



O‘ZBEK SHEVALARI LEKSIKASINI KORPUSGA JOYLASHTIRISHNING DASTURIY TA‘MINOTI

Ashirboyev Samixon

F.f.d., ToshDO‘TAU professori

Turdialiyev Anvarbek

ToshDO‘TAU o‘qituvchisi

Annotatsiya. Jahon tilshunosligida shevalarni o‘rganishga e‘tibor kuchaydi. To‘g‘ri, o‘zbek tilshunosligida shevalarni o‘rganish masalasi uzoq tarixni bosib o‘tgan bo‘lsa-da, keyingi 30 yil ichida bu yo‘nalishdagi ishlar ancha sustlashdi, lekin shevashunosligimizning bir tarmog‘i bo‘lgan lingvistik geografiya bo‘yicha qator ishlar paydo bo‘ldi. Ayniqsa, o‘zbek shevalarining areal tadqiqotlari amalga oshirilganligi va bu sohaga doir darsliklari yaratilganligini alohida ta‘kidlash mumkin. O‘zbek tilshunosligida bugungi kunda dialektologik tadqiqotlar bo‘yicha qilingan bir qancha tadqiqotlar bajarilyapti. O‘zbek shevalarini o‘rganish va asrab qolish yo‘llarida amaliy loyihalar amalga ishirilyapti. Dialektal korpus yaratish yo‘lida qilingan ishlar ham bunga misol.

Abstract. In world linguistics, attention to the study of dialects has increased. True, although the issue of studying dialects in Uzbek linguistics has a long history, in the next 30 years, the work in this direction slowed down, but linguistic geography, which is a branch of our dialectology a number of works appeared. In particular, it is possible to emphasize that area studies of Uzbek dialects have been carried out and textbooks in this field have been created. In Uzbek linguistics, a number of dialectological researches are currently being promoted. Practical projects are being implemented to learn and preserve Uzbek dialects. An example of this is the work done to create a dialectal corpus.

Аннотация. В мировом языкознании возросло внимание к изучению диалектов. Правда, хотя вопрос изучения диалектов в узбекском языкознании имеет давнюю историю, в последующие 30 лет работа в этом направлении замедлилась, но по лингвогеографии, являющейся отраслью нашей диалектологии, появился ряд работ. В частности, можно подчеркнуть, что проведены региональные исследования узбекских диалектов и созданы учебники в этой области. В узбекском языкознании в настоящее время продвигается ряд диалектологических исследований. Реализуются практические проекты по изучению и сохранению узбекских диалектов. Примером тому является работа по созданию диалектного корпуса.

Kalit so‘zlar: *dialekt, dialektal sheva, areal lingvistika, razmetkalash, subdialektal korpus.*

Dunyo tilshunosligi zamon bilan hamnafas rivojlanish bosqichida bo‘lib, zamon bilan hisoblashmagan til egalari shu tilning kelajagini yo‘qotib qo‘yadi. Buni



dunyo tilshunosligida qilinadigan tadqiqotlardan ham kuzatish mumkin. Har qanday tilning kelib chiqish asosi va uning milliyligini saqlab turgan lingvistik bazasi mavjud bo'ladi. Lingvistik bazalar esa bu har bir tilning o'z shevasi (shevalari) hisoblanib, sheva sofligi qanchalik yuqori saqlansa, tilning ham yashovchanligi o'sib boradi. Hozirgi axborot texnologiyalar rivojlangan dunyoda tillar xilma-xil bo'lishiga qaramay u doimiy kamayib bormoqda, chunki mahalliy nutq shakllari kichik til oilalari, rivojlanmagan davlatlar tili asosiy muloqot vositasi bo'lgan rivojlangan (davlatlar) tillar ta'sirida o'lib bormoqda. Shu nuqtayi nazardan Amerika tilshunoslik maktabi (LSA) saytida tillarning yo'qolishi haqida quyidagicha jumlar keltiriladi, “*Tilni yosh bolalar o'rganishdan to'xtasa, uning (o'lim)kuni aniq sanaladi*”¹. Demak, bundan shunday xulosaga kelish mumkinki, har qanday tilni asrab qolish uchun til ma'lumotlarini bir joyga to'plash uni qayta ishlash tahlil qilish va uni keng omma foydalanishi uchun qulay tizimga tushirish kerak. Dialektologik korpuslar yo'qolish arafasidagi shevalarni asrashga mo'ljallangan dasturlardan biridir.

