RAQAMLI MUHITDA TIL VA NUTQ EKOLOGIYASINI O‘RGANISHNING AMALIY VA KOMPYUTER-LINGVISTIK IMKONIYATLARI

Abduvaxabova Dilnoza Nurmaxamatovna
O‘zDJTU, dotsent

Annotatsiya. *Ko‘p tilli raqamli matnlarni o‘rganishda autentik korpus tuzish, annotatsiya, kompyuter yordamida hisoblash va kontekstual-lingvoekologik talqin bosqichlari asoslanadi.*

Kalit so‘zlar: *korpus lingvistikasi; kompyuter lingvistikasi; autentik material; annotatsiya; raqamli diskurs; lingvoekologik tahlil.*

Аннотация. *Рассматриваются теоретические и методологические аспекты экологии языка и речи в многоязычной цифровой коммуникации. Предложены принципы контекстуальной оценки языковых явлений.*

Ключевые слова: *экология языка, экология речи, многоязычие, цифровая коммуникация, лингвистический анализ.*

Abstract. *The paper examines theoretical and methodological aspects of language and speech ecology in multilingual digital communication and proposes context-sensitive principles for linguistic analysis.*

Keywords: *language ecology, speech ecology, multilingualism, digital communication, linguistic analysis.*

Raqamli kommunikatsiya tilshunoslik uchun yangi kuzatish maydonini yaratdi. Ijtimoiy tarmoq izohlari, messenjer yozishmalari, forum muhokamalari va elektron ta’lim platformalaridagi matnlar real nutqiy amaliyotni o‘rganish imkonini beradi. Shu bilan birga, bunday materiallar hajmi, janri, til tarkibi va kommunikativ vazifasi jihatidan xilma-xildir. Natijada til va nutq ekologiyasiga oid xulosalarni alohida tanlangan misollar asosida emas, balki tizimli yig‘ilgan hamda kontekst bilan ta’minlangan materiallar asosida ishlab chiqish ehtiyoji kuchayadi.

Til ekologiyasi an’anaviy ravishda til va uning ijtimoiy muhiti munosabatini o‘rgangan bo‘lsa, raqamli davr bu munosabatga texnologik infratuzilma, platforma algoritmlari va avtomatik matn ishlab chiqarish omillarini ham qo‘shmoqda. Haugenning til muhiti haqidagi qarashlari [1] bugungi sharoitda raqamli makonni tilning faoliyat yuritish muhitlaridan biri sifatida ko‘rishga asos bo‘ladi. Biroq zamonaviy tadqiqot faqat nazariy tavsif bilan cheklanmasdan, raqamli materialdagi real tendensiyalarni empirik aniqlashni ham talab qiladi. Bu vazifani bajarishda korpus lingvistikasi muhim metodologik imkoniyatga ega. Korpus muayyan ilmiy

mezonlar asosida tanlangan, elektron shaklda saqlanadigan va qidiruv hamda tahlil uchun tayyorlangan matnlar majmuasidir. T. McEnery va A. Hardie korpus tadqiqotida ma’lumot tanlash, reprezentativlik va tahlil metodlarining o‘zaro bog‘liqligini alohida ko‘rsatadi [2]. Ko‘p tilli raqamli muloqot uchun tuziladigan korpus esa faqat so‘zlar ro‘yxati emas, balki kommunikativ vaziyat haqidagi ma’lumotlarni ham qamrab olishi kerak.

