

DIALEKTAL KORPUSNING ARXITEKTURASI VA MA'LUMOTLAR BAZASI

Muftillayeva Ruxsora Toshmuhammad qizi

ruxsoramuftillayeva851@gmail.com

ToshDO'TAU magistranti

Annotatsiya. Dialektal korpus bu tilning turli mintaqaviy yoki ijtimoiy turlarini ifodalovchi matnlar va og'zaki til namunalari to'plami. U lingvistik tadqiqotlar, til hujjatlari va madaniy merosni saqlash uchun muhim manba hisoblanadi. Dialektal korpusni boshqarish va tahlil qilish uchun yaxshi ishlangan arxitektura, ma'lumotlar bazasi tizimi zarur. Dialektal yozuvlari to'plami keng ko'lamli fonologik, morfologik, leksik va sintaktik tadqiqotlar uchun ham qimmatli ma'lumotlar manbayi hisoblanadi. Ushbu maqolada biz dialektal korpus uchun arxitektura va ma'lumotlar bazasi talablarini o'rganamiz. Bu xususiyatlar quyidagilarni o'z ichiga oladi: ma'lumot to'plash, ma'lumotlarni qayta ishlash, saqlash, qidiruv tizimi, xavfsizlik va maxfiylik, ma'lumot beruvchi tafsilotlarini ko'rsatish (so'zlar soni va axborot beruvchilar haqidagi boshqa ma'lumotlar).

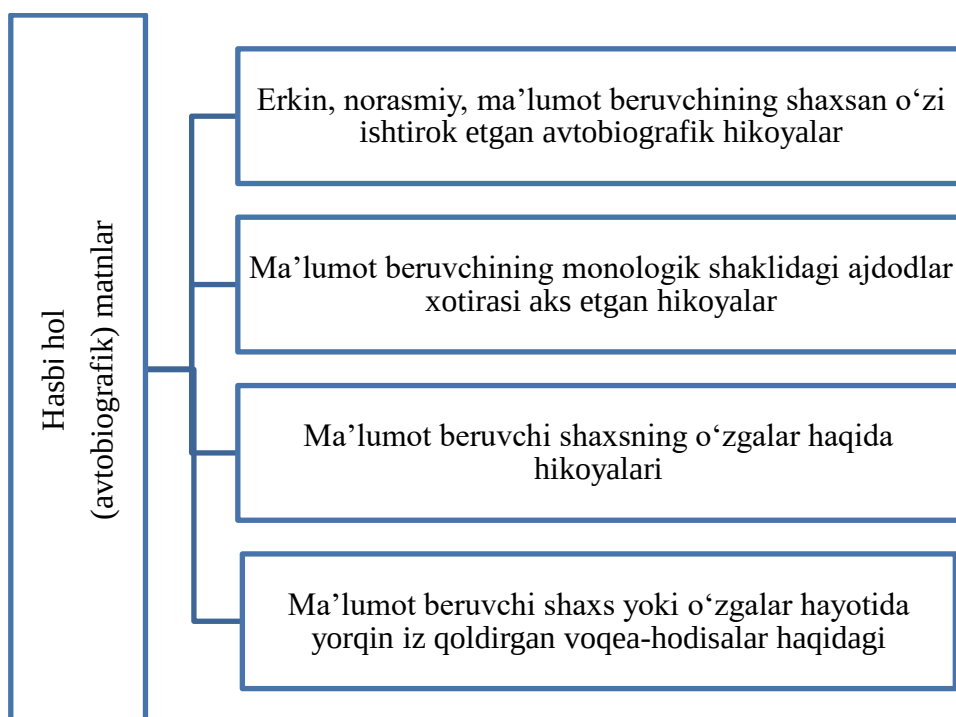
Abstract. A dialectal corpus is a collection of texts and spoken language samples representing different regional or social varieties of a language. It is an important resource for linguistic research, language documentation and cultural heritage preservation. A well-designed architecture and database system is necessary for the management and analysis of the dialectal corpus. The collection of dialect records is also a valuable source of information for large-scale phonological, morphological, lexical and syntactic studies. In this article, we explore the architecture and database requirements for a dialectal corpus. These features include: data collection, data processing, storage, search engine, security and privacy, disclosure of informant details (word count and other information about informants 'data).

