Available online at https://refocus.uz/

RESEARCH FOCUS
INTERNATIONAL SCIENTIFIC JOURNAL

Vol. 5, Issue 10, 2026
ISSN 2181-3833 | Uzbekistan
https://refocus.uz/

RESEARCH ARTICLE • PEER-REVIEWED • OPEN ACCESS • SERIES D • HAC/OAK 10.00.00 — PHILOLOGICAL SCIENCES • UDC: 811.512.133’28:81’32

Integrating a written dialectological source into a digital dialect dictionary: method and quality control (the case of F. Musayeva’s dictionary “O‘zbek xalq so‘zlari”)

Yozma dialektologik manbani raqamli sheva lug‘atiga kiritish: metodika va sifat nazorati (F. Musayevaning «O‘zbek xalq so‘zlari» lug‘ati misolida)

Включение письменного диалектологического источника в цифровой словарь говоров: методика и контроль качества (на материале словаря Ф. Мусаевой «O‘zbek xalq so‘zlari»)

SALOYDINOVA Maxliyo Abduvali qizi[1]

[1] *Scientific and Technical Journal of “Uzbekhydroenergo” JSC, Tashkent, Uzbekistan*

ARTICLE INFO

ARTICLE TYPE
Research Article

LANGUAGE OF FULL TEXT
O‘zbekcha (UZ)

SCIENTIFIC FIELD / OAK
☑ 10.00.00 – Philological Sciences
Filologiya fanlari
Series D

UDC / UDK
811.512.133’28:81’32

ARTICLE HISTORY
Received: 2026-10-08
Revised: 2026-10-21
Accepted: 2026-10-24
Published: 2026-10-31

DOI
https://doi.org/10.66073/researchfocus.v5i10.1019

CONTACT
ReFocusJournal@gmail.com

LICENSE
Creative Commons Attribution 4.0 International (CC BY 4.0). This is an open access article distributed under the terms of the CC BY 4.0 licence.

AUTHOR, TITLE AND ABSTRACT METADATA (EN / UZ / RU)

ENGLISH

SALOYDINOVA Maxliyo Abduvali qizi[1]
[1] *Scientific and Technical Journal of “Uzbekhydroenergo” JSC, Tashkent, Uzbekistan*

Abstract. The article describes a method for bringing a printed dialectological dictionary into a digital dialect dictionary and atlas, together with the quality control of this process. The material is F. Musayeva’s dictionary “O‘zbek xalq so‘zlari” (2023); the target system is the O‘zSheva platform (umidliyoshlar.uz). The method comprises parsing the entries, classifying the regional labels and linking them to provinces, formulating the meanings independently, matching them with existing concepts, citing the source with page-level precision for each variant and assigning a confidence level. Of the 2,084 entries extracted from the dictionary, 548 carried a label in parentheses; 454 of them were linked to at least one province, while 94 labels could not be linked to any province (Latin biological names, places outside Uzbekistan, place names shared by several locations and one district name missing from the district list). The import produced 423 new entries (394 concepts), added 17 words to existing entries as regional variants and created 460 regional variants in total. Quality control revealed 43 errors of five types in Latin-to-Cyrillic transliteration, which were eliminated through rules and tests. The approach allows a written source to be included in a digital atlas without guessing locations or copying definitions.

Keywords: dialect lexicography, Uzbek dialects, digital dictionary, linguistic atlas, regional label, source citation, transliteration, quality control, open data

O‘ZBEKCHA

SALOYDINOVA Maxliyo Abduvali qizi[1]
[1] *«O‘zbekgidroenergo» aksiyadorlik jamiyati ilmiy-texnik jurnali, Toshkent, O‘zbekiston*

