

NUROTA SHEVASI (QARLUQ LAHJASI) KORPUSINING MA'LUMOTLAR BAZASI

Ruxsora Muftillayeva Toshmuhammad qizi,
Kompyuter lingvistikasi mutaxassisligi magistranti
ruxsoramuftillayeva@gmail.com
ToshDO'TAU

Annotatsiya. Ushbu maqolada milliy korpusda asta-sekin har xil korpuslarning paydo bo'lishi, sheva korpuslarini nazariy asoslanishi, yig'ilgan shevalardan foydalanish va tadqiqotlarni yengillashtirish yo'lini bosqichma-bosqich bosib o'tish, zamonaviy tilshunoslikning ushbu yo'nalishidagi yangi korpuslarning paydo bo'lishiga bir qadar poydevor bo'lib xizmat qiladi. Shevashunoslikda so'zlashuv jarayonlarini transkripsion shaklda yozib olish, shevaga oid materiallarni izohlash orqali dialektning o'ziga xos xususiyatlarini tadqiq etish, shevalar tadqiqotiga bag'ishlangan monografiyalar, lug'atlar, yaratiladi hamda dialektal xaritalar tuziladi.

Annotation: This article discusses the gradual emergence of various corpora within the national corpus, the theoretical foundations of dialect corpora, and the step-by-step process of utilizing collected dialects to facilitate research. It serves as a partial foundation for the development of new corpora in this modern field of linguistics. In dialectology, the transcription of spoken language, the annotation of dialect materials, and the study of distinctive features of dialects are emphasized. Additionally, monographs, dictionaries dedicated to dialect research are created, and dialectal maps are compiled.

Kalit so'zlar: *shevashunoslik, lug'at, milliy korpus, dialektal korpus*

KIRISH

Dialektal korpus yaratish, yig'ilgan ma'lumotlardan samarali foydalanish va lingvistik tahlil uchun zaruriy vosita hisoblanadi. Ushbu jarayon, foydalanuvchilarga kerakli ma'lumotlarni qidirishda vaqtni sezilarli darajada qisqartirish imkonini beradi. Nurota tumani dialektal korpusining ma'lumotlar bazasini tuzishda, ijtimoiy jihatdan tashqi dunyodan qisman uzilgan va internet tarmoqlaridan foydalanish ko'nikmasiga ega bo'lmagan keksa avlod vakillari ishtirok etishdi. Ushbu jarayon, shahardan uzoqda joylashgan hududlarda yashovchilar o'rtasidagi o'zaro suhbatlarni yozib olish orqali amalga oshirildi. Dialektal korpus uchun yig'ilgan audio fayllar dastavval matn ko'rinishiga keltiriladi. Audio faylni matn holatiga keltirish uchun biz har xil dasturlarni ishlatib ko'rdik. Xususan, “speech-to-text” dasturi. Bu dastur imkoniyatlari ancha yuqori. Dasturga audio faylni MP3 shaklida yuklashimiz yoki mikrofon orqali to'g'ridan

to'g'ri ovoz yozishimiz mumkin. Birinchi qadam, yuklangan faylga ishlov beradi. Ya'ni ovozda ortiqcha shovqinlarni olib tashlab, ovoz tiniqligini oshiradi. Keyingi bosqichda audioning matn ko'rinishi taqdim etiladi. Matnni fayl ko'rinishida saqlab olish imkoniyati mavjud. Dasturda 100 dan ortiq til mavjud, ammo o'zbek tili yo'qligi uchun bu dasturdan foydalana olmadik. Foydalangan keyingi dasturimiz "Aisha" o'zbek tilida mukammal muloqat qiladigan sun'iy intellekt. Bu dastur ovozni matnga, matnni ovozga kabi xizmatlarini taqdim etadi. Dasturdan foydalanish barcha uchun ochiq. SST faqat 30 soniyagacha bo'lgan MP3 formatdagi fayllarni qabul qiladi. Shuning uchun audio fayllarni 30 soniyadan qilib kesib chiqdik. Audio fayllarni dasturga yuklaganimizda bizga matn shaklida, transkripsiyasi bilan taqdim etdi. Bu dasturlar har doim ham yuqori aniqlikda ishlamaydi, ayniqsa, mahalliy lahjalarda. Xato bo'lgan joylari qo'lda to'g'irlandi. Audio fayllar matn ko'rinishiga kelgandan so'ng matnlar bilan ishlendi. Matnlarni alohida satrlarga bo'lib chiqdik. Har bir satrni yonidan matnning audio variantini joylab chiqdik. Audio fayllarni satrlarga moslab kesib, satrlar bilan yonma-yon joylashtirdik. Audiolarni kesish uchun maxsus "Simple MP3 Cutter" programmasidan foydalanib audiolarni satrlarga moslab kesib chiqdik. Dasturdan foydalanish bepul. Dasturdan ovozda shovqinlarni kesishda ham foydalandik.

