# SUN’IY INTELLEKT TOMONIDAN YARATILGAN VA INSON TOMONIDAN YOZILGAN MATNLARDAGI LEKSIK XUSUSIYATLARNING QIYOSIY TAHLILI

**Achilova Shahrizabonu Qobil qizi**
Qashqadaryo viloyati, Qarshi shahri
Iqtisodiyot va pedagogika universiteti
Xorijiy til va adabiyoti (ingliz tili) fakulteti
https://doi.org/10.5281/zenodo.24142192

## KIRISH

Raqamli texnologiyalarning jadal rivojlanishi, xususan, generativ sun’iy intellekt tizimlarining keng ommalashuvi matn yaratish jarayoniga yangicha yondashuvlarni shakllantirmoqda. Bugungi kunda sun’iy intellekt vositalari yordamida turli mavzularda maqolalar, ilmiy matnlar, tarjimalar, tahliliy materiallar va boshqa yozma nutq namunalarini yaratish imkoniyati mavjud. Ushbu jarayon matnning grammatik va sintaktik xususiyatlari bilan bir qatorda, uning leksik tarkibini ham ilmiy jihatdan o‘rganish zaruratini yuzaga keltirmoqda. Inson tomonidan yozilgan matnlar muallifning individual nutqiy tajribasi, bilim darajasi, kommunikativ maqsadi va uslubiy tanlovlari bilan chambarchas bog‘liq. Muallif matn yaratish davomida mavzuga mos so‘zlarni tanlaydi, sinonimik birliklardan foydalanadi, ayrim tushunchalarni izohlash yoki ta’kidlash maqsadida leksik vositalarni takror qo‘llaydi. Sun’iy intellekt tomonidan yaratilgan matnlarda esa leksik birliklar til modelining o‘rgatilgan ma’lumotlari va ehtimollik asosida ishlaydigan mexanizmlari orqali tanlanadi. Shu sababli ushbu ikki turdagi matnlarda so‘z tanlash, leksik xilma-xillik, takrorlanish va so‘zlarning kontekstual qo‘llanishida muayyan farqlar kuzatilishi mumkin.

Matnning leksik xususiyatlarini o‘rganishda faqat undagi so‘zlar sonini aniqlash bilan cheklanib qolish yetarli emas. Bunda turli so‘zlarning takrorlanish chastotasi, umumiy lug‘aviy birliklar tarkibidagi mazmunli so‘zlarning ulushi, leksik xilma-xillik darajasi hamda so‘zlarning o‘zaro birikish xususiyatlarini hisobga olish lozim. Aynan shu ko‘rsatkichlar matnning lug‘aviy tuzilishini chuqurroq tahlil qilish va turli mualliflik xususiyatlarini aniqlash imkonini beradi.

D. Biberning matn turlarini lingvistik jihatdan tasniflashga oid tadqiqotlarida yozma va og‘zaki nutq o‘rtasidagi farqlarni aniqlashda leksik birliklarning taqsimlanishi, funksional xususiyatlari va qo‘llanish chastotasiga alohida e’tibor qaratilgan [1, 2]. Mazkur yondashuv matnlarni faqat mazmuniga ko‘ra emas, balki ularning lingvistik tarkibi asosida ham qiyoslash mumkinligini ko‘rsatadi. Leksik xilma-xillikni baholash masalasi ham zamonaviy korpus lingvistikasining muhim yo‘nalishlaridan biri hisoblanadi. P. McCarthy va S. Jarvis tomonidan olib borilgan tadqiqotlarda MTLD, vocd-D va HD-D kabi ko‘rsatkichlarning matndagi lug‘aviy xilma-xillikni aniqlashdagi imkoniyatlari tahlil qilingan [6]. Ushbu ko‘rsatkichlar oddiy turli so‘zlar soniga nisbatan matn uzunligining ta’sirini kamaytirishga yordam beradi.

