

O'ZBEK TILI ELEKTRON KORPUSIDA DIALEKTAL KORPUS LOYIHASI

Rustamova Zulfiya Xolmirzayevna

Nizomiy nomidagi Toshkent davlat pedagogika universiteti

Boshlang'ich ta'limda ona tili va uni o'qitish metodikasi kafedrasida dotsenti

<https://doi.org/10.53885/edinres.2024.04.2.087>

Annotasiya: Turkiy tillar oilasiga mansub o'zbek tili O'zbekistonda va qo'shni mamlakatlarda keng tarqalgan. Aloqa vositasi sifatida til doimo rivojlanib bormoqda va o'zbek tili ham bundan mustasno emas. Texnologiyalar paydo bo'lishi bilan o'zbek tilining elektron korpusi tilshunoslar, tadqiqotchilar va til o'rganuvchilar uchun muhim vositaga aylandi. Ushbu raqamli ombor doirasida dialektal korpus o'zbek tilining boy til xilma-xilligini saqlash va targ'ib qilishda muhim rol o'ynaydi. Ushbu maqolada, biz o'zbek tilining elektron korpusida dialektal korpusi loyixasi haqida ma'lumot beramiz.

Kalit so'zlar: dielektal korpus, lahjalar, o'zgarishlar, elektron ombor, elektron tizimlar, integratsiya

Аннотация: Узбекский язык, относящийся к тюркской языковой семье, широко распространён в Узбекистане и соседних странах. Язык как средство общения постоянно развивается, и узбекский язык не является исключением. С появлением технологий электронный корпус узбекского языка стал важным инструментом для лингвистов, исследователей и изучающих язык. В этом цифровом хранилище диалектный корпус играет важную роль в сохранении и развитии богатого языкового разнообразия узбекского языка. В этой статье мы предоставляем информацию о проекте диалектного корпуса в электронном корпусе узбекского языка.

Ключевые слова: диалектный корпус, диалекты, изменения, электронное хранилище, электронные системы, интеграция.

Abstract: The Uzbek language, belonging to the Turkic language family, is widely spoken in Uzbekistan and neighboring countries. Language as a means of communication is constantly evolving, and the Uzbek language is no exception. With the advent of technology, the electronic corpus of the Uzbek language has become an important tool for linguists, researchers and language learners. In this digital repository, the dialect corpus plays an important role in preserving and developing the rich linguistic diversity of the Uzbek language. In this article we provide information about the dialect corpus project in the electronic corpus of the Uzbek language.

Key words: dialect corpus, dialects, changes, electronic storage, electronic systems, integration.

Kirish: Dialektal korpus-bu o'zbek tilining turli lahjalari va mintaqaviy o'zgarishlarini ifodalovchi matnlar, audio yozuvlar va boshqa multimedia materiallari to'plamidir. Ushbu korpus dialektal farqlar ko'pincha mintaqaviy, ijtimoiy va madaniy farqlarni aks ettiradigan O'zbekistonning lingvistik manzarasini tushunish uchun juda muhimdir. O'zbek tilining elektron korpusiga dialektal korpusni kiritish orqali tadqiqotchilar va tilshunoslar tilning murakkab dinamikasini chuqurroq tushinishlari mumkin.

Dialektal korpusning asosiy afzalliklaridan biri uning o'zbek tilining til xilma-xilligini namoyish etish qobiliyatidir. O'zbekiston boy madaniy merosga ega mamlakat bo'lib, uning dialektlari millatni shakllantirgan murakkab tarix va madaniy almashinuvni aks ettiradi. Dialektal korpus standart o'zbek tili foydasiga ko'pincha e'tibordan chetda qoladigan ushbu dialektlarni hujjatlashtirish va tahlil qilish uchun platforma yaratadi. Ushbu dialektlarga ovoz berish orqali dialektal korpus lingvistik xilma-xillikni targ'ib qiladi va standart tilning ustunligiga qarshi chiqadi.

