

## O‘ZBEK TILI MATNLARIDA IMLO XATOLARINI ANIQLASH VA TUZATISHNING ZAMONAVIY USULLARI

**Elov Botir Boltayevich,**  
Texnika fanlari falsafa doktori, dotsent  
[elov@navoiy-uni.uz](mailto:elov@navoiy-uni.uz)  
ToshDO‘TAU

**Sobirova Nazira G‘anijon qizi,**  
Tayanch doktorant (PhD)  
[sobirovanazira134@gmail.com](mailto:sobirovanazira134@gmail.com)  
ToshDO‘TAU

**Annotatsiya.** Ushbu maqolada imloviy xatolar bir necha guruhlariga ajratilib, ularning yuzaga kelish sabablari izohlanadi. Shuningdek, xatolarni aniqlashda qo‘llaniladigan turli yondashuvlar – oddiy lug‘aviy tekshiruvdan tortib, tahrir masofasiga asoslangan algoritmlar, statistik modellar va kontekstni hisobga oluvchi zamonaviy sun‘iy intellekt yechimlari qiyosiy jihatdan ko‘rib chiqiladi. Tahlillar shuni ko‘rsatadiki, alohida olingan metodlar barcha turdagi xatolarni to‘liq qamrab ololmaydi. Shu sababli, bir nechta usullarni o‘zida birlashtirgan integratsiyalashgan yondashuv amaliy jihatdan samaraliroq natija beradi. Tadqiqot natijalari o‘zbek tilida ishlovchi imlo tekshiruv tizimlarini ishlab chiqish, matnlarni avtomatik qayta ishlash jarayonlarini optimallashtirish hamda sun‘iy intellekt asosidagi lingvistik texnologiyalarni rivojlantirish uchun nazariy va amaliy asos bo‘lib xizmat qiladi.

**Kalit so‘zlar:** *imlo xatolari, avtomatik aniqlash, tabiiy tilni qayta ishlash (NLP), agglyutinativ til, Levenshteyn masofasi, N-gram modeli, neyron tarmoqlar, gibrid model, matnni avtomatik tahrirlash.*

**Abstract.** In this study, spelling errors are classified into several categories, and their underlying causes are examined. Furthermore, different approaches to error detection are comparatively analyzed, ranging from basic dictionary-based methods to edit distance algorithms, statistical models, and modern artificial intelligence techniques that take contextual information into account. The analysis demonstrates

that individual methods are insufficient to fully address all types of errors. Therefore, an integrated approach that combines multiple techniques provides more effective results in practice. The findings of this research offer both theoretical and practical contributions to the development of spelling correction systems for the Uzbek language, optimization of automatic text processing, and advancement of AI-based linguistic technologies.

**Keywords:** *spelling errors, automatic detection, natural language processing (NLP), agglutinative language, Levenshtein distance, N-gram model, neural networks, hybrid model, automatic text correction.*

**Kirish.** Avtomatik imlo tekshiruv masalasi kompyuter lingvistikasining asosiy yo'nalishlaridan biri sifatida shakllangan bo'lib, dastlabki tadqiqotlar oddiy lug'aviy solishtirishga asoslangan edi. Biroq real til amaliyoti shuni ko'rsatdiki, barcha xatolarni faqat lug'at yordamida aniqlash mumkin emas. Ayniqsa, semantik jihatdan noto'g'ri, lekin grammatik jihatdan to'g'ri yozilgan birliklarni aniqlash yanada murakkab vazifa hisoblanadi.

O'zbek tilining agglyutinativ tuzilishi mazkur muammoni yanada murakkablashtiradi. So'zlarga ketma-ket qo'shimchalar qo'shilishi natijasida juda ko'p shakllar hosil bo'ladi va bu esa so'zlarning barcha variantlarini oldindan lug'atga kiritishni amalda imkonsiz darajaga olib keladi. Natijada an'anaviy yondashuvlar o'z samaradorligini yo'qotadi va yangi, moslashuvchan usullarni ishlab chiqish zarurati yuzaga keladi.

