

## SURXONDARYO SHEVALARI KORPUSIGA SHEVA MATNLARINI ICHKI TIZIMGA KIRITISH PRINSIPLARI

*Xolova Muyassar Abdulhakimovna*

*Termiz davlat universiteti, O'zbek tilshunosligi kafedrasida katta o'qituvchisi, f.f.f.d.  
(PhD)*

*Annotatsiya. Ushbu maqolada o'zbek milliy korpusi interfeysi va ichki tizimiga oid ma'lumotlar, korpus bazasini tadrijiy to'ldirish, taxlashda va ma'lumotlarni kiritish ketma-ketligiga doir lingvistik ko'rsatmalar berilgan va bu jarayon Surxondaryo viloyati shevalari misolida ko'rsatib o'tilgan. Bunda shevalar korpusining ichki tizimda dialektal matnlarning content, topic, genre, source, locations kabi maydonlar to'ldirilishiga doir qarashlar aks etgan.*

*Kalit so'zlar: Surxondaryo viloyati shevalari, o'zbek milliy korpusi, o'zbek shevalari ma'lumotlar ombori, ma'lumotlarni taxlashda va kiritish, dialektal matnlar, content, topic, genre, source, locations.*

## ПРИНЦИПЫ ВНЕДРЕНИЯ ДИАЛЕКТНЫХ ТЕКСТОВ В КОРПУСЕ СУРХАНДАРЬИНСКИХ ДИАЛЕКТОВ

*Холова Муяссар Абдулхакимовна*

*кандидат филологических наук (PhD) старший преподаватель  
кафедры узбекского языкознания, Термезского государственного  
университета*

*Аннотация. В данной статье изложено лингвистические указания относительно последовательности периодического заполнения и ввода база данных в узбекском национальном корпусе и показан этот процесс на примере диалектов Сурхандарьинской области. Это отражало взгляды диалектного корпуса на то, как во внутренней системе диалектных текстов заполняются такие поля, как содержание, тематика, жанр, источник, местоположение*

*Ключевые слова: диалекты Сурхандарьинской области, узбекский национальный корпус, словарь узбекских диалектов, вставка и ввод данных, диалектные тексты, содержание, тематика, жанр, источник, места.*

## PRINCIPLES OF THE INTRODUCTION OF DIALECT TEXTS IN THE CORPUS OF SURKHANDARYA DIALECTS

*Kholova Muyassar Abdulhakimovna*

*f.f.f.d. (PhD) Senior lecturer, departments of Uzbek Linguistics, Termez State  
University*

*Annotation. This article presents linguistic instructions regarding the sequence of periodic filling and entering of the data warehouse in the Uzbek National Corpus and shows this process by the example of the dialects of Surkhandarya region. This reflected the views of the dialect corpus on how fields such as content, subject, genre, source, location are filled in the internal system of dialect texts.*

*Keywords: dialects of Surkhandarya region, Uzbek national corpus, dictionary of Uzbek dialects, insertion and data entry, dialect texts, content, subject, genre, source, locations*

**KIRISH.** Bugungi kunda ijtimoiy-iqtisodiy hayotning barcha jabhalari rivojlanishini kompyuter texnologiyalarisiz tasavvur qilish qiyin va shu bilan birga lingvistika oldida turgan masalalari yechimida ham asta-sekin kompyuter texnologiyalari yordamida tuzilgan elektron matn korpuslariga murojaat qilish boshlangan [13], amaliy lingvistikada o'zbek milliy korpusini tuzish muammolari asta-sekin o'z yechimini topib kelmoqda.

**TADQIQOT METODOLOGIYASI.** O'zbek shevalar korpusi ichki tizimiga ma'lumotlar kiritishga doir tadqiqotlarni o'tkazishda "kuzatish", "suhbat", "hujjatlarni o'rganish va tahlil qilish", "ketma-ket joylash" metodlaridan foydalangan holda amalga oshirildi.