Annotatsiya. Maqolada bosma dialektologik lug‘atni raqamli sheva lug‘ati va atlasiga kiritish metodikasi hamda bu jarayonning sifat nazorati bayon qilinadi. Material sifatida F. Musayevaning «O‘zbek xalq so‘zlari» (2023) lug‘ati, maqsadli tizim sifatida O‘zSheva platformasi (umidliyoshlar.uz) olindi. Metodika lug‘at yozuvlarini ajratish, hudud yorliqlarini tasniflab viloyatlarga bog‘lash, ma’noni mustaqil ifodalash, mavjud tushunchalar bilan moslashtirish, har bir variantga sahifa aniqligidagi iqtibos qo‘shish va ishonch darajasini belgilash bosqichlaridan iborat. Lug‘atdan ajratib olingan 2084 ta yozuvdan 548 tasida qavs ichida yorliq aniqlandi; ulardan 454 tasi kamida bitta viloyatga bog‘landi, 94 tasi esa hech bir viloyatga bog‘lanmadi (lotincha biologik nomlar, O‘zbekistondan tashqaridagi joylar, bir nechta joyga to‘g‘ri keladigan nomlar va ro‘yxatda yo‘q bitta tuman nomi). Natijada 423 ta yangi yozuv (394 ta tushuncha) yaratildi, 17 ta so‘z mavjud yozuvlarga hududiy variant sifatida qo‘shildi, jami 460 ta hududiy variant hosil bo‘ldi. Sifat nazorati lotin–kirill transliteratsiyasida besh turdagi 43 ta xatoni aniqladi, ular qoidalar va testlar bilan bartaraf etildi. Yondashuv yozma manbani hududni taxmin qilmasdan va ta’riflarni

How to cite / Iqtibos keltirish / Как цитировать: SALOYDINOVA, M. A. (2026). Integrating a written dialectological source into a digital dialect dictionary: method and quality control (the case of F. Musayeva’s dictionary “O‘zbek xalq so‘zlari”). Research Focus International Scientific Journal, 5(10), 98–102. https://doi.org/10.66073/researchfocus.v5i10.1019

© 2026 The Author(s). Published by Research Focus International Scientific Journal. Open access article under the CC BY 4.0 licence.

HAC / OAK: On the HAC list of national scientific publications (Academy of Sciences of the Republic of Uzbekistan) for 06.00.00 Agricultural Sciences and 16.00.00 Veterinary Sciences.
O‘zbekiston Respublikasi FA huzuridagi OAK ro‘yxatida: 06.00.00 – Qishloq xo‘jaligi fanlari; 16.00.00 – Veterinariya fanlari.

ko‘chirmasdan raqamli atlasga kiritish imkonini beradi.

Kalit so‘zlar: dialektal leksikografiya, o‘zbek shevalari, raqamli lug‘at, lingvistik atlas, hudud yorlig‘i, manba iqtibosi, transliteratsiya, sifat nazorati, ochiq ma’lumotlar

РУССКИЙ

САЛОЙДИНОВА Махлиё Абдували кизи[1]
[1] *Научно-технический журнал АО «Узбекгидроэнерго», Ташкент, Узбекистан*

Аннотация. В статье описывается методика включения печатного диалектологического словаря в цифровой словарь и атлас говоров, а также контроль качества этого процесса. Материалом послужил словарь Ф. Мусаевой «O‘zbek xalq so‘zlari» (2023), целевой системой — платформа O‘zSheva (umidliyoshlar.uz). Методика включает разбор словарных статей, классификацию территориальных помет и их привязку к областям, самостоятельную формулировку значений, сопоставление с имеющимися понятиями, указание источника с точностью до страницы для каждого варианта и присвоение уровня достоверности. Из 2084 извлечённых из словаря статей в 548 выявлена помета в скобках; 454 из них привязаны хотя бы к одной области, а 94 пометы ни к одной области привязать не удалось (латинские биологические названия, названия мест за пределами Узбекистана, топонимы, относящиеся к нескольким местам, и одно название района, отсутствовавшее в справочнике). В результате создано 423 новые статьи (394 понятия), 17 слов добавлены к имеющимся статьям как региональные варианты, всего получено 460 региональных вариантов. Контроль качества выявил 43 ошибки пяти типов в транслитерации с латиницы на кириллицу; они устранены с помощью правил и тестов. Подход позволяет включать письменный источник в цифровой атлас, не домысливая территориальную привязку и не копируя толкования.

Ключевые слова: диалектная лексикография, узбекские говоры, цифровой словарь, лингвистический атлас, территориальная помета, ссылка на источник, транслитерация, контроль качества, открытые данные

1. KIRISH
Introduction

O‘zbek shevalari leksikasini hujjatlashtirishning asosiy shakli hozirgacha bosma lug‘at bo‘lib kelmoqda. Bir jildli «O‘zbek xalq shevalari lug‘ati» [1] va shevalarning qarluq, qipchoq, o‘g‘uz guruhlariga tasnifi [2] bu sohaning tayanch manbalaridir, yangi tadqiqotlar esa ilmiy to‘plamlarda e’lon qilinmoqda [3]. So‘nggi yillarda shevalarda qo‘llanib kelgan so‘zlarni jamlagan yangi lug‘atlar ham nashr etildi [4].