Hozirgi kunda imlo xatolarini aniqlashda bir nechta yondashuvlar parallel ravishda qo'llanilmoqda. Ular orasida tahrir masofasiga asoslangan algoritmlar, ehtimollik modellariga tayanuvchi statistik usullar hamda kontekstni chuqur tahlil qiluvchi neyron tarmoqlar alohida ahamiyatga ega. Ushbu metodlar turli darajada samaradorlik ko'rsatadi, biroq ularning har biri ma'lum cheklolarga ega bo'lib, universal yechim sifatida qaralishi qiyin.



Shu sababli zamonaviy ilmiy izlanishlarda turli yondashuvlarni integratsiyalash orqali yanada barqaror va aniq ishlaydigan tizimlar yaratish ustuvor yo'nalishlardan biriga aylangan. Mazkur maqolada imloviy xatolarni aniqlash usullari tizimli ravishda tahlil qilinadi hamda o'zbek tili sharoitida samarali ishlashi mumkin bo'lgan kompleks yondashuv asoslab beriladi.

Imlo xatolarini tuzatish – bu tabiiy tilda yaratilgan hujjatdagi noto'g'ri yozilgan birlikni uning eng ehtimoliy to'g'ri shakli bilan almashtirish jarayonidir [2:377–380]. Ushbu muammo tabiiy tilni qayta ishlash (NLP) sohasida uzoq yillardan buyon faol o'rganilib kelinayotgan muhim tadqiqot yo'nalishlaridan biri hisoblanadi [3:1–5]. NLPning asosiy maqsadi inson tilini avtomatik tarzda tahlil qila oladigan, mazmunini tushunadigan va yangi matnlar hosil qilaoladigan dasturiy tizimlarni ishlab chiqishdan iborat [4:15–18]. Ushbu soha doirasida matnni avtomatik qisqartirish (text summarization), savol-javob tizimlari (question answering), mashina tarjimasi (machine translation), sintaktik tahlil (parsing), axborotni qidirish (information retrieval) hamda optik belgilarni tanish (optical character recognition) kabi ko'plab amaliy yo'nalishlar shakllangan [1:7–12; 4:20–25]. Imlo tekshiruv tizimlarining asosiy vazifasi yozma matnda uchraydigan xatolarni aniqlashdan iborat. Xatolarni aniqlashning eng keng qo'llaniladigan yondashuvlari qatoriga lug'atga asoslangan tekshiruv (dictionary look up) hamda statistik modellarga tayanuvchi N-gram tahlili kiradi [2:377–380; 3:190–195]. Xatoni aniqlash jarayoni, odatda, kiritilgan birlikning lug'atda mavjud yoki mavjud emasligini tekshirishga asoslanadi. Ushbu usul ayniqsa non-word turidagi xatolarni aniqlashda samarali hisoblanadi [2:381–385].

Zamonaviy imlo tekshiruv tizimlarida lug'at va ehtimollik modellari o'zaro integratsiyalashgan holda qo'llaniladi. Agar matnda noto'g'ri yozilgan so'z aniqlansa, tuzatish algoritmlari ushbu birlik uchun eng ehtimoliy variantlarni generatsiya qiladi va foydalanuvchiga tavsiya etadi [4:1–3]. Bunda turli

yondashuvlar – qoidaviy usullar (rule-based techniques), tahrirlash masofasi (edit distance), N-gram modellari hamda chuqur o'rganish (deep learning) asosidagi metodlar keng qo'llaniladi [1:7–12; 3:210–215].

