SUN’IY INTELLEKT YORDAMIDA O‘ZBEK TILIDAGI NUTQIY XATOLARNI ANIQLASH VA ULARNING LINGVISTIK TASNIFI

Dushamova Nodirabegim *– mustaqil tadqiqotchi*

Annotatsiya. Mazkur maqolada sun’iy intellekt texnologiyalari yordamida o‘zbek tilidagi nutqiy xatolarni avtomatik aniqlash, ularni lingvistik jihatdan tasniflash va tuzatish imkoniyatlari tahlil qilinadi. O‘zbek tilining agglutinativ xususiyatlari, morfologik shakllarning ko‘pligi, so‘z tartibining o‘zgaruvchanligi, leksik-semantik munosabatlar hamda nutq vaziyatiga bog‘liq pragmatik omillar xatolarni aniqlash jarayoniga ta’sir etuvchi asosiy omillar sifatida ko‘rib chiqiladi. Shuningdek, imloviy, morfologik, sintaktik, leksik-semantik, uslubiy va pragmatik xatolarni aniqlashda qoidaga asoslangan algoritmlar, statistik modellar va neyron tarmoqlardan foydalanish masalalari yoritiladi. Tadqiqot doirasida o‘zbek tilidagi nutqiy xatolarni tizimli tasniflash va ularni avtomatik tahlil qilish uchun lingvistik mezonlarga asoslangan yondashuv taklif etiladi.

Kalit so‘zlar: sun’iy intellekt, o‘zbek tili, nutqiy xatolar, kompyuter lingvistikasi, tabiiy tilni qayta ishlash, avtomatik tahlil, morfologik xato, sintaktik xato, semantik tahlil, pragmatika.

Bugungi kunda sun’iy intellekt texnologiyalarining rivojlanishi tilni o‘rganish, matnlarni tahrirlash, avtomatik tarjima qilish va inson nutqini tahlil etish imkoniyatlarini kengaytirmoqda. Tabiiy tilni qayta ishlash texnologiyalari matndagi grammatik va imloviy xatolarni aniqlash, gap tuzilishini tahlil qilish hamda matn mazmunini baholashda muhim ahamiyat kasb etmoqda. Shu bilan birga, o‘zbek tilining o‘ziga xos grammatik va semantik xususiyatlari xatolarni avtomatik aniqlash tizimlarini ishlab chiqishda alohida yondashuvni talab qiladi. Nutqiy xato tushunchasi til birliklarining adabiy til me’yorlari, grammatik qoidalar yoki

kommunikativ vaziyat talablariga mos kelmasligi bilan bog‘liq hodisalarni qamrab oladi. Bunday xatolar yozma matnda imloviy va grammatik buzilishlar ko‘rinishida, og‘zaki nutqda esa talaffuz, so‘z tanlash, grammatik shakllantirish yoki fikrni ifodalashdagi nomuvofiqliklar tarzida namoyon bo‘lishi mumkin.

Sun’iy intellekt yordamida nutqiy xatolarni aniqlash faqat noto‘g‘ri yozilgan so‘zni topish bilan cheklanmaydi. Tizim xatoning qaysi lingvistik sathga tegishli ekanini, uning yuzaga kelish sababini va kontekstda qanday tuzatish mumkinligini ham aniqlashi kerak. Masalan, “Men maktabga bordim kecha” jumlasi ayrim kontekstlarda grammatik jihatdan mumkin bo‘lsa-da, neytral yozma bayonda so‘z tartibini “Men kecha maktabga bordim” shaklida berish uslubiy jihatdan tabiiyroq bo‘lishi mumkin. Demak, barcha farqli ifodalarni xato deb baholash to‘g‘ri emas.

O‘zbek tili turkiy tillar oilasiga mansub bo‘lib, agglutinativ tuzilishi bilan ajralib turadi. So‘zlarga turli qo‘shimchalarning ketma-ket qo‘shilishi orqali grammatik va yangi leksik ma’nolar hosil qilinadi. Shu sababli so‘zning asosini, qo‘shimchalar tartibini va grammatik vazifasini to‘g‘ri aniqlash avtomatik tahlil uchun muhim hisoblanadi. Masalan, “kitoblarimizdan” so‘zi asos, ko‘plik, egalik va kelishik ma’nolarini birlashtiradi. Ushbu tarkibdagi qo‘shimchalarning noto‘g‘ri qo‘llanishi morfologik xatoga olib kelishi mumkin. Biroq tizim kam uchraydigan so‘z shaklini avtomatik ravishda xato deb belgilamasligi kerak, chunki u to‘g‘ri yasalgan, ammo korpusda kam uchraydigan birlik bo‘lishi mumkin. O‘zbek tilidagi xatolarni aniqlashni murakkablashtiradigan yana bir omil — lotin va kirill yozuvlarining parallel qo‘llanishidir. Bundan tashqari, apostrof va tutuq belgilarining turli belgilar bilan yozilishi, o‘zlashma so‘zlarning variantlari, shevaga xos shakllar va norasmiy internet yozishmalari ham avtomatik tekshiruv natijalariga ta’sir qiladi. Nutqiy xatolarni tahlil qilishda tilning me’yoriy qoidalarini, nutq uslubini va kommunikativ vaziyatni birgalikda hisobga olish zarur. Chunki og‘zaki suhbatda tabiiy bo‘lgan ayrim shakllar rasmiy hujjat yoki ilmiy matnda maqsadga muvofiq bo‘lmasligi mumkin.