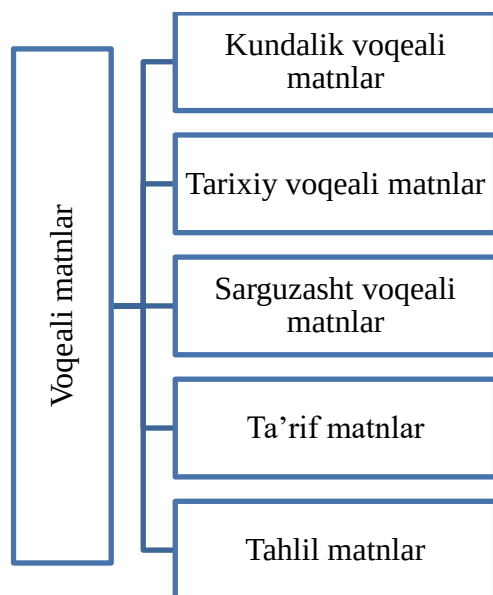
Аннотация. Диалектный корпус — это совокупность текстов и образцов разговорного языка, представляющих различные региональные или социальные разновидности языка. Это важный ресурс для лингвистических исследований, языковой документации и сохранения культурного наследия. Для управления и анализа диалектного корпуса необходимы хорошо продуманная архитектура и система баз данных. Коллекция диалектных записей является также ценным источником информации для масштабных фонологических, морфологических, лексических и синтаксических исследований. В этой статье мы исследуем требования к архитектуре и базе данных для диалектного корпуса. Эти функции включают в себя: сбор данных, обработку данных, хранение, поисковую систему, безопасность и конфиденциальность, раскрытие данных информанта (количество слов и другая информация о данных информаторов).

Kalit soʻzlar: *dialektal korpus, transkripsiya, dialekt, integratsiya, qidiruv tizimi.*

Dialektal korpus tizimining arxitekturasini dialektal ma'lumotlarni saqlash, olish, tahlil qilish uchun birgalikda ishlaydigan turli komponentlar va qatlamlarni o'z ichiga oladi. Shevalar korpusining texnik tomoniga ishlov berish uchun juda katta mablag' va ilmiy salohiyatga ega guruh shakllantirish eng asosiy vazifa. Dialektal korpus yaratish jarayoni uzoq va qiyin mehnatni talab etishini hisobga olgan holda korpusning dasturiy ta'minotini ishlab chiquvchi dasturchi texniklar guruhini shakllantirish lozim bo'ladi. Birinchi navbatda korpusdan foydalanish qulay va ommaviy bo'lishini ta'minlash kerak. Dialektal korpus arxitekturasining ba'zi asosiy komponentlari:

Ma'lumot to'plash. Dialektal korpusni qurishda birinchi qadam turli dialekt mintaqalari yoki hududlardan ma'lumotlarni yig'ishdir. Bunda asosan intervyu manbalardan foydalaniladi. Qishloq aholisi o'rtasidagi muloqatlar yozib olinadi. Intervyular dialog yoki polilog shaklida bo'ladi. Respondentlarning asosiy qismi 55 yoshdan oshgan odamlar. Tadqiqot uchun ataylab o'rta ma'lumotga ega respondentlar tanlanadi, chunki aynan ularning nutqida dialektal xususiyatlar yaxshiroq saqlangan bo'ladi. Har bir matn uchun “pasport” ma'lumotlari yozuvning yili, joyini va respondent haqida ma'lumotlarni o'z ichiga oladi. Yozilgan dialekt ma'lumotlarining mavjudligi lingvistik va tarixiy tadqiqotlar uchun bebaho manba bo'lib xizmat qiladi. Ma'lumot to'plash suhbat shaklida o'tkaziladi.