Shu nuqtayi nazardan, avtomatik so'z tuzatish muammosi odatda uchta asosiy bosqich doirasida ko'rib chiqiladi:

1. *lug'atda mavjud bo'lmagan (non-word) xatolarni aniqlash;*
2. *alohida so'z darajasidagi xatolarni tuzatish;*
3. *kontekstga asoslangan holda so'zlarni to'g'rilash [2:390–395; 3:220–225].*

### **Imlo xatolari turlari**

Ko'plab ilmiy manbalarga ko'ra, imlo xatolari, asosan, ikki asosiy guruhga ajratiladi: *tipografik xatolar* (harf tushib qolishi, ortiqcha harf yozilishi, harflarning almashishi) hamda *sintaktik xatolar* (so'zlarning noto'g'ri bog'lanishi yoki kontekstga mos kelmasligi) [1:7–12; 2:385–390].

*Tipografik xatolar* matn yozish jarayonida inson omili ta'sirida yuzaga keladi. Bunday xatolar, odatda, e'tiborsizlik, noto'g'ri tugmani bosish yoki so'zning yozilishini yaxshi bilmaslik natijasida paydo bo'ladi [2:377–380].

*1-Jadval. Imlo xatolari turlari*

<b>№</b>	<b>Xato turi</b>	<b>Tavsifi</b>	<b>Noto'g'ri misol</b>	<b>To'g'ri shakli</b>
<b>1</b>	<b>Non-word error</b>	Lug'atda mavjud bo'lmagan so'z	Ktob	Kitob
<b>2</b>	<b>Real-word error</b>	Lug'atda bor, lekin kontekstga mos emas	Men kitob ichdim	Men kitob o'qidim
<b>3</b>	<b>Fonologik xato</b>	Talaffuzi o'xshash so'zlar chalkashadi	yuz (100) ↔ yuz (bet)	Kontekstga mos tanlanadi
<b>4</b>	<b>Transpozitsiya xatosi</b>	Harflar o'zni almashadi	Dras	Dars
<b>5</b>	<b>O'chirish xatosi</b>	Harf tushib qoladi	Maktb	Maktab
<b>6</b>	<b>Qo'shish xatosi</b>	Ortiqcha harf qo'shiladi	Kitobb	Kitob
<b>7</b>	<b>Almashtirish xatosi</b>	Harf boshqa harf bilan almashtiriladi	bola → vola	Bola

Lug‘atda mavjud bo‘lmagan xatolarni aniqlash imlo tekshiruv tizimlarining asosiy bosqichlaridan biri hisoblanadi. Ushbu turdagi xatolarni aniqlashda odatda **lug‘atga asoslangan tekshiruv** va **N-gram tahlili** qo‘llaniladi [2:377–380; 3:190–195].

2-Jadval. Non-word xatolarni aniqlash usullarining taqqoslanishi

№	Yondashuv	Ishlash prinsipi	O‘zbekcha misol	Afzalligi	Kamchiligi
1	Lug‘atga asoslangan (Dictionary lookup)	So‘z lug‘atda bor-yo‘qligi tekshiriladi	<i>ktob</i> ✗ → <i>kitob</i> ✓	Tez va sodda	Yangi so‘zlarni aniqlayolmaydi
2	Unigram	Har bir harf yoki so‘z alohida tekshiriladi	<i>k, t, o, b</i>	Oddiy model	Kontekstni hisobga olmaydi
3	Bigram	2 harfli ketma-ketliklar tekshiriladi	<i>kt, to, ob</i>	Xatoni yaxshiroq aniqlaydi	Korpus talab qiladi
4	Trigram	3 harfli ketma-ketliklar tekshiriladi	<i>kto, tob</i>	Aniqligi yuqori	Hisoblash murakkab
5	N-gram (umumiy)	Statistik ehtimollik asosida tekshiradi	<i>maktb</i> → kam ehtimol	Kontekstga yaqinlashadi	Katta ma’lumot talab qiladi

3-Jadval. N-gram tahlili

So‘z	Bigramlar	Korpusdagi holati	Natija
<b>Kitob</b>	ki, it, to, ob	Ko‘p uchraydi	To‘g‘ri
<b>Ktob</b>	kt, to, ob	“kt” juda kam	Xato
<b>maktab</b>	ma, ak, kt, ta, ab	Normal	To‘g‘ri
<b>Maktb</b>	ma, ak, kt, tb	“tb” kam	Xato

Jadvaldan ko‘rinib turibdiki, lug‘atga asoslangan usul xatolarni aniqlashda tezkor bo‘lsa-da, u faqat lug‘atda mavjud bo‘lmagan birliklarni aniqlay oladi. N-

gram yondashuvi esa harflar yoki so'zlar ketma-ketligining ehtimolini hisoblash orqali xatolarni aniqlash imkonini beradi [3:200–205].

