ISSN 2181-922X

UZBEKISTAN LANGUAGE & CULTURE

O‘ZBEKISTON TIL VA MADANIYAT

KOMPYUTER LINGVISTIKASI

2023 Vol. 4 (6)

www.compling.tsuull.uz

ISSN 2181-922X

O‘ZBEKISTON

TIL VA MADANIYAT

KOMPYUTER LINGVISTIKASI

2023 Vol. 4 (6)

compling.tsuull.uz

Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti

Bosh muharrir: **Botir Elov**

Bosh muharrir o‘rinbosari: **Shahlo Hamroyeva**

Mas’ul kotib: **Oqila Abdullayeva**

Tahrir kengashi

Jurnal haqida ma’lumot

“O‘zbekiston: til va madaniyat. Kompyuter lingvistikasi” seriyasi – Oliy attestatsiya komissiyasi ilmiy nashrlar ro‘yxatidagi “O‘zbekiston: til va madaniyat” akademik jurnalining ilovasi hisoblanib, unda professor-o‘qituvchilar, doktorantlar, stajor-tadqiqotchilar, mustaqil izlanuvchilar, magistrantlarning kompyuter lingvistikasi, jumladan, tabiiy tilga ishlov berish (NLP), o‘zbek tilining formal grammatikasi, korpus lingvistikasi, mashina tarjimasi, nutqni qayta ishlash tizimlari, intellektual tizimlar, kompyuter leksikografiyasi hamda lingvistik ontologiyalar kabi sohalarga oid tadqiqotlari nashr qilinadi.

Jurnal ilovasi bir yilda to‘rt marta chop etiladi.

O‘zbek, turk, rus va ingliz tillarida yozilgan maqolalar qabul qilinadi.

Jurnalda kitoblarga yozilgan taqrizlar, adabiyotlar sharhi, konferensiyalar hisobotlari va tadqiqot loyihalari natijalari ham e‘lon qilinadi.

Mualliflar fikri tahririyat nuqtayi nazaridan farq qilishi mumkin.

“O‘zbekiston: til va madaniyat. Kompyuter lingvistikasi” seriyasi 2023-yildan chiqa boshlagan.

Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti. O‘zbekiston, Toshkent, Yakkasaroy tumani, Yusuf Xos Hojib ko‘chasi, 103-uy.

E-mail: kompling@navoiy-uni.uz
Website: compling.tsuull.uz

Alisher Navo'i Tashkent State University of the Uzbek Language and Literature

Information about the magazine

"Uzbekistan: language and culture. "Computer Linguistics" series is an appendix of the academic journal "Uzbekistan: Language and Culture" in the list of scientific publications of the Higher Attestation Commission, in which computer linguistics, including natural language processing (NLP) of professors-teachers, doctoral students, intern-researchers, independent researchers, master's students, researches related to formal grammar of the Uzbek language, corpus linguistics, machine translation, speech processing systems, intelligent systems, computer lexicography and linguistic ontologies are published.

The magazine supplement is published four times a year.

Articles written in Uzbek, Turkish, Russian and English languages are accepted.

The journal also publishes book reviews, literature reviews, conference reports, and research project results.

The opinion of the authors may differ from the editorial point of view.

"Uzbekistan: language and culture. "Computer Linguistics" series has been published since 2023.

Tashkent State University of Uzbek Language and Literature named after Alisher Navoi. Yusuf Khos Hajib street, 103, Yakkasaray district, Tashkent, Uzbekistan.

