INNOVATIVE PUBLICATION
Journal of Effective Learning and Sustainable Innovation
Vol.4 №5 (2026). MAY
innovativepublication.uz

DIALEKTOLOGIK KORPUS YARATISHNING XALQARO TAJRIBASI VA UNING O‘ZBEK TILIGA TATBIQI

Qurbonnazarova Aisla O‘ktamjonovna
O‘zbekiston Milliy universiteti,
Jurnalistika va o‘zbek filologiyasi fakulteti,
Kompyuter lingvistikasi mutaxassisligi I bosqich magistranti
asilaqurbonnazarova74@gmail.com

Annotatsiya

Mazkur maqolada dialektologik korpuslarni yaratish bo‘yicha xalqaro tajribalar, xususan Nordik Dialekt Korpusi hamda ManDi korpusi tahlil qilinadi. Ushbu korpuslar zamonaviy lingvistik tadqiqotlar uchun mo‘ljallangan ilg‘or resurslar bo‘lib, ular turli metodologik yondashuvlarni o‘zida jamlagan kompleks tizim sifatida shakllangan. Korpuslarda nutq materiallari web-platforma asosida yig‘ilgan bo‘lib, audio va video yozuvlar transkripsiyalar bilan integratsiya qilingan holda taqdim etiladi. Mazkur yondashuv til materiallarini multimodal formatda tahlil qilish imkonini yaratadi. Ushbu tajribalar asosida o‘zbek dialektologik korpusini yaratishda qo‘llanishi mumkin bo‘lgan asosiy metodologik tamoyillar aniqlanadi.

Kalit so‘zlar: dialektologik korpus, korpus lingvistikasi, Nordik Dialekt Korpusi, ManDi Korpusi, o‘zbek tili, multimodal korpus, masofaviy ma’lumot yig‘ish, fonetik tahlil, transkripsiya, lingvistik resurslar, NLP, sotsiolingvistika.

Abstract

This paper analyzes international experiences in the development of dialectological corpora, in particular the Nordic Dialect Corpus and the ManDi Corpus. These corpora are advanced linguistic resources designed for modern linguistic research and represent complex systems that integrate various methodological approaches. In these corpora, speech data are collected through web-based platforms, where audio and video recordings are fully integrated with transcriptions. This approach enables the analysis of language data in a multimodal format. Based on these experiences, the study identifies key methodological principles that can be applied to the development of an Uzbek dialectological corpus.

Keywords: dialectological corpus, corpus linguistics, Nordic Dialect Corpus, ManDi Corpus, Uzbek language, multimodal corpus, remote data collection, phonetic analysis, transcription, linguistic resources, NLP, sociolinguistics.

Аннотация

В данной статье анализируется международный опыт создания диалектологических корпусов, в частности Нордического диалектного корпуса и корпуса ManDi. Эти корпуса представляют собой современные лингвистические ресурсы, предназначенные для научных исследований в области языкознания, и являются сложными системами, объединяющими различные методологические подходы. В данных корпусах речевой материал собирается через веб-платформы, при этом аудио- и видеозаписи интегрируются с транскрипциями. Такой подход позволяет анализировать языковые данные в мультимодальном формате. На основе данного опыта определяются ключевые методологические принципы, которые могут быть использованы при создании узбекского диалектологического корпуса.

Ключевые слова: диалектологический корпус, корпусная лингвистика, Нордический диалектный корпус, корпус ManDi, узбекский язык, мультимодальный корпус, удалённый сбор данных, фонетический анализ, транскрипция, лингвистические ресурсы, NLP, социолингвистика.

Kirish

So‘nggi yillarda tilshunoslikda empirik ma’lumotlarga asoslangan tadqiqotlarga bo‘lgan ehtiyoj keskin ortib bormoqda. Ayniqsa, korpus lingvistikasi yo‘nalishining rivojlanishi natijasida til birliklarini real nutq materiallari asosida o‘rganish imkoniyati kengaydi. Bu jarayon dialektologiya sohasiga ham bevosita ta’sir ko‘rsatib, hududiy til variantlarini tizimli ravishda yig‘ish, saqlash va tahlil qilish uchun yangi metodologik asoslarni yuzaga keltirdi. Dialektlar tilning dinamik va o‘zgaruvchan qatlamini tashkil etgani sababli ularni faqat yozma manbalar asosida o‘rganish yetarli emas, balki tabiiy og‘zaki nutq namunalariga tayanish zarur hisoblanadi.