O‘zbek tilidagi nutqiy xatolarni avtomatik aniqlash tizimida ularni lingvistik sathlar bo‘yicha tasniflash muhim ahamiyatga ega. Quyidagi tasnif amaliy tahlil uchun taklif etiladi.

Imloviy xatolar so‘zlarning yozilish me’yorlariga rioya qilmaslik natijasida yuzaga keladi. Bunga harflarni tushirib qoldirish, ortiqcha harf yozish, harflarni almashtirish, qo‘shib yoki ajratib yozish qoidalarini buzish kiradi.

Misollar:

“kitobhon” — “kitobxon”;

“ma’sul” — “mas’ul”;

“ma’lumotlar bazasi” birikmasidagi so‘zlarni imlo qoidalariga muvofiq yozish va ajratish.

Sun’iy intellekt imloviy xatolarni lug‘atlar, belgilar ketma-ketligini solishtirish, tahrir masofasi algoritmlari va kontekstual til modellari yordamida aniqlashi mumkin. Bunda faqat so‘zning shakli emas, balki gapdagi ma’nosi ham hisobga olinishi kerak.

Morfologik xatolar so‘zning grammatik shakllarini noto‘g‘ri hosil qilish yoki qo‘llash bilan bog‘liq. Ular kelishik, egalik, ko‘plik, zamon, shaxs-son va boshqa grammatik kategoriyalarda kuzatilishi mumkin.

Masalan:

“Men kitobni o‘qidi” — “Men kitobni o‘qidim”;

“Biz maktabga bordim” — “Biz maktabga bordik”.

Bunday xatolarni aniqlash uchun tizim ega va kesim o‘rtasidagi shaxs-son mosligini, fe’l shakllarini hamda qo‘shimchalarning grammatik vazifasini tekshirishi lozim. O‘zbek tilining agglutinativ xususiyati sababli morfologik tahlil faqat tayyor lug‘at shakllariga tayanmasligi kerak. So‘z tarkibini ajratish va grammatik belgilarni aniqlash xatolarni ishonchliroq topishga yordam beradi.

Sintaktik xatolar gap bo‘laklari o‘rtasidagi grammatik munosabatlar buzilganda yoki gap tuzilishi mazmunni ifodalashga xalaqit berganda yuzaga keladi.

Misol:

“Men kitob o‘qiyapti” — “Men kitob o‘qiyapman”.

Birinchi jumlada ega va kesim o‘rtasidagi shaxs-son mosligi buzilgan. Avtomatik tizim bunday xatoni sintaktik tahlil va morfologik belgilarni o‘zaro solishtirish orqali aniqlashi mumkin.

Shu bilan birga, o‘zbek tilida so‘z tartibi kommunikativ maqsadga qarab o‘zgarishi mumkin. Shuning uchun gapdagi so‘zlarning odatiy tartibdan farq qilishi o‘z-o‘zidan sintaktik xato hisoblanmaydi. Tizim grammatik jihatdan mumkin bo‘lgan variantlarni uslubiy afzalliklardan farqlashi zarur.

Leksik-semantik xatolar so‘zning ma’nosini noto‘g‘ri tanlash, o‘zaro mos kelmaydigan birliklarni biriktirish yoki kontekstga mos bo‘lmagan ifodadan foydalanish natijasida yuzaga keladi. Masalan, “U baland ovozda kitobni ko‘rdi” jumlasida “baland ovozda” birikmasi “ko‘rdi” fe’li bilan mazmunan mos kelmaydi. Muallifning maqsadiga qarab “U kitobni diqqat bilan ko‘rdi” yoki “U kitobni baland ovozda o‘qidi” kabi variantlar mos kelishi mumkin. Bunday xatolarni aniqlash imloviy xatolarni topishdan murakkabroq, chunki gapdagi barcha so‘zlar alohida olinganda to‘g‘ri yozilgan bo‘lishi mumkin. Tizim ularning ma’no munosabatlarini, fe’lning semantik talablarini va umumiy kontekstni tahlil qilishi lozim.