E-mail: kompling@navoiy-uni.uz
Website: compling.tsuull.uz

MUNDARIJA

CONTENT

ISSN 2181-922X 38-48

DIALEKTAL KORPUSLARNING UMUMIY TAVSIFI: TAJRIBA VA TAHLIL

Ruxsora Muftillayeva[1]

Annotatsiya. Tildagi dialektal o‘zgarishlar - bu inson nutqining xilma-xilligini yoritib beradigan qiziqarli tadqiqot sohasi. Ushbu abstrakt dialektal shevalarning umumiy tavsifini taqdim etadi, bu til hodisalarini o‘rganish uchun eksperimental usullar va analitik yondashuvlarga e’tibor beradi. Dialektal shevalarni tekshirish orqali tadqiqotchilar dialektlarning tarixiy evolyutsiyasi va geografik taqsimoti haqida qimmatli fikrlarni ochib berishlari mumkin. Ushbu abstrakt dialektal shevalarni o‘rganishda keng qo‘llaniladigan eksperimental texnikalar, ma’lumotlarni yig‘ish usullari va statistik tahlillar haqida umumiy ma’lumot beradi. Bundan tashqari u tilshunoslikdagi dialektal o‘zgaruvchanlikni tushunishning muhimligini va uning tilni saqlash, madaniy merosga ta’sirini ta’kidlaydi.

Kalit so‘zlar: *dialektal korpus, lingvokulturologik, intervyu, sintaktik, morfologik.*

Kirish

Dialektal korpus deganda, ma’lum dialektlardan yoki tilning mintaqaviy turlaridan lingvistik ma’lumotlarni to‘playdigan matnlar yoki yozuvlar to‘plami tushuniladi. Ushbu korpuslar ko‘pincha keng ko‘lamli ma’lumotlar to‘plash harakatlari natijasida, ma’lum bir dialektdagi ona tilida so‘zlashuvchilar bilan o‘zaro aloqalarni o‘z ichiga olgan holda yaratiladi. Dialektal korpus bilan ishlash tajribasi ma’lum bir mintaqaning til boyligini o‘rganishni o‘z ichiga oladi. Ushbu korpuslar bilan ishlaydigan tilshunoslar va tadqiqotchilar bir dialektni boshqasidan ajratib turadigan o‘ziga xos xususiyatlar, lug‘at, grammatika, talaffuz va boshqa jihatlar haqida tushunchaga ega bo‘ladi. Dialektal korpusni tahlil qilish orqali tadqiqotchilar dialektal o‘zgaruvchanlik, vaqt o‘tishi bilan til o‘zgarishi, til aloqasi va tildan foydalanishga ta’sir qiluvchi ijtimoiy omillar kabi lingvistik hodisalarni o‘rganishlari mumkin. Ushbu tadqiqot tilning rivojlanishi, madaniy o‘ziga xoslik til va jamiyat o‘rtasidagi munosabatlarni tushu-

[1]*Muftillayeva Ruxsora Toshmuhammad qizi* – Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti Kompyuter lingvistikasi mutaxassisligi magistranti
E-pochta: *ruxsoramuftillayeva851@gmail.com*
ORCID: 0009-0007-3046-3953

nishimizga yordam beradi. Koʻpgina Yevropa tillarining dialekt korpuslari hozirda mavjud boʻlib, ular odatda ma'lum bir mamlakatning turli mintaqalaridagi materiallarni oʻz ichiga oladi. Portugaliya korpusi –The syntax-oriented Corpus of Portuguese dialects -Portugal lahjalarining sintaksisiga yoʻnaltirilgan korpus (COSYPOR) turli portugal lahjalari sintaksisiga qaratilgan lingvistik manba. U portugal tilida soʻzlashadigan turli mintaqalardan sintaktik ma'lumotlarning keng qamrovli toʻplamini taqdim etishga qaratilgan boʻlib, tadqiqotchilarga dialektlardagi sintaktik tuzilmalarni oʻrganish va solishtirish imkonini beradi.

Asosiy qism

Oʻzbek shevalari bazasining milliy korpusini yaratishda hali tajribamiz yoʻqligini inobatga olsak, bu jarayonda tilning tabiatidan qat'iy nazar dunyo tilshunos olimlari bilan birgalikda yaratilgan korpuslarni koʻzdan kechirishga toʻgʻri keladi [Xolova, 2022. 31].