Mazkur maqolaning dolzarbligi sun’iy intellekt tomonidan yaratilgan matnlar soni ortib borayotgan sharoitda ularning inson tomonidan yozilgan matnlardan leksik jihatdan qanday farqlanishini aniqlash, matn muallifligini baholashda lingvistik mezonlarning ahamiyatini asoslash bilan belgilanadi. Maqolaning asosiy maqsadi sun’iy intellekt tomonidan yaratilgan va

inson tomonidan yozilgan matnlarning leksik xususiyatlarini ilmiy manbalar asosida qiyosiy tahlil qilish, ularning o‘xshash va farqli jihatlarini aniqlash hamda leksik xilma-xillik ko‘rsatkichlarining matn tahlilidagi ahamiyatini yoritishdan iborat.

**ASOSIY QISM**

Sun’iy intellekt va inson tomonidan yaratilgan matnlarning leksik xususiyatlarini qiyosiy o‘rganish, avvalo, leksik tarkib, so‘z chastotasi, leksik xilma-xillik va kontekstual moslik kabi tushunchalarni izohlashni talab qiladi. Leksik tarkib matnda qo‘llanilgan barcha lug‘aviy birliklarning umumiy yig‘indisini ifodalaydi. Leksik xilma-xillik esa matnda turli so‘zlarning qay darajada faol va rang-barang qo‘llanganini ko‘rsatadi. Ushbu ikki tushuncha o‘zaro bog‘liq bo‘lsa-da, mazmunan bir xil emas: matnda so‘zlar soni ko‘p bo‘lishi mumkin, biroq ularning takrorlanish darajasi yuqori bo‘lsa, leksik xilma-xillik nisbatan past bo‘lishi ehtimoli mavjud.

Inson tomonidan yozilgan matnlarda so‘z tanlash jarayoni muallifning individual uslubi, mavzuga munosabati va kommunikativ maqsadi bilan belgilanadi. Masalan, ilmiy matnlarda terminologik birliklar, aniq tushunchalarni ifodalovchi otlar va mantiqiy bog‘lovchi vositalar faol qo‘llanadi. Badiiy matnlarda esa obrazlilik, sinonimik almashuv, ko‘chma ma’noli birliklar va emotsional-ekspressiv vositalarning ulushi ortishi mumkin. Demak, inson matnidagi leksik xilma-xillikni baholashda uning janri, mavzusi va kommunikativ vazifasini hisobga olish zarur.

Sun’iy intellekt tomonidan yaratilgan matnlarda leksik tanlov boshqacha mexanizm asosida shakllanadi. Til modeli mavjud matnlar tarkibidagi qonuniyatlarni o‘rganadi va berilgan kontekstga mos kelishi ehtimoli yuqori bo‘lgan til birliklarini ketma-ket hosil qiladi. Natijada grammatik jihatdan ravon, mazmunan izchil ko‘rinadigan matnlar yuzaga keladi. Biroq ayrim holatlarda umumiy va ko‘p qo‘llaniladigan so‘zlarning ustunligi, bir-biriga yaqin ma’noli birliklarning takrorlanishi yoki bir xil sintaktik-leksik qoliplarning qayta qo‘llanishi kuzatilishi mumkin.

Bu xususiyatlarni o‘rganishda korpus lingvistikasi muhim metodologik asos hisoblanadi. T. McEnery va A. Hardie korpus lingvistikasini katta hajmdagi tabiiy til materiallarini tizimli o‘rganish va lingvistik xulosalarni real nutq ma’lumotlariga tayangan holda shakllantirish imkonini beruvchi yo‘nalish sifatida izohlaydi [7]. Korpusga asoslangan tahlil inson va sun’iy intellekt matnlaridagi so‘zlarning qo‘llanish chastotasi, takrorlanishi hamda leksik taqsimotini bir xil mezonlar asosida solishtirishga yordam beradi.

