QO‘QON DAVLAT PEDAGOGIKA INSTITUTI ILMIY XABARLARI
(2026-yil 10-son)
FILOLOGIYA
PHILOLOGY
https://doi.org/10.70728/b.series.fil.v08.i10.088

UDK: 81'42:004.8

INGLIZ VA O‘ZBEK TILLARIDA INSON VA SUN’IY INTELLEKT YARATGAN MATNLARNI FARQLASHNING LINGVISTIK MEZONLARI

L.Iskanova
Samarqand davlat chet tillar instituti mustaqil tadqiqotchisi

Annotatsiya. Ushbu maqolada inson tomonidan yaratilgan matnlar bilan sun’iy intellekt tizimlari tomonidan hosil qilingan matnlarni lingvistik jihatdan farqlash masalasi ingliz va o‘zbek tillari materiallari asosida yoritiladi. Tadqiqotda matn muallifligini aniqlashda muhim bo‘lgan leksik, morfologik, sintaktik, punktuatsion belgilar tizimlashtiriladi.

Kalit so‘zlar: sun’iy intellekt, inson matni, AI-yaratilgan matn, lingvistik mezon, stilometriya, mualliflik, leksik xilma-xillik, sintaksis, pragmatika, o‘zbek tili, ingliz tili, matn tahlili.

Abstract. This article examines the issue of linguistically distinguishing between texts written by humans and texts generated by artificial intelligence systems based on materials from the English and Uzbek languages. The study systematizes a set of lexical, morphological, syntactic, and punctuation features that are important for determining text authorship.

Keywords: artificial intelligence, human-written text, AI-generated text, linguistic criteria, stylometry, authorship, lexical diversity, syntax, pragmatics, Uzbek language, English language, text analysis.

Аннотация. В статье рассматривается проблема лингвистического разграничения текстов, созданных человеком, и текстов, сгенерированных системами искусственного интеллекта, на материале английского и узбекского языков. В исследовании систематизируется совокупность лексических, морфологических, синтаксических и пунктуационных признаков, имеющих значение для определения авторства текста.

Ключевые слова: искусственный интеллект, человеческий текст, текст, созданный искусственным интеллектом, лингвистические критерии, стилометрия, авторство, лексическое разнообразие, синтаксис, прагматика, узбекский язык, английский язык.

Kirish. So‘nggi yillarda matn yaratishga qodir sun’iy intellekt tizimlarining tez rivojlanishi matnning muallifligini aniqlash masalasini tilshunoslikning amaliy va nazariy jihatdan muhim yo‘nalishlaridan biriga aylantirdi. Avvallari mualliflikni aniqlash asosan badiiy asarlar, anonim qo‘lyozmalar yoki bahsli ilmiy va publitsistik matnlar bilan bog‘liq

bo‘lgan bo‘lsa, bugungi sharoitda inson va algoritm yaratgan matn o‘rtasidagi farqni aniqlash zarurati yuzaga keldi. Ayniqsa, ta’lim, ilmiy yozuv, jurnalistika va raqamli kommunikatsiyada ushbu muammo alohida ahamiyat kasb etmoqda.

Stilometriya doirasida mualliflikni aniqlash uchun so‘zlarning takrorlanish chastotasi, funksional so‘zlar, gap uzunligi, lug‘at tarkibining xilma-xilligi va boshqa ko‘rsatkichlardan foydalanish an’anaviy metodlardan hisoblanadi. R.Chen, Xiong, He va Ross stilometrik yondashuvning asosiy afzalligi sifatida til uslubini o‘lchash mumkin bo‘lgan birliklarga ajratish imkoniyatini ko‘rsatadilar [1, p. 1]. Ularning fikricha, inson matni individual o‘zgaruvchanlik bilan ajralib tursa, sun’iy intellekt yaratgan matnlarda nisbatan bir xil stilistik ko‘rinish kuzatilishi mumkin [1, p. 1]. Bu farqni faqat “inson matni xilma-xil, sun’iy intellekt matni bir xil” tarzida talqin qilish yetarli emas. Chunki inson ham qoliplangan rasmiy uslubda yozishi, sun’iy intellekt esa turli uslubni taqlid qilishi mumkin. Shuning uchun masalani matnning tashqi ko‘rinishidan ko‘ra, uning ichki lingvistik tashkil topish usulidan kelib chiqib o‘rganish maqsadga muvofiq.