Raqamli sheva lug‘ati va atlasi bosma lug‘atdan bir muhim talab bilan farq qiladi: unda har bir so‘z shakli aniq hududga bog‘langan bo‘lishi kerak. Busiz so‘zni xaritada ko‘rsatib ham, hududlararo qiyoslab ham bo‘lmaydi. Bosma lug‘atlarda esa hudud turlicha belgilanadi: viloyat nomi to‘liq yoki qisqartirib yoziladi, ba’zan tuman yoki qishloq nomi keltiriladi, ba’zan O‘zbekistondan tashqaridagi joy ko‘rsatiladi. Ko‘p yozuvlarda hudud umuman berilmaydi. Masalan, o‘zbek tili sinonimlarining izohli lug‘atlarida [5; 6] ayrim so‘zlar shevaga xos deb izohlangan, ammo ularning qaysi hududda qo‘llanishi ko‘rsatilmagan.

Bunday manbani raqamli lug‘atga o‘tkazishda uchta xavf bor: hududni taxmin qilib, atlasni buzib qo‘yish; ta’riflarni ko‘chirib, mualliflik huquqini buzish; avtomatik o‘girishda, masalan, lotin yozuvidan kirillga o‘tkazishda sezilmaydigan xatolarga yo‘l qo‘yish. Ushbu maqolaning maqsadi — shu xavflarni hisobga oladigan import metodikasini bayon qilish va uni F. Musayevaning «O‘zbek xalq so‘zlari» lug‘ati [4] misolida baholash. Metodika O‘zSheva raqamli lug‘at va atlas platformasida [7] ishlab chiqilgan va qo‘llangan.

2. MATERIALLAR VA METODLAR
Materials and methods

Material. F. Musayevaning lug‘ati [4] 392 betdan iborat, lug‘at yozuvlari 54–390-betlarda joylashgan. Undan 2084 ta yozuv ajratib olindi; har bir yozuv uchun bet, bosh so‘z, omonim raqami, etimologik belgi, qavs ichidagi yorliq va ta’rifning birinchi jumlasi qayd etildi. Bosh so‘zdan keyin qavs ichida kelgan izoh yorliq deb olindi, faqat lotincha tur nomidan iborat izohlar (13 ta) bundan mustasno.

Maqsadli model. O‘zSheva platformasida ma’lumot to‘rt darajada saqlanadi: tushuncha, lug‘at yozuvi, hududiy variant va manba. Tushunchada qisqa ma’no (o‘zbekcha lotin va kirill yozuvida, ruscha va inglizcha) va mavzu guruhi, lug‘at yozuvida sheva shakli, adabiy muqobil va so‘z turkumi, hududiy variantda joy, shakl, IPA transkripsiyasi [8], qo‘llanish ulushi va ishonch darajasi, manbada esa so‘z qayerda va qachon qayd etilgani hamda iqtibos saqlanadi. Ishonch darajasi dalil kuchiga qarab belgilanadi: «yuqori» — kamida uchta mustaqil informant va dialektolog tasdig‘i; «o‘rta» — kamida ikkita informant yoki yozma manba va dialektolog tasdig‘i bilan quvvatlangan bitta informant; «past» — qo‘shimcha dalilsiz bitta informant. Yagona yozma manbadan olingan, dala tekshiruvidan o‘tmagan yozuvlarga ham «past» daraja beriladi. Nashr etiladigan yozuvda hududiy variant, manba va IPA transkripsiyasi bo‘lishi ma’lumotlar bazasi darajasida tekshiriladi. Manba yozuvi maydonlarining modelga o‘tkazilishi 1-jadvalda ko‘rsatilgan.