Korpusni shakllantirishning dastlabki bosqichida tadqiqot savoli aniqlanadi. Masalan, o‘zbek, rus va ingliz tillari ishtirokidagi yozishmalarda kod almashinuvi qanday kommunikativ vazifalarni bajaradi? Qaysi turdagi platformalarda transliteratsiya faolroq uchraydi? Qanday sharoitlarda xorijiy termin muloqotni aniqlashtiradi, qanday vaziyatlarda esa tushunarlilikka xalaqit beradi? Bunday savollar material yig‘ish strategiyasini belgilaydi va tasodifiy misollar to‘plash xavfini kamaytiradi. Material tanlashda autentiklik tamoyili asosiy o‘rin tutadi. Autentik material real kommunikativ jarayonda tabiiy ravishda yuzaga kelgan, tadqiqotchi tomonidan sun’iy tuzilmagan matndir. Tadqiqotchi misol tariqasida «faylni upload qildim» kabi jumla yaratishi mumkin, ammo uning real yozishmadan olingani hujjatlashtirilmagan bo‘lsa, uni korpusdagi empirik dalil sifatida taqdim etish to‘g‘ri emas. Shuningdek, so‘rovnoma orqali olingan javoblar empirik ma’lumot bo‘lsa-da, ular har doim tabiiy raqamli diskurs namunasi hisoblanmaydi. Bu farq natijalarni ilmiy talqin qilishda muhimdir.

Ikkinchi tamoyil — korpusning muvozanatliligi. Faqat bitta platforma yoki bitta yosh guruhiga mansub yozishmalar asosida barcha raqamli muloqot haqida umumiy xulosa chiqarish qiyin. Materiallar platforma, janr, kommunikativ vaziyat va til tarkibiga ko‘ra saralanishi kerak. Shu bilan birga, mutlaq reprezentativlikka erishish murakkab ekanini tan olish zarur. Shuning uchun tadqiqot chegaralari ochiq ko‘rsatilishi, qaysi turdagi matnlar qamrab olingani va qaysilari tahlildan tashqarida qolgani izohlanishi lozim.

Uchinchi bosqich — metama’lumotlar va annotatsiya. Har bir matn uchun platforma, mavjud bo‘lsa sana, janr, asosiy va qo‘shimcha tillar, kommunikativ vaziyat kabi belgilar qayd etiladi. Annotatsiya jarayonida kod almashinuvi, kod aralashuvi, transliteratsiya, internet slengi, qisqartmalar, o‘zlashgan leksika hamda emoji kabi elementlar alohida belgilanadi. N. Ide va J. Pustejovsky tahriridagi qo‘llanma lingvistik annotatsiyaning izchil va qayta tekshiriladigan tizimini yaratish ahamiyatini ko‘rsatadi [3]. Annotatsiya nafaqat birliklarni sanash, balki ularni kontekstga bog‘lash imkonini ham beradi. Masalan, inglizcha terminning uchrashi kod almashinuvi bilan bir xil hodisa emas: ayrim terminlar tilning o‘zlashgan leksik qatlamiga kirgan bo‘lishi mumkin. Shu sababli avtomatik belgilar qo‘yishdan oldin tushunchalarning operatsion ta’riflari ishlab chiqilishi zarur. Kod almashinuvi,

o‘zlashma va transliteratsiya chegaralari aniq belgilanmasa, miqdoriy natijalar bir-biriga aralashib ketadi.

Kompyuter lingvistikasi vositalari katta hajmdagi matnlar bilan ishlashni yengillashtiradi. AntConc kabi dasturlarda chastota ro‘yxatlari, konkordanslar va kollokatsiyalarni aniqlash mumkin. Zaruratga qarab Python vositasida matnlarni tozalash, belgilangan kategoriyalarni hisoblash va natijalarni jadvalga joylashtirish amalga oshiriladi. Biroq avtomatik aniqlashning chegaralari mavjud. O‘zbek tilidagi morfologik shakllar, lotin-kirill yozuvi almashinuvi, imlo variantlari va aralash til birliklari dasturiy tahlilda xatolarga olib kelishi mumkin. Shuning uchun miqdoriy tahlil lingvistik ekspertiza bilan to‘ldirilishi lozim. Masalan, ma’lum so‘zning yuqori chastotasi uning ekologik xavf tug‘dirishini o‘z-o‘zidan isbotlamaydi. So‘zning kimga qaratilgani, qanday vaziyatda ishlatilgani, hazil yoki haqorat vazifasini bajarayotgani alohida o‘rganiladi. Kompyuter vositasida olingan natijalarni kontekstual va pragmatik tahlil bilan bog‘lash metodologik jihatdan ishonchliroq yondashuvdir.