K.Schaaff, T.Schlippe va L.Mindner tomonidan olib borilgan tadqiqotda inson va sun’iy intellekt yaratgan matnlarni tasniflash uchun sakkiz guruhga birlashtirilgan 37 ta belgi tahlil qilingan [2, p. 941]. Tadqiqotchilar ularni matnning bashorat qilinuvchanligi, semantik xususiyatlari, funksional birliklar, hujjatning struktur xususiyatlari, xatolar, o‘qish qulayligi, sun’iy intellekt bahosi va vektor xususiyatlari kabi guruhlarga ajratganlar [2, pp. 941–945].

Mazkur yondashuv ingliz tili uchun katta amaliy imkoniyatga ega. Biroq o‘zbek tili masalasida ayni belgilarni to‘g‘ridan-to‘g‘ri qo‘llash muayyan cheklovlarga ega. Chunki o‘zbek tili agglutinativ tuzilishga ega bo‘lib, grammatik ma’no ko‘pincha ketma-ket qo‘shimchalar orqali ifodalanadi. Demak, ingliz tilidagi so‘z chastotasi yoki gap uzunligi bilan bog‘liq ko‘rsatkich o‘zbek tilida aynan bir xil lingvistik mazmun bermasligi mumkin.

So‘nggi tadqiqotlarda o‘zbek tilida katta til modellari yaratgan matnlarning morfologik, sintaktik va pragmatik jihatdan o‘ziga xos muammolari ham qayd etilmoqda. D.Solidjonov va Najmiddinov o‘zbek tilidagi sun’iy intellekt matnlarini tahlil qilib, qo‘shimchalar ketma-ketligi, morfofonologik jarayonlar, kelishuv, sintaksis, leksika va nutq darajasidagi farqlarga e’tibor qaratganlar [3]. Bu holat o‘zbek tilini inson va sun’iy intellekt matnlarini farqlash bo‘yicha alohida lingvistik model sifatida tadqiq etish zarurligini ko‘rsatadi.

Tadqiqot metodologiyasi. Tadqiqotda qiyosiy, tavsifiy, stilistik, distributiv va lingvostatistik tahlil usullaridan foydalaniladi. Inson va sun’iy intellekt yaratgan matnlar quyidagi asosiy sathlar bo‘yicha qiyoslanadi:

1. leksik sath;
2. morfologik sath;
3. sintaktik sath;
4. punktuatsion-grafik sath;

Bunday yondashuv matnni faqat so‘zlar yig‘indisi sifatida emas, balki turli til birliklari o‘zaro bog‘langan murakkab kommunikativ birlik sifatida ko‘rishga imkon beradi.

Tadqiqotning muhim tamoyili shundan iboratki, inson va sun’iy intellekt matnini ajratishda bitta belgiga tayanish mumkin emas. Masalan, grammatik xatoning yo‘qligi

avtomatik ravishda matnning sun’iy intellekt tomonidan yaratilganini anglatmaydi. Xuddi shuningdek, badiiy ifodaning mavjudligi ham matn inson tomonidan yozilganini isbotlamaydi. Shu sababli mezonlar majmuasi asosida tahlil qilish zarur.