1-jadval

Manba yozuvi maydonlarining O‘zSheva modeliga o‘tkazilishi / **Table 1.** Mapping of the source entry fields to the O‘zSheva model

Manba yozuvidagi maydon	O‘zSheva modelidagi o‘rni	Izoh
Bosh so‘z	Lug‘at yozuvining sheva shakli va doimiy manzili	omonimlar raqam bilan ajratiladi (adoqlamoq-2)
Hudud yorlig‘i	Bitta yoki ikkita hududiy variant	asl yorliq iqtibosda o‘zgarmagan holda saqlanadi
Bet raqami	Har bir variantdagi iqtibos	«…Toshkent, 2023, 134-bet. Manbadagi hudud: „Surx.-Qashq.“»
Ta’rif	Tushunchaning qisqa ma’nosi (o‘zbek, rus, ingliz tillarida; kirillchasi transliteratsiya bilan hosil qilinadi)	matn ko‘chirilmaydi, ma’no mustaqil ifodalanadi
—	Adabiy muqobil, so‘z turkumi, mavzu guruhi	ma’noga qarab belgilanadi
—	IPA transkripsiyasi	yozma shakldan qoidalar asosida hosil qilinadi
—	Ishonch darajasi	«past»: yagona yozma manba, dala tekshiruvi yo‘q

Import bosqichlari. Birinchi bosqichda kitobning PDF matnidan yozuvlar jadvali tuziladi. Kitob matni serverdagi yopiq papkada faqat tahlil uchun saqlanadi, loyiha repozitoriyiga esa faqat faktlar (bet, so‘z, yorliq) va mustaqil yozilgan qisqa ma’nolar (bundan keyin — izohlar) tushadi.

Ikkinchi bosqichda yorliqlar tasniflanadi. Har bir yorliq qismlarga ajratiladi: vergul bilan ajratilgan qismlar va «Surx.-Qashq.» kabi qo‘shma qisqartmalar alohida olinadi. Har bir qism viloyatlar, tumanlar, O‘zbekistondan tashqaridagi joylar va bir nechta joyga to‘g‘ri keladigan nomlar ro‘yxatlari bilan solishtiriladi. Tuman yoki shahar o‘zi joylashgan viloyatga bog‘lanadi: Olot — Buxoroga, Forish — Jizzaxga, Chiroqchi — Qashqadaryoga, Qo‘shrabot va Poyariq (rasmiy nomi Payariq) — Samarqandga, Mang‘it — Qoraqalpog‘iston Respublikasiga. O‘zbekistondan tashqaridagi (Iqon, Qarnoq, Turkiston, Tojikiston) va bir nechta joyga to‘g‘ri keladigan (Toshkent — shahar yoki viloyat, Toshqo‘rg‘on, Qarnob, Qorachiq) nomlar bog‘lanmaydi. Aralash yorliqda faqat aniq bog‘lanadigan qism olinadi.

Uchinchi bosqichda ma’no ifodalanadi. Ta’rif kitobdan ko‘chirilmaydi: har bir so‘z uchun qisqa ma’no o‘zbek, rus va ingliz tillarida mustaqil yoziladi, mavzu guruhi va so‘z turkumi belgilanadi, mos adabiy so‘z bo‘lsa, u adabiy muqobil sifatida ko‘rsatiladi.

To‘rtinchi bosqichda tushunchalar moslashtiriladi. Ma’nosi lug‘atdagi mavjud tushunchaga to‘g‘ri keladigan so‘z yangi yozuv sifatida emas, shu tushunchaga tegishli yozuvning hududiy varianti sifatida qo‘shiladi. Import ichidagi sinonimlar bitta tushunchaga birlashtiriladi.

Beshinchi bosqichda iqtibos va ishonch darajasi belgilanadi. Har bir variantga kitob, bet va manbadagi asl yorliq ko‘rsatilgan iqtibos hamda «Dala tekshiruvi kutilmoqda» izohi qo‘shiladi. Ishonch darajasi «past» deb belgilanadi, manbada berilmagani uchun qo‘llanish ulushi bo‘sh qoldiriladi.

Oltinchi bosqichda yozuvlar bazaga yuklanadi. Faktlar va izohlar fayllaridan SQL skript dastur yordamida hosil qilinadi, maydonlar soni va majburiy qiymatlar qat’iy tekshiriladi. Skript avval zaxira nusxasidan tiklangan bazada sinaladi, so‘ng ishchi bazaga bitta tranzaksiyada qo‘llanadi.