Jahon tilshunosligida dialektlarga oid elektron korpuslar yaratish bo‘yicha sezilarli tajriba to‘plangan. Xususan, Shimoliy Yevropa hududida shakllangan Nordik dialekt korpusi hamda Xitoy tilining mintaqaviy variantlarini qamrab olgan ManDi korpusi bu boradagi ilg‘or loyihalar sirasiga kiradi. Mazkur korpuslar o‘zining tuzilishi va qo‘llanilgan metodlari bilan bir-biridan farq qiladi: bir tomondan, Nordik korpusda spontan nutq, multimodal materiallar (audio va video) hamda keng qidiruv tizimi asosiy o‘rin tutsa, ikkinchi tomondan, ManDi korpusida masofaviy tarzda yig‘ilgan audio yozuvlar va akustik-fonetik tahlil metodlari ustuvor hisoblanadi. Har ikkala yondashuv ham dialektlarni ilmiy asosda o‘rganishda muhim natijalar berayotganini ko‘rsatadi.

Shu bilan birga, o‘zbek tilshunosligida dialektologik tadqiqotlar ancha boy tarixga ega bo‘lishiga qaramay, ularni zamonaviy raqamli korpus shaklida jamlash va

tizimlashtirish masalasi yetarli darajada hal etilmagan. Turli hududlarga xos nutq namunalarini yagona platformada to‘plash, ularni fonetik, leksik va grammatik jihatdan belgilash hamda qiyosiy tahlil qilish imkonini beruvchi integrallashgan tizimning yo‘qligi ushbu yo‘nalishda ilmiy izlanishlarni cheklab kelmoqda. Bu holat esa zamonaviy texnologiyalarga asoslangan o‘zbek dialektologik korpusini yaratish zaruratini yanada dolzarblashtiradi.

Mazkur maqolaning asosiy maqsadi xorijiy tajribani tahlil qilish orqali dialekt korpuslarini yaratishda qo‘llanilayotgan samarali metod va usullarni aniqlashdan iborat. Shu maqsadda Nordik dialekt korpusi va ManDi korpusining tuzilishi, ma’lumot yig‘ish jarayoni, annotatsiya tizimi hamda tahlil imkoniyatlari o‘rganiladi, ularning o‘xshash va farqli jihatlari ilmiy asosda solishtiriladi. Tadqiqot natijasida o‘zbek dialektologik korpusini yaratishda qo‘llash mumkin bo‘lgan amaliy tavsiyalar ishlab chiqiladi.

Adabiyotlar sharhi

Zamonaviy lingvistik tadqiqotlarda korpus texnologiyalari tilni tizimli o‘rganishning eng muhim vositalaridan biri sifatida e’tirof etilmoqda. Ayniqsa, dialektologik korpuslar hududiy til variantlarini ilmiy asosda tahlil qilish, ularning fonetik, leksik va grammatik xususiyatlarini aniqlash hamda til o‘zgarish jarayonlarini kuzatishda katta ahamiyat kasb etadi. So‘nggi yillarda turli tillar uchun yaratilgan dialekt korpuslari ushbu yo‘nalishning jadal rivojlanayotganini ko‘rsatadi.

So‘nggi yillarda o‘zbek tilshunosligida ham korpus lingvistikasi jadal rivojlanayotgan istiqbolli yo‘nalishlardan biri bo‘lib, ushbu soha doirasida qator ilmiy tadqiqotlar amalga oshirilgan. Xususan, A. Po‘latov, S. Muhamedova, A. Rahimov, Z. Xolmanova hamda N. Abduraxmonova kabi olimlar tomonidan mazkur soha doirasida darsliklar, qo‘llanmalar va monografik tadqiqotlar yaratildi. Ushbu ilmiy ishlarda tilni kompyuter modellashtirish, lingvistik dasturiy ta’minot ishlab chiqish hamda kompyuter texnologiyalari asosida lingvistik muammolarni hal etish masalalari yoritilgan. Xususan, O‘zbekiston Milliy Universiteti professori N. Z. Abdurahmonova o‘zbek tili elektron korpusini yaratish g‘oyasini ilgari surib, uning konseptual asoslarini ishlab chiqdi hamda ushbu yo‘nalishda amaliy loyihani yo‘lga qo‘ydi[1]. Shu bilan birga, ToshDO‘TAU olimasi Sh. Hamroyeva mualliflik korpusini yaratish mezonlarini tadqiq etib, milliy korpus shakllanishiga muhim hissa qo‘shdi[2]. Tadqiqotchilar o‘z ishlarida

1Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (содда гаплар мисолида). филол.фан.бўйича фалсафа доктори (PhD)…дисс. – Тошкент, 2018. –165 б.

2Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: филол. Фанлари бўйича

milliy korpusni yaratish tilni raqamlashtirish, katta hajmdagi matnlarni tizimlashtirish hamda zamonaviy til holatini aks ettiruvchi elektron baza shakllantirish uchun zarur omil ekanini ta’kidlaydilar. Shuningdek, Qarshiyev, Tursunov va Maxmidov (2022) tadqiqotida o‘zbek tili milliy korpusini yaratishning strukturaviy, lingvistik va texnologik tamoyillari yoritilib, uni multidisiplinar yondashuv asosida shakllantirish, jumladan, dialektal materiallarni ham korpus tarkibiga kiritish zarurligi alohida qayd etilgan.

Mazkur yo‘nalishda amalga oshirilgan ilmiy tadqiqotlar til birliklarini zamonaviy texnologiyalar asosida tahlil qilish, lingvistik resurslarni yaratish hamda tilni raqamlashtirish jarayonlarini rivojlantirishga xizmat qilmoqda. Biroq o‘zbek tilshunosligida dialektologik korpuslarni yaratish masalasi hanuzgacha yetarli darajada o‘rganilmagan va kam tadqiq etilgan yo‘nalishlardan biri hisoblanadi. Shu sababli, xorijiy tajribalarni chuqur tahlil qilish hamda ularning samarali metod va tamoyillarini o‘zbek dialektologik korpusini shakllantirish jarayoniga tatbiq etish ilmiy jihatdan ham, amaliy jihatdan ham maqsadga muvofiq bo‘lib, ushbu yo‘nalishning dolzarbligini yanada asoslaydi. Ushbu tadqiqot doirasida Nordik Dialekt Korpusi hamda ManDi korpusi tajribalari tahlil qilinadi va ularning metodologik yondashuvlari o‘zbek dialektologik korpusini yaratish nuqtayi nazaridan ko‘rib chiqiladi.

Nordik mamlakatlarda yaratilgan dialekt korpusi ushbu sohadagi eng mukammal loyihalardan biri hisoblanadi. Mazkur korpus Daniya, Norvegiya, Shvetsiya, Islandiya va Farer orollari kabi hududlarda so‘zlashiladigan shimoliy german tillari dialektlarini qamrab oladi. Uning asosiy xususiyati shundaki, u nafaqat matn shaklidagi ma’lumotlarni, balki audio va video yozuvlarni ham o‘z ichiga olgan multimodal tizim sifatida tashkil etilgan. Bunday yondashuv dialekt nutqining real fonetik va kommunikativ xususiyatlarini saqlab qolish imkonini beradi.

Mazkur korpusda ikki xil transkripsiya tizimi — orfografik va fonetik shakllar qo‘llanilishi alohida e’tiborga loyiqdir. Bu esa tadqiqotchilarga bir vaqtning o‘zida ham standart yozuv asosida, ham real talaffuz darajasida izlanish olib borish imkonini yaratadi. Bundan tashqari, korpusda morfosintaktik belgilash (annotatsiya), kengaytirilgan qidiruv interfeysi, statistik tahlil vositalari va foydalanuvchi uchun qulay vizualizatsiya imkoniyatlari mavjud. Shu jihatlari bilan u nafaqat dialektologiya, balki umumiy tilshunoslik, sotsiolingvistika va kompyuter lingvistikasida ham keng qo‘llanilmoqda.

Xitoy tilshunosligida esa dialektlarni o‘rganish uchun yaratilgan ManDi korpusi

фалсафа докт. дисс. – Бухоро, 2018. – 165 б.

alohida e’tiborga loyiqdir. Ushbu korpus mandarin tilining oltita hududiy varianti hamda standart mandarin asosida shakllantirilgan bo‘lib, asosan fonetik va akustik tahlillarga yo‘naltirilgan. Korpusning o‘ziga xos jihati shundaki, undagi nutq ma’lumotlari an’anaviy laboratoriya sharoitida emas, balki masofaviy tarzda — ishtirokchilarning shaxsiy smartfonlari orqali yig‘ilgan.