COSYPOR yozma va ogʻzaki matnlarni, shuningdek, portugal lahjalarida soʻzlashuvchilardan toʻplangan audio yozuvlarni oʻz ichiga oladi. Korpus turli mavzular va janrlarni, jumladan suhbatlar, intervyular, hikoyalar va boshqalarni qamrab oladi. Tilshunoslik tadqiqoti va tahlilini osonlashtirish uchun matnlar nutq qismi teglari va sintaktik tahlil kabi lingvistik ma'lumotlar bilan izohlanadi. Korpus tilshunoslar, tadqiqotchilar va portugal lahjalari sintaksisini oʻrganishga qiziqqan talabalar tomonidan foydalanish uchun moʻljallangan. U sintaktik oʻzgarishlarni oʻrganish va turli dialektlarning oʻxshash va farqlarini tushunish uchun qimmatli manba boʻlib xizmat qiladi. COSYPOR ma'lumotlarini tahlil qilish orqali tadqiqotchilar portugal lahjalarining sintaktik tuzilmalari haqida tushunchaga ega boʻlishlari va tilshunoslik sohasiga hissa qoʻshishlari mumkin. Resurs turli mazmundagi korpuslarga (asosan, Portugal va Braziliya gazetalari matnlari, shuningdek, portugal fantastika asarlari toʻplami, Braziliya elektron pochta xabarlari va boshqalar) kirish imkonini beradi. Umuman olganda, toʻplamning aksariyat qismini portugal tilining brazil tilidagi versiyasini aks ettiruvchi matnlar egallaydi. Deyarli barcha matnlar morfologik belgilar bilan ta'minlangan. Korpusning umumiy hajmi 70,8 million soʻzdan foydalanishni tashkil etadi. Ulardan 69,8 millioni morfologik, mantiqiy izohlangan. Korpusga kirish bepul. Umuman olganda, Portugal lahjalarining sintaksisiga yoʻnaltirilgan korpusi portugal lahjalari sintaksisini oʻrganish uchun qimmatli vosita boʻlib, tadqiqotchilarga portugal tilida soʻzlashadigan turli mintaqalardagi sintaktik oʻzgarishlarni oʻrganish va tahlil qilish imkonini beradi.

Corpus Oral y Sonoro del Español Rural- Ispaniya korpusi umumiy nom bilan COSER deb nomlanuvchi korpus dialektal korpus bo'lsa ham, lekin u an'anaviy dialektologiyaga qiziqish obyekti bo'lgan ma'lumot beruvchilarning nutqi bilan cheklangan: qishloqda yashovchi aholi asosan yoshi katta, maktab ma'lumoti o'rtacha va ular suhbatlashgan hududda tug'ilgan. Bugungi kunda (2022-yil dekabr) ma'lumotlariga ko'ra 2961 nafar ma'lumot beruvchi ro'yxatga olingan.

1-jadval. COSER korpusining qatnashuvchilari haqidagi ma'lumotlar

Axborot beruvchilar	**Soni**	**O'rtacha yoshi**
Erkaklar:	1.415 (47,8%)	75 yosh
Ayollar:	1.546 (52,1%)	73,6 yosh
Jami:	2.961	74,2 yosh

COSERni tashkil etuvchi yozuvlar 1990-yildan 2022-yil dekabrigacha bir qator so'rov kompaniyalaridan olingan. Ushbu korpus ishi bir nechta tadqiqot loyihalari ko'magida va "Dialectologia Hispanica", "El español hablado. Variantes peninsulares", Madrid avtonom universitetida ispan filologiyasi bakalavriatiga tegishli ixtiyoriy fanlar (Universidad Autónoma de Madrid, UAM). 2011-yildan hozirgi kunga qadar ular ushbu universitetda ispan tilini o'rganish darajasining "Lengua española. Variedades de la lengua" (3-kurs) fanining ixtiyoriy faoliyati sifatida birlashtirilgan.