Lotin va kirill yozuvlari. Kirillcha izohlar lotincha izohdan avtomatik transliteratsiya bilan hosil qilinadi. Qidiruvda apostrofning to‘rt xil shakli — U+2018, U+2019, U+02BB va oddiy U+0027 [9] — bir xil hisoblanadi, qidiruv kalitida esa apostrof olib tashlanadi. Shu sababli apostrofning istalgan shakli bilan yoki apostrofsiz yozilgan so‘rov bir xil natija beradi.

3. NATIJALAR

Results

Yorliqlar tahlili. Ajratib olingan 2084 ta yozuvdan 548 tasida qavs ichida yorliq aniqlandi. Ulardan 454 tasi kamida bitta viloyatga bog‘landi, 94 tasi bog‘lanmadi (2-jadval). Bog‘lanmagan yorliqlarning eng katta qismi (58 ta) hudud emas, balki o‘simlik va boshqa organizmlarning lotincha nomlari bo‘lib chiqdi. Manbadagi «Surxandaryo» yozilishi Surxondaryo deb olindi. Tahlil davomida bitta kamchilik ham aniqlandi: Jizzax viloyatidagi G‘allaorol tumanining nomi tumanlar ro‘yxatida yo‘qligi sababli bitta yozuv bog‘lanmay qoldi. Yorliq aniqlanmagan 1536 ta yozuv (ulardan 13 tasida qavs ichida faqat lotincha tur nomi bor) kiritilmadi.

2-jadval

«O‘zbek xalq so‘zlari» lug‘atidagi yorliq turlari / **Table 2.** Types of labels in the dictionary “O‘zbek xalq so‘zlari”

Yorliq turi	Misol	Yozuvlar	Qaror
Viloyatga bog‘langan yorliqlar			
Bitta viloyat nomi	Xorazm; Jizzax	419	o‘sha viloyat
Ikki viloyat, to‘liq yozilgan	Jizzax, Qashqadaryo	8	ikkala viloyat
Ikki viloyat, qisqartirilgan	Surx.-Qashq.	10	ikkala viloyat
Tuman nomi	Olot; Forish	10	tuman joylashgan viloyat
Viloyat va tuman	Xorazm, Olot	1	Xorazm va Buxoro
Viloyat va O‘zbekistondan tashqaridagi joy	Xorazm, Iqon	6	faqat viloyat
Jami bog‘langan		**454**	
Bog‘lanmagan yorliqlar			
O‘zbekistondan tashqaridagi joy	Qarnoq; Iqon	18	bog‘lanmaydi
Bir nechta joyga to‘g‘ri keladigan nom	Toshkent; Toshqo‘rg‘on	11	bog‘lanmaydi
Lotincha biologik nom	Achillea millefolium L.	58	hudud emas
Boshqa izoh va qisqartmalar	—	6	hudud emas
Ro‘yxatda yo‘q tuman nomi	G‘allaorol	1	keyin tuzatiladi
Jami bog‘lanmagan		**94**	
Hammasi		**548**	

Import natijasi. Bog‘langan 454 ta yozuvdan 13 tasi qo‘lda ko‘rib chiqilganda kiritilmadi, asosan ta’rifi matndan aniq ajralmagani yoki havola qilingan so‘z ajratib olingan yozuvlar orasida topilmagani sababli. Qolgan 441 ta yozuvdan 423 tasi yangi lug‘at yozuviga aylandi va 394 ta tushunchaga birlashtirildi: 24 ta tushuncha 53 ta sinonim so‘zni o‘z ichiga oladi (masalan, «g‘ildirak» ma’nosida digil, digirchak, do‘ngalak). 17 ta so‘z lug‘atdagi mavjud yozuvlarga hududiy variant sifatida qo‘shildi: barak — chuchvara (Xorazm), buvi — ona (Farg‘ona), ro‘d — ariq (Buxoro), doyi — tog‘a (Xorazm, Buxoro), checha — yanga (Surxondaryo), jilob — paypoq (Jizzax) va boshqalar. Yana bitta so‘z import qilingan boshqa yozuvga variant bo‘ldi. Jami 460 ta hududiy variant yaratildi (3-jadval).