Masofaviy ma’lumot yig‘ish usuli qisqa vaqt ichida keng geografik hududni qamrab olish imkonini beradi va turli ijtimoiy guruhlardan nutq namunalarini olishni osonlashtiradi. Shu bilan birga, ManDi korpusida nutq materiallari maxsus tanlangan o‘qilgan matnlar asosida yig‘ilgan bo‘lib, bu fonetik parametrlarni aniq nazorat qilish va solishtirish imkonini beradi. Korpusda so‘z va fonema darajasida segmentatsiya ishlari maxsus dasturlar yordamida amalga oshirilgan bo‘lib, bu uning akustik tahlil uchun yuqori darajada moslashganini ko‘rsatadi.

Boshqa tillarda ham dialekt korpuslarini yaratish bo‘yicha bir qator loyihalar mavjud bo‘lsa-da, ularning aksariyati cheklangan funksionallikka ega. Masalan, ayrim korpuslarda faqat audio yozuvlar mavjud bo‘lib, lingvistik annotatsiya yetarli darajada rivojlanmagan. Ba’zilarida esa qidiruv tizimi sodda bo‘lib, murakkab lingvistik tahlillarni amalga oshirish imkoniyati cheklangan. Shu sababli, Nordik dialekt korpusi va ManDi korpusi zamonaviy talablar asosida yaratilgan eng ilg‘or tizimlar sifatida alohida ajralib turadi.

Yuqoridagi adabiyotlar tahlili shuni ko‘rsatadiki, zamonaviy dialektologik korpuslar yaratishda multimodal yondashuv, masofaviy ma’lumot yig‘ish, avtomatik annotatsiya va rivojlangan qidiruv tizimlari muhim o‘rin tutadi. Mazkur tajribalar o‘zbek dialektologik korpusini yaratishda metodologik asos sifatida xizmat qilishi mumkin.

Tadqiqot metodologiyasi

Mazkur tadqiqotda dialektologik korpuslarni qiyosiy o‘rganish orqali o‘zbek dialektologik korpusini yaratish uchun samarali metod va yondashuvlarni aniqlash maqsad qilib qo‘yildi. Tadqiqot jarayonida sifat jihatidan tahlil (qualitative analysis) va qiyosiy metod (comparative method) asosiy ilmiy usullar sifatida qo‘llanildi.

Tadqiqot obyekti sifatida ikki zamonaviy dialektologik korpus — Nordik dialekt korpusi va ManDi korpusi tanlab olindi. Ushbu korpuslar turli lingvistik maktablar va texnologik yondashuvlarni aks ettirgani sababli ular o‘zaro qiyoslash uchun mos namunalar hisoblanadi. Tadqiqot predmeti esa mazkur korpuslarda qo‘llanilgan metodlar, texnologiyalar va tashkiliy tamoyillarni tahlil qilishdan iborat bo‘ldi.

Tahlil jarayonida korpuslar bir xil mezonlar asosida o‘rganildi. Xususan, quyidagi parametrlar asosiy tahlil mezonlari sifatida belgilandi: ma’lumot yig‘ish modeli, nutq materialining turi, transkripsiya va annotatsiya tizimi, texnologik vositalar, qidiruv va

qayta ishlash imkoniyatlari hamda foydalanuvchi interfeysi. Har bir mezon bo‘yicha korpuslarning funksional imkoniyatlari alohida baholandi va ularning ustun hamda zaif jihatlari aniqlashtirildi.

Tadqiqot davomida ma’lumotlar ochiq ilmiy manbalardan, jumladan, konferensiya maqolalari va korpus tavsiflariga oid rasmiy materiallardan olindi. Ushbu ma’lumotlar tizimlashtirilib, konseptual tahlil asosida umumlashtirildi. Tahlil natijalari asosida korpuslar o‘rtasidagi o‘xshashlik va farqlar aniqlanib, ularning ilmiy ahamiyati baholandi.

