O‘zbek shevalarining raqamli lug‘ati va atlasi (O‘zSheva): yaratish tamoyillari va dastlabki natijalar

Saloydinova Maxliyo Abduvali qizi
«O‘zbekgidroenergo» aksiyadorlik jamiyati ilmiy-texnik jurnali mas’ul kotibi

UO‘K 811.512.133’28

IQTIBOS UCHUN
Saloydinova M.A. O‘zbek shevalarining raqamli lug‘ati va atlasi (O‘zSheva): yaratish tamoyillari va dastlabki natijalar // Zamonaviy fan, texnologiya va innovatsiyalar: xalqaro ilmiy-amaliy konferensiya materiallari (ZFTI–2026, 5-noyabr 2026-yil). – 2026. – B. 42–45. – https://doi.org/10.66073/ReFocus-Conf/051108
QR-kod — DOI sahifasiga havola

ANNOTATSIYA

Tadqiqotda o‘zbek shevalari leksikasini raqamli lug‘at va atlas ko‘rinishida jamlovchi O‘zSheva platformasi (umidliyoshlar.uz) tavsiflanadi. Platformada har bir sheva so‘zi tushunchaga bog‘lanadi, hududiy variantlari IPA transkripsiyasi, manbasi va ishonch darajasi bilan saqlanadi, o‘zgarishlar versiyalanadi. Qidiruv lotin va kirill yozuvida ishlaydi, ma’lumotlar CSV, JSON-LD va TEI formatlarida CC BY 4.0 litsenziyasi bilan ochiq beriladi. Yozma manbadan import qilish tartibi F. Musayevaning «O‘zbek xalq so‘zlari» lug‘ati misolida sinaldi: lug‘at 88 tadan 511 ta yozuvga, hududiy variantlar 200 tadan 660 taga yetdi. Natijalar yozma manba qamrovni tez kengaytirishini, lekin dala tekshiruvisiz ishonch darajasi past bo‘lib qolishini ko‘rsatadi.

Kalit so‘zlar: o‘zbek shevalari, dialektologiya, raqamli lug‘at, lingvistik atlas, hududiy variant, IPA, transliteratsiya, ochiq ma’lumot, TEI.

АННОТАЦИЯ

ЦИФРОВОЙ СЛОВАРЬ И АТЛАС УЗБЕКСКИХ ГОВОРОВ (O‘ZSHEVA): ПРИНЦИПЫ СОЗДАНИЯ И ПЕРВЫЕ РЕЗУЛЬТАТЫ

Салойдинова Махлиё Абдували кизи, ответственный секретарь научно-технического журнала АО «Узбекгидроэнерго»

В работе описывается платформа O‘zSheva (umidliyoshlar.uz), объединяющая лексику узбекских говоров в цифровом словаре и атласе. Каждое диалектное слово связано с понятием, его региональные варианты хранятся вместе с транскрипцией МФА, источником и уровнем достоверности, а все изменения версионируются. Поиск работает на латинице и кириллице, данные открыто распространяются в форматах CSV, JSON-LD и TEI по лицензии CC BY 4.0. Порядок импорта из письменного источника апробирован на словаре Ф. Мусаевой «O‘zbek xalq so‘zlari»: число словарных статей выросло с 88 до 511, региональных вариантов — с 200 до 660. Результаты показывают, что письменный источник быстро расширяет охват, однако без полевой проверки уровень достоверности остаётся низким.

Ключевые слова: узбекские говоры, диалектология, цифровой словарь, лингвистический атлас, региональный вариант, МФА, транслитерация, открытые данные, TEI.

ABSTRACT

A DIGITAL DICTIONARY AND ATLAS OF UZBEK DIALECTS (O‘ZSHEVA): DESIGN PRINCIPLES AND FIRST RESULTS

Saloydinova Maxliyo Abduvali qizi, Executive Secretary, Scientific and Technical Journal of “Uzbekhydroenergo” JSC

The paper describes O‘zSheva (umidliyoshlar.uz), a platform that brings the vocabulary of Uzbek dialects together in a digital dictionary and atlas. Each dialect word is linked to a concept; its regional variants are stored with an IPA transcription, a source and a confidence level, and every change is versioned. Search works in both Latin and Cyrillic script, and the data are openly available in CSV, JSON-LD and TEI under the CC BY 4.0 licence. The procedure for importing a written source was tested on F. Musayeva’s dictionary “O‘zbek xalq so‘zlari”: the number of entries grew from 88 to 511 and the number of regional variants from 200 to 660. The results show that a written source extends coverage quickly, but without fieldwork the confidence level remains low.

Keywords: Uzbek dialects, dialectology, digital dictionary, linguistic atlas, regional variant, IPA, transliteration, open data, TEI.