Dialektologik korpuslarning maqsadi dialektologiya, sotsiolingvistika, nutq madaniyati, fonetika-fonologiya, leksikologiya, grammatikaga oid tadqiqotlar hamda mahalliy aholining urf-odatlarini, tarixi va boshqa ko'plab sohalar uchun tadqiqotlar maydonini qulaylashtirish, kelajak avlodga qimmatli ma'lumotlarni yetkazish vazifasini bajarishdir.

Dialektologik korpuslarni tuzish ahamiyati shundaki, dunyoda tan olingan tillardan tashqari bir necha 10 ming so'zlovchiga ega bo'lsada iqtisodiy rivojlanishning pastligi sababli yo'qolib borayotgan til va ularning shevasini saqlab qolish hamda tadqiq qilishdir. Dialektlarni aniqlash va tasniflashning turli usullari mavjud bo'lib, bugungi kunda dunyo aholisi 7000 ga yaqin tillarda so'zlashayotgani taxmin qilinmoqda². Ushbu tillarning har biri ijtimoiy guruh yoki boshqa omillarga qarab farq qilishi mumkin bo'lgan bir nechta dialekt (sheva)lar bo'lishi mumkin.

Dunyoda qancha dialekt borligini aniq aytish qiyin. Bu qiyinchilik qisman tilshunoslarning dialekt nimadan iborat ekanligi haqidagi kelishmovchiligidan kelib chiqadi – ba'zi tilshunoslar tilning ikki shaklini turli dialekt deb bilishadi, boshqalari esa bunday fikrda emas. Shu o'rinda aytish joizki, dunyodagi dialektlarning miqdori juda katta bo'lib ana shu taxmin qilingan 7000 tildan faqat bittasi bo'lgan ingliz tilida 160ga yaqin lahjalar borligi taxmin qilinadi³. Ba'zi ma'lumotlarga ko'ra, dialektlarning umumiy soni 10 000 dan ortiq. Ammo shuni ta'kidlash kerakki, bu lahjalarning aksariyati rasman tan olinmagan yoki ro'yxatga olinmagan. Bunga misol sifatida Hindiston davlatini olishimiz mumkin. 2001-yildagi aholini hisoblash davrida Hindistonda 1 mlngacha so'zlashuvchiga ega bo'lgan 30ta til, 10 000 nafargacha so'zlashuvchiga ega bo'lgan 122ta til, 1599ta kam sonli so'zlashuvchiga

¹ <https://www.linguisticsociety.org/content/how-many-languages-are-there-world>

² <https://www.quora.com/How-many-languages-and-dialects-are-there>

³ <https://www.translateday.com/how-many-dialects-does-english-have/>



ega tillar borligi qayd qilinadi. 2011-yildagi hisob kitoblarda esa 121ta til borligi aniqlanadi va shu tillardan 22tasi Hindiston konstitutsiyasida til sifatida rasman tan olinadi⁴. Dialekt va lahjalarning miqdorini aniqlashning imkoni yo'qligi birgina hind tili misolida ham ko'rishimiz mumkin. Turkiy tillarda mavjud lahja va dialektlar ham shular sirasiga kiradi, chunki aholi migratsiyasi, iqtisodiy va ijtimoiy omillar bu farqlarni o'sishida hamda tillarning yo'qolishiga sabab bo'ladi. Korpus dialektologiyasining ikkinchi bir ahamiyati shu yo'qolish arafasidagi til(sheva)larni saqlab qolishdan iboratdir.

Jahon tilshunosligida yaratilgan dialektal korpuslarning yuzaga kelishiga nazar tashlaydigan bo'lsak: Xelsinki Britaniya ingliz dialekti korpusi (HD) Sharqiy Angliya, Janubiy-g'arbiy Lankashir orollari aholisidan to'plangan orfografik transkripsiyalangan audio yozuvlar to'plamini uchratishimiz mumkin.