Metodologik yondashuvning muhim jihatlaridan biri sifatida induktiv umumlashtirish usuli qo‘llanildi. Ya’ni, alohida korpuslarga xos xususiyatlar umumiy ilmiy xulosalarga keltirildi va ular asosida o‘zbek dialektologik korpusini yaratish bo‘yicha tavsiyalar shakllantirildi. Bu yondashuv mavjud tajribani to‘g‘ridan-to‘g‘ri ko‘chirish emas, balki uni mahalliy til xususiyatlariga moslashtirish imkonini beradi. Shuningdek, tadqiqotda tizimli yondashuv (system approach) ham qo‘llanildi. Bunda korpuslar faqat ma’lumotlar to‘plami sifatida emas, balki o‘zaro bog‘liq komponentlardan tashkil topgan murakkab tizim sifatida ko‘rib chiqildi. Natijada, har bir elementning (ma’lumot yig‘ish, annotatsiya, qidiruv, tahlil) umumiy tizimdagi o‘rni aniqlashtirildi. Mazkur metodologiya tadqiqotning izchilligi va ilmiy asoslanganligini ta’minlab, olingan natijalar asosida amaliy tavsiyalar ishlab chiqish imkonini berdi.

Tahlil va muhokama

Nordik dialekt korpusi va ManDi korpusi zamonaviy dialektologik tadqiqotlarda qo‘llanilayotgan ikki turli, ammo o‘zaro to‘ldiruvchi metodologik yondashuvni ifodalaydi. Ularning har biri til materiallarini yig‘ish, qayta ishlash va tahlil qilish jarayonida o‘ziga xos ustunliklarga ega bo‘lib, mazkur jihatlar ularni qiyosiy o‘rganishni dolzarb qiladi.

Avvalo, ma’lumot yig‘ish usullariga e’tibor qaratilsa, Nordik dialekt korpusida asosan an’anaviy dala tadqiqotlari asosida to‘plangan spontan nutq materiallari ustunlik qiladi. Bu yondashuv tilning tabiiy qo‘llanishini aks ettirish imkonini beradi hamda sotsiolingvistik omillarni hisobga olish uchun qulay sharoit yaratadi. Aksincha, ManDi korpusida nutq ma’lumotlari masofaviy tarzda, ya’ni ishtirokchilarning shaxsiy smartfonlari orqali yig‘ilgan. Ushbu usul qisqa vaqt ichida keng geografik hududni qamrab olish imkonini berishi bilan ajralib turadi. Shunga qaramay, masofaviy yozuvlarda tashqi shovqin va texnik farqlar sababli ma’lumot sifati notekis bo‘lishi mumkinligi kuzatiladi.

Nutq materialining tabiati jihatidan ham sezilarli farqlar mavjud. Nordik korpusda spontan, ya’ni erkin suhbat nutqi asosiy o‘rinni egallasa, ManDi korpusi maxsus

tanlangan o‘qilgan materiallarga tayangan. Spontan nutq tilning real kommunikativ xususiyatlarini ochib berishda muhim bo‘lsa, o‘qilgan nutq fonetik va eksperimental tahlillar uchun zarur bo‘lgan nazoratni ta’minlaydi. Demak, har ikkala yondashuv turli ilmiy maqsadlar uchun samarali hisoblanadi.

Transkripsiya va annotatsiya tizimlari ham ushbu korpuslarning muhim farqlovchi jihatlaridan biridir. Nordik dialekt korpusida ikki bosqichli transkripsiya — ortografik va fonetik shakllar qo‘llanilishi tadqiqotchilarga dialektal xususiyatlarni chuqurroq tahlil qilish imkonini beradi. Bundan tashqari, unda grammatik belgilash (POS tagging) tizimi mavjud bo‘lib, morfosintaktik tadqiqotlar uchun keng imkoniyat yaratadi. ManDi korpusida esa asosiy e’tibor fonetik segmentatsiya va akustik tahlilga qaratilgan bo‘lib, Praat va Montreal Forced Aligner dasturlari yordamida so‘z va fonema darajasida aniqlik bilan belgilash amalga oshirilgan. Bu esa ayniqsa, tonal tizimlarni o‘rganishda yuqori darajadagi aniqlikni ta’minlaydi.