Korpus interfeysi. Interfeys – piegasse so'zidan olingan bo'lib, texnik va programmalash vositalari yig'ilmasi hisoblanib, boshqarish yoki o'lchash, hisoblashdagi turli vazifaviy qurilmalarning o'zaro ta'sirini ta'minlab beradi. Berilayotgan axborotlarni kodlash, sinxronlash hamda nisbatan belgilangan qoida hamda kelishuvga muvofiq tayyor modullardan sistemani, qurilmalarning mexanik va elektr birikmalarini signal, ya'ni axborot shaklida to'plashga imkon beradi[1]. Korpusni loyihalash jarayonida avvalo uning dizayniga alohida e'tibor qaratish talab etiladi. Korpus foydalanuvchilarida ijobiy taassurot qoldirishi uchun shevaga xos elementlar va Nurota tumaniga aloqador bo'lgan vizual oynalardan foydalanish zarur (1-rasm).



1-rasm. Nurota tumani dialektal korpusi interfeysi

Korpus dizaynida shu belgilarni inobatga olish juda muhim omillardan. Milliy til va uning shevalarini informatsion dunyoga olib chiqish, zamonaviy tadqiqotlar uchun muhim bir asos bo'lib xizmat qiladi. Bu jarayon nafaqat tilshunoslik, balki madaniyatshunoslik va antropologiya sohalarida ham yangi imkoniyatlarni ochadi. Dastlabki shevalar bazasini yaratish jarayoni bir qator muhim talablar va maqsadlar asosida amalga oshiriladi. Ushbu bazaning asosiy maqsadi o'zbek milliy shevalarining xilma-xilligini, ularning o'ziga xos leksik va grammatik xususiyatlarini chuqur o'rganishdir. Buning uchun quyidagi talablar ko'zda tutiladi: (Qarang: 1-jadval)

1-jadval. Dialektal korpusda so'zlar tasnifi

Shevaga oid so'z	Shipirgi
Adabiy til	Supurgi
Turkum	Ot
So'zning izohi	Supurish-sidirish, tozalash uchun burgan, shuvoq, makkasupurgi va shu kabilardan dastlab bog'lab yasaladigan ro'zg'or buyumi
Sinonimi	Sinonimi mavjud emas

Turli xil dialektal so'zlardan foydalanish, respondentlarning turli ijtimoiy va madaniy kontekstlarini aks ettirishga yordam beradi. Bu esa tadqiqotchilarga aniq va ishonchli xulosalar chiqarishga imkon yaratadi. Bir so'zni har xil yo'l bilan so'rovnoma berish manbadan keng foydalanish imkoniyatini beradi. Ushbu metodologik yondashuvlar orqali so'rovnoma natijalari yanada aniqroq bo'ladi. Turli xil dialektal so'zlardan foydalanish, respondentlarning turli ijtimoiy va madaniy kontekstlarini aks ettirishga yordam beradi. Bu esa tadqiqotchilarga aniq va ishonchli xulosalar chiqarishga imkon yaratadi. Boshqird dialekti xonqizi (kamka)ning 148 ta dialektal variantlarini berib o'tadi, shulardan 17 ta varianti faqat sharq dialektlari tarkibida uchraydi[2]. Nurota shevalar elektron lug'ati internetda onlayn ish yurituvchi bazadir. Bazadagi ma'lumotlar doirasini kengaytirish xususiyatiga ega bo'lgan har tomonlama qulay dialektal bazalardan biridir. Ushbu elektron baza keng imkoniyatlar taqdim etadi. Baza foydalanuvchilar auditoriyasi uchun quyidagi imkoniyatlarni taqdim etadi.