3-jadval

Import natijalari / **Table 3.** Import results

Ko‘rsatkich	Soni
Lug‘atdan ajratib olingan yozuvlar	2084
Qavs ichida yorlig‘i bor yozuvlar	548
Kamida bitta viloyatga bog‘langan yozuvlar	454
shundan ikki viloyatga bog‘langanlari	19
Bog‘langan, ammo kiritilmagan yozuvlar	13
Yangi lug‘at yozuvlari	423
Yangi tushunchalar	394
Mavjud yozuvlarga variant sifatida qo‘shilgan so‘zlar	17
Import qilingan boshqa yozuvga variant bo‘lgan so‘zlar	1
Yaratilgan hududiy variantlar	460
Yorliq aniqlanmagani sababli kiritilmagan yozuvlar	1536

Yaratilgan variantlar viloyatlar bo‘yicha quyidagicha taqsimlandi: Xorazm — 255 (55%), Qashqadaryo — 68, Jizzax va Surxondaryo — 44 tadan, Buxoro — 17, Farg‘ona va Samarqand — 12 tadan, Namangan — 4, Andijon — 3, Qoraqalpog‘iston Respublikasi — 1. Yangi yozuvlarning 310 tasi ot, 49 tasi fe’l, 47 tasi sifat, 9 tasi ravish, 6 tasi ibora, bittadan undov va yuklama. Eng katta mavzu guruhlari — inson (50), tabiat (36), o‘simliklar (32), dehqonchilik (26), chorvachilik (25) va hayvonlar (24). 195 ta yozuvga (46%) adabiy muqobil ko‘rsatildi. Import natijasida platformadagi lug‘at 88 tadan 511 ta yozuvga, hududiy variantlar 200 tadan 660 taga yetdi. Import qilingan variantlar ochiq ma’lumotlar to‘plamiga iqtibosi bilan birga kiradi: CSV eksportidagi 653 ta qatordan 460 tasining manba ustunida Musayeva lug‘atiga havola bor, TEI formatida [10] esa bu havola alohida izoh (note) elementida beriladi.

Sifat nazorati. Import qilingan yozuvlar saytda e’lon qilingandan keyin kirillcha izohlarda transliteratsiya xatolari aniqlandi. Avvalgi transliteratorda kontekstga bog‘liq qoidalar yo‘q edi, uning testlari esa faqat oddiy so‘zlarni tekshirgan, shuning uchun xatolar testlarda ko‘rinmadi. Yangi izohlardan 41 tasida besh turdagi 43 ta xato topildi (4-jadval). Bundan tashqari, avvalgi sinov to‘plamidagi 82 ta tushunchaning kirillcha maydoniga lotincha matn ko‘chirilgani ma’lum bo‘ldi. Transliterator tuzatilgach, har bir qoida uchun real misollardan test yozildi. Izohlar qayta hosil qilinganda faqat avtomatik yaratilgan qiymatlar yangilandi, qo‘lda yozilgan izohlarga tegilmadi: jami 123 ta kirillcha izoh tuzatildi. Shu jarayonda loyiha doirasida yozilgan bitta o‘zbekcha izohdagi imlo xatosi ham topildi: «(em)» o‘rniga «(yem)» bo‘lishi kerak edi.

4-jadval

Kirill transliteratsiyasida aniqlangan xato turlari / **Table 4.** Types of errors found in Cyrillic transliteration

Xato	To‘g‘ri qoida	Misol: xato → to‘g‘ri	Soni
ye → йе	ye → e	йенгил танбеҳ → енгил танбеҳ	12
so‘z boshidagi va unlidan keyingi e → e	e → э	ертапишар → эртапишар	20
yo‘ → ё‘	yo‘ → йў	тоғ ё‘ли → тоғ йўли	5
tutuq belgisi → ‘	’ → ъ	на‘матак → наъматак	5
o‘zakdagi t va qo‘shimchadagi s → ц	ts → тс	қат‘ияциз → қатъиятсиз	1
Jami (41 ta izohda)			**43**

4. MUHOKAMA

Discussion

Bitta yozma manba lug‘atni qariyb olti barobar kengaytirdi, ammo yangi variantlarning barchasi «past» ishonch darajasida: ular dala tekshiruvidan o‘tmagan, qo‘llanish ulushi esa noma’lum. Shu sababli platformada ulushi noma’lum hudud xaritada «0%» deb emas, xiraroq rang bilan ko‘rsatiladi, ulush ustunida esa chiziqcha turadi. Bu yozma manbadan olingan ma’lumotni dala ma’lumotidan ajratib ko‘rsatishga imkon beradi.