Qidiruv va natijalarni qayta ishlash imkoniyatlari nuqtai nazaridan ham Nordik korpus ancha rivojlangan tizimga ega. Unda foydalanuvchi lingvistik birliklarni turli mezonlar asosida izlash, natijalarni statistik jihatdan tahlil qilish va vizual shaklda ko‘rish imkoniyatiga ega. ManDi korpusida esa asosiy urg‘u ma’lumot yig‘ish va fonetik tahlilga berilgan bo‘lib, keng qamrovli qidiruv interfeysi yetarli darajada rivojlanmagan.

Shuningdek, ikkala korpusning ilmiy yo‘nalishi ham farqlanadi. Nordik dialekt korpusi ko‘proq umumiy lingvistik, xususan, sintaktik va leksik tadqiqotlarga mo‘ljallangan bo‘lsa, ManDi korpusi fonetik va akustik tahlillarga ixtisoslashgan. Bu esa ularning metodologik asoslari turlicha shakllanganini ko‘rsatadi.

Yuqoridagi tahlillar shuni ko‘rsatadiki, har ikkala korpus alohida holda muayyan ilmiy vazifalarni muvaffaqiyatli bajaradi. Biroq ularning kuchli tomonlarini integratsiya qilish orqali yanada mukammal dialektologik korpus yaratish mumkin. Xususan, Nordik korpusning multimodal va qidiruvga yo‘naltirilgan tizimi hamda ManDi korpusining masofaviy ma’lumot yig‘ish va akustik tahlil metodlarini birlashtirish istiqbolli natijalarga olib keladi.

O‘zbek dialektologik korpusini yaratishning konseptual modeli

Dunyo tilshunosligida dialektologik korpuslarni yaratish bo‘yicha shakllangan tajribalar shuni ko‘rsatadiki, zamonaviy korpus tizimlari oddiy ma’lumotlar yig‘indisi sifatida emas, balki ko‘p qatlamli lingvistik infratuzilma sifatida qaraladi. Xususan, Nordik dialekt korpusi hamda ManDi korpusining ilmiy-tashkiliy yechimlari tahlili asosida o‘zbek dialektologik korpusini yaratishda qo‘llanishi mumkin bo‘lgan bir qator ustuvor metodologik yo‘nalishlarni ajratib ko‘rsatish mumkin.

1. Multimodal korpus yaratish tamoyili

Zamonaviy dialektologik tadqiqotlar nutqni faqat matn ko‘rinishida qayd etish bilan cheklanib qolmasdan, uni audio va video shakllarda kompleks tarzda hujjatlashtirishni talab qiladi. Shu bois o‘zbek dialektologik korpusini yaratishda multimodal yondashuvni asosiy prinsip sifatida qabul qilish maqsadga muvofiqdir. Bunday yondashuv dialektal nutqning fonetik, prosodik va paralingvistik xususiyatlarini to‘liq aks ettirish imkonini beradi. Shuningdek, nutq jarayonidagi imo-ishoralar, kommunikativ kontekst va ijtimoiy omillarni ham tahlil qilish imkoniyati kengayadi. Nordik korpus tajribasi multimodal integratsiya lingvistik tahlil sifatini sezilarli darajada oshirishini ko‘rsatadi.

2. Masofaviy ma’lumot yig‘ish tizimini joriy etish

Zamonaviy texnologiyalar rivoji dialekt materiallarini an’anaviy dala tadqiqotlaridan tashqari masofaviy usulda yig‘ish imkonini bermoqda. ManDi korpusi tajribasi asosida o‘zbek dialektologik korpusini yaratishda masofaviy (remote) ma’lumot yig‘ish modelini qo‘llash samarali hisoblanadi. Ushbu yondashuv respublikaning turli hududlaridan keng ko‘lamli nutq materiallarini qisqa vaqt ichida yig‘ish, moliyaviy va tashkiliy resurslarni tejash hamda turli ijtimoiy guruhlarni qamrab olish imkonini beradi. Mazkur tizimni amalga oshirish uchun maxsus mobil ilova yoki veb-platforma ishlab chiqilishi zarur bo‘lib, unda ishtirokchilar o‘z nutqlarini mustaqil ravishda yozib, markazlashgan bazaga yuborishlari mumkin bo‘ladi.