2-jadval. COSER korpusining audio yozuvlari haqida m'lumot

Yozib olingan hududlar	Viloyat va orollar	Yozib olingan ma'lumotlarning umumiy hajmi	Har bir intervyu uchun o'rtacha yozib olish	Suhbatlar soni	Matn va audio shaklidagi intervyular (may 2022)
1,415	55	1,910 soat	1soat, 4 min.	1,772	218

2022-yilgacha suhbatlar Pireney yarim oroli va ikkita arxipelagdagi 55 provinsiya yoki orolga tegishli 1415 ta qishloq aholi punktlarida o'tkazilgan. Ularning geografik joylashuvi xaritada ko'rsatilgan, bu yerda ularni viloyat va hududni alifbo tartibida umumlashtiruvchi raqamli kod yordamida aniqlash mumkin (masalan, Alava provinsiyasidagi Berganzo shahrida 0101 kodi mavjud). Ovoz materiallari Pireney yarim orolining katta qismini qamrab oladi. Umuman olganda, COSER hozirda 1910 soatlik yozuvlarga ega. Ularning aksariyati analog formatda yozilgan. Materiallarning yarmi

turli xil ilmiy loyihalar tomonidan qo‘lga kiritilgan va o‘zlarining akademik kurs ishlarining bir qismi sifatida o‘zlari to‘plagan yozuvlarni transkripsiya qilgan UAM bakalavriat talabalarining bir necha avlodlari ishtiroki tufayli, turli xarakterdagi va aniqlikdagi transkripsiyalarga ega. Korpus 2015-yilda BConcord muharriri bilan qayta ko‘rib chiqilgan va standartlashtirilgan 141 ta hududga (taxminan 183 soat) mos keladigan 147 ta transkripsiya ushbu veb-saytda nashr etildi va qidiruv tizimi orqali qidirish mumkin bo‘ldi. 2017-yildan beri korpusga oddiy qidiruv va kengaytirilgan qidiruv rejimlarida ham kirish mumkin (bu lemmalar va morfosintaktik teglar bo‘yicha so‘rovlar o‘tkazish imkonini beradi). 2020-yilda ushbu so‘rovda geografik koordinatalar va joylarning pochta indeksi yoqilgan bo‘lib, ma’lumotlar geografik axborot tizimlarida tahlil qilinishi mumkin va matnni audio bilan sinxronlashtirish tugallangan.

3- jadval. COSER korpusining transkripsiya qilingan so‘zlar haqida ma’lumot

Transkripti bor hududlar	Viloyatlar va orollar	Transkripsiya qilingan soatlar	Umumiy transkripsiya qilingan so‘zlar	Umumiy birliklar (tokenlar)
218	55	295 soat, 48 minut	3,596,205 so‘zlar	4,591,828

Deutsches Referenzkorpus (DeReKo) nemis tilidagi yozma matnlarning katta korpusidir. DeReKo dunyodagi eng yirik korpuslardan biri bo‘lib, 24 milliarddan ortiq so‘zdan iborat. U turli janr va sohalardagi yozma matnlarning keng doirasini o‘z ichiga oladi. Korpus 20-asrdan to hozirgi kungacha bo‘lgan turli davrlardagi matnlarni qamrab oladi. U gazetalar, jurnallar, kitoblar, veb-saytlar va boshqa manbalardan olingan matnlardan tashkil topgan. DeReKo matnlar uchun batafsil lingvistik izohlarni taqdim etadi, jumladan nutq qismlarini teglash, lemmatizatsiya va sintaktik tahlil qilish. Bu uni lingvistik tadqiqotlar va til tahlili uchun qimmatli manbaga aylantiradi. DeReKoga kirish Germaniyaning Mannheim shahridagi Nemis tili instituti (Institut für Deutsche Sprache, IDS) orqali mavjud. Tadqiqotchilar korpusga kirish uchun ariza topshirishlari va undan tadqiqot loyihalari uchun foydalanishlari mumkin.

FRED - ingliz dialektlarining FREIBURG korpusi FRED - bu bir tilli og‘zaki dialekt korpusi bo‘lib, u Angliya, Shotlandiya, shuningdek (to‘liq versiyada) Uels, Hebridlar va Men orolidagi ona tilida

so‘zlashuvchilar bilan to‘liq metrajli audio suhbatlardan tashkil topgan. Korpus audio yozuvlardan (wav formatda aksariyat intervyular to‘liq formatda berilgan) va orfografik transkriptlardan (txt fayllari) iborat.