1.1-jadval. Sheva suhbat janrining ichki qismi

Dialektal korpus tuzishda ishtirok etgan respondentlar haqida ma'lumotlar beriladi.

T /p	So'rovnoma	Javoblar
1	Yozib olingan joy (viloyat, tuman)	Yangihayot mahallasi, Nurota tumani, Navoiy viloyati
2	Yozib olingan vaqt (yil, kun, oy)	2023- yil 23- avgust
3	Axborot beruvchi (familiya, ism, otasining ismi)	Muftillayeva Ruxsora Toshmuhammad qizi
4	Axborot beruvchining jinsi (ayol, erkak)	Ayol
5	Axborot beruvchining tug'ilgan joyi va yili	Nurota tumani, Yangihayot mahallasi
6	Axborot beruvchining ma'lumoti (o'rta, oliy)	O'rta ma'lumotli
7	Axborot beruvchining kasbi	Uy bekasi
8	Ma'lumot yig'uvchi (familiya, ism, otasining ismi)	Muftillayeva Ruxsora Toshmuhammad qizi
9	Ish olib borilayotgan muassasa	Toshkent davlat o'zbek tili va adabiyoti unversiteti

10	Matn shakli	Og‘zaki
11	Janr	Suhbat
12	Yozuv turi (qo‘lyozma, audio, kamera, diktafon, fotoapparat)	Diktafon, kamera
13	Mavzu	Dehqonchilik
14	Voqealik davri Mustaqillikka qadar Mustaqillikdan keyin	Mustaqillikdan keyin

1.2-jadval. Shevaning pasport metarazmetkasi

Suhbat matnining transkripsiyasi. Yuqoridagi so‘rovlar asosida qidiruv o‘tkazilsa, quyidagicha natijaga erishiladi: Dehqonchilik bilan bog‘liq suhbat. Matn haqidagi ma‘lumotlar aks etgan matn metarazmetkasi.

Bu yozuv (diktafon, kamera) 2023-yil 23-avgustda Yangihayot mahallasida dehqonchilik bilan shug‘ullangan Muftillayeva Ruxsora Toshmuhammad qizidan yozib olindi; nutq maqsadi: o‘tgan kunlarni eslash, kasbiy bilimlar berish orqali shevaga xos so‘zlarni aniqlash va tadqiq etish; nutq jarayonida tilga olinadigan shaxslar; turmush o‘rtog‘i, qaynota, onasi, dadasi; voqea-hodisaning mustaqillikdan oldin yoki keyin ro‘y berganligi; voqea-hodisaning mustaqillikdan oldin 1954-yillarda, kelin bo‘lganidan keyin dehqonchilik bilan bog‘liq voqealar. [Xolova, 2022:25]

Ma‘lumotlarni oldindan qayta ishlash. Ma‘lumotlar to‘planganidan keyin, izchillik va sifatni ta‘minlash uchun uni oldindan qayta ishlash kerak. Bu dialektal matnlar yoki nutq namunalarini transkripsiya qilish, normallashtirish, izohlash va moslashtirishni o‘z ichiga olishi mumkin. Ma‘lumotlarni oldindan qayta ishlashda qo‘llaniladigan ba‘zi umumiy usullar mavjud:

Ma‘lumotni tozalash: to‘plangan ma‘lumotlardagi takroriy yozuvlar, chetga chiqish kabi xatolardan tozalash yoki tuzatish.

Ma‘lumotlarni o‘zgartirish: ma‘lumotlardan erkin foydalanish va tahlil qilish uchun ularni mos formatga aylantirish.

Ma‘lumotlar hajmini qisqartirish: O‘lchamlarni kichraytirish yoki ma‘lumotlarning muhim xususiyatlarini saqlab qolgan holda ma‘lumotlar hajmini kamaytirish.