Shevalar korpusini yaratish bosqichida shevashunosning tadqiqot maydoni doirasida matnlarni fonetik, leksik, grammatik jihatdan razmetkalash bosqichi dastlabki vazifa hisoblanadi. Razmetkalash deganda matn va uning komponentlariga maxsus teglarni biriktirib chiqish nazarda tutiladi. Maxsus teglar ikki xil bo'ladi: lingvistik teglar va ekstralingvistik (tashqi) teglar. Lingvistik teglar matn elementlarining leksik, grammatik va shunga o'xshash boshqa xususiyatlarini tavsiflaydigan ma'lumotlardan iboratdir. Ekstralingvistik teglar esa muallif haqidagi va matn haqidagi ma'lumotlarni (muallif, nomlanishi, nashr yili va joyi, janr, tematika va hokazo) tavsiflaydi⁵. Yangi tuzilgan korpuslar uchun razmetkalash qo'l mehnati asosida bajariladigan ish hisoblanadi. Razmetkalash uch yo'nalishda olib boriladi: kuzatish, fonetik razmetkalash, shevaga oid (dialektal) matnlarni razmetkalash. Kuzatish bosqichida matn olinadigan shahar, qishloq, aholining etnik tarkibi, respondentning kelib chiqishi, kasbi, ma'lumoti (o'rta, oliy) kabilar o'rin olsa, fonetik razmetkalashda unli (old qator, orqa qatorlik, singarmonizm, urg'u olish, olmaslik), undosh (portlovchi-sirg'aluvchi, jarangli-jarangsiz) tovushlar doirasidagi o'zgarishlar; dialektal matnlarni razmetkalashda⁶ sheva materiallarining leksik metarazmetkasi ham muhim hisoblanib har bir shevadagi so'zlar alohida teglanib adabiy til bilan bog'lab chiqiladi, dastlabki dialektal korpuslarda har qanday tildagi korpuslarda bo'lganidek har bir so'z (variantlari) sinonimi, omonimi, antonimi, paronimlari alohida teglanadi. Shevalar o'rtasidagi so'zlarni leksik-semantik teglash murakkab jarayon hisoblanadi. Grammatik razmetkalash yagona standartlashtirilgan grammatik birliklar mavjudligi shevalararo bog'liqliklar sabab ozmi ko'pmi bir-birini tushunishga olib keladi va bu jarayon yuqoridagi uch bosqishdan yengilroq sanaladi. Bundan tashqari, avtomatik teglash (razmetkalash) ham mavjud bo'lib bunda teglashni “tegger [Abdullayeva, 2022: 103] (teglovchi

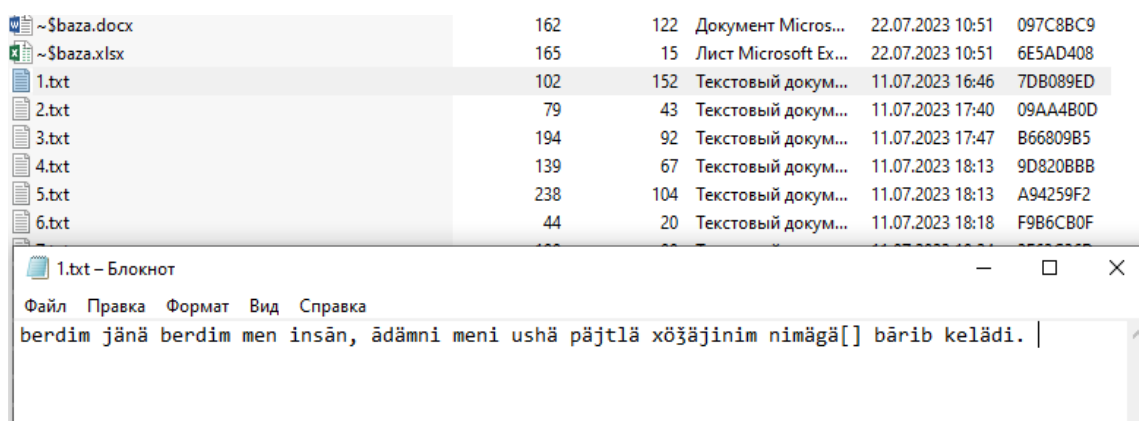
⁴https://en.wikipedia.org/wiki/Eighth_Schedule_to_the_Constitution_of_India

⁵ <http://compling.navoiy-uni.uz/index.php/conferences2/article/view/167/160>

⁶ Matnlarni razmetkalashda: nasriy matn, she'riy matn va aralash (she'riy-nasriy) matnlar ko'proq tadqiq qilingan bo'lib Dialektal korpusda tabiiy va ishonchliligini ta'minlash uchun dialog va monolog (audio) matnlarni tanlab oldik va shularni izohladik.

dastur)” bajaradi. Yuqorida sanalgan korpuslarda qisman razmetkalash amalga oshirilganligi bizning ishimizda ham o‘z o‘rnida foydalanildi. Til korpuslarida sheva ma’lumotlar bazasi shakllantirildi. Bunda avval Denov shevalari haqidagi ma’lumotlar jamlanib tahlil qilindi. Denov tumanida o‘zbek shevalarining qipchoq, qarluq lahjasida so‘zlashuvchi aholi va tojik tilida so‘zlashuvchi aholi vakillari yashaydi. Denov tumanida joylashgan 98ta (MFY) aholi yashash joylarini kuzatib, yashaydigan odamlar bilan suhbatlashish chog‘ida yig‘ilgan audio matnlar tahlil qilindi.