Leksik tahlilning asosiy ko‘rsatkichlaridan biri token va type tushunchalaridir. Token matndagi barcha so‘z qo‘llanishlarini, type esa takrorlanishidan qat’i nazar, turli so‘z shakllarini ifodalaydi. Masalan, bir matnda “language” so‘zi besh marta uchrasa, tokenlar soniga beshta birlik qo‘shiladi, type ko‘rsatkichida esa u bitta alohida lug‘aviy birlik sifatida hisoblanadi. Token va type nisbatidan foydalanish matndagi lug‘aviy xilma-xillikni dastlabki baholash imkonini beradi. Biroq matn uzunligi ortishi bilan ushbu nisbatning o‘zgarishi mumkinligi sababli uni boshqa ko‘rsatkichlar bilan birgalikda qo‘llash maqsadga muvofiq.

Leksik xilma-xillikni chuqurroq baholashda MTLD (Measure of Textual Lexical Diversity), vocd-D va HD-D kabi metodlardan foydalaniladi. MTLD matn bo‘ylab leksik xilma-xillikning saqlanish darajasini baholashga qaratilgan bo‘lib, oddiy token/type nisbatiga nisbatan matn uzunligi ta’sirini kamaytiradi. vocd-D va HD-D esa matndagi so‘zlar xilma-xilligini statistik yondashuvlar asosida tavsiflashga xizmat qiladi. McCarthy va Jarvis ushbu

ko‘rsatkichlarning qo‘llanish imkoniyatlarini qiyosiy o‘rganib, leksik xilma-xillikni baholashda ularning metodologik ahamiyatini asoslagan [6].

Sun’iy intellekt va inson matnlarini qiyoslashda quyidagi asosiy mezonlarga e’tibor qaratish mumkin:

**1-jadval.** Inson va sun’iy intellekt tomonidan yaratilgan matnlarning leksik xususiyatlari

| Tahlil mezoni | Inson tomonidan yozilgan matn | Sun’iy intellekt tomonidan yaratilgan matn |
|---|---|---|
| Leksik tanlov | Muallifning individual uslubi va kommunikativ maqsadiga bog‘liq | Model o‘rgangan til qonuniyatlari va kontekst asosida shakllanadi |
| So‘z takrorlanishi | Mavzu va uslubga qarab turlicha namoyon bo‘ladi | Ayrim holatlarda umumiy birliklar va qoliplarning takrorlanishi kuzatilishi mumkin |
| Sinonimik almashuv | Muallifning uslubiy tanlovi asosida keng qo‘llanishi mumkin | Kontekst va model imkoniyatlariga bog‘liq holda yuzaga keladi |
| Leksik xilma-xillik | Janr, mavzu va muallifning lug‘aviy boyligiga bog‘liq | Matn mavzusi, hajmi va modelning ishlash xususiyatlariga bog‘liq |
| Kollokatsion moslik | Til tajribasi va individual nutqiy malaka bilan belgilanadi | O‘rganilgan til birikmalari va ehtimoliy kontekstual munosabatlarga asoslanadi |
| Uslubiy moslashuv | Muallifning individual uslubiy maqsadi bilan belgilanadi | Berilgan topshiriq va ko‘rsatmalarga muvofiq shakllantiriladi |

Jadvaldagi xususiyatlar umumiy nazariy tavsif bo‘lib, barcha inson yoki sun’iy intellekt matnlariga bir xil darajada tegishli deb qaralmasligi kerak. Chunki leksik xususiyatlar matnning janri, uzunligi, mavzusi, tili va yaratilish sharoitiga qarab o‘zgaradi. Shu sababli matn muallifligini aniqlashda birgina leksik belgiga asoslanish ilmiy jihatdan yetarli emas. Leksik tahlilda so‘zlarning qo‘llanish chastotasi ham alohida ahamiyatga ega. Chastota tahlili matnda qaysi birliklar ko‘proq ishlatilgani, qaysi so‘zlar kam uchrashi va mazmuniy birliklarning umumiy matn tarkibidagi ulushini aniqlashga yordam beradi. Bunda matn hajmi turlicha bo‘lganda normallashtirilgan chastota ko‘rsatkichlaridan foydalanish muhim. Masalan, ikki matnning hajmi bir-biridan farq qilsa, so‘zlarning oddiy takrorlanish sonini bevosita solishtirish noto‘g‘ri xulosaga olib kelishi mumkin. Shu sababli chastotani ma’lum miqdordagi so‘zlar soniga nisbatan hisoblash maqsadga muvofiq.