*Qoidaviy yondashuvlar* imlo tekshiruv jarayonida tilning ichki qonuniyatlariga – xususan, morfologik tuzilma, so'z turkumlari (POS – Part of Speech) va grammatik qoidalarga tayanadi. Ushbu usullarda har bir so'z va uning shakllanishi tilga xos qoidalar asosida tekshiriladi va xatolar aniqlanadi [3:210–215].

Turli tillar uchun qoidaviy yondashuv asosida ishlab chiqilgan tizimlar mavjud. Masalan, ayrim tadqiqotlarda morfologik tahlil va leksik qidiruvga asoslangan imlo tekshiruvchilar taklif qilingan bo'lsa, boshqa ishlarda POS teglash orqali xatolarni aniqlash usullari ishlab chiqilgan [2:380–385; 3:220–225]. Bundan tashqari, Hidden Markov Model (Yashirin Markov modeli) kabi ehtimollik modellari ham qoidaviy yondashuv bilan birgalikda qo'llanilib, tizim samaradorligini oshirishga xizmat qilgan [3:230–235].

4-Jadval. Qoidaviy usullar asosiy xususiyatlari

№	Yondashuv	Asosiy tamoyil	O'zbekcha misol	Afzalligi	Kamchiligi
1	Morfologik tahlil	So'z tarkibini tekshiradi	<i>borish ni</i> ✗ → <i>borishni</i> ✓	Aniq natija	Ko'p qoida talab qiladi
2	POS teglash	So'z turkumini aniqlaydi	<i>kitob ichdi</i> ✗ → <i>kitob o'qidi</i> ✓	Kontekstni hisobga oladi	Murakkab
3	Qoidaviy tekshiruv	Grammatik qoidalar asosida	<i>men boradi</i> ✗ → <i>men boraman</i> ✓	Tilga mos	Moslashuvchan emas
4	HMM modeli	Ehtimollik + qoidalar	noto'g'ri so'z tanlash	Samarali	Hisoblash murakkab

Qoidaviy usullar yuqori aniqlikni ta'minlasa-da, ular katta hajmdagi lingvistik bilimlar bazasini talab qiladi. Ayniqsa, o'zbek tilidagi ko'p qo'shimchali shakllar ushbu usullarni yaratishni murakkablashtiradi.

*Statistik yondashuvlar* til birliklarining ehtimollik xususiyatlariga asoslanadi va ular aniq grammatik qoidalarga kamroq bog‘liq bo‘ladi. Bu usullar til korpuslaridan olingan chastota va ehtimollik ma’lumotlari asosida ishlaydi [3:200–205].

5-Jadval. Statistik usullar taqqoslanishi

№	Usul	Ishlash prinsipi	O‘zbekcha misol	Afzalligi	Kamchiligi
1	N-gram	Ehtimollik asosida	<i>maktb</i> → kam ehtimol	Tez	Kontekst cheklangan
2	FSA	Holatlar orqali tekshiradi	so‘z ketma-ketligi	Barqaror	Murakkab
3	Chastota usuli	Eng ko‘p uchraydigan variant	<i>kitob</i> > <i>ktob</i>	Oddiy	Xatolik ehtimoli bor
4	Edit distance	Eng yaqin so‘zni topadi	<i>ktob</i> → <i>kitob</i>	Samarali	Sekin bo‘lishi mumkin