KIRISH

O‘zbek shevalari an’anaviy ravishda qarluq, qipchoq va o‘g‘uz guruhlariga ajratiladi [1]. Ularning leksikasi bosma lug‘atlarda [2; 3] va ilmiy to‘plamlarda [4] ko‘p yillar davomida to‘plangan. Biroq bu materiallar tarqoq: so‘z qaysi hududda qayd etilgani xaritada ko‘rinmaydi, lotin va kirill yozuvidagi manbalarni birga qidirib bo‘lmaydi, ma’lumotni boshqa tadqiqotlarda qayta ishlatish uchun mashina o‘qiy oladigan shakl yo‘q. Ushbu tadqiqotda shu muammoga yechim sifatida yaratilgan O‘zSheva platformasining [9] tamoyillari va uni to‘ldirishning dastlabki natijalari bayon qilinadi.

MATERIALLAR VA METODLAR

Platformaning ma’lumotlar modeli to‘rt bo‘g‘indan iborat: tushuncha – sheva so‘zi (lug‘at yozuvi) – hududiy variant – manba. Sheva shakli va adabiy muqobil alohida maydonlarda saqlanadi, ular bir-birining o‘rnini bosmaydi. Har bir hududiy variant uchun joy (viloyat, tuman yoki aholi punkti), IPA transkripsiyasi [7], qo‘llanish ulushi va ishonch darajasi (past, o‘rta, yuqori) qayd etiladi. Nashr etiladigan yozuvda hududiy variant, manba va IPA transkripsiyasi bo‘lishi shart. Bu talab ma’lumotlar bazasi darajasida tekshiriladi, shuning uchun to‘liq bo‘lmagan yozuv saytga chiqmaydi. Yozuv tahrirlanganda avvalgi holati o‘chirilmaydi: versiyalar tarixi saqlanadi va so‘z sahifasida ko‘rsatiladi.

Interfeys to‘rt tilda ishlaydi: o‘zbek (lotin va kirill), rus va ingliz. Lotin–kirill transliteratsiyasi yagona modulda bajariladi va avtomatik testlar bilan tekshiriladi. Qidiruv apostrofning turli shakllarini bir xil hisoblaydi, shu sababli «cho‘pon», «chopon» va «чўпон» so‘rovlari bitta yozuvni topadi. Xaritada so‘zning viloyatlar bo‘yicha tarqalishi ko‘rsatiladi, xarita ostida esa JavaScript o‘chirilgan bo‘lsa ham o‘qiladigan jadval beriladi. Foydalanuvchilar yangi so‘z yuborishi mumkin: har bir hissa moderatordan o‘tadi, informant roziligi uch darajada (faqat tadqiqot uchun, faqat saytda ko‘rsatish, ochiq dataset) hisobga olinadi, shaxsga doir ma’lumotlar shifrlangan holda saqlanadi [8]. Ochiq ma’lumot FAIR tamoyillariga [5] tayanadi: lug‘at CSV, JSON-LD (schema.org) va TEI P5 [6] formatlarida CC BY 4.0 litsenziyasi bilan beriladi, muzlatilgan relizlar SHA-256 nazorat yig‘indisi bilan chiqariladi, dasturiy interfeys (API) ham mavjud.

Yozma manbadan import qilish tartibi F. Musayevaning «O‘zbek xalq so‘zlari» lug‘ati [3] misolida ishlab chiqildi. Qoidalar quyidagicha: faqat hududi aniq ko‘rsatilgan so‘z kiritiladi,

hudud taxmin qilinmaydi; tuman nomi o‘z viloyatiga bog‘lanadi, manbadagi asl yorliq izohda saqlanadi; O‘zbekistondan tashqaridagi va ikki xil talqin qilinadigan yorliqlar bog‘lanmaydi; ta’rif kitobdan ko‘chirilmaydi, ma’no qisqa va o‘z so‘zimiz bilan beriladi; har bir variant yonida kitob, bet va manbadagi yorliq ko‘rsatiladi; ishonch darajasi «past» deb belgilanadi. Lug‘atda bor tushunchaga mos so‘z yangi yozuv sifatida emas, mavjud yozuvning hududiy varianti sifatida qo‘shiladi.

NATIJALAR VA MUHOKAMA

Importdan oldin lug‘atda umumiy leksikografik ma’lumotlar asosida sinov uchun kiritilgan 88 ta yozuv va 200 ta hududiy variant bor edi. Musayeva lug‘atidagi 2084 ta yozuvdan 548 tasida qavs ichida yorliq bor edi, ularning bir qismi o‘simliklarning lotincha nomlari bo‘lib chiqdi. 454 ta yozuv hududga aniq bog‘landi, shulardan 13 tasi kiritilmadi (ta’rifi ajratilmagan yoki havolasi topilmagan). Natijada 423 ta yangi yozuv yaratildi va 18 ta mavjud yozuvga hududiy variant qo‘shildi, masalan: chuchvara – barak (Xorazm), ona – buvi (Farg‘ona), ariq – ro‘d (Buxoro), tog‘a – doyi (Xorazm, Buxoro), yanga – checha (Surxondaryo). Hududi ko‘rsatilmagan 1536 ta yozuv nashr etilmadi.