1. Bosh sahifa. 2. So'zlar jadvali. 3. Loyihasi haqida. 4. Transkripsiya. 5. Aloqa.

Bosh sahifa shevalar bazasining markaziy qismi bo'lib, u foydalanuvchilarga bazadagi ma'lumotlarga oson kirish imkonini beradi. Ushbu sahifa yuqoridagi beshta asosiy bo'limdan iborat bo'lib, har bir bo'lim boshqa sahifalarga yo'naltiruvchi aloqa vositasi sifatida xizmat qiladi. Bazadagi so'zlarning turli

kesimlarda tahlil qilinishi foydalanuvchilarga shevaning leksik boyligini to'liq anglash imkonini beradi. So'z turkumlari asosan ikkita kategoriya bo'yicha bo'linadi: Umumiy tashqi so'z turkumlari: Mustaqil so'z turkumi, yordamchi so'z turkumi alohida so'z turkumi. Xususiy ichki so'z turkumlari: ot, son, fe'l, ravish, sifat, olmosh, ko'makchi, bog'lovchi, yuklama, undov, modal va taqlid so'zlar. Bu statistika foydalanuvchilarga shevalarning strukturasini va ularning foydalanish doirasini aniqlashda yordam beradi. Bazadagi shevaga xos leksemalar foydalanuvchilar uchun alohida ahamiyatga ega. Ushbu so'zlar bazaning keng auditoriyaga mo'ljallanganligini hisobga olib, foydalanuvchilar uchun qulaylik tug'dirish maqsadida transkripsiya shaklida taqdim etilgan. Bu esa shevalarni o'rganish jarayonini osonlashtiradi. Shevalar bazasining bosh sahifasi foydalanuvchilar uchun shevalarni o'rganish jarayonini osonlashtiradi, ularning leksik boyligi, strukturası va tarixiy konteksti haqida aniq va to'liq ma'lumotlar taqdim etadi. Bu o'z navbatida, shevalarning saqlanishi va rivojlanishiga xizmat qiladi.

Shevalar Bazasi: Shevalar bazasida so'zlar jadvali foydalanuvchilarga turli xil imkoniyatlarni taqdim etadi. Ushbu sahifa barcha qatorlar soni, izlash, turkum bo'yicha va shevaga xos so'zlarning etimologik saralash imkoniyatlarini o'z ichiga oladi[3]. Jadval quyidagi ma'lumotlarni o'z ichiga oladi: adabiy til, transkripsiya (mavjud jahon standartlariga muvofiq), turkum (so'zning turkumini ko'rsatish, saralash imkoniyati bilan), izoh (so'zning ma'nosi yoki qo'llanilishi haqida tushuntirish), sinonimi (ushbu so'zga o'xshash ma'noga ega so'zlar)[4]. Shevalar bazasini yangi dialektal so'zlar bilan to'ldirish jarayonlari o'z qoidalariga ega bo'lib, quyidagi bosqichlarni o'z ichiga oladi:

Yangi dialektal so'zlarni aniqlash va ularni bazaga kiritish, har bir so'zni tekshirib, uning to'g'riligini va mosligini aniqlash, so'zlarni turkumlar bo'yicha saralash va tegishli joylarga joylashtirish, so'zlarning transkripsiya qoidalariga mos kelishini ta'minlash. Ushbu jarayonlar shevalar bazasining boyligini oshirishga va foydalanuvchilarga yanada qulay xizmat ko'rsatishga yordam beradi.

XULOSA

Shevalar bazasi foydalanuvchilarga turli xil dialektal so'zlar bilan ishlash imkonini beruvchi muhim resursdir. So'zlar jadvali, adabiy til, transkripsiya, turkum, izoh va sinonimlar kabi ma'lumotlarni o'z ichiga olgan holda, foydalanuvchilarga so'zlarni oson topish va ularning ma'nolarini tushunishga yordam beradi. Yangi dialektal so'zlarni bazaga kiritish jarayoni aniq qoidalar va bosqichlarga asoslangan bo'lib, bu jarayonlar so'zlarning to'g'riligini va mosligini ta'minlashga, shuningdek, bazaning boyligini oshirishga xizmat qiladi. Natijada, shevalar bazasi foydalanuvchilarga yanada qulay va samarali xizmat ko'rsatishni maqsad qiladi. Bu resurs tilshunoslik va dialektologiya sohalarida muhim



ahamiyatga ega bo'lib, mahalliy tillarning saqlanishi va rivojlanishiga hissa qo'shadi.

Foydalanilgan adabiyotlar:

1. <https://uz.wikipedia.org/wiki/Interfeys>
2. Sirazitdinov Z.A., Buskunbaeva L.A., Karimova R.N. – P. 212–219, DOI 10.22162/2075- 7794-2016-23-1-212-219
3. 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.
4. Иванцова. Е.В. Томский диалектный корпус: обоснование концепции и перспективы развития. Вопросы лексикографии. 2017. № 11. – С. 62.
5. Tregubova E. N. Multi-level thematic markup as a tool for ethnolinguistic representation of dialect discourse in an e-text corpusю Vestnik. Tomsk State. University. Phil. . Vol.№ 1 (33). 2015. –P. 66-77.