3. Standart va dialekt nutqni parallel yig‘ish modeli

Dialektologik korpusning ilmiy qiymatini oshirish maqsadida nutq materiallarini ikki parallel shaklda yig‘ish tavsiya etiladi: adabiy o‘zbek tili va hududiy dialekt varianti. Ushbu yondashuv dialekt va standart til o‘rtasidagi fonetik, leksik va grammatik tafovutlarni tizimli ravishda qiyosiy tahlil qilish imkonini beradi. Bundan tashqari, tilning ichki rivojlanish jarayonlarini, dialektlarning standartlashuvga ta’sirini hamda lingvistik o‘zgarish dinamikasini aniqlash uchun muhim empirik asos yaratadi. ManDi korpusida qo‘llangan ushbu model o‘zining yuqori ilmiy samaradorligini namoyon etgan.

4. Ikki bosqichli transkripsiya tizimini joriy etish

Dialekt nutqini to‘liq va aniq ifodalash uchun ikki qatlamli transkripsiya tizimini joriy etish zarur: ortografik va fonetik transkripsiya. Ortografik qatlam matnni umumiy foydalanish uchun qulay shaklda taqdim etsa, fonetik transkripsiya real talaffuz xususiyatlarini aks ettiradi. Bunday yondashuv hududiy talaffuz farqlarini aniqlash, fonologik tizimni chuqur o‘rganish hamda avtomatlashtirilgan lingvistik tahlilni rivojlantirish uchun muhim asos bo‘lib xizmat qiladi. Nordik korpus tajribasi ushbu

modelning amaliy jihatdan samarali ekanini tasdiqlaydi.

5. Avtomatik annotatsiya va lingvistik belgilash tizimi

Zamonaviy korpus lingvistikasi ma’lumotlarni avtomatik qayta ishlash va tahlil qilish imkoniyatlarini kengaytirishni talab qiladi. Shu sababli o‘zbek dialektologik korpusida so‘z turkumlarini aniqlash (POS tagging), morfologik tahlil, fonetik segmentatsiya va sintaktik belgilash kabi annotatsiya qatlamlarini joriy etish zarur. Bunda tabiiy tilni qayta ishlash (NLP) texnologiyalari hamda mashinaviy o‘rganish algoritmlaridan foydalanish korpusning funksional imkoniyatlarini sezilarli darajada kengaytiradi.

6. Akustik-fonetik tahlil modulini integratsiya qilish

Dialektlarni chuqur o‘rganishda akustik parametrlar muhim ahamiyat kasb etadi. ManDi korpusi tajribasi asosida o‘zbek korpusida F0 (asosiy chastota), intonatsiya va urg‘u kabi ko‘rsatkichlarni tahlil qilish imkonini beruvchi maxsus modul joriy etilishi lozim. Praat kabi dasturlar integratsiyasi hamda avtomatik segmentatsiya vositalari yordamida nutqning akustik xususiyatlarini aniq va tizimli tahlil qilish imkoniyati yaratiladi.

7. Qidiruv va statistik tahlil tizimini ishlab chiqish

Korpusdan samarali foydalanish uchun qulay va funksional qidiruv interfeysi muhim ahamiyatga ega. Nordik korpus tajribasiga tayangan holda lemma, affiks, fonetik shakl va geografik parametrlar asosida qidiruv imkoniyatlarini joriy etish tavsiya etiladi. Shuningdek, natijalarni statistik jihatdan qayta ishlash, chastota tahlili, kollokatsiyalarni aniqlash va grafik vizualizatsiya vositalarini qo‘llash ilmiy tadqiqotlarning samaradorligini oshiradi.

8. Informantlar bazasini tizimli shakllantirish

Dialektologik tadqiqotlarda sotsiolingvistik omillar muhim rol o‘ynaydi. Shu sababli har bir yozuv bilan birga informantga oid demografik va ijtimoiy ma’lumotlarni yig‘ish zarur. Jumladan, yosh, jins, ta’lim darajasi, yashash hududi va tilni qo‘llash chastotasi kabi ko‘rsatkichlar keyingi tahlillar uchun muhim empirik asos yaratadi.

9. Ma’lumotlarni standartlashtirish va xavfsizlikni ta’minlash

Korpusning ilmiy ishonchliligini ta’minlash uchun ma’lumotlarni yig‘ish va saqlash jarayonida yagona standartlarga amal qilish zarur. Audio yozuvlar uchun bir xil texnik parametrlar (masalan, WAV formati, 44.1 kHz)ni belgilash, fayl nomlash tizimini standartlashtirish, shaxsiy ma’lumotlarni anonimlashtirish va xavfsiz serverlarda saqlash muhim talablar hisoblanadi.