Kollokatsion tahlil ham inson va sun’iy intellekt matnlarini o‘rganishda muhim o‘rin tutadi. Kollokatsiya — til birliklarining nutqda odatiy va barqaror birikish xususiyatidir. Masalan, ingliz tilidagi “make a decision”, “strong evidence” va “deep analysis” kabi birikmalar muayyan leksik birliklarning tabiiy qo‘llanishini ifodalaydi. Bunday birikmalarni tahlil qilish matnning lug‘aviy mosligi, tabiiyligi va uslubiy xususiyatlarini baholash imkonini beradi. Statistik tilni qayta ishlash bo‘yicha tadqiqotlarda ham so‘zlar o‘rtasidagi ehtimollik munosabatlari va ularning kontekstda qo‘llanishi muhim tahlil yo‘nalishi sifatida qaraladi [4, 5].

Sun’iy intellekt tomonidan yaratilgan matnlarda kollokatsion birliklar ko‘pincha grammatik jihatdan to‘g‘ri shakllanishi mumkin. Biroq ularning har doim ham muayyan janr yoki kommunikativ vaziyatga to‘liq mos kelishini kafolatlab bo‘lmaydi. Inson tomonidan yozilgan matnlarda esa muallifning til tajribasi, madaniy bilimi va individual uslubi so‘z

birikmalarining tanlanishiga ta’sir ko‘rsatadi. Shu jihatdan kollokatsion moslikni leksik xilma-xillik va chastota ko‘rsatkichlari bilan birgalikda tahlil qilish zarur.

Leksik xususiyatlarni qiyosiy o‘rganish uchun metodologik jihatdan quyidagi ketma-ketlikka amal qilish mumkin:

- bir xil mavzu va janrga mansub inson hamda sun’iy intellekt matnlarini tanlash;
- matnlarning umumiy hajmini imkon qadar tenglashtirish;
- token va type birliklarini aniqlash;
- yuqori chastotali va takrorlanuvchi so‘zlarni ajratish;
- MTLD, vocd-D yoki HD-D ko‘rsatkichlari orqali leksik xilma-xillikni baholash;
- kollokatsion birliklar va mazmuniy so‘zlarning qo‘llanishini qiyoslash;
- olingan natijalarni matn janri va kommunikativ vaziyatni hisobga olgan holda izohlash.

Bunday yondashuv matnlar o‘rtasidagi leksik farqlarni tizimli aniqlash imkonini beradi. Bunda matnlar bir xil sharoitda tanlanishi, tahlil mezonlari oldindan belgilanishi va statistik ko‘rsatkichlar asosida xulosa chiqarilishi muhimdir. Mazkur maqolada keltirilgan qiyosiy xususiyatlar ilmiy adabiyotlar tahliliga asoslangan nazariy umumlashtirish bo‘lib, muayyan hajmdagi korpusda o‘tkazilgan eksperimental natijalar sifatida talqin qilinmasligi lozim. Umuman olganda, sun’iy intellekt va inson tomonidan yaratilgan matnlarning leksik xususiyatlarini qiyoslashda so‘zlarning soni yoki takrorlanish darajasigina emas, balki ularning matn ichidagi taqsimlanishi, semantik vazifasi, kollokatsion mosligi va uslubiy qo‘llanishini ham hisobga olish zarur. Bu esa matn muallifligini aniqlash, akademik yozuv sifatini baholash va korpus lingvistikasi tadqiqotlarini rivojlantirishda muhim nazariy-amaliy ahamiyat kasb etadi.