Uslubiy xatolar til birliklarining matn turi, kommunikativ maqsad yoki adabiy uslub talablariga mos kelmasligi bilan bog‘liq. Masalan, ilmiy maqolada haddan tashqari so‘zlashuv uslubidagi iboralar yoki rasmiy hujjatda noaniq, hissiy ifodalar qo‘llanishi matn sifatiga salbiy ta’sir qilishi mumkin. Sun’iy intellekt matn uslubini aniqlash, so‘z tanlovini baholash va muqobil ifodalarni tavsiya etish orqali bunday muammolarni kamaytirishga yordam beradi. Biroq uslubiy tavsiyalarni mutlaq grammatik qoida sifatida ko‘rsatmaslik kerak. Pragmatik xatolar til birliklarining muayyan vaziyat, suhbatdoshlar o‘rtasidagi munosabat yoki kommunikativ maqsadga mos kelmasligi bilan tavsiflanadi. Masalan, rasmiy vaziyatda suhbatdoshga murojaat qilishda “Siz” shakli o‘rniga noo‘rin tarzda “sen” shaklidan

foydalanish muloqot me’yorlariga mos kelmasligi mumkin. Lekin do‘stlar o‘rtasidagi suhbatda “sen” shakli mutlaqo tabiiy bo‘lishi mumkin. Demak, pragmatik xatoni aniqlash uchun tizim suhbat konteksti, muloqot ishtirokchilari va nutq maqsadini hisobga olishi kerak. Bunday xatolarni faqat grammatik qoidalar asosida to‘liq aniqlash qiyin.

O‘zbek tilidagi nutqiy xatolarni avtomatik aniqlashda bir nechta texnologik yondashuvlardan foydalanish mumkin.

1. Birinchi yondashuv — qoidalarga asoslangan tizimlar. Bunday tizimlar imlo lug‘atlari, grammatik qoidalar va oldindan belgilangan andozalar asosida ishlaydi. Ular aniq grammatik xatolarni aniqlashda samarali bo‘lishi mumkin, biroq murakkab kontekst va uslubiy variantlarni tahlil qilishda cheklovlarga ega.
2. Ikkinchi yondashuv — statistik va ehtimollik modellaridan foydalanish. Ushbu usulda so‘zlar va grammatik shakllarning korpusdagi uchrash chastotasi hamda kontekst bilan bog‘liqligi hisobga olinadi. Biroq kam uchraydigan, lekin to‘g‘ri shakllarni xato deb belgilash xavfi mavjud.
3. Uchinchi yondashuv — neyron tarmoqlar va kontekstual til modellari. Ular gapning umumiy mazmunini, so‘zlar o‘rtasidagi munosabatlarni va uzoq masofadagi grammatik bog‘lanishlarni o‘rganishi mumkin. Bunday modellar katta hajmdagi sifatli ma’lumotlar bilan o‘qitilganda murakkab xatolarni aniqlash imkoniyatini kengaytiradi.
4. To‘rtinchi yondashuv — gibrid tizimlar. Ular lingvistik qoidalar, morfologik tahlil, lug‘atlar va neyron modellarni birlashtiradi. O‘zbek tili uchun bunday yondashuv istiqbolli hisoblanadi, chunki u grammatik bilimlardan foydalanish bilan birga kontekstual tahlil imkoniyatlarini ham saqlab qoladi.

Sun’iy intellekt yordamida o‘zbek tilidagi nutqiy xatolarni aniqlash va ularni lingvistik jihatdan tasniflash zamonaviy kompyuter lingvistikasining muhim yo‘nalishlaridan biridir. Nutqiy xatolarni imloviy, morfologik, sintaktik, leksik-

semantik, uslubiy va pragmatik turlarga ajratish avtomatik tahlil tizimlarining vazifalarini aniq belgilashga yordam beradi. O‘zbek tilining agglutinativ tuzilishi, grammatik shakllarning xilma-xilligi, yozuv variantlari va kontekstga bog‘liq ma’no xususiyatlari xatolarni aniqlashda maxsus yondashuvlarni talab qiladi. Shu sababli qoidalarga asoslangan usullarni kontekstual neyron modellar va lingvistik bilimlar bilan birlashtirish maqsadga muvofiqdir. Tadqiqotning amaliy ahamiyati shundaki, taklif etilgan tasnif va metodologiya o‘zbek tilidagi matnlarni avtomatik tekshirish, tahrirlash va lingvistik tahlil qilish tizimlarini ishlab chiqishda foydalanilishi mumkin. Kelgusida sifatli belgilangan korpuslar yaratish, sheva va uslubiy variantlarni hisobga olish hamda modellarni real nutqiy vaziyatlarda sinash ushbu sohaning rivojlanishiga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Hojiyev A. Tilshunoslik terminlarining izohli lug‘ati. — Toshkent: O‘zbekiston Milliy ensiklopediyasi, 2002.
2. Ne’matov H., Rasulov R. O‘zbek tili sistem leksikologiyasi asoslari. — Toshkent: O‘qituvchi, 1995.
3. Sayfullayeva R. va boshqalar. Hozirgi o‘zbek adabiy tili. — Toshkent: Fan va texnologiya, 2009.
4. Abdujalilova Sh. A. O‘zbek tili kompyuter lingvistikasi asoslari. — Toshkent: Noshir, 2021.