10. Bosqichma-bosqich rivojlantirish strategiyasi

Dialektologik korpusni yaratish uzluksiz va bosqichma-bosqich amalga oshiriladigan jarayon sifatida qaralishi lozim. Dastlab pilot loyiha asosida kichik hudud qamrab

olinadi, keyinchalik esa regional kengayish orqali ma’lumotlar bazasi boyitiladi. Yakuniy bosqichda to‘liq milliy korpus shakllantirilib, ilmiy platformalar bilan integratsiya qilinadi. Bunday strategiya tizimning barqaror rivojlanishini ta’minlaydi.

Tahlillar shuni ko‘rsatadiki, zamonaviy dialektologik korpuslar ko‘p qatlamli va texnologik jihatdan rivojlangan tizim sifatida shakllanadi. Nordik dialekt korpusi multimodal va qidiruvga yo‘naltirilgan yondashuvni ifodalasa, ManDi korpusi masofaviy ma’lumot yig‘ish hamda akustik-fonetik tahlil imkoniyatlarini namoyon etadi. Ushbu ikki modelning integratsiyasi o‘zbek dialektologik korpusini yaratishda ilmiy asoslangan, samarali va istiqbolli yo‘nalish sifatida qaralishi mumkin.

Xulosa

Xulosa qilib aytganda, zamonaviy tilshunoslikda dialektologik korpuslar tilning hududiy variativligini tizimli o‘rganish, lingvistik ma’lumotlarni raqamlashtirish hamda avtomatlashtirilgan tahlil jarayonlarini rivojlantirishda muhim ilmiy vosita sifatida xizmat qilmoqda. Nordik Dialekt Korpusi va ManDi korpusi tajribalari shuni ko‘rsatadiki, multimodal ma’lumot yig‘ish, masofaviy yozuvlar olish, standart va dialekt nutqni parallel korpus shaklida tashkil etish hamda avtomatik annotatsiya tizimlaridan foydalanish zamonaviy korpus lingvistikasining asosiy metodologik tamoyillarini tashkil etadi.

Ushbu tajribalar tahlili o‘zbek dialektologik korpusini yaratishda ham ilmiy va amaliy jihatdan samarali yondashuvlarni aniqlash imkonini beradi. Xususan, multimodal korpus tuzilishi, masofaviy ma’lumot yig‘ish tizimi, ikki bosqichli transkripsiya va akustik-fonetik tahlil modullarini joriy etish kabi yondashuvlar o‘zbek dialekt materiallarini tizimli o‘rganish uchun muhim asos bo‘lib xizmat qiladi. Shu bilan birga, informantlarga oid sotsiolingvistik ma’lumotlarni integratsiya qilish korpusning ilmiy qiymatini yanada oshiradi.

Demak, xorijiy tajribalarni o‘rganish va ularning samarali metodologik yechimlarini moslashtirish o‘zbek dialektologik korpusini yaratishda eng muhim bosqichlardan biri hisoblanadi hamda bu yo‘nalish kelajakdagi lingvistik tadqiqotlar uchun keng ilmiy istiqbollarni ochib beradi.

Foydalanilgan adabiyotlar:

1. Абдурахмонова, Н. З. (2018). *Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (содда гаплар мисолида)*. Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси. Тошкент. – 165 б.

2. Hagen, K., Johannessen, J. B., & Nøklestad, A. (2000). *A Constraint-based Tagger for Norwegian*. In C. E. Lindberg & S. N. Lund (Eds.), 17th Scandinavian Conference of Linguistics, Odense Working Papers in Language and Communication 19, 31–48.
3. Ho, D. A. (2003). *The characteristics of Mandarin dialects*. In **The Sino-Tibetan languages** (pp. 126–130).
4. Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). *Perception and production of Mandarin Chinese tones*. (pp. 209–217).
5. Johannessen, J. B., Hagen, K., Priestley, J., & Nygaard, L. (2007). *An Advanced Speech Corpus for Norwegian*. In NODALIDA Proceedings, 29–36.
6. Nivre, J., & Grönqvist, L. (2001). *Tagging a Corpus of Spoken Swedish*. International Journal of Corpus Linguistics, 6(1), 47–78.
7. Хамроева, Ш. (2018). *Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари*. Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси. Бухоро. – 165 б.