**TAHLIL VA NATIJALAR.** Dunyoda sun'iy intellektning dialektal korpuslar tuzish jarayonidagi o'rniga alohida e'tibor berilishini Britaniya, Xelsinki, Gruziya, Rus tili korpusi

(NRKY) [7], Hindiston (dialektlari korpusi) [10], Xitoy Mandarin dialektlari korpusi [5], arab dialektlari sotsial Mediya korpusi [4] misolida ko‘rib o‘tish mumkin. O‘zbek tili amaliy tilshunosligi doirasida milliy shevalar korpuslarini tuzish, uning lingvistik va texnologik asoslari doirasida olib borilgan tadqiqotlar va tahlillar va bu jarayonni o‘zbek tili korpusi (elektron resurslar) [3], o‘zbek tili ta’limiy korpusi [11], o‘zbek tili mualliflik korpusi [12], o‘zbek-ingliz, turk, koreys tili parallel korpusi [6], ustida olib borilayotgan amaliy tadqiqotlarda ko‘rib o‘tishimiz mumkin.

O‘zbek milliy shevalar korpusi interfeysi korpus, jadval, loyiha haqida, aloqa kabi oynalardan iborat bo‘lib, qidiruv va saralov quyidagicha amalga oshiriladi:

leksemaning yoki so‘zlarning ikki fonemasini kiritish orqali kontekstdagi gap kesimida taqdimoti va statistikasi;

leksemani to‘liq kiritish orqali kontekstdagi gap kesimida taqdimoti va statistikasi;

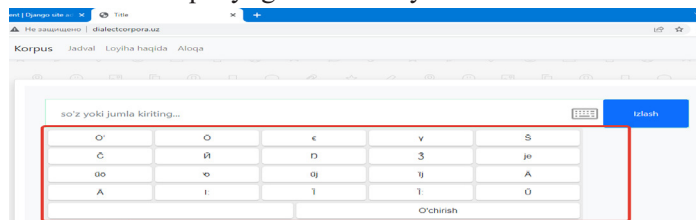
so‘zlarni kiritish orqali kontekstdagi gap kesimida taqdimoti va statistikasi;

so‘z birikmasi kiritish orqali kiritish orqali kontekstdagi gap kesimida taqdimoti va statistikasi;

fraza kiritish orqali kontekstdagi gap kesimida taqdimoti va statistikasi;

gap kiritish orqali kontekstdagi gap kesimida taqdimoti va statistikasi keltirilib o‘tiladi.

Shevalar korpusi tashqi interfeysida  $\epsilon/\text{y}/\text{z}/\text{a}/\text{a}/\text{u}/\text{u}/\text{y}/\text{y}$  kabi fonetik transkriptiv [1], [14] belgilarning joylashtirilganligi, belgi (simvollardan)lar uchun alohida vaqt ajratmaslik, sheva birliklari qidiruvda izlash qulayligini ta’minlaydi.