**XULOSA**

Sun’iy intellekt tomonidan yaratilgan va inson tomonidan yozilgan matnlarning leksik xususiyatlarini qiyosiy tahlil qilish natijasida matnning lug‘aviy tarkibi, so‘zlarning qo‘llanish chastotasi, leksik xilma-xillik va kollokatsion moslik kabi mezonlar muhim ilmiy ahamiyatga ega ekani aniqlandi. Ushbu ko‘rsatkichlar matnlarning lingvistik tuzilishini chuqurroq o‘rganish va ularning o‘ziga xos xususiyatlarini tavsiflash imkonini beradi.

Inson tomonidan yaratilgan matnlarda leksik tanlov muallifning individual uslubi, mavzu, janr va kommunikativ maqsadi bilan chambarchas bog‘liq. Sun’iy intellekt matnlarida esa lug‘aviy birliklarning tanlanishi til modelining o‘rgatilgan ma’lumotlari va kontekstual ehtimollik munosabatlariga asoslanadi. Shu sababli ayrim matnlarda takrorlanish, umumiy leksik birliklarning ustunligi va bir xil qoliplarning qo‘llanishi kuzatilishi mumkin. Biroq bu xususiyatlarni barcha sun’iy yoki inson matnlariga xos mutlaq belgi sifatida baholash to‘g‘ri emas.

Leksik xilma-xillikni aniqlashda MTLD, vocd-D va HD-D kabi ko‘rsatkichlardan foydalanish oddiy token/type nisbatiga qaraganda kengroq tahlil imkoniyatini yaratadi. Chastota va kollokatsion tahlil bilan birgalikda qo‘llanganda ushbu metodlar matnning leksik tuzilishi haqida yanada to‘liq tasavvur hosil qilishga ko’maklashadi.

Tadqiqotning nazariy umumlashtirishlari asosida matn muallifligini aniqlashda faqat bitta lingvistik belgiga tayanish yetarli emasligi ta’kidlanadi. Ishonchliroq baholash uchun leksik, sintaktik, semantik va pragmatik ko‘rsatkichlarni o‘zaro uyg‘unlashtirish lozim. Ayniqsa, matnlarning mavzusi, hajmi, janri va yaratilish sharoitini hisobga olish qiyosiy tahlilning

xolisligini ta’minlaydi. Kelgusidagi tadqiqotlarda ingliz tilidagi inson va sun’iy intellekt matnlaridan iborat muvozanatlangan korpus shakllantirish, ularni maxsus lingvistik dasturlar yordamida tahlil qilish va leksik ko‘rsatkichlarni statistik usullar bilan tekshirish maqsadga muvofiq. Bunday tadqiqotlar avtomatik yaratilgan matnlarni lingvistik jihatdan baholash, akademik yozuv madaniyatini rivojlantirish va zamonaviy korpus lingvistikasining amaliy imkoniyatlarini kengaytirishga yordam beradi.

**FOYDALANILGAN ADABIYOTLAR**

1. Biber D. A Typology of English Texts // Linguistics. – 1989. – Vol. 27, No. 1. – P. 3–44. – DOI: https://doi.org/10.1515/ling.1989.27.1.3.
2. Biber D. Variation across Speech and Writing. – Cambridge: Cambridge University Press, 1988.
3. Gries S. Th. Statistics for Linguistics with R: A Practical Introduction. – Berlin; Boston: De Gruyter Mouton, 2013.
4. Jurafsky D., Martin J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. – Upper Saddle River, NJ: Prentice Hall, 2000.
5. Manning C. D., Schütze H. Foundations of Statistical Natural Language Processing. – Cambridge, MA: MIT Press, 1999.
6. McCarthy P. M., Jarvis S. MTLD, vocd-D, and HD-D: A Validation Study of Sophisticated Approaches to Lexical Diversity Assessment // Behavior Research Methods. – 2010. – Vol. 42, No. 2. – P. 381–392. – DOI: https://doi.org/10.3758/BRM.42.2.381.
7. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. – Cambridge: Cambridge University Press, 2012.