Bundan tashqari, dialektal korpus lingvistik tadqiqotlar va tahlillar uchun noyob imkoniyatni taqdim etadi. Dialektal o'zgarishlarni o'rganish orqali tadqiqotchilar tilning tarixiy rivojlanishi, til bilan aloqasi va til o'zgarishi haqida tushunchaga ega bo'lishlari mumkin. Korpus tilshunoslar uchun har bir dialektning fonologik, leksik va grammatik xususiyatlarini o'rganish, o'zbek tilini shakllantirgan lingvistik jarayonlarni yoritib berish uchun qimmatli manba bo'lib xizmat qiladi. Bundan tashqari, dialektal korpus tilni o'qitish va til siyosatini xabardor qilishi mumkin, chunki u o'zbek tilining lingvistik xilma-xilligi haqida to'liq tushuncha beradi.

Bundan tashqari, dialektal korpusni o'z ichiga olgan o'zbek tilining elektron korpusi til o'rganuvchilar va o'qituvchilar uchun bir qator imtiyozlarni taqdim etadi. Korpus tilni o'rganish materiallari va

mashqlarini yaratish uchun ishlatilishi mumkin bo'lgan matnlar, audio yozuvlar va videolarni o'z ichiga olgan ko'plab haqiqiy til materiallarini taqdim etadi. Ushbu korpus til o'qituvchilari uchun qimmatli manba bo'lib xizmat qilishi mumkin, bu ularga o'zbek tilining til xilma-xilligini hisobga olgan holda yanada samarali til o'qitish metodikasini ishlab chiqish imkonini beradi.

Dialektal korpus tilni saqlash va madaniy merosni saqlashda hal qiluvchi rol o'ynashi mumkin. Dialektal korpus ushbu dialektlarni hujjatlashtirish va saqlash, ularning kelajak avlodlarga yo'qolmasligini ta'minlash vositasini taqdim etadi. Ushbu korpus madaniy meros va lingvistik xilma-xillikni targ'ib qilish, lingvistik xilma-xillikni saqlash va madaniy tushunishni rivojlantirish muhimligi to'g'risida xabardorlikni oshirish uchun vosita bo'lib xizmat qilishi mumkin.

Lingvistik asos-bu sizning matningizni keyingi tahlilga tayyorlaydigan va tokenlar, lemmalar, nutq qismlari va boshqa ko'p narsalarni samarali qidirish yoki qayta ishlashga imkon beradigan asosiy morfologik qurilish bloklari. Tilning asosiy vositalari tabiiy tilni eng yaxshi qayta ishlash, tezligi va ravshanligi uchun ona tilidagi asl matnni boyitadi. Dunyo internetda ishlaydigan aqlli, muvaffaqiyatli qidiruv semantikasini ishlab chiqdi. Biror kishi kompyuter qidiruv tizimiga so'rov yuboradi va javobni xohlaydi. Hozirgi vaqtda ochiq manbali platformalar teskari to'liq matnli qidiruv tizimlari uchun asosiy tizimni taqdim etadi, ammo aniq qidiruv muammolari qiyinlashadi, chunki siz so'rovlar va natijalarga ko'proq til (siz izlayotgan til) qo'shasiz. Lingvistik ma'lumotlar bazasiga kiritilgan so'zlar qidiruv tizimiga asoslangan holda so'ralgan tilda qidiruvni amalga oshiradi. Masalan, Arab tilshunosligi allaqachon Arab tilida yozilgan hujjatlarni tahlil qilishni osonlashtirish uchun asosiy qidiruv tizimlari va ma'lumotlarni qidirish dasturlariga ulanish uchun mo'ljallangan lingvistik ma'lumotlar bazasiga Arabcha matnni kiritgan. Biroq, an'anaviy ravishda gapiradigan Arab yozuvini hisobga oladigan standart avtomatlashtirilgan tahlil usullari uchun mos emas. Bunday lingvistik asosda Arabcha so'zlar ko'pincha fe'l yo'nalishi, ob'ekt, shaxs, raqam, jins va hokazo kabi xususiyatlarni bildiruvchi grammatik elementlarni o'z ichiga oladi. Ushbu tizim avtomatik ravishda quyidagi harakatlarni amalga oshiradi:

- 1) so'zning lingvistik shaklini shakllantiradi;
- 2) nutq qismlarini aniqlaydi;
- 3) imlo normalarini, shu jumladan unli va otlarni olib tashlashni, hamda shakllarini birlashtirishni va doimiy singan tartibsiz «singan» ko'plik shakllarini normallashtiradi;
- 4) Fors (fors va Dari), Pashto va Urdu tillarida ishlashi mumkin.

