

XORAZM SHEVASIDAGI MATNLARNI O'ZBEK ADABIY TILIGA O'GIRISH DASTURINI YARATISHNING AMALIY AHAMIYATI

Xusainova Zilola Yuldashevna,

Filologiya fanlari bo'yicha falsafa doktori

xusainovazilola@navoiy-uni.uz

ToshDO'TAU

Bekchanova Munira Oybek qizi,

Kompyuter lingvistikasi yo'nalishi 3-kurs talabasi

bekchanovamunir@gmail.com

ToshDO'TAU

Annotatsiya. Bugungi kunda tilshunoslikda hududiy shevalarning tizimli o'rganilishi va elektron resurslar shaklida saqlanishi katta ahamiyat kasb etmoqda. Xorazm shevasi o'zining fonetik, morfologik va leksik xususiyatlari bilan umumiy adabiy o'zbek tilidan sezilarli farq qiladi. Xorazm shevasi korpusining yaratilishi o'zbek tilining dialektologik xilma-xilligini saqlash va o'rganishda muhim qadam hisoblanadi. Korpus yaratish jarayonida audio yozuvlar, transkripsiyalar va matnli ma'lumotlar tizimli ravishda yig'ilib, ularni annotatsiya qilish muhim ahamiyat kasb etadi. Matnlar turli manbalardan (suhbatlar, folklor namunalari, she'riy va nasriy asarlar, audio yozuvlar) to'planadi, ularning lingvistik va etnografik jihatlari belgilanadi. Ushbu maqolada Xorazm shevasidagi matnlarni o'zbek adabiy tiliga avtomatik tarzda o'girish imkoniyatlarini tadqiq etish va bu yo'nalishda dasturiy ta'minot yaratish masalasi ko'rib chiqiladi.

Аннотация. Сегодня систематическое изучение региональных диалектов и сохранение их в виде электронных ресурсов имеет большое значение в языкознании. Хорезмский диалект существенно отличается от общелитературного узбекского языка по своим фонетическим, морфологическим и лексическим признакам. Создание корпуса хорезмских диалектов является важным шагом в сохранении и изучении диалектологического многообразия узбекского языка. В процессе создания корпуса важно систематически собирать и аннотировать аудиозаписи, транскрипции и текстовые данные. Тексты собираются из различных источников (интервью, образцы фольклора, поэтические и прозаические произведения, аудиозаписи), определяются их лингвистические и этнографические аспекты. В данной статье рассматривается возможность автоматического перевода текстов на хорезмском диалекте на узбекский литературный язык и разработка программного обеспечения в этом направлении.

Abstract. Today, the systematic study of regional dialects and their preservation in the form of electronic resources is of great importance in linguistics. The Khorezm dialect is significantly different from the general literary Uzbek language with its phonetic, morphological and lexical features. The creation of a corpus of the Khorezm dialect is an important step in preserving and studying the dialectological diversity of the Uzbek language. In the process of creating a corpus, audio recordings, transcriptions and textual data are systematically collected and annotated. Texts are collected from various sources (interviews, folklore samples, poetic and prose works, audio recordings), their linguistic and ethnographic aspects are determined. This article considers the issue of researching the possibilities of automatically translating texts in the Khorezm dialect into the Uzbek literary language and creating software in this direction.

Kalit soʻzlar. *Dialektal korpus, Xorazm shevasi, NLP, mashinali oʻqitish, lingvistik oʻzgartirish.*

Dialektal nutqni adabiy tilga moslashtirish kompyuter lingvistikasi va tabiiy tilni qayta ishlash (NLP) sohasida dolzarb yoʻnalishlardan biri hisoblanadi. Dialektal nutqning adabiy til bilan muvofiqlashtirilishi til birligini mustahkamlash, raqamli til resurslarini kengaytirish va mashinali tarjima tizimlarining sifatini oshirish uchun muhim ahamiyat kasb etadi. Xorazm shevasidagi matnlarni oʻzbek adabiy tiliga oʻgirish dasturini yaratish oʻzbek tilining dialektal xilma-xilligini saqlash va kengroq auditoriyaga yetkazishda muhim ahamiyatga ega. Xorazm shevasining oʻziga xos fonetik, morfologik va sintaktik xususiyatlari uning adabiy tilga taʼsirini va oʻziga xosligini yanada kuchaytiradi. Xorazm viloyatining turli hududlarida yashovchi xalqning tiliga oid toʻplangan maʼlumotlarni tizimli ravishda tahlil qilish, oʻziga xos dialektal korpus yaratish zarurati paydo boʻlgan. Xorazm shevasida yaratilgan matnlarni oʻzbek adabiy tiliga oʻgirish uchun zamonaviy texnologiyalar va lingvistik yondashuvlarni uygʻunlashtirish muhim hisoblanadi. Ushbu tadqiqot natijasida amaliy dastur ishlab chiqiladi. Xorazm shevasidagi matnlarni oʻzbek adabiy tiliga oʻgirish dasturini yaratish sheva matnlarini avtomatik ravishda adabiy tilga oʻtkazish orqali taʼlim materiallarini tayyorlashni osonlashtiradi. Shevadagi matnlarni adabiy tilga oʻgirish matnlarning kengroq kontekstda tushunarli boʻlishini taʼminlaydi. Xorazm aholisi, ayniqsa yoshlar, adabiy til bilan muloqotda boʻlishda qiyinchiliklarga duch kelishi, ushbu dastur ularga kommunikatsiya vositasini yengillashtiradi.

Shuningdek, maqolada Xorazm shevasining fonetik, morfologik va sintaktik xususiyatlari, bu xususiyatlarning dialektal korpusga qanday aks etishi va uning Xorazm shevasida qanday qoʻllanishi haqida soʻz boradi. Xorazm shevasining

dialektal korpusini yaratish, o'g'uz she'valarining o'ziga xosligini aniqlash va adabiy tilda ularning rolini tahlil qilishga yordam beradi.

Dunyo tilshunosligida XXI asrga kelib, korpus muammolarini o'rganish harakati kuchaydi. Hozirda til korpuslarining ijtimoiy ahamiyati, tuzish tamoyillari, lingvistik teglar tizimini ishlab chiqish, avtomatik tahlil qilish masalalari bo'yicha maqsadli tadqiqotlar olib borilmoqda. Shu nuqtai nazardan olib qaraganda globallashtirish jarayonida ayrim tillar qo'llanishining kengayishi boshqa tillarning saqlanib qolishiga xavf solmoqda, shuning uchun mazkur korpuslar tilning yashab qolishi uchun imkoniyatlardan biri sifatida qarasak bo'ladi. Korpus borasida fanda o'ziga xos ilmiy ishlar qilinmoqda. Jahon tilshunosligida raqamli axborot vositalaridan foydalanish qulayligi hamda internet orqali milliy tillarni o'zlashtirishga bo'lgan harakat ta'sirida milliy korpuslarda shevalar bazasini yaratish talabi o'smoqda. Natijada tilda mavjud dialektlarni nafaqat lingvistik, balki amaliy jihatdan tadqiq etish, bu orqali sheva birliklarini saqlab qolish, avloddan-avlodga yetkazish imkoniyatining ko'lami kengaymoqda.

Shevadagi so'zlar jamlanmasi va ularning elektronlangan varianti dialektal korpus deyiladi. Dialektal korpusning arxitekturasi, umuman olganda, bir nechta asosiy qism va bosqichlardan tashkil topgan bo'lib, ularning har biri ma'lumotlar yig'ish, tasniflash va tahlil qilish jarayonlarini o'z ichiga oladi. 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 Xorazm shevasiga xos matnlar o'zbek adabiy tiliga o'girish dasturini yaratish uchun shevadagi so'zlar yozma manbalardan to'plandi. Matnlar raqamlashtirildi, ortiqcha belgilardan tozalandi hamda morfosintaktik tahlil uchun tayyorlandi. Sheva-adabiy til o'rtasidagi morfologik farqlarni aniqlash uchun qoidaga asoslangan yondashuvlar ham modelga integratsiya qilindi. Annotatsiya jarayonida matnlar shevaga xos fonetik va morfologik xususiyatlarga asoslanib belgilandi. O'zbek adabiy tiliga o'girish uchun dastlab qoidaviy (rule-based) yondashuv asosida o'girish qoidalari ishlab chiqildi.

Dialektal korpusning amaliy ahamiyati: umumiy qilib olganda, shevaga oid so'zlarni saqlab qolish uchun elektron bazaga qo'shish hisoblanadi. Ham so'z saqlanib qolinadi, ham leksemalar soni bittaga ko'payadi. O'zbek xalq shevalari juda ko'p va xilma-xil bo'lib, ularning ayrim qismlarigina fonetik va grammatik jihatdan o'rganilgan. O'tgan asrlarda yaratilgan ilmiy asarlar bunga yaqqol misol

bo'la oladi. Jumladan, F.Abdullayev “O'zbek tilining Xorazm she'valari”, N.Rajabov “O'zbek she'vashunosligi”, V.V.Reshetov “Тюркологический сборник” va boshqa olimlarning ilmiy ishlarini aytishimiz mumkin. Shu jumladan, F.Abdullayev Xorazm shevalariga oid namunalarni xalq orasidan, ko'chalarda yurib jonli she'valardan tuzgan. O'g'uz lahjasiga oid Xorazm shevalari ham yana bir qancha mayda sheva guruhlariga bo'lingan. Bunga turlicha sabablar bo'lishi mumkin. Ular bir-biridan nafaqat geografik joylashuv jihatidan, balki sheva tomondan ham birmuncha farq qiladi. Demak, shevalar umumiy bir lahja tarkibida birlashsa ham, fonetik, morfologik jihatlaridan turlicha bo'lishi mumkin. Na'muna qilib Urganch va Xiva hududlari shevalarini olishimiz mumkin. Professor E.D. Polivanov tasniflashicha, “o'g'uz lahjasi” o'z ichiga ikki guruh shevalarini olar ekan:

1. Janubiy Xorazm guruhi shevalari (Xiva, Urganch, G'azavot, Shovot, Xazorasp, Yangiariq, Xonqa kabi) kiritilgan.
2. Shimoliy o'g'uz guruhi shevalari (Iqon-Qorabuloq kabi) kiritilgan ekan.

Yuklamalar Xorazm shevasida tubdan farq qiladi. Misol uchun “-da” yuklamasi kuchaytiruv-ta'kid ma'nosini bildiradi. Xorazm shevasida “don” qo'shimchasiga to'g'ri keladi: *Sovg'a bermaganingdan keyin xafa bo'ladi-da. Sovg'a barmaganingdan keyin xofo bo'ladidon.*

“-ku” qo'shimchasi ham kuchaytiruv-ta'kid yuklamasi turiga kiradi, Xorazmda bu qo'shimcha “-qu” shaklida keladi: *Hamma gaplarimni senga aytdim-ku! Hamma gaplarimi songo oytdm-qu.*

Ko'makchilarga keladigan bo'lsak, ular ham Xorazm she'vasida turlicha ifodalanadi: 1) “so'ngra” ko'makchisi vazifadosh ko'makchilar sirasiga kirib, Xorazm she'vasida “o'sin” shaklida keladi: *Tushda ishdan kelaman, so'ng uyda dam olaman. Abetda ishdan galaman, o'sin uyda dam olomon.*

Xorazm shevasini korpusini yaratish uchun avvalo ma'lumot yig'ildi ya'ni 600ta shevaga oid so'zlar yig'ib chiqilgan. Ushbu jadvalda to'rtta tumanda so'zlarning qanday aytilishi alohida ajratilib ko'rsatilgan. Korpusda foydalanilgan dataset turli sohalarga doir va turli so'z turkumlari oid so'zlardan foydalanib yig'ilgan.

№	Shevadagi soʻz	Adabiy shakl	soʻz turkumi	Urganch	Xiva	Qoʻshkoʻpir	Gurlan
1	yimirta	tuxum	ot	tuxum	yimirta	yimirta	tuxum
2	gashir	sabzi	ot	gashir	gashir	gashir	gashir
3	timittin	binoyiday	sifat	timittin	timittin	timittin	timittin
4	parahatchiliq	tinchlik	ot	parahatchiliq	parahatchiliq	parahatchiliq	parahatchiliq
5	navlin	bilmadim	fe'l	navlin	navlin	navlin	navlin
6	zangi	narvon	ot	zangi	zangi	zangi	zangi
7	opo	ona	ot	opo	opo	opo	opes
8	aka	ota	ot	aka	aka	aka	akes
9	ejashdi	hazillashdi	fe'l	ejashdi	ejashdi	ejashdi	ejashdi
10	chapak	qarsak	ot	chapak	chapak	chapak	chapak
11	nerda	qayerda	olmosh	nerda	nerda	nerda	nerda
12	man	men	olmosh	man	man	man	men
13	eshik	eshik	ot	eshik	qopi	qopi	eshik
14	qarinja	chumoli	ot	qarinja	qarinja	qarinja	qarinja
15	pitta	ozgina	ravish	pitta	pitta	pitta	pitta
16	borosonmi	borasanmi	fe'l	borosonmi	borosommi	borosomo	borosomo
17	go'sht	go'sht	ot	go'sht	et	go'sh	go'sht

Yuqoridagi ma'lumotlardan kelib chiqib Pythonida kod yozilgan.

```
print("\tO'zbek shevalari lug'ati(Xorazm shevasi)")
soz = input("So'z kiriting=").lower()
soz1 = ["angak", "aka", "apka", "arka", "adik", "alashtirma", "arik", "asirak", "barak", "bajik", "bangi",
"biyi", "bon", "bidirdamoq", "bordi-galdi", "bo'kchimoq", "bol", "bug'oz", "chapak", "chitmak", "chiya",
"cholzur", "chug'ilchi", "digirjak", "doyi", "dim", "do'ddi", "duvalang", "dish", "dingan", "dapa", "domar",
"deshon", "diri", "dobbi", "do'shik", "do'shak", "dik", "elot", "ejashmoq", "guch", "galishmoq", "garak",
"galin", "gashir", "gal", "guyov", "guppi", "get", "gechchi", "gizlanmajak", "go'z", "go'rmoq", "go'shanga",
"gurung", "guldirmomo", "halpa", "horvuz", "hovor", "hurkak", "ijjon", "ich qismoq", "ilik", "ino",
"inna", "ishu", "indi", "ichon", "jakka-jukka", "jinnoddin", "jo'nqi", "jo'ro", "jo'gi", "jolli", "katman",
"ko'pin", "kichchi", "kirpi", "kisa", "kadi", "kat", "loppa at", "lo'ppi", "mongloyi qoro", "mo'chchi",
"mijina", "navlin", "novvi", "nishatin", "nera", "nerda", "nichik", "ojji", "opo", "oston", "orva",
"ongiyo'q", "pishik", "potya", "polos", "po'm", "qorinja", "qumursqa", "qopoq", "qopshirma", "sollanmoq",
O'zbek shevalari lug'ati(Xorazm shevasi)
So'z kiriting=barak
○ Adabiy tarjimasini: chuchvara
So'z turkumi: ot
```

Yuqoridagi kodni ishga tushirganimizdagi Xorazm shevasiga oid soʻz kiritiladi natijada shevadagi soʻzning adabiy tilda qaysi soʻzga teng kelishi va uning qaysi soʻz turkumiga tegishli ekanini chiqarib beradi. Olingan natijalar Xorazm shevasidan oʻzbek adabiy tiliga avtomatik oʻgirish mumkinligini isbotladi.

Shunga qaramay, tadqiqot davomida baʼzi cheklovlar ham aniqlangan: shevada mavjud boʻlgan leksik arxazimlar va lokal semantik oʻzgarishlar modelni chalkashtirishi mumkin. Xorazm viloyatining oʻzida ham turli shevalar mavjud. Bunga Gurlan va Yangibozor shevalarini misol qilishimiz mumkin. Gurlan tumanining Amudaryo bilan chegaradosh boʻlgan joylarida “j” lashishlarga duch kelinadi. Shimoliy Xorazmdan Xiva tomonga qarab “j”lashish kamayib boradi. Bu

shevada ayrim leksikologik o'ziga xosliklar ham mavjud. Xorazm shevasida asosan olmoshlarning variantlari ko'p va adabiy tildan tubdan farq qiladi.

Shevalarda ham omonimlikni kuzatishimiz mumkin. Masalan “*kadi*” so'zini ikki xil ma'nosi bor. Birinchi ma'nosi “*qovoq*” bo'lsa, ikkinchi ma'nosi “*sevgan qizi(yigiti)*” ma'nosida keladi. Korpusni yaratishda frazeologiyalarni ham kiritish muhim. Adabiy tildagi frazeologiyalarning ma'nosi Xorazm shevasida umuman boshqa ma'noda keladigan holatlarni kuzatishimiz mumkin.

bunga	binga
bunga	minga
unga	vunga
undan	vundan
ana shu	onoshu
mana shu	inashu
ana	ono
mana	ina
menga	monga
senga	songa

Og'zaki nutqda juda charchamoq ma'nosini bildiruvchi “*Og'zini suvi oqdi*” iborasi mavjud bo'lib, bu ibora o'zbek adabiy tilidagi “*Holdan toymoq*” iborasiga teng keladi. Shuningdek, ushbu iboraning Xorazm shevasida “*Oyoqinnan o't chiqdi*” sinonim iborasi ham uchrab turadi. Bu iboraning birinchi iboradan deyarli farqi yo'q. Misol uchun, “*to'y o'yda ish atavarip, oyoqimnon o't chiqdi*” yoki “og'zimni suvi oqdi” birikmasini ham qo'llash mumkin.

Jumladan, Qo'shko'pir aholi shevasi nutqida bu iboralar bilan yana bitta sinonimlik hosil qiluvchi ibora borki, bu ibora “*Go'zzini go'ki so'kildi*” tarzida qo'llanadi. Bu ibora ham “*qattiq charchadi*” ma'nosini bildiradi. Garchi uning ma'nosi ham yuqoridagi iboralar bilan deyarli bir xil bo'lsa-da, bu iboraning ahamiyatli jihati ham borki, ibora sal tor ma'noda ishlatilib, faqat ko'z orqali qilinadigan jismoniy harakatlarga nisbatangina(kompyuterda ishlash, tikuv mashinasida ishlash) ishlatiladi. Misol uchun, *Dinmin moshinda tikavarip go'zzimni go'ki so'kildi* (Qo'shko'pir sheva og'zaki nutqidan).

Shuningdek ushbu shevalarda “*Borg'ono boldiz, galgana yengga*” kerakli va keraksiz joylarga ham boraveradigan, hamma joyda topiladigan, odamga nisbatan qo'llaniladi.

“*O'Imagay bi bolo hammeda bo bi borg'ono boldiz, galgana yenggachilik atip.*”

Xulosa.

Xorazm shevasidagi matnlarni o'zbek adabiy tiliga avtomatik tarzda o'girish dasturini ishlab chiqish bo'yicha olib borilgan tadqiqot natijalari dialektal korpuslar va tabiiy tilni qayta ishlash sohasida muhim ilmiy va amaliy ahamiyatga ega ekanligini ko'rsatdi. Yaratilgan dastur dialektal farqlarni sezilarli darajada aniqlay olgan va adabiy til me'yorlariga moslasha olgan. Biroq, yuqori aniqlikka erishish uchun kengroq, sifatli parallel korpuslar va tilga xos xususiyatlarni yanada chuqurroq o'rganishga ehtiyoj mavjud. Tadqiqot natijalari O'zbek tilining hududiy variantlarini standartlashtirish, madaniy merosni raqamlashtirish va mashinali tarjima tizimlarini takomillashtirish yo'lida mustahkam asos yaratdi.

Foydalanilgan adabiyotlar:

1. Elov B.B., Xusainova Z.Y. Katta til modellari / “O'zbekiston: til va madaniyat. Kompyuter lingvistikasi”. 2024. Vol 2(6). – P. 78-92.
2. Elov B., Abdullayeva O. Dialektal korpusni qurish va uning ahamiyati. // O'zbek folklori va shevalari tadqiqotlari: amaliyot, metodologiya, yangicha yondashuv.
3. Palvonnazarova H., Shamuratova S. O'g'uz shevalarida yordamchi so'z turkumlarini qo'llanilishi. Eurasian Journal of Social Sciences, Philosophy and Culture.
4. Ibragimov Y., Ibragimova Zamira. O'zbek tilining dialektal korpusini yaratish tamoyillari. // O'zbek Tilshunosligi va Adabiyotshunosligi Taraqqiyoti, Xalqaro Hamkorlik Masalalari.
5. Elov B., Hamroyeva Sh., Alayev R., Xusainova Z., Yodgorov U. O'zbek tili korpusi matnlarini qayta ishlash usullari / Raqamli Transformatsiya va Sun'iy Intellekt ilmiy jurnali, October 2023. Volume 1, Issue 3.
6. Xusainova, Z., & Yangibayeva, S. (2024). Til korpusi turlari. / Uzbekistan: Language and Culture, 2(2).
7. Abdullayev F. O'zbek tilining Xorazm she'valari.
8. Elov, B. B., Hamroyeva, S., & Xusainova, Z. (2022). NLP (tabiiy tilga ishlov berish) ning vazifalari va zamonaviy yondashuvlar. / TerDU, filologik tadqiqotlar: til, adabiyot, ta'lim, 5-6.
9. Paquot M., Th.Gries S. Corpus Architecture. A Practical Handbook of Corpus Linguistics.
10. Boberg Ch., Nerbonne J., Watt D. Corpus-Based Approaches to Dialect Study. The Handbook of Dialectology.