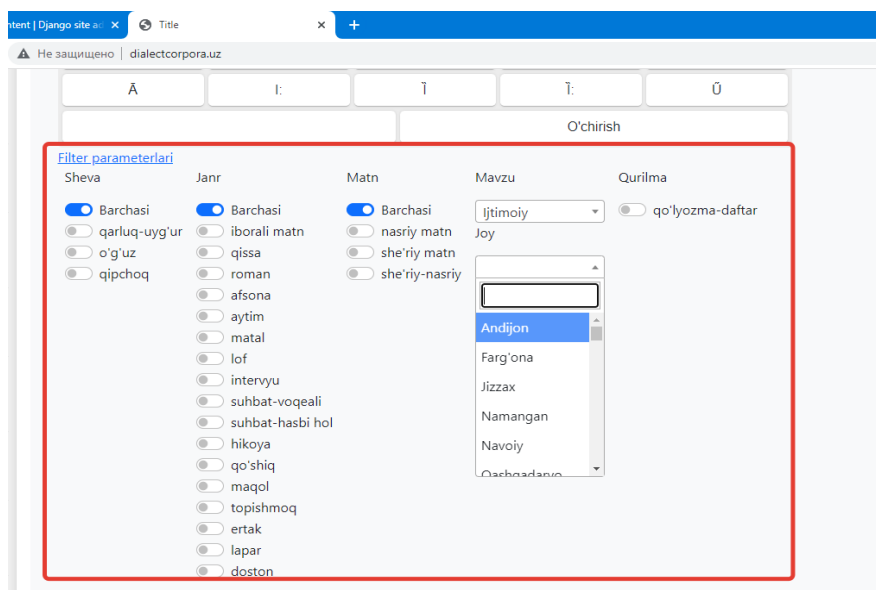


1-rasm.

Izoh: o‘zbek milliy shevalar korpusi interfeysi

Korpus interfeysida qipchoq, qarluq, o‘g‘uz shevalari ichida she’riy, nasriy, she’riy-nasriy matnlara; iborali matn, qissa, roman, afsona, aytim, matal, lof, intervyu, suhbat, hikoya, qo‘shiq, maqol, topishmoq, ertak, lapar, doston kabi janrlararo; ijtimoiy, iqtisodiy, siyosiy-ma’rifiy mavzulararo filtrlash tugmalari kiritilgan bo‘lib, ushbu yuklangan ma’lumotlardan qipchoq, qarluq, o‘g‘uz shevalari qo‘llanish chastotasi, shevalarning mavzuviy qamrovi, umumiy va xususiy funktsionalligi aniqlanadi [2].

2-rasm.

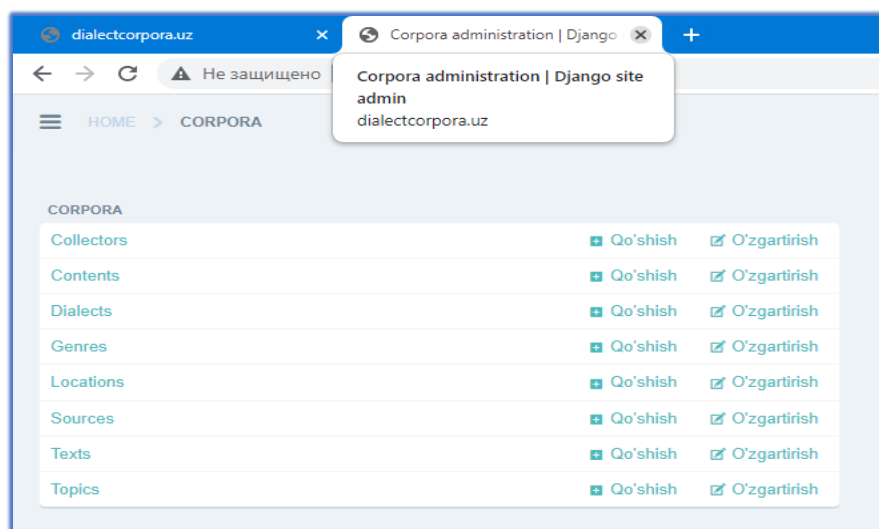


Izoh: o‘zbek milliy shevalar korpusi interfeysida filtr parametrlari.

Shevalar korpusining ichki oynasi, avvalo, autentifikatsiya (bu o'zini tekshirish jarayoni), avtorizatsiya (bu kirish huquqini tekshirish jarayoni) [16] maydonida groups (guruhlar) va users (korpus boshqaruvchisi) qismalaridan iborat va users maydoni subdialektalkorpusi admini tomonidan kengaytirilishi mumkin.

Subdialektalkorpus ichki tizimi quyidagi maydonlarni o'z ichiga oladi. Bular: collectors, dialect, jenre, locations, sources, texts, topis kabilar va har bir tizim o'z ichki qismida ma'lumotlar ketma-ketligi saqlaydi.

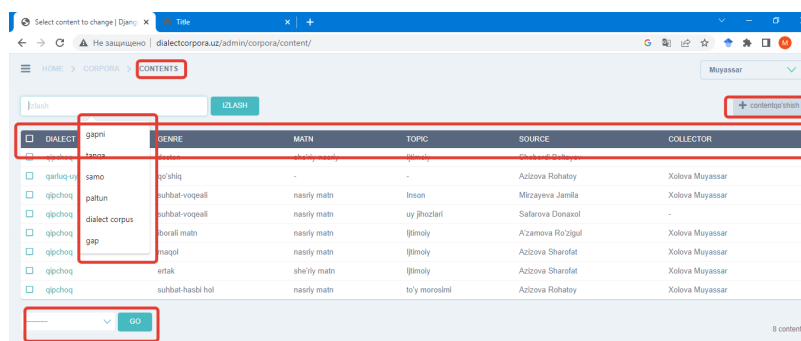
3-rasm.



Izoh: O'zbek milliy korpusining ichki tizimi

O'zbek milliy korpusining ichki tizimida collectors – shevaga oid matnlarni yig'uvchi shaxslar haqidagi to'liq ma'lumot bersa, dialect – sheva matning qaysi lahjaga mansubligi haqidagi ma'lumotni saqlaydi. Jenre maydoni orqali sheva matni qaysi janrga tegishliligi aniqlanadi va joylashtiriladi, Sourse maydoni orqali esa sheva matni yozib olinadigan shaxs haqida yoki yozma manbalar kiritiladi, agar xatoliklar bo'lsa, o'zgartirishlar tugmasi orqali o'zgartiriladi. Topics maydoni orqali sheva matni qaysi mavzuga oidligi haqidagi ma'lumot kiritiladi. Locations maydoni shevaga doir matnlarni O'zbekiston Respublikasining barcha viloyatlar, shaharlar, tumanlar, qishloqlar, mahallalar kesimida kiritish imkonini beruvchi tizim hisoblanib, unda barcha hududlardan yig'ilgan ma'lumotlarning hududiy mansubligiga ko'ra ajratiladi. Hamma maydonlarni o'z ichida jamlagan maydon content nomi ostida – dialect, jenre, sourse, topic, collectors kabi submaydonlarni o'z ichida jamlagan bo'lib, yuklanadigan matnning qipchoq, qarluq, o'g'uz shevalari tegishliligi, ayni davrga qadar 14 ta janrga, she'riy, nasriy, aralashliligi, manbasi, mavzulardan biriga, sheva matnlarini jamlovchilar haqida to'la ma'lumotlarni saqlaydi.

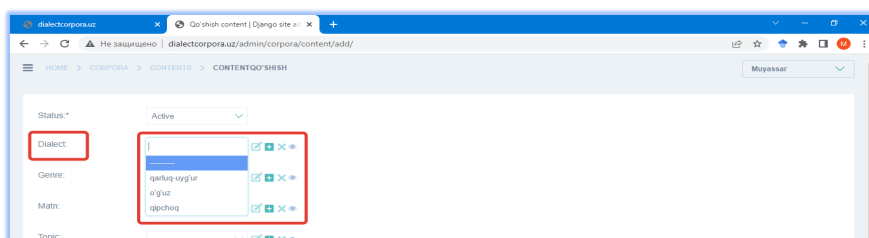
4-rasm.



Izoh: o'zbek milliy shevalar korpusining ichki contenti

Sheva matnini yuklash ketma-ketligi. <http://dialectcorpora.uz/admin/corpora/content/> ga kiriladi va content qo'shish maydoni ochiladi. Har bir maydon uchun alohida tayyor yacheykalar ajratilgan va ularni tanlash imkoniyati mavjud. Masalan, dialect yacheykasi bosilganda uch: qarluq, qipchoq, o'g'uz lahjalari yacheykalari taqdim etiladi va yuklovchi sheva matni mansub bo'lgan lahja tanlanadi: dialect – qipchoq, dialect – qarluq, dialect – o'g'uz va hokazo shaklida.