4- jadval. FRED korpusining umumiy tuzilishi haqida ma’lumot

Loyiha rahbari	Tuzilish vaqti:	Hajmi:	Til:	Matnlar soni:	Davr:	Holati:
Prof. Dr. Bernd Kortmann	2000–2005 (loyiha guruhi “Tipologik nuqtai nazardan ingliz shevasi sintaksisi”)	1 011 396 so‘z (to‘liq versiya 2 496 763, intervyu oluvchining so‘zlaridan tashqari)	Ingliz (Britaniya, Shotlandiya, [Uels, Manx] navlari)	121 intervyu/ transkript	1970–1999	2005- yilda tugallangan

FRED korpusi yana quyidagilarni o‘z ichiga oladi: 1 011 396 ta o‘zgaruvchi so‘z, 123 soat yozib olingan nutq, 121 ta intervyu, 144 ta dialekta so‘zlashuvchi 57 ta turli joylarda, 18 ta okrugda, 5 ta asosiy dialekt hududida yozib olingan.

5-jadval. Yozuv sanasi bo‘yicha matn taqsimoti, FRED-S

Qayd etilgan sana	Matnlar soni	Matn materialining %
1970-1979 yillar	47	42,2%
1980-1989 yillar	56	43,9%
1990-1999 yillar	15	10,3%
Noma’lum	3	3,6 %
Jami	121	100,0%

6- jadval. Matnni dialekt maydoni bo‘yicha taqsimlash, FRED-S.

Dialekt maydoni	Matnlar soni	O‘zgaradigan so‘zlar	Matn materialining %
Janubiy-g‘arbiy (SW)	38	264 863	26,2%
Janubiy-sharqiy (SE)	17	260 643	25,8%
Midlans (o‘rta)	16	152 535	15,1 %
Shimoliy (N)	30	266 955	26,4%
Shotlandiya pasttekisligi (ScL)	20	66 400	6,6%
Jami	121	1 011 396	100,0 %

Har bir matndan oldin matn sarlavhasi mavjud bo‘lib, unda matn identifikatori, shuningdek dialekt hududi, okrug va suhbat bo‘lib, o‘tgan joy (ya’ni, ma’ruzachi qayerdan kelgan), mavjud sotsiolingvistik ma’lumotlar va suhbat sanasi haqidagi ma’lumotlar mavjud.

Og‘zaki tarix bo‘yicha suhbatning o‘ziga xos xususiyatlaridan biri shevalarni o‘zining eng asl ko‘rinishida o‘rganishning muhim sharti bo‘lib, so‘zlovchilar o‘z umrlarining ko‘p qismini ma’lum bir geografik hududda o‘tkazgan bo‘lishi va bu hudud bilan mustahkam aloqada bo‘lishi kerak. Ma’ruzachilarning aksariyati birinchi jahon urushidan oldin tug‘ilgan va suhbat vaqtida 60 va undan katta yoshda bo‘lgan (eng keksa ma’ruzachi 1877-yilda tug‘ilgan; barcha ma’ruzachilarning deyarli 90 foizi 1920-yilgacha tug‘ilgan). FRED-S dagi matnli materialning qariyb 70% 60+ yosh guruhi tomonidan yozib olingan intervyulardir.

7-jadvalda. Intervyu beruvchilar va yosh guruhlari bo‘yicha ma’lumot berilgan

Yosh guruhlari	Ma’ruzachilar soni	O‘zgaradigan so‘zlar	Matn materialining %
0-44 yosh	4	12 287	1,2 %
45-59 yosh	5	40, 258	4,0 %
60+ yosh	71	726,134	71,8 %
Yoshi noma’lum	64	232 558	23,0 %