Inson va sun’iy intellekt matnlarini ajratishda eng avvalo lug‘aviy tanlov e’tiborga olinadi. Inson muayyan fikrni ifodalash jarayonida o‘zining bilim darajasi, kasbiy tajribasi, ijtimoiy muhiti va kommunikativ maqsadiga mos ravishda so‘z tanlaydi. Shu bois bir mavzu doirasida yozilgan inson matnlarida ham leksik tanlov individual xarakterga ega bo‘lishi mumkin. R.Chen va hammualliflar stilometrik tahlilda funksional so‘zlarning chastotasini muhim ko‘rsatkich sifatida baholaydilar. Ular funksional so‘zlar mazmuniy yuklamasi nisbatan kam bo‘lsa-da, muallifning yozish odatini aks ettirishi mumkinligini ta’kidlaydilar [1, p. 2]. Mazkur mezon o‘zbek tilida yanada murakkab tus oladi. Chunki o‘zbek tilida grammatik ma’no ko‘pincha mustaqil yordamchi so‘zlar bilangina emas, balki affikslar orqali ham ifodalanadi. Masalan, “kitoblarimizdan”, “o‘qituvchilarimizga”, “kelmaganliklari” kabi shakllarda bir so‘z tarkibida bir necha grammatik ma’no mujassamlashadi. Shuning uchun o‘zbek tilidagi avtomatik matn tahlilida faqat alohida so‘zlar chastotasini emas, morfemalar va affikslarning taqsimlanishini ham hisobga olish lozim.

Ikkinchi muhim belgi – “so‘z takrorining tabiati”. Sun’iy intellekt yaratgan matnlarda mavzu nomi yoki asosiy kalit birliklarning ortiqcha takrorlanishi kuzatilishi mumkin. K.Schaaff va hammualliflar AI matnlarida sarlavha yoki mavzuni ifodalovchi birliklarning takrorlanishini alohida ko‘rsatkich sifatida ajratganlar [2, p. 943].

Masalan:

“*Artificial intelligence is changing education. Artificial intelligence provides new opportunities for students. Artificial intelligence also changes the role of teachers.*” Bu yerda birikmaning grammatik jihatdan xato emasligi muhim emas. Asosiy masala - mazmunni rivojlantirish o‘rniga ayni birlikning qayta-qayta qo‘llanishidir.

O‘zbek tilida esa bunday holat quyidagicha ko‘rinishi mumkin:

“*Sun’iy intellekt ta’limni o‘zgartirmoqda. Sun’iy intellekt o‘quvchilar uchun yangi imkoniyatlar yaratmoqda. Sun’iy intellekt o‘qituvchining roliga ham ta’sir ko‘rsatmoqda*.”

Inson bunday vaziyatda ikkinchi yoki uchinchi gapda “ushbu texnologiya”, “mazkur tizim”, “yangi texnologik vosita” kabi birliklardan foydalanishi ehtimoli yuqori. Albatta, bu ham mutlaq mezon emas. Biroq takror va uning matndagi kommunikativ vazifasini birgalikda kuzatish muhimdir.

Matnning leksik xilma-xilligi ham mualliflikni aniqlashda ahamiyatli. Inson yozuvida bir fikrni turli birliklar yordamida ifodalash, sinonimik almashtirish, qisqa va uzun birliklarni navbatlab qo‘llash, ba’zan esa muallifga xos so‘z birikmalarini yaratish holatlari uchraydi.

K.Schaaff va hammualliflar inson matnlarida so‘z tanlovining o‘zgaruvchanligi yuqoriroq bo‘lishini, AI matnlarida esa statistik jihatdan bashorat qilinadigan birliklar ko‘proq uchrashini ko‘rsatadilar [2, p. 941]. Ular perplexity ko‘rsatkichini ham shu jarayon bilan bog‘laydilar: AI yaratgan matnlarda keyingi so‘zning ehtimoliy tanlovi ko‘proq bashorat qilinadigan bo‘lishi mumkin [2, p. 942].

Bu yerda bir muhim jihatni ajratib ko‘rsatish lozim. Bashorat qilinuvchanlikni bevosita “sun’iylik belgisi” deb olish ilmiy jihatdan to‘g‘ri emas. Masalan, rasmiy hujjat, darslik yoki ilmiy ta’rif ham juda bashorat qilinadigan bo‘lishi mumkin. Demak, bu mezon faqat boshqa belgilar bilan birga qo‘llangandagina qiymat kasb etadi.