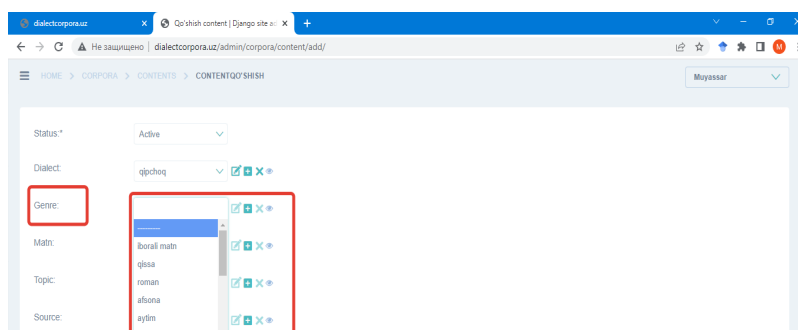
5-rasm.



Izoh: shevalar korpusida matnning shevaga oidligiga ko'ra ajratilgan maydon.

Genre maydoni uchun ajratilgan yacheykasi tanlanganda avtomatik tayyorlangan janrlar taqdim etiladi va sheva matni uchun mos janr tanladi, ya'ni masalan, genre – iborali matn, genre – suhbat, genre – qissa, genre – afsona va hokazolar. Shevalar korpusi ustida ishlashning boshlang'ich nuqtasida barcha janrlar qamrab olinmagan bo'lsa, korpus tizimi kiritilmagan janrlarni keyinchalik ham qo'shib borishni ta'minlaydi va bu holat korpusning kengayish imkoniyati cheksizligini ifodalaydi.

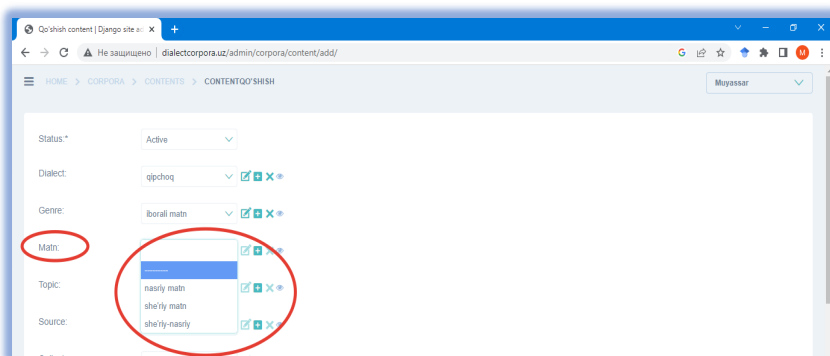
6-rasm.



Izoh: shevalar korpusida matnning janrga oidligiga ko'ra ajratilgan maydon.

Keyingi bosqich sheva matnining she'riy, nasriy va aralash yacheykalaridan birini tanlashdir.

7-rasm.