Statistik yondashuvlarning asosiy ustunligi – ular tilga qat’iy bog‘liq emas va katta ma’lumotlar asosida samarali ishlaydi. Biroq ular chuqur lingvistik qoidalarni hisobga olmaydi. Shu sababli ko‘plab zamonaviy tizimlarda qoidaviy va statistik metodlar integratsiyasi, ya’ni gibril model qo‘llaniladi [2:390–395]

So‘nggi yillarda *chuqur o‘rganish (deep learning)* texnologiyalarining rivojlanishi imlo tekshiruv tizimlarida yangi bosqichni boshlab berdi. Ushbu yondashuvlar ayniqsa kontekstga bog‘liq xatolarni aniqlashda yuqori aniqlik ko‘rsatadi [3:240–245].

6-Jadval. Deep Learning usullar xususiyatlari

№	Model	Tavsifi	O‘zbekcha misol	Afzalligi	Kamchiligi
1	RNN	Ketma-ketlikni o‘rganadi	<i>kitob ichdi</i> ✗	Kontekstni tushunadi	Sekin
2	LSTM	Uzoq bog‘lanishni eslab qoladi	gap konteksti	Aniqligi yuqori	Resurs talab qiladi
3	Transformer	To‘liq kontekstni hisobga oladi	real-word xato	Eng samarali	Juda katta model

Chuqur o‘rganish asosidagi yondashuvlar an’anaviy usullarga nisbatan ancha yuqori natija ko‘rsatadi. Biroq ular katta hajmdagi ma’lumot va hisoblash



resurslarini talab qiladi. Ayniqsa, o'zbek tili kabi kam resursli tillarda bu usullarni qo'llash hali ham rivojlanish bosqichida.

Xulosa qilib aytganda, o'zbek tilining agglyutinativ xususiyati, ya'ni so'zlarning ko'plab qo'shimchalar orqali hosil bo'lishi imlo xatolarini aniqlash jarayonini murakkablashtiradi. Shu bois, o'zbek tili uchun maxsus moslashtirilgan, morfologik va kontekstual xususiyatlarni hisobga oluvchi imlo tekshiruv tizimlarini ishlab chiqish dolzarb vazifa bo'lib qolmoqda. Umuman olganda, mazkur tadqiqot natijalari o'zbek tilida ishlovchi avtomatik imlo tekshiruv tizimlarini takomillashtirish, matnni qayta ishlash jarayonlarini optimallashtirish hamda sun'iy intellekt asosidagi lingvistik texnologiyalarni rivojlantirish uchun muhim ilmiy va amaliy asos bo'lib xizmat qiladi.

### **Foydalanilgan adabiyotlar ro'yxati**

1. Bopche L., Dhopavkar G., Kshirsagar M. Grammar checking system using rule-based morphological process for an Indian language. In International Conference on Computing and Communication Systems, 2011. Pp. 524–531.
2. Caryappa B. C., Hulipalled V. R., Simha J. B. Kannada grammar checker using LSTM neural network. In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020. Pp. 332–337.
3. Dereza O. Lemmatization for ancient languages: Rules or neural networks? In Conference on Artificial Intelligence and Natural Language, 2018. Pp. 35–47.
4. Etoori P., Chinnakotla M., Mamidi R. Automatic spelling correction for resource-scarce languages using deep learning. In Proceedings of ACL 2018, Student Research Workshop. Pp. 146–152.
5. Islam S., Sarkar M. F., Hussain T., Hasan M. M., Farid D. M., Shatabda S. Bangla sentence correction using deep neural network based sequence-to-



sequence learning. In 2018 21st International Conference of Computer and Information Technology (ICCIT), 2018. Pp. 1–6.

6. Li H., Wang Y., Liu X., Sheng Z., Wei S. Spelling error correction using a nested RNN model and pseudo training data, 2018.

7. Singh S., Singh S. HINDIA: A deep-learning-based model for spell-checking of Hindi language. Neural Computing and Applications, 33(8), 2021. Pp. 3825–3840.