Biz taklif qilayotgan mezon quyidagicha ifodalanishi mumkin:

“*leksik xilma-xillik + takrorlarning funksional tabiati + muallifga xos birliklar + kontekstga mos sinonimik tanlov*.”

Aynan shu to‘rtlik matnning individual yoki standartlashtirilgan xarakterini aniqlashda samaraliroq bo‘lishi mumkin.

Ingliz va o‘zbek tillari o‘rtasidagi eng muhim farqlardan biri grammatik ma’noning ifodalanish usulidir. Ingliz tilida analitik vositalar, yordamchi fe’llar va qat’iyroq sintaktik tartib muhim o‘rin tutsa, o‘zbek tilida agglutinatsiya orqali grammatik ma’nolarni ketma-ket ifodalash imkoniyati yuqori. Shu sababli o‘zbek tilidagi AI matnlarni tekshirishda morfologik mezon alohida ahamiyatga ega. D.Solidjonov va M.Najmiddinov tadqiqotida o‘zbek tilida yaratilgan matnlarda qo‘shimchalar tartibi, unli uyg‘unligi, grammatik moslashuv, morfologik shakllanish va nutq darajasiga oid muammolar kuzatilgani ko‘rsatiladi [3]. Tadqiqotchilar o‘zbek tilining morfologik murakkabligi sun’iy intellekt uchun alohida sinov ekanini ta’kidlaydilar. Bu holat matnni farqlash uchun ikki tomonlama imkoniyat yaratadi. Bir tomondan, sun’iy intellekt yaratgan matn grammatik jihatdan juda silliq bo‘lishi mumkin. Ikkinchi tomondan esa, ayniqsa murakkab morfologik shakllarda, qo‘shimchalarning noo‘rin tanlanishi yoki grammatik ma’noning sun’iy ravishda ko‘chirilishi kuzatilishi ehtimoli mavjud.

Masalan:

Inson nutqiga yaqin “*Talabalarning ilmiy izlanishlari natijalari ushbu masalaning yanada chuqurroq o‘rganilishini talab etad*”.

Shubhali avtomatlashtirilgan shaklda “*Talabalarlarning ilmiy izlanishlari natijalari...”*

Ikkinchi misoldagi qo‘shimcha ortiqchaligi oddiy xato bo‘lib ko‘rinsa-da, katta hajmdagi korpusda bunday morfologik og‘ishlarning takrorlanishi muayyan modelning tilga munosabatini ko‘rsatishi mumkin. Shu nuqtadan kelib chiqib, o‘zbek tili uchun quyidagi morfologik ko‘rsatkichlarni taklif qilish mumkin:

- qo‘shimchalar ketma-ketligi;
- kelishik qo‘shimchalarining qo‘llanishi;
- egalik shakllari;
- shaxs-son ko‘rsatkichlari;
- fe’l nisbatlari;
- zamon va mayl shakllari;
- hurmat shakllari;
- morfemalarning noto‘g‘ri segmentatsiyasi.

Gap qurilishi inson va AI matnlari o‘rtasidagi farqni aniqlashda eng istiqbolli yo‘nalishlardan biridir. Inson matnida gap uzunligi, sodda va qo‘shma gaplarning nisbati, ergash gaplarning joylashuvi, inversiya, izohlovchi konstruksiyalar va ellipsis kabi

vositalarning taqsimlanishi muallifning individual uslubini aks ettirishi mumkin. K.Schaaff va hammualliflar hujjatning struktur xususiyatlari qatorida gaplar soni, gap uzunligi, paragraf hajmi, noyob so‘zlarning taqsimlanishi va sintaktik ko‘rsatkichlardan foydalanadilar [2, pp. 944–945]. Ingliz tili uchun gap uzunligi va sintaktik murakkablik muhim bo‘lsa, o‘zbek tilida “so‘z tartibining variantliligi” alohida mezon sifatida qaralishi kerak.

Masalan, o‘zbek tilida:

“*Bugun talabalar yangi loyihani muhokama qildilar*” va “*Yangi loyihani talabalar bugun muhokama qildilar”* kabi variantlar kommunikativ urg‘uga qarab ishlatilishi mumkin.