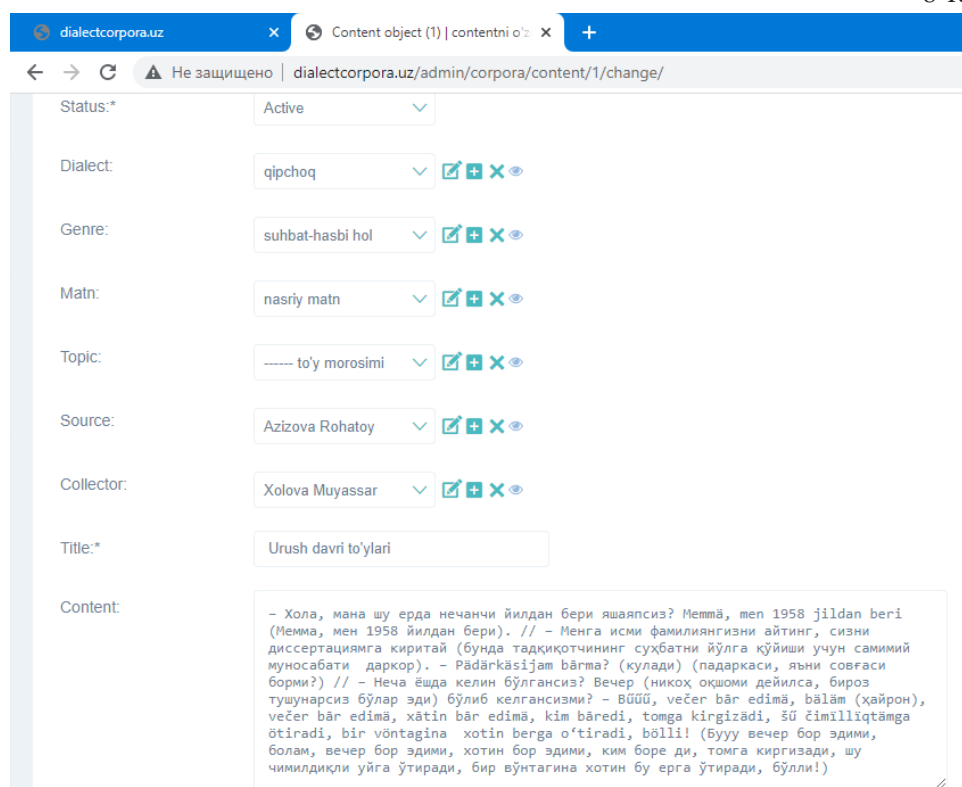


Izoh: shevalar korpusida matn shakli uchun ajratilgan yacheyka.

Kollektor va respondent orasida olib borilgan suhbatning mavzusini kiritish uchun topic (sheva matni mavzusi) yacheykasi tugmasi bosiladi va ajratilgan mavzuviy maydonlar xolonimik va meronimik namoyon bo'ladi. Kollektor mavzuviy

xususiylik mavjud bo‘lganda meronimik mavzularni tanlaydi, mavzularda umumiylik kuzatilganda xolonim mavzular tanlanadi. Ya’ni suhbat mavzusi o‘z ichki uzvlaridan suhbat- hasbi-hol va suhbat –voqeali qismlardan iborat bo‘lib, respondent A zizova Rohatoy tomonidan berilgan urush davrida bo‘lib o‘tgan to‘ylar haqidagi xotiralari bilan bog‘liq ma’lumotlar meronimik mavzuga biriktirilgan holda (topic: {ijtimoiy} {[marosim]} {[to‘y marosimi]}/ title (sarlavha): urush davri to‘ylari) kiritiladi: – Хола, мана шу ерда нечанчи йилдан бери яшяяпсиз? Мемма, мен 1958 жилдан бери (Мемма, мен 1958 йилдан бери). //– Pädärkäsijam bärma? (кулади) (падаркаси, яъни совғаси борми?) // – Неча ёшда келин бўлгансиз? Вечер (никоҳ оқшоми дейилса, бироз тушунарсиз бўлар эди) бўлиб келгансизми? – Вүүү, вечер бәр едимә, бәләм (хайрон), вечер бәр едимә, хәтин бәр едимә, ким бәреді, томга киргизәді, шү чиміллиқтәмга өтирадi, бир вөнтагина хотин берга о‘тирадi, бөлли! (Бууу вечер бор эдимі, болам, вечер бор эдимі, хотин бор эдимі, ким боре ді, томга киргизади, шу чимилдиқли уйга ўтиради, бир вўнтагина хотин бу ерга ўтиради, бўлли!) [12]

8-rasm.