Ma‘lumotlar integratsiyasi: bir nechta manbalar va usullar orqali olingan ma‘lumotlarni tahlil qilish uchun ularni yagona ma‘lumotlar to‘plamiga birlashtirish. Ma‘lumotlarni oldindan qayta ishlashni amalga oshirish orqali tahlilchilar va ma‘lumot oluvchilar ma‘lumotlarning toza, izchil, tahlilga tayyor bo‘lishini ta‘minlashi mumkin, bu esa kompyuter, raqamli texnologiyalarga asoslangan ilovalarda aniq va ishonchli natijalarga olib keladi.

Saqlash. Ma'lumotlar kelajakda foydalanish uchun raqamli texnologiyada saqlash tushuniladi. Dialektal ma'lumotlarning katta hajmini saqlash uchun ishonchli va kengaytiriladigan saqlash tizimi talab qilinadi. Barcha natijalar saqlanishi yoki yuklab olinishi mumkin.¹¹ Saqlash orqali natijalarni qayta ko'rish imkoniyatiga ega bo'lamiz. Boshqa tomondan yuklab olingan natijalar fayl yoki matn ko'rinishida import qilinishi mumkin. Bundan tashqari saqlash tizimi kompyuter tizimida hal qiluvchi rol o'ynaydi, chunki u turli maqsadlar uchun ma'lumotlarni saqlash va olish imkonini beradi, masalan, dastur ma'lumotlari, fayllar, ma'lumotlar bazalari va tizimning zaxira nusxalari. Saqlashni samarli boshqarish ma'lumotlar mavjudligi, ishonchligi va xavfsizligini ta'minlash uchun zarur.

Qidirish tizimi. Dialektal korpus kontekstida qidirish va izlash ma'lum bir tadqiqot savoliga yoki tahliliga tegishli bo'lgan ma'lum lingvistik ma'lumotlarni yoki korpus ichidagi ma'lumotlarni topish va ularga kirish jarayonini anglatadi. Dialektal korpus bu dialektlardagi yoki mintaqaviy tillardan foydalanishdagi o'zgarishlarni aks ettiruvchi til ma'lumotlari to'plamidir¹². Tadqiqotchilar ko'pincha o'z tadqiqotlari uchun ushbu korpus ichidagi o'ziga xos lingvistik xususiyatlarni izlashlari va olishlari kerak. Tadqiqotchilar dialektal korpus ichida o'rganishga qiziqqan muayyan lingvistik xususiyatlar, atamalar yoki shevalar asosida qidiruv so'rovlarini tuzadi. Ushbu so'rovlar kerakli til hodisalarni qamrab oluvchi kalit so'zlar, iboralar yoki muntazam iboralar yordamida tuzilishi mumkin. Tadqiqotchilar dialektal korpusga kirish va o'z so'rovlarini bajarish uchun qidiruv vositalari yoki dasturiy ilovalardan foydalanadilar. Qidiruv vositasi korpus ma'lumotlarni skanerlaydi va belgilangan mezonlarga mos keladigan tegishli lingvistik misollarni oladi. Tadqiqotchilar olingan lingvistik ma'lumotlarni to'g'riligini baholash uchun qidiruv natijalarini ko'rib chiqadi. Bu ma'lumotlar matndan parchalar, audio yozuvlar, transkripsiyalar yoki dialektal o'zgarishlarni aks ettiruvchi boshqa lingvistik izohlarni o'z ichiga olishi mumkin. Tadqiqotchilar olingan lingvistik ma'lumotlarni tahlil qilib, korpus ichidagi dialektal o'zgarishlar yoki tildan foydalanish bilan bog'liq xususiyatlarni, tushunchalarni aniqlashadi. Ular ma'lumotlarni sharhlash va tadqiqot maqsadlari uchun xulosalar chiqarish sifat yoki miqdoriy usullardan foydalanishlari mumkin. Foydalanuvchilar uchun qulay qidiruv interfeysi tadqiqotchilarning tadqiqot qiziqishlari asosida aniq dialektal ma'lumotlarga kirishlari uchun juda muhim. Korpus tarixchilar, etnologlar, filolog olimlar uchun emas balki oddiy odamlar uchun ham ochiq bo'ladi, shu sababli, matnlar fonetik alifbolarni yoki mahalliy dialektlarning barcha o'ziga xos xususiyatlarini yaxshi bilmaydiganlar uchun ham tushunarli bo'lishi kerak. So'rovnoman quyidagicha amalga oshirish maqsadga muvofiqdir. Korpus