Inson muallif matn davomida bunday sintaktik variantlardan vaziyatga qarab foydalanishi mumkin. Sun’iy intellekt esa ko‘pincha statistik jihatdan ko‘proq uchraydigan, neytral va xavfsiz sintaktik modelni tanlashi mumkin.

Shuning uchun o‘zbek matnlarida quyidagi sintaktik ko‘rsatkichlar muhim:

1. SOV tartibining chastotasi;
2. inversiya holatlari;
3. gaplarning o‘rtacha uzunligi;
4. ergash gaplar nisbati;
5. uyushiq bo‘laklar;
6. qisqa va uzun gaplarning almashinuvi;
7. ellipsis;
8. sintaktik parallelizm.

Punktuatsiya ham matn muallifligi haqida ma’lum darajada ma’lumot beradi. Insonlar vergul, tire, ikki nuqta, qavs, qo‘shtirnoq va nuqtali verguldan foydalanishda individual odatlarga ega bo‘lishlari mumkin.

K.Schaaff va hammualliflar maxsus belgilar, punktuatsiya va qo‘shtirnoq ishlatilishini alohida tahlil qilganlar. Ularning ma’lumotlariga ko‘ra, ingliz tilidagi AI matnlarida qo‘shtirnoq belgilarining uchrashi inson matnlariga nisbatan kamroq bo‘lishi mumkin [2, p. 944]. Bu natijani o‘zbek tiliga ko‘chirishda ehtiyotkorlik zarur. O‘zbek tilida badiiy, ilmiy va publitsistik matnlarning punktuatsion me’yorlari bir-biridan farq qiladi. Masalan, dialogik matnda chiziqcha, ilmiy matnda qavs va ikki nuqta, publitsistik matnda qo‘shtirnoq faol ishlatilishi mumkin. Demak, punktuatsiyani alohida emas, “janr bilan bog‘langan holda” tahlil qilish kerak. O‘zbek tili uchun punktuatsion mezonlarning eng foydali ko‘rinishi quyidagicha:

“belgi chastotasi + belgilar kombinatsiyasi + gap tuzilishi + janr”.

Masalan, ketma-ket bir necha gaplarda bir xil punktuatsion qolipning takrorlanishi matnning standartlashtirilganligini ko‘rsatishi mumkin.

Xulosa. Inson va sun’iy intellekt yaratgan matnlarni farqlash bugungi tilshunoslikda yangi, lekin tez rivojlanayotgan yo‘nalishlardan biridir. Mavjud tadqiqotlar matn muallifligini aniqlashda stilometriya, leksik chastota, funksional so‘zlar, gap uzunligi, punktuatsiya, diskurs markerlari, grammatik xatolar, o‘qish murakkabligi va matnning bashorat qilinuvchanligi kabi belgilar samarali bo‘lishi mumkinligini ko‘rsatmoqda.

FOYDALANILGAN ADABIYOTLAR

1. Chen R., Xiong S., He J., Ross G.J. Stylometric detection of AI-generated texts: evidence from human and machine-written essays // Digital Scholarship in the Humanities. – 2026. – DOI: 10.1093/llc/fqag064.

2. Schaaff K., Schlippe T., Mindner L. Classification of human- and AI-generated texts for different languages and domains // International Journal of Speech Technology. – 2024. – Vol. 27. – P. 935–956. – DOI: 10.1007/s10772-024-10143-3.

3. Solidjonov D., Najmiddinov M. Investigating linguistic errors in large language model generation of Uzbek text // Cogent Arts & Humanities. – 2026. – Vol. 13, No. 1. – Article 2600519. – DOI: 10.1080/23311983.2025.2600519.

4. Mindner L., Schlippe T., Schaaff K. Classification of Human- and AI-Generated Texts: Investigating Features for ChatGPT // Artificial Intelligence in Education Technologies: New Development and Innovative Practices. – Singapore: Springer, 2023. – P. 152–170. – DOI: 10.1007/978-981-99-7947-9_12.