The screenshot shows a web browser window with the URL `dialectcorpora.uz/admin/corpora/content/1/change/`. The page displays a form for editing content with the following fields:

- Status: Active
- Dialect: qipchoq
- Genre: suhbat-hasbi hol
- Matn: nasriy matn
- Topic: ----- to'y morosimi
- Source: Azizova Rohatoy
- Collector: Xolova Muyassar
- Title: Urush davri to'ylari
- Content: - Хола, мана шу ерда нечанчи йилдан бери яшяяпсиз? Мемма, мен 1958 жилдан бери (Мемма, мен 1958 йилдан бери). // – Менга исми фамилиянгизни айтинг, сизни диссертациямга киритай (бунда тадқиқотчининг суҳбатни йўлга қўйиши учун самимий муносабати даркор). – Pädärkäsijam bärma? (кулади) (падаркаси, яъни совғаси борми?) // – Неча ёшда келин бўлгансиз? Вечер (никоҳ оқшоми дейилса, бироз тушунарсиз бўлар эди) бўлиб келгансизми? – Вүүү, вечер бәр едимә, бәләм (хайрон), вечер бәр едимә, хәтин бәр едимә, ким бәреді, томга киргизәді, шү чиміллиқтәмга өтирадi, бир вөнтагина хотин берга о‘тирадi, бөлли! (Бууу вечер бор эдимі, болам, вечер бор эдимі, хотин бор эдимі, ким боре ді, томга киргизади, шу чимилдиқли уйга ўтиради, бир вўнтагина хотин бу ерга ўтиради, бўлли!)

Izoh: shevalar korpusida content yacheykalarining to‘ldirilgan holati.

Shevalar korpusining ichki tizimida barcha ycheykalarga qo‘shish, o‘zgartirish, va olib tashlash mumkin. Bu holat, ayniqsa, source (manba: respondent) va collector (sheva ma’lumotlari yig‘uvchi) bilan bog‘liq holatlarda yaqqol ko‘zga tashlanadi, ya’ni korpus tizimida oldingi respondentlar va kollektor haqidagi ma’lumotlar saqlanadi. Sheva matnlarini yuklash jarayonida undan foydalanish imkoni mavjud. Agar yangi respondentlar haqidagi ma’lumotlarni yuklash kerak bo‘lsa, source yacheykasiga tegishli ikkinchi tugma bilan yangi oynaga o‘tiladi va kollektor tomonidan yig‘ilgan respondentlar haqidagi yangi ma’lumotlar ma’lumotlar korpusining ichki tizimiga kiritiladi. Ya’ni biz uchun source qo‘shish oynasi taqdim etiladi va unda quyidagicha

yacheykalar: fullname (respondent haqidagi to'liq ismi familiyasi va otasining ismi); sex (respondentning jinsi: female (ayol), male (erkak)); education (bunda respondentning ma'lumotiga doir ma'lumotlar aks etadi: school (maktabning 1-8 sinf oralig'i), high school (maktabning 9-11 sinfi), college (kollej), bachelor (bakalavr), master (magistr), locations (respondent yashovchi hudud (butundan qismga qarab): viloyat, shahar, tuman, qishloq, mahalla va hokazo) ko'rinadi.

9-rasm.

Izoh: shevalar korpusida source (respondent) haqida ma'lumot kiritish maydoni.

Collector haqida yangi ma'lumotlar joylash uchun ochilgan yangi oynada fullname (shevaga oid matn yig'uvchining ismi, sharifi, otasining ismi), equipment (matn yoki intervyu yozib oligan uskuna: diktafon, qo'lyozma, fonografik apparat va hokazo).

10-rasm.

Izoh:Kollektor haqida ma'lumotlar kiritish oynasi.

Xulosa shuki, korpus tizimida umumiy ma'lumotlar bazasi bir yoki ikki collector tomonidan shakllantirilishi mumkin emas, nafaqat Surxondaryo shevalari, balki O'zbekistonning barcha hududlarida mavjud sheva birliklarini bir korpusga solish uchun juda katta dialektolog va amaliy filologiya sohasida tadqiqot olib boruvchi olimlar hamda dialektologik amaliyot o'tab sheva matnlarini yig'ish orqali tadqiqotlar uchun asos yig'uvchi talabalar guruhi kerak bo'ladi.

#### FOYDALANILGAN ADABIYOTLAR:

Abdulkhakovna K. M. Transcription in the corpus of the uzbek national boysun dialect (on the example of Baysun district» j» dialects) //Journal of Critical Reviews. – 2020. – T. 7. – №. 5. – C. 844-847./

Abdulkhakovna K. M. Software and data entry principles of the online dialect database // current research journal of philological sciences. – 2022. – T. 3. – №. 11. – C. 46-53.

Абдурахмонова Н. “О‘zbek tili korpusini yaratishda lingvistik annotatsiyalash tamoyillari.” Международный журнал искусство слова 4.1-1 (2021). /https://uzbecorpus.uz/

Areej Alshutayri, Eric Atwell. A Social Media Corpus of Arabic Dialect Text. Mustafa Jarrar, Fadi A. Zaraket, Tymaa Hammouda, Daanish Masood Alavi, Martin Wählisch Lîsan: Yemeni, Iraqî, Libyan, and Sudanese Arabic Dialect Corpora with Morphological Annotations.

<https://www.researchgate.net/publication/327976527>.

Liang Zhao, Eleanor Chodroff. The ManDi Corpus: A Spoken Corpus of Mandarin

Regional Dialects. /Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)/ – P. 1985–1990 Marseille, 20-25. 07. 2022.

Karimov R., Mengliev B. Theoretical fundamentals of uzbek-english parallel corpus / Journal of critical reviews. 2020. Vol. № 7. Is. №17. – P. 907-911. /<https://uzbekcorpus.uz/paralel/>

Ковалевская К. В. Возможности текстового поиска в шведском подкорпусе НРКЯ. Минск. <https://elib.bsu.by/bitstream/123456789/275383/1/.pdf>

Mahmaraimova S. T. Lingvokul'turologija. Tashkent: Ch'ylpon nomidagi nashrijot-matbaa izhodij uji. – 2017.

Mengliyev B. et al. The morphological analysis and synthesis of word forms in the linguistic analyzer //Journal of Language and Linguistic Studies. – 2021. – Т. 17. – №. 1. – С. [558]-564. /<https://uzbekcorpus.uz/autorCorpus/>

Niladri Sekhar Dash. Documentation of a Dialect through the Use of a Dialect Corpus. /<https://www.researchgate.net/publication/348482915/>

Халияров У. М. О'zbek ta'limiy korpusida dialektologik birliklarining ifodasi // Международный журнал искусство слова. – 2021. – Т. 4. – №. 4. /<https://uzschoolcorpora.uz/>

Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси. – 2018./

Холова М. А. Шевалар корпусининг жанрий теглашнинг лингвистик асослари // Международный журнал искусство слова. – 2021. – Т. 4. – №. 4.

Холова М. А. О'zbek shevalarining standart tilga munosabati va til tizimidagi o'rni (Boysun tumani “j” lovchi shevalar misolida) //Международный журнал искусство слова. – 2020. – Т. 3. – №. 5.

Холова М. А. Boysun shevasi korpusi: Ibtido (Boysun tumani “j” lovchi shevalar misolida) // Инновации в педагогике и психологии. – 2020. – №. SI-2№ 5.

Xidirova I., Xidirova N. Gender Characteristics of Family Speech Speech (On the Example of the Uzbek Family) //EUROPEAN JOURNAL OF INNOVATION IN NONFORMAL EDUCATION. – 2021. – Т. 1. – №. 2. – С. 196-199.

Shodiyevna K. N. Linguistic study of some Uzbek ethnographisms expressing wedding ceremonies //International Journal of Psychosocial Rehabilitation. – 2020. – Т. 24. – №. 5. – С. 3735-3740.

<https://www.texnoman.uz/cgi-sys/suspendedpage.cgi>