8-jadval. Intervyu beruvchilarning tug‘ilgan yili bo‘yicha matn taqsimoti

Tug‘ilgan yili	Ma’ruzachilar soni	O‘zgaradigan so‘zlar	Matn materialining %
1870-1879-yillar	1	6899	0,7 %
1880-1889-yillar	5	89 615	8,9%
1890-1899-yillar	28	260 909	25,8 %
1900-1909-yillar	30	281 068	27,8 %
1910-1919-yillar	21	176 507	17,5 %
1920-1929-yillar	7	74 365	7,4 %
1930-1939-yillar	3	23 494	2,3 %
1940-1949-yillar	1	1684	0, 2 %
Noma’lum	48	96 696	9,5 %

Ma’ruzachilarning uchdan ikki qismidan ko‘prog‘i NORM deb ataladi, ya’ni mobil bo‘lmagan keksa qishloq erkaklari bo‘lib, ular odatda o‘n to‘rt yoki undan kichik yoshda maktabni tark etishgan. Intervyu beruvchilarning jami 87 erkak va 52 ayol ma’lumot

beruvchidan iboratdir. Korpusda ayol va erkak ismlari va taxalluslari, bosh harflari va familiyalari uchun turli teglar ishlatilgan (obyektlar, kompaniyalar, brendlar va boshqalar nomlari o‘zgarishsiz qoladi). Nashr etilgan va davom etayotgan tadqiqotlar, jumladan, magistrlik va doktorlik dissertatsiyalari haqidagi dolzarb ma’lumotni loyiha veb-saytidan topishingiz mumkin: http://www2.anglistik.uni-freiburg.de/institut/lskortmann/FRED/.

Nordic Dialect corpus (NDC) – korpus Daniya, Islandiya, Norvegiya, Farer va Shvetsiya kabi davlatlarning og‘zaki so‘zlashuv tillaridan tashkil topgan. U shimoliy german tillarining barcha shimoliy mamlakatlaridagi dialektlarning spontan nutq ma’lumotlaridan iborat. Korpusdagi lingvistik ma’lumotlar turli manbalardan olingan. Korpusda dialektlarda so‘zlashuvchilarning suhbatlari va intervyularidan olingan 2,75 milliondan ortiq so‘z mavjud. U transkripsiya qilingan audio va videoga bog‘langan, xarita funksiyasiga ega va uni turli yo‘llar bilan qidirish mumkin. Korpusning maqsadi shimoliy sintaksis tadqiqoti bo‘lsa ham, korpus umumiy, Norvegiya dialekti korpusi, shved dialekti korpusi va boshqalar bo‘lib, fonologiya, morfologiya va leksikografiya kabi keng ko‘lamli tadqiqot sohalarida qo‘llaniladi. Ma’lumotlar bazasi turli sintaktik hodisalarni aks ettiruvchi jumlalar ro‘yxatiga 207 joydan 924 ta shimoliy lahjada so‘zlashuvchilarning nutqlaridan iborat. Ko‘pgina ma’ruzachilar ma’lumotlar bazasida ham, korpusda ham bir xil. Korpusda joy, yosh, ma’lumot beruvchilarning jinsi yoki sintaktik hodisa turiga qarab saralangan. Nordic dialektal korpusi yaratilgandan boshlab hozirgi vaqtga qadar takomillashtirilib borilmoqda, xususan korpusga bir qancha yangi funksiyalar kiritilgan.

Nordic dialect Corpus v. 4.0: Faqat 1998-2015-yillardagi dialekt yozuvlari va transkripsiyalari. (2019-yil sentyabr)

Yangi qidiruv interfeysida Nordic dialect Corpus uchun foydalanuvchi qo‘llanmasi (2019-yil iyun)

Nordic dialect Corpus v. 3.0: kengaytirilgan Islandiya va Shvetsiya qismi. Islandiyadan 16 nafar, Shvetsiya va Swedia 2000dan 24 nafar yangi informator (2017-yil sentyabr)

Nordic Dialect Corpus v. 2.0 va Nordic Syntax Database uchun yangi qidiruv interfeyslari. (2017)

Nordic Atlas of Language Structures (NALS) jurnali chop etildi (2014)