Subkorpusda Denov shevalari ma’lumotlar bazasining shakllantirilishi



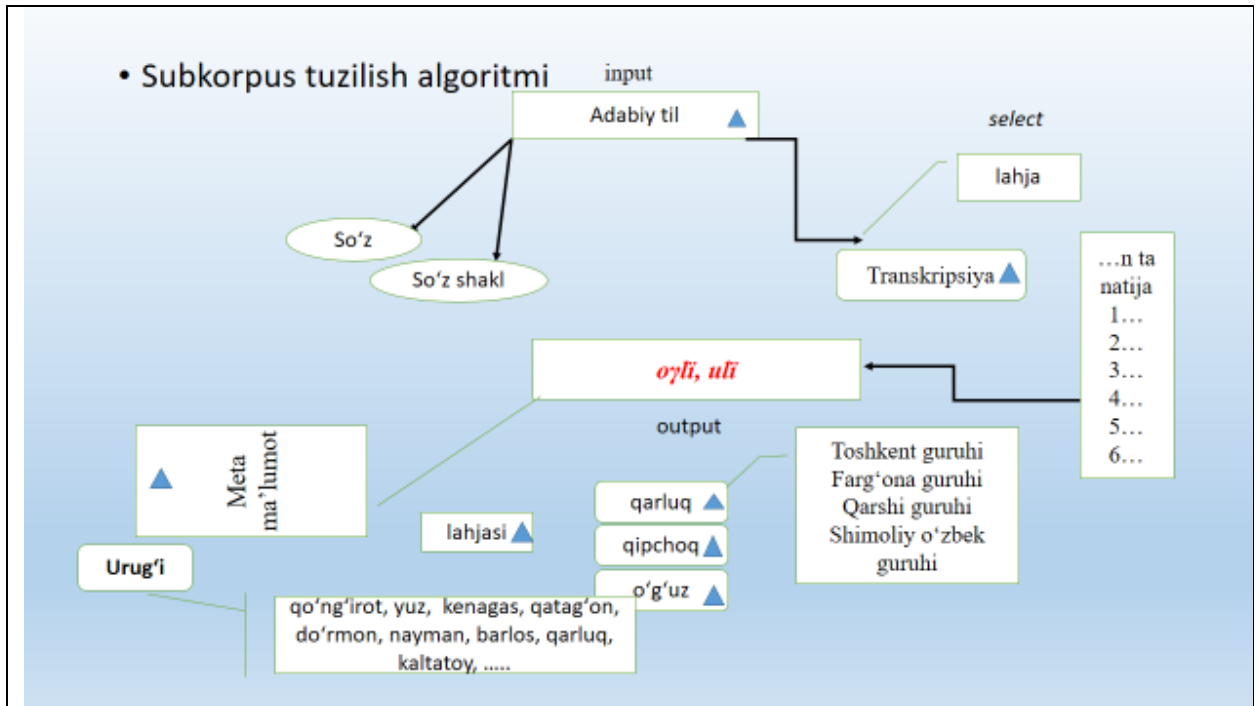
quyidagicha amalga oshirildi. Birinchi bosqichda diktafon, telefon so‘zlashuvlari so‘rovnoma o‘tkazish, anketa to‘ldirish usullari orqali to‘plangan dala ma’lumotlari (audio materiallar) transkripsiyada⁷ yozib chiqildi.

To‘plangan materiallar asosida dialektologik lug‘at tuzildi. Lug‘atda o‘zbek adabiy tilda uchramaydigan, adabiy tilda talaffuzi bilan farqlanadigan so‘zlar alohida meta ma’lumotlari bilan joylashtirib chiqildi. Lug‘atdan o‘rin olgan har bir so‘z uchun maxsus kompyuter dasturlari o‘qishi uchun mo‘ljallangan yagona yunikot tizimiga ega bo‘lgan *UTF-8 (txt)* faylda matnlar tayyorlandi (1-rasm)

Denov shevasining sheva ma’lumotlarini til korpuslari qidiruv tizimida joylanishi va uni amalda qo‘llanish bosqichlarini ishlab chiqishda shu kungacha yaratilgan dunyo korpusshunoslik va o‘zbek tilshunoslari tomonidan yaratilgan til korpuslari metodologiyasiga asoslanildi.