¹¹ <https://www.helsinki.fi/varieng/CoRD/corpora/Dialects/fieldwork.html>

¹² <http://www.tekstlab.uio.no/scandiasyn/index.html>

kengaytirilgan qidiruv tizimi, natijalarni o'chirish, izohlash va saqlash funksiyasiga ega bo'ladi.

- 1) aniq dialektal so'z kiritish orqali;
- 2) gap oxirida, boshida va istalgan qismida kelgan dialektal so'zni kiritish orqali amalga oshiriladi.

Bir so'zni turli xil yo'l bilan qidirish bazadan keng foydalanish imkoniyatini beradi va xulosaning aniqligi uchun xizmat qiladi. Bazaga kiritilgan shevaga oid ma'lumotlar bazasidagi lingvistik tahlillar bir necha bosqich va prinsiplarga asoslangan holda tuziladi. Dastlabki shevalar bazasi quyidagi talablar asosida quriladi.

Qidiruv tizimi parametrlari	Misol
1. Shevaga oid so'z	Jang'i
2. Adabiy til	Yangi
3. Izoh	Ilgari bo'lmagan, endi yuzaga kelgan
4. Turkum	Sifat/ravish
5. Omonim	Yangi - ravish (yaqindagina, hozirgina) Yangi - sifat
6. Sinonim	Mavjud emas
7. Antonim	Eski

1.3-jadval. Korpusning qidiruv turlari

Bazada qidiruv tizimi ayni shu 7 maydonda amalga oshiriladi. Shevalar bazasida so'zlarning turkumlar bo'yicha statistikasi beriladi. Bunda bazaga kiritilgan jami kiritilgan so'zlarning tashqi va ichki statistikasi keltiriladi. Tashqi statistikada jami so'zlar soni, ya'ni jami mustaqil so'z turkumi, yordamchi so'z turkumi va alohida olingan so'zlar soni berilishi. Ichki statistikada turkumlarning xususiyatiga e'tibor beriladi. Unga ko'ra jami shevaga oid so'zlarning turkumga oidligi qayd etiladi.

T/R	Turkum	So'zlar soni
1.	Ot	323
2.	Sifat	48
3.	Son	12
4.	Fe'l	101
5.	Olmosh	4
6.	Ravish	6
7.	Ko'makchi	6

8.	Bog'lovchi	7
9.	Yuklama	9
10.	Modal	2
11.	Undov	8
12.	Taqlid	9

1.4- jadval. Turkumlar statistikasi

Moslashuvchanlik. Moslashuvchan dialektal korpus tadqiqotchilarga o'zlarining tahlil usullarini, tadqiqot savollarini va ma'lumotlarni o'rganish usullarini o'zlarining aniq tadqiqot maqsadlari va qiziqishlariga mos ravishda sozlash va moslashtirish qobiliyatini taklif qiladi. Moslashuvchan dialektal korpus turli dialektal o'zgarishlarni, mintaqaviy til xususiyatlarini va tildan foydalanish kontekstlarini ifodalovchi keng ko'lamli lingvistik ma'lumotlarni o'z ichiga olishi kerak. Korpus tadqiqotchilarga til o'zgarishi va turli til o'lchovlari bo'yicha o'zgarishlarni o'rganishga imkon berish uchun bir nechta dialektlardan, sotsiolektlardan, tarixiy davrlardan yoki janrlardan olingan ma'lumotlarni o'z ichiga olishi kerak. [Sirazitdinov, Buskunbayeva, Karimova, 2008: 212] Ma'lumotlar bazasi har xil turdagi ma'lumotlar jumladan, matn, audio, video tasvirlar va izohlar bilan ishlash uchun yetarlicha moslashuvchan bo'lishi kerak. Korpus bir nechta ma'lumotlar formatlarini qo'llab-quvvatlashi va boshqa vositalar yoki tizimlar bilan oson integratsiyalashuviga imkon berishi kerak.

Dialektal korpus vaqt o'tishi bilan sezilarli darajada o'sishi mumkin, shuning uchun ma'lumotlar bazasi katta hajmdagi ma'lumotlarni samarali boshqarish uchun kengaytirilishi kerak.

Xavfsizlik va maxfiylik. Dialektal korpusda shaxslar yoki jamoalar haqida nozik ma'lumotlar bo'lishi mumkin. Korpusga kirishni boshqarish mexanizmlari va shifrlashni himoya qilish uchun mustahkam xavfsizlik choralari ega bo'lishi kerak.

Dialektal korpus ko'pincha lingvistik tahlil uchun maxsus vositalarni talab qiladi. Ushbu vositalar asosiy statistik ma'lumotdan tortib tabiiy tilni qayta ishlashning ilg'or usullarigacha bo'lishi mumkin. Bunday vositalarni korpus arxitekturasiga integratsiyalashuvi dialektal ma'lumotlarni samarali tahlil qilishga yordam beradi. Dialektal korpus tadqiqotida ishlatiladigan ba'zi umumiy tahlil vositalari quyidagilardan iborat: Konkordans tadqiqotchilarga ma'lum kalit so'zlar, iboralar asosida korpus ichida lingvistik ma'lumotlarni qidirish va tahlil qilish imkonini beruvchi dasturiy vositalar. Muvofiqlashtiruvchilar kontekstda qidiruv so'zlari misollarni aks ettiruvchi muvofiqlik qatorlarini taqdim etadi, bu esa tadqiqotchilarga turli dialektal kontekstlarda o'ziga xos lingvistik xususiyatlar qanday ishlatilishini tekshirishga imkon beradi.



Xulosa qilib aytganda, yaxshi ishlab chiqilgan arxitektura va ma'lumotlar bazasi tizimi dialektal korpusni samarali boshqarish va tahlil qilish uchun muhimdir. Arxitektura ma'lumotlarni yig'ish, oldindan qayta ishlash, saqlash, metama'lumotlarni boshqarish, qidirish, tahlil qilish vositalarini o'z ichiga olishi kerak. Ma'lumotlar bazasi moslashuvchan, kengaytirilgan, ma'lumotlar xavfsizligi va maxfiyligini ta'minlashi kerak.

Foydalanilgan adabiyotlar:

1. Холова М. ўзбек миллий шевалари корпусини тузишнинг лингвистик асослари (Бойсун тумани “ж”ловчи шевалари мисолида): Фил. фан. бўйича фалсафа доктори (PhD)...дисс. Автореф. – Термиз, 2022. -12 Б
2. <https://www.helsinki.fi/varieng/CoRD/corpora/Dialects/fieldwork.html>
3. <http://www.tekstlab.uio.no/scandiasyn/index.html>
4. Sirazitdinov Z.A., Buskunbaeva L.A., Karimova R.N. Dialectological Base of the Machine Fund of the Bashkir Language as a Tool for Research Bashkir Dialect Continuum.\\ Russian Federation Bulletin of the Kalmyk Institute for Humanities of the Russian Academy of Sciences Has been 149 issued since. 2008. Vol. №23, Is. №1. –P. 212–219. DOI: 10.22162/2075- 7794-2016-23-1-212-219