O'zbek tilining milliy korpusi leksik birliklar, shu jumladan o'zbek tilida mavjud bo'lgan sinonimlar, antonimlar, omonimlar, so'zlar ierarxiyasi; so'zning morfologik tuzilishini, so'z qurilishini, so'zlarning ma'nosini, morfologik xususiyatlarini avtomatik tahlil qila olishi kerak. Ya'ni, korpusni tuzish, lemmalash, belgilash jarayonida matnlarda korpusning bir qismi bo'lgan bunday so'zlarni individual izlanishlar asosida topish va ularni maxsus talqin qilish kerak. Buning uchun algoritmik, lingvistik modellashtirish ishlarini bajarish kerak. «Interfeys» so'zi ingliz tilidan olingan va «tashqi ko'rinish» ma'nosida ishlatiladi. Bu so'z ko'pincha kompyuter texnologiyalarida qo'llaniladi. Kompyuter-bu odam va mashina o'rtasida turli xil ma'lumot almashinuvini ta'minlaydigan yagona aloqa tizimi. Interfeys-bu bitta tizimning ikkita elementi va ushbu tizim yordamida ishlaydigan bog'lovchi bo'g'in. Interfeys-bu turli xil tugunlar va murakkab apparat bloklari, shuningdek texnologiya va foydalanuvchi o'rtasidagi aloqa tizimi. U mantiqiy (axborotni taqdim etish tizimi) va rasmiy (axborot xususiyatlari) shaklida ifodalanadi. U muayyan vazifalar uchun buyruqlar berish uchun ishlatiladi. Bunday interfeys foydalanuvchi interfeysi deb ataladi. Har qanday qurilmaning interfeysi bajaradigan funktsiyalariga qarab tashqi va ichki ko'rinishlarga bo'linadi. Foydalanuvchi ichki interfeysga to'g'ridan-to'g'ri kirish huquqiga ega bo'lmaydi, u shaxsiy variantga ega. Tashqi interfeys yordamida foydalanuvchi to'g'ridan-to'g'ri aloqa o'rnatishi va undan qurilmani boshqarish uchun foydalanishi mumkin. Ushbu ikki turdagi interfeyslar har doim bitta qurilmaga mos keladi va uning ishlashini ta'minlaydi, ular alohida mavjud bo'lolmaydi.

Foydalanuvchi interfeysini 2 qismga bo'lish mumkin. Masalan, bu qurilmaga ma'lumotlarni kiritish uchun mas'ul bo'lgan qism va foydalanuvchi uning chiqishi uchun javobgardir. Agar biz oddiy ishchi kompyuter haqida gapiradigan bo'lsak, unda birinchi toifada bizda kompyuterda ishlaydigan hamma narsa mavjud. Shunga ko'ra, hamma narsa ikkinchi toifaga kiradi, bu orqali kompyuter bir xil klaviatura, sichqoncha va boshqa kirish qurilmalari, ya'ni monitorlar, karnaylar, minigarnaturalar, printerlar, plutorlar va boshqalar tomonidan berilgan buyruqlarga javoban foydalanuvchiga ma'lumot uzatadi. Kompyuter texnologiyalarida ishlatiladigan interfeyslar quyidagi turlarga bo'linadi:

Vizual. Monitorlarda ko'rsatilgan vizual tasvirlar yordamida ma'lumotlarni uzatuvchi standart kompyuter interfeysi.