Yaratilgan subkorpus interfeysining **Input** qismida so‘zning adabiy tildagi varianti masalan: *o‘g‘li* so‘zi kiritilganda qidiruv tizimida joylashtirilgan ma’lumotlar (**select**) lahja bo‘limiga buyruqni uzatadi va **transkripsiyada** berilgan sheva materiallaridan nta (cheklanmagan) qidiruv natijalarini mavjud bo‘lgan matnlarni interfeys (**output**) ga chiqaradi. Olingan natija *o‘g‘li* so‘zining qipchoq va qarluq shevalaridagi barcha variantlarini so‘z shakllari bilan birgalikda aks ettiradi. Olingan natija haqida yuqorida ta’kidlaganimizdek meta ma’lumotlarni ham olishimiz mumkin.

⁷ Lotin alifbosidag transkripsiyasidan foydalandik.



Denov kichik dialektal zonasi shevalarining asosiy xususiyatlari, shevalarning bir-biriga ta'siri alohida o'rganildi va baza lug'atiga kiritildi. Subdialektal korpus tuzishda so'zlarning transkripsiyasi va adabiy tildagi transliteratsiyasi alohida izohlar bilan berildi. Har bir so'zning shevadagi varianti va o'zbek adabiy tilidagi sinonim va omonimlari alohida izohlar orqali teglandi. Bundan tashqari korpus imkoniyatidan kelib chiqib ma'lumotlar bazasi doimiy ravishda to'ldirib borishga imkoniyat yaratildi. Keyingi o'rinlarda korpusga joylangan audio materiallar avtomatik ravishda razmetkalanish ham imkoniyati ishlab chiqildi.

Xulosa o'rnida shuni aytish mumkinki, o'zbek shevalar korpusining dastlabki tajribasida quyidagi vazifalar bajarilganini ko'rishimiz mumkin. Korpusni sintaktik tegrash va annotatsiyalashda universal tobelik nazariyasi metodologiyasi asosida o'zbek tilining sintaktik modellari aniqlanganligini hisobga olish zarur. Shuningdek, dialektal zona materiallari asosida subkorpus ma'lumotlar bazasi shakllantirilib, qidiruv oynasida natijalarni aks ettirilishi tajribalarda isbotlandi. Korpusni grammatik tegrash bosqichida ontologik modellashtirish usulining ilmiy asoslari isbotlangan. Korpusda qidirilayotgan so'zning imloviy xatolarini aniqlash va unga yaqin bo'lgan so'zlar koeffitsentini belgilashda qidiruv interfeysida ilova qilishda yuqorida ta'kidlangan algoritmlarni qo'llashning maqsadga muvofiq ekanligi ta'kidlandi.

Denov shevasi ma'lumotlarining til korpuslari qidiruv tizimida joylashtirilishi va uning amalda qo'llanish bosqichlarini ishlab chiqish shu kungacha yaratilgan dunyo korpusshunoslik va o'zbek tilshunoslari tomonidan yaratilgan til korpuslari metodologiyasiga asoslanishi maqsadga muvofiq deb topildi.



Foydalanilgan adabiyotlar:

1. Ashirboyev S. O‘zbek dialektologiyasi. – Toshkent: Nodirabegim, 2021. – B. 176.
2. Abdullayeva O. O‘zbek tilining internet axborot matnlari korpusini shakllantirishning nazariy va amaliy asoslari: Filol. fan. bo‘yicha falsafa doktori (PhD) diss. – Toshkent, 2022. – B. 103.
3. Abdurahmonova N. Kompyuter lingvistikasi. – Toshkent: Nodirabegim, 2021. –394 b.
4. Холийёров Ў. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) дис. автор. – Термиз, 2021. – 52 б.
5. Xidraliyeva Z. O‘zbek tilining iqon shevasi. Filol. fan. bo‘yicha falsafa doktori (PhD) diss. – Toshkent, 2023. – 166 b.
6. Холова М. Ўзбек миллий шевалари корпуси тадқиқи (Бойсун тумани “ж” ловчи шевалари мисолида). –Термиз: ТерДУ НММ, 2022. – 141 б.
7. Khidralieva Z. R. Vowel harmony features in ikan dialect. // ACADEMICIA: An International Multidisciplinary Research Journal, Vol. 10, Issue 12, December 2020. pp. 1177-1182.
8. Turdaliyev A. Denov kichik dialektal zonasining xususiyatlariga doir. Oltin bitiglar — Golden cript. ISSN 2181-9238// 2022/3