Nordic dialekt korpusi: kengaytirilgan Islandiya qismi - 6 joydan 20 ta yangi ma’lumot beruvchi (2013)

Nordic Dialect Corpus v. 4.0 uchun new qidiruv interfeyslari (2023-yil noyabr)

9- jadval. Norvegiya sheva korpusi jamlanmasi

Mamlakat	Ma'lumot beruvchilar soni	Hududlar	Identifikatsion belgilar
Daniya	81	15	220, 360
Farer	20	5	64,803
Islandiya	48	8	94,338
Norvegiya	438	111	1,997,920
Shvetsiya (jumladan ovdalian)	150	44	376,868

Ushbu korpus sistemasiga 3 ta kirish tizimi mavjud: 1. EDUGAIN. 2. CLARIN (Skandinaviya shevalar korpusi). 3. FEIDE (Norveg millatiga mansub aholiga foydalanish imkonini beruvchi). Korpusdagi ba'zi yozuvlar ikki tomonlama transkripsiya - orfografiya yoki transkripsiya fonetik yig'ilmalar jamlanmasiga ega. Transkripsiyalarni eshitish yoki ko'rish mumkin bo'lgan audio va video fayl taqdimotlari tayyorlangan.

Helsenki korpusi - Britaniya ingliz dialektlari korpusi bo'lib, asosan, Sharqiy Angliya va Janubiy-G'arbiy hududlardan, Lankashirdan kichik to'plamga ega bo'lgan orfografik transkripsiyalangan audio yozuvlar to'plamidir. Yozuvlar 1970 va 1980-yillarda Finlyandiya aspirantlari tomonidan yig'ilgan. Dialektologik korpusning maqsadi nafaqat dialektologiya, balki sotsiolingvistika, nutq tahlili, morfologiya, sintaksis va fonologiya sohalarida lingvistik tadqiqotlar uchun material taqdim etishdan iborat. Korpus shuningdek, aloqa etnografiyasi, mahalliy odatlar va tarix kabi tilga oid bo'lmagan, ko'p tarmoqli tadqiqotlar uchun material beradi.

10- jadval. Helsenki korpusi haqida ma'lumot

Loyiha rahbarlari	Hajmi	Vaqt davrlari	Til
Ossi Ihalainen Kristi Peitsara Anna-Liisa Vasko	**1.008.641** so'zdan iborat jami **187** ta fayl.	1970-1980-yillar	Ingliz qishloq (Kembridgeshire, Devon, Ile of Ely, Somerset, Suffolk) va shahar (Esseks, Lankashir).

Helsenki korpusining birinchi bosqichi 2006-yilda yakunlangan, ikkinchi bosqichi esa hali ham davom etmoqda. Korpusning asosiy materiallari og'zaki dialekt nutqining audio yozuvlardan iborat. Helsenki korpusi Finlandiya madaniyat jamg'armasi, Finlandiya akademiyasi, Helsenki universiteti tomonidan moliyalashtirilgan.

11- jadval. Korpusga kiritilgan geografik mintaqalar haqida ma'lumot

Viloyat (qishloq)	Qishloqlar soni	Ma'lumot beruvchilar soni
Kembridjeshire	26	38 (+6)
Devon	9	33 (+8)
Ely oroli	20	52 (+6)
Somerset	15	24 (+5)
Suffolk	16	47 (+4)
Hudud (shahar)		
Essex	3	6 (+1)
Lancashir	3	6 (+1)
Jami	92	206 (+31)

Korpusda asosan ma'lumot beruvchilar sifatida erkaklarning ulushi ko'proq. Korpusda ayollar nutqi taxminan beshdan bir qismni tashkil etadi. 1970 va 1980-yillardagi suhbat chog'ida qishloq xabarchilarining aksariyati nafaqaga chiqqan va yoshi 70 dan oshgan yoki undan katta bo'lgan, shuning uchun XX asrning birinchi o'n yilligida ta'lim olgan. 1980-yillarning oxirida to'plangan Essex va Lancashire shahar korpuslari yuqