtirish	Soundex, Metaphone (o'zbekchaga moslashtirilgan), fonetik qoidalar bazasi
Morfologik xatolar	So'zni morfemalarga ajratib tekshirish	Morfologik analizator, chekli avtomatlar (FST), qoidaviy yondashuv
Klaviatura xatolari	Lug'at bilan solishtirib eng yaqin variantni topish	Damerau-Levenshtein, klaviatura tartibiga asoslangan og'irlikli masofa, BK-tree
Orfografik xatolar	Imlo qoidalari bo'yicha tekshirish va istisnolar lug'ati	Qoidalarga asoslangan tizim, bigram va trigram modellari
Transliteratsiya xatolari	Unicode normalashtirish va yozuvlararo xarita	NFKC/NFKD, lotin-kirill konvertor, gomogliif aniqlash
Real-word xatolar	Kontekstual tahlil va til modeli	N-gram til modellari, BERT, neyron til modellari, semantik vektorlar

Demak, samarali tizim bir emas, bir nechta algoritmnining birgalikda ishlashiga — “hibrid” yondashuvga asoslanishi kerak. Kelgusi tadqiqotda uch bosqichli tizim — xatolarni aniqlash, variant shakllantirish va kontekstga ko'ra eng to'g'ri nomzodni tanlash — ishlab chiqiladi.

Xulosa va takliflar. Maqolada imlo xatolari kelib chiqish manbalariga ko'ra tahlil qilindi; asosiy xulosalar quyidagicha.

Birinchiidan, imlo xatolari yagona kategoriya emas, balki tilning turli sathlariga (fonetik, morfologik, leksik) tegishli murakkab hodisalar bo'lib, ularni yagona algoritm bilan tuzatib bo'lmaydi.

7. Tahlil natijalarining algoritmik amaliyotga tatbiqi
Tahlil natijalari imlo tuzatuvchi tizim arxitekturasini loyihalashda metodologik baza yaratadi [3]. Har bir guruh uchun tavsiya etilgan algoritmik yondashuvlar 1-jadvalda umumlashtirilgan.

Ikkinchiidan, agglyutinativ tabiat, ikki yozuvning parallel qo'llanilishi va shevalar ta'siri boshqa tillar uchun yaratilgan tizimlarni to'g'ridan-to'g'ri ko'chirishga imkon bermaydi.

Uchinchiidan, har bir xato turi maqbul algoritmi talab qiladi: klaviatura xatolariga masofaviy algoritmlar, fonetik xatolarga fonetik kodlash, morfologik xatolarga analizator, real-word xatolarga kontekstual til modellari.

To'rtinchiidan, sifatli tizim uchun keng qamrovli lug'at, xatolar korpusi va morfologik analizator zarur [2]; mazkur tahlil keyingi tadqiqotlarga nazariy poydevor bo'ladi.

ADABIYOTLAR

1. Abduraxmonova N. Kompyuter lingvistikasi (darslik). — Toshkent: Nodirabegim, 2021. — 398 b.
2. Abduraxmonova N. Mashina tarjimasining lingvistik ta'minoti (monografiya). — Toshkent, 2018.
3. Abduraxmonova N. O'zbek tili elektron korpusining kompyuter modellari (monografiya). — Toshkent, 2021. — 202 b.
4. Abduraxmonova N., Isroilov J., Matlatipov S. O'zbek tili lotin alifbosi uchun klaviatura yoki “unicode” masalasi // Til va adabiyot ta'limi. — 2019. — № 6.
5. Abdurakhmonova N., Ismailov A., Sayfullayeva R. MorphUz: Morphological analyzer for the Uzbek language // 2022 7th International Conference on Computer Science and Engineering (UBMK). — IEEE, 2022. — P. 61–66.
6. Agostini A., Usmanov T., Khamdamov U., Abdurakhmonova N., Mamasaidov M. UzWordNet: A lexical-semantic database for the Uzbek language // Proceedings of the 11th Global WordNet Conference (GWC). — 2021. — P. 8–19.
7. Damerau F.J. A technique for computer detection and correction of spelling errors // Communications of the ACM. — 1964. — Vol. 7, No. 3. — P. 171–176.
8. Jurafsky D., Martin J.H. Speech and Language Processing. 3rd ed. — Stanford, 2024. — 600 p.
9. Kukich K. Techniques for automatically correcting words in text // ACM Computing Surveys. — 1992. — Vol. 24, No. 4. — P. 377–439.
10. Levenshtein V.I. Binary codes capable of correcting deletions, insertions, and reversals // Soviet Physics Doklady. — 1966. — Vol. 10. — P. 707–710.
11. Mansurov B., Mansurov A. UzBERT: pretraining a BERT model for Uzbek // arXiv preprint arXiv:2108.09814. — 2021.
12. Mengliev D., Abdurakhmonova N., Hayitbayeva D., Barakhnin V.B. Automating the transition from dialectal to literary forms in Uzbek language texts: an algorithmic perspective // 2023 IEEE XVI International Scientific and Technical Conference (APEIE). — 2023. — P. 1440–1443.
13. Mengliev D., Barakhnin V., Abdurakhmonova N. Development of intellectual web system for morph analyzing of Uzbek words // Applied Sciences. — 2021. — Vol. 11, No. 19. — Art. 9117.
14. Mengliev D., Abdurakhmonova N., Barakhnin V., Ibragimov B. va b. Integrating morphological stemming and syntactic parsing for low-resource Uzbek texts // AIP Conference Proceedings. — 2025. — Vol. 3377, No. 1. — Art. 040003.
15. Norvig P. How to write a spelling corrector [Elektron resurs]. — URL: <https://norvig.com/spell-correct.html> (murojaat sanasi: 01.04.2026).
16. O'zbek tilining imlo lug'ati / Tuzuvchilar: A.Madvaliyev, N.Mahkamov, E.Begmatov. — Toshkent: Akademnashr, 2013. — 480 b.
17. O'zbekiston Respublikasi Prezidentining “Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida” 2020-yil 20-oktabrdagi PF-6084-son Farmoni.
18. O'zbekiston Respublikasi Vazirlar Mahkamasining “O'zbek tilining asosiy imlo qoidalari tasdiqlash haqida” 1995-yil 24-avgustdagi 339-sonli qarori.
19. Salaev U., Kuriyozov E., Matlatipov S., Gomez-Rodriguez C. SimRelUz: similarity and relatedness scores as a semantic evaluation dataset for the Uzbek language. — 2022.
20. O'zbek tili elektron korpusi [Elektron resurs]. — URL: <https://uzbekcorpus.uz> (murojaat sanasi: 01.04.2026).