Hozir lug‘atda 511 ta yozuv, 482 ta tushuncha va 660 ta hududiy variant bor. IPA transkripsiyasi barcha variantlarda mavjud, adabiy muqobil 283 ta yozuvda (55%) ko‘rsatilgan. So‘z turkumlari bo‘yicha: ot – 393, sifat – 52, fe’l – 49, ravish – 9, boshqalar – 8. Eng katta mavzu guruhlari – tabiat (44), inson (43), uy-ro‘zg‘or (37), o‘simliklar (33) va oziq-ovqat (32). 101 ta yozuv ikki va undan ortiq viloyatda qayd etilgan. Ma’lumotlar viloyat darajasidagi 14 ta hududdan 13 tasini qamraydi, variantlarning taqsimoti 1-jadvalda berilgan.

1-jadval. Hududiy variantlarning viloyatlar bo‘yicha taqsimoti

Hudud	Variantlar	Shundan yozma manbadan [3]
Xorazm viloyati	275	255
Qashqadaryo viloyati	89	68
Jizzax viloyati	59	44
Namangan viloyati	54	4
Surxondaryo viloyati	50	44
Buxoro viloyati	43	17
Samarqand viloyati	29	12
Farg‘ona viloyati	22	12
Toshkent viloyati	12	0
Andijon viloyati	9	3
Qoraqalpog‘iston Respublikasi	8	1
Navoiy viloyati	7	0
Sirdaryo viloyati	3	0
Jami	**660**	**460**

Manba: O‘zSheva ma’lumotlar bazasi, 2026-yil 9-noyabr holati.

https://doi.org/10.66073/ReFocus-Conf/051108

Taqsimot notekis: variantlarning 42 foizi Xorazmga to‘g‘ri keladi. Bu shevalarning haqiqiy boyligini emas, manbadagi yorliqlar tarkibini aks ettiradi – yozma manbadan kelgan 460 ta variantning 255 tasi Xorazmga tegishli. Yozma manba qamrovni tez kengaytiradi, lekin bunday variantlarning qo‘llanish ulushi noma’lum va ishonch darajasi past. Hozir barcha variantlarning 84 foizi (554 tasi) shu darajada. Shuning uchun xaritada ulushi noma’lum hudud «0%» deb emas, xiraroq rang bilan ko‘rsatiladi. Import jarayonida transliteratsiya bo‘yicha ham saboq olindi: kirill imlosining kontekstga bog‘liq qoidalari (yo‘ → йў, so‘z boshidagi е → э, tutuq belgisi → ъ) testlarda tekshirilmagan ekan, natijada 124 ta kirillcha izohni tuzatishga to‘g‘ri keldi. Endi har bir qoida alohida test bilan nazorat qilinadi. Platforma hozircha sinov rejimida ishlaydi va ma’lumotlar dala tekshiruvidan o‘tmaguncha qidiruv tizimlari uchun yopiq turadi.

XULOSA

O‘zSheva sheva so‘zini hudud, talaffuz, manba va ishonch darajasi bilan birga saqlaydi hamda ochiq formatlarda beradi, shuning uchun bu ma’lumotlardan boshqa tadqiqotlarda ham foydalanish mumkin. Musayeva lug‘ati misolida sinalgan import tartibi boshqa yozma manbalarga ham qo‘llanadi. Navbatdagi vazifalar: yozuvlarni dala ekspeditsiyalarida tekshirib, ishonch darajasini oshirish; «O‘zbek shevalari tadqiqotlari» to‘plamidagi [4] hududiy materiallarni kiritish; qamrovi kam viloyatlarda informantlar bilan ishlash.

ADABIYOTLAR

1. Решетов В.В., Шоабдураҳмонов Ш. Ўзбек диалектологияси. – Тошкент: Ўқитувчи, 1978.
2. Ўзбек халқ шевалари луғати. – Тошкент: Фан, 1971.
3. Musayeva F. O‘zbek xalq so‘zlari. – Toshkent: Innovatsion rivojlanish nashriyot-matbaa uyi, 2023. – 392 b.
4. 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.
5. 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.
6. TEI Consortium. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.4.0. – 2022. – DOI: 10.5281/zenodo.6482461.
7. International Phonetic Association. Handbook of the International Phonetic Association. – Cambridge: Cambridge University Press, 1999. – DOI: 10.1017/9780511807954.
8. Шахсга доир маълумотлар тўғрисида: Ўзбекистон Республикасининг Қонуни, 02.07.2019, ЎРҚ-547. – URL: https://lex.uz/docs/4396419
9. O‘zSheva – o‘zbek shevalarining raqamli lug‘ati va atlasi. – URL: https://umidliyoshlar.uz (murojaat sanasi: 09.11.2026).

https://doi.org/10.66073/ReFocus-Conf/051108