Raqamli nutq ekologiyasini baholashda kamida uch turdagi indikatorni farqlash maqsadga muvofiq: lingvistik indikatorlar, kommunikativ-pragmatik indikatorlar va ekologik xavf indikatorlari. Birinchi guruhga me’yoriy variantlar, til aralashuvi va transliteratsiya kiradi. Ikkinchi guruh tushunarlilik, adresatga moslik va kommunikativ vazifani qamrab oladi. Uchinchi guruhda esa haqorat, diskriminativ ifoda, manipulyativ nutq kabi hodisalar kontekst bilan birga ko‘rib chiqiladi. Bunday tasnif bir hodisani boshqa hodisaning o‘rniga qo‘yish xavfini kamaytiradi.

Generativ sun’iy intellekt vositalari ham til ekologiyasi uchun yangi savollarni yuzaga keltirmoqda. Avtomatik tarjima va matn yaratish tizimlari raqamli kontent ishlab chiqarishni osonlashtiradi, biroq ayrim hollarda uslubiy birxillik, noaniq terminlar yoki tabiiy bo‘lmagan konstruksiyalarni ham ko‘paytirishi mumkin. Bu ta’sirlarni umumiy taxmin sifatida emas, balki inson yozgan va avtomatik yaratilgan matnlarni aniq mezonlar asosida qiyoslash orqali o‘rganish zarur. Hozircha bunday taqqoslash natijalari mavjud bo‘lmagan holatda uning oqibatlarini qat’iy tasdiqlash ilmiy jihatdan to‘g‘ri bo‘lmaydi.

Tadqiqot etikasiga rioya qilish ham metodologiyaning ajralmas qismidir. Ochiq platformadagi matnning mavjudligi undan istalgan shaklda foydalanish mumkinligini anglatmaydi. Shaxsga oid ma’lumotlarni anonimlashtirish, yopiq guruh materiallari uchun tegishli ruxsatlarni olish va matnni qayta e’lon qilishning mumkin bo‘lgan oqibatlarini baholash kerak. Internet tadqiqotlari bo‘yicha AoIR tavsiyalari kontekstga sezgir etik qarorlarning muhimligini ko‘rsatadi [4]. Taklif etilayotgan metodologik izchillik quyidagicha umumlashtiriladi: tadqiqot savolini

aniqlash, autentik materialni tanlash, korpusni tizimlashtirish, metama'lumotlarni qayd etish, annotatsiya, miqdoriy hisoblash, kontekstual talqin va lingvoekologik baholash. Bu ketma-ketlik nazariy konsepsiyalar bilan amaliy kuzatish o'rtasida bog'lanish yaratadi. Ayniqsa, til hodisalarini avtomatik ravishda «ijobiy» yoki «salbiy» deb tasniflamasdan, ularning vazifasi va kommunikativ oqibatini tekshirish imkonini beradi.

Xulosa qilib aytganda, amaliy va kompyuter lingvistikasi til va nutq ekologiyasini o'rganishda bir-birini to'ldiruvchi vositalardir. Korpus texnologiyalari hodisalarning tarqalishi va kontekstini ko'rsatadi; avtomatik ishlov katta hajmdagi materialni tartibga soladi; diskursiv hamda pragmatik tahlil esa miqdoriy natijalarning mazmunini ochadi. Mazkur integrativ yondashuv keyinchalik o'zbek, rus va ingliz tillari ishtirokidagi autentik raqamli matnlar asosida tekshirilishi mumkin.

FOYDALANILGAN ADABIYOTLAR

1. Haugen E. The Ecology of Language. — Stanford: Stanford University Press, 1972.

2. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. — Cambridge: Cambridge University Press, 2012.

3. Ide N., Pustejovsky J. (eds.). Handbook of Linguistic Annotation. — Dordrecht: Springer, 2017.

4. Association of Internet Researchers. Internet Research: Ethical Guidelines 3.0. — 2020.