Variantlarning 55 foizi Xorazmga to‘g‘ri kelishi, birinchi navbatda, manbadagi yorliqlar tarkibi bilan bog‘liq: «Xorazm» yorlig‘ining o‘zi 257 marta uchraydi. Navoiy va Sirdaryo viloyatlari bu manbaning hudud yorliqlarida umuman uchramaydi, Toshkent esa faqat ikki ma’noli yorliq sifatida keladi. Demak, atlasdagi zichlikni shevalarning leksik boyligi o‘lchovi deb talqin qilib bo‘lmaydi; har bir variantni uning manbasi bilan birga ko‘rish kerak.

Hududni taxmin qilmaslik qamrovni kamaytiradi: yorliq aniqlanmagan 1536 ta yozuv va sinonimlar lug‘atlaridagi [5; 6] hududi ko‘rsatilmagan sheva so‘zlari atlasga kiritilmadi. Ammo aynan shu qoida xaritaning ishonchliligini saqlaydi. Bunday yozuvlarni nashr etmasdan qoralama sifatida saqlab, keyinchalik dala tadqiqotida tekshirish rejalashtirilgan.

Manba har bir variantda sahifa aniqligida ko‘rsatiladi, bu FAIR tamoyillaridagi ma’lumotning kelib chiqishini qayd etish talabiga mos keladi [11]. Ta’riflar ko‘chirilmagani uchun esa ma’lumotlar to‘plamini CC BY 4.0 litsenziyasi [12] bilan tarqatish mumkin bo‘ldi.

Transliteratsiya bo‘yicha asosiy saboq shuki, testlarning hammasi o‘tgani to‘g‘rilikni kafolatlamaydi: avvalgi testlar real ma’lumotdagi kontekstli holatlarni qamramagan edi. Testlar lug‘atning o‘zidan olingan misollar asosida tuzilishi kerak. Rus imlosidan olingan «ts → ц» kabi qoidalar o‘zbekcha so‘zlarda yonma-yon kelgan t va s harflarini noto‘g‘ri o‘giradi (aytsa, ketsa, qat’iyatsiz). Yana bir cheklov bor: import qilingan variantlardagi IPA transkripsiyasi yozma shakldan qoidalar asosida hosil qilingan. U taxminiy fonologik ifoda bo‘lib, dala yozuvlaridagi talaffuzning o‘rnini bosmaydi.

5. XULOSA

Conclusion

Yozma dialektologik manbani raqamli lug‘at va atlasga kiritishda quyidagi qoidalar o‘zini oqladi. Birinchidan, hudud yorlig‘ini qat’iy tasniflash kerak: faqat aniq bog‘lanadigan yorliq olinadi, qolganlari sababi bilan qayd etiladi. Ikkinchidan, ma’noni mustaqil ifodalash va har bir variantga sahifa aniqligidagi iqtibos berish mualliflik huquqini buzmasdan ochiq ma’lumot tarqatish imkonini beradi. Uchinchidan, yangi so‘zni mavjud tushunchaga moslashtirish dublikatlarning oldini oladi va bir so‘zning hududiy variantlarini bir sahifada jamlaydi. To‘rtinchidan, avtomatik transliteratsiya real ma’lumotdagi misollar bilan sinalishi kerak.

F. Musayevaning lug‘ati asosida O‘zSheva lug‘atiga 423 ta yangi yozuv va 460 ta hududiy variant qo‘shildi. Keyingi vazifalar: G‘allaorol kabi tushib qolgan tuman nomlarini ro‘yxatga qo‘shish, yorliqsiz yozuvlarni qoralama sifatida saqlash, «O‘zbek shevalari tadqiqotlari» to‘plamidagi [3] hududiy materiallarni kiritish va yozuvlarni dala ekspeditsiyalarida tekshirish.

DECLARATIONS

Declarations

Author contributions *CRediT*

Conceptualization; Methodology; Investigation; Writing – original draft; Writing – review & editing. The author has read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Ethics approval

Not applicable. This study did not involve humans, animals or other regulated materials.

Informed consent

Not applicable.

Data availability statement

All data supporting the findings of this study are included in the article.

Conflict of interest

The author declares no conflict of interest.

Generative AI disclosure

Not used.

ADABIYOTLAR RO‘YXATI

References (original language)

[1] Ўзбек халқ шевалари луғати. – Тошкент: Фан, 1971.

[2] Решетов В.В., Шоабдураҳмонов Ш. Ўзбек диалектологияси. – Тошкент: Ўқитувчи, 1978.

[3] O‘zbek shevalari tadqiqotlari: amaliyot, metodologiya va yangicha yondashuv: II Respublika ilmiy-nazariy konferensiyasi materiallari / mas’ul muharrir Sh. Sirojiddinov. – Toshkent: Donishmand ziyosi, 2022. – 216 b.

[4] Musayeva F. O‘zbek xalq so‘zlari. – Toshkent: Innovatsion rivojlanish nashriyot-matbaa uyi, 2023. – 392 b.

[5] O‘zbek tili sinonimlarining izohli lug‘ati / N. Mahmudov tahriri ostida. – Toshkent: G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2023. – 512 b.

[6] O‘zbek tili sinonimlarining katta izohli lug‘ati. I jild / N. Mahmudov tahriri ostida. – Toshkent: G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2022. – 504 b.

[7] O‘zSheva – o‘zbek shevalarining raqamli lug‘ati va atlasi. – URL: https://umidliyoshlar.uz (murojaat sanasi: 09.11.2026).

[8] International Phonetic Association. Handbook of the International Phonetic Association. – Cambridge: Cambridge University Press, 1999. – DOI: 10.1017/9780511807954.

[9] The Unicode Consortium. The Unicode Standard, Version 16.0.0. – 2024. – URL: https://www.unicode.org/versions/Unicode16.0.0/

[10] TEI Consortium. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.4.0. – 2022. – DOI: 10.5281/zenodo.6482461.

[11] Wilkinson M.D., Dumontier M., Aalbersberg I.J. et al. The FAIR Guiding Principles for scientific data management and stewardship // Scientific Data. – 2016. – Vol. 3. – Art. 160018. – DOI: 10.1038/sdata.2016.18.

[12] Creative Commons. Attribution 4.0 International (CC BY 4.0). – URL: https://creativecommons.org/licenses/by/4.0/

REFERENCES (ROMAN SCRIPT)

Transliterated list — required by Scopus and Web of Science

[1] O‘zbek xalq shevalari lug‘ati. – Toshkent: Fan, 1971. (in Uzbek)

[2] Reshetov V.V., Shoabdurahmonov Sh. O‘zbek dialektologiyasi. – Toshkent: O‘qituvchi, 1978. (in Uzbek)

[3] O‘zbek shevalari tadqiqotlari: amaliyot, metodologiya va yangicha yondashuv: II Respublika ilmiy-nazariy konferensiyasi materiallari / mas’ul muharrir Sh. Sirojiddinov. – Toshkent: Donishmand ziyosi, 2022. – 216 b.

[4] Musayeva F. O‘zbek xalq so‘zlari. – Toshkent: Innovatsion rivojlanish nashriyot-matbaa uyi, 2023. – 392 b.

[5] O‘zbek tili sinonimlarining izohli lug‘ati / N. Mahmudov tahriri ostida. – Toshkent: G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2023. – 512 b.

[6] O‘zbek tili sinonimlarining katta izohli lug‘ati. I jild / N. Mahmudov tahriri ostida. – Toshkent: G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2022. – 504 b.

[7] O‘zSheva – o‘zbek shevalarining raqamli lug‘ati va atlasi. – URL: https://umidliyoshlar.uz (murojaat sanasi: 09.11.2026).

[8] International Phonetic Association. Handbook of the International Phonetic Association. – Cambridge: Cambridge University Press, 1999. – DOI: 10.1017/9780511807954.

[9] The Unicode Consortium. The Unicode Standard, Version 16.0.0. – 2024. – URL: https://www.unicode.org/versions/Unicode16.0.0/

[10] TEI Consortium. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.4.0. – 2022. – DOI: 10.5281/zenodo.6482461.

[11] Wilkinson M.D., Dumontier M., Aalbersberg I.J. et al. The FAIR Guiding Principles for scientific data management and stewardship // Scientific Data. – 2016. – Vol. 3. – Art. 160018. – DOI: 10.1038/sdata.2016.18.

[12] Creative Commons. Attribution 4.0 International (CC BY 4.0). – URL: https://creativecommons.org/licenses/by/4.0/