Imo-ishora. Qoida tariqasida, u telefonlar yoki planshetlar uchun interfeys bo'lib xizmat qiladi. Ko'pgina hollarda, bu tizimni boshqaradigan odamning barmoqlari harakatlariga javob beradigan va har bir aniq harakatga ma'lum darajada javob beradigan sensorli panel. Buni oddiy vizual interfeysning soddalashtirilgan versiyasi deb atash mumkin.

Ovoz. Ushbu turdagi interfeys nisbatan yaqinda paydo bo'ldi. Ovozli buyruqlar yordamida tizimni boshqarishga imkon beradi. Tizim, o'z navbatida, foydalanuvchi aloqasi orqali javob beradi. Qizig'i shundaki, zamonaviy texnologiyalar nafaqat telefonlar yoki kompyuterlarning ovozini, balki maishiy texnika va hatto bort kompyuterlarining ovozini ham boshqarish imkonini beradi. Milliy korpusning interfeysi boshqa dizaynga, tuzilishga ega bo'lib, uning mukammalligi korpusni yaratgan muallifga ishonib topshirilgan. Chunki interfeys insonda birinchi taassurot qoldiradigan jozibali umumiy ko'rinishdir. Interfeysni yaratishda milliy koloritni aks ettiruvchi bezaklar, shuningdek klassik yoki zamonaviyni aks ettiruvchi belgilar hisobga olinishi kerak. Demak, milliy korpusdan foydalanishda o'zining qulay va samarali imkoniyatlarini namoyish etish uchun interfeys mukammal va tizimli ravishda ishlab chiqilgan bo'lishi muhimdir. Shuning uchun interfeys zamonaviy dasturiy ta'minot dizayni talablariga javob beradigan foydalanuvchilarga qulay, ishlatish uchun qulay shaklda yaratilgan.

Xulosa.

Xulosa qilib aytganda, dialektal korpus o'zbek tilining elektron korpusining ajralmas tarkibiy qismidir. U o'zbek tilining lingvistik xilma-xilligini namoyish etish uchun zamin yaratadi, lingvistik tadqiqotlar va tahlillarni targ'ib qiladi, tillarni o'qitish va til siyosatini xabardor qiladi, tilni saqlash va madaniy merosni saqlashga hissa qo'shadi. Raqamli ombor sifatida dialektal korpus tilshunoslar, tadqiqotchilar va til o'rganuvchilar uchun o'zbek tili bilan yanada kengroq va nozikroq aloqada bo'lish uchun noyob imkoniyatni taqdim etadi. O'zbek tili rivojlanishda davom etar ekan, dialektal korpus O'zbekistonning boy lingvistik merosini tushunish va targ'ib qilishda muhim vosita bo'lib qoladi.

Foydalanilgan adabiyotlar ro'yxati:

1. Sh. Hamroeva, O'zbek tilining mualliflik korpusini yaratishning lingvistik asoslari: Filologiya fanlari nomzodi dissertatsiyasining avtoreferati, Toshkent (2018)
2. R. Karimov, Tilshunoslik va o'zbek va ingliz tillarining parallel korpusini yaratishning dasturlash masalalari, PhD dissertatsiyasining avtoreferati, Buxoro, 151 (2021)
3. R. Qo'ldoshev va boshqalar, Pearson muvofiqlik mezonlari asosida boshlang'ich chap qo'l o'quvchilarning o'zlashtirish darajalarining matematik statistik tahlili, E3S Web of Conferences EDP Sciences, 371 (2023)
4. R. G. Kotov, Avtomatlashtirilgan boshqaruv tizimlarining lingvistik jihatlari. Moskva, Nauka (1977) 13. A. B. Kutuzov, Korpus tilshunosligi (2017) <http://lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf> (So'nggi kirish 14.05.2023)
5. G. Lich, Korpus lingvistikasidagi san'at darajasi, ingliz korpus tilshunosligi, London (1991)
6. B. Mengliev, O'zbek tili korpusi yaratilmoqdam? "Ma'rifat" gazetasi. 3 aprel (2018-yil) marifat.uz/marifat/v_pomosh_uchitelu-marifat/savol/1142.htm. (So'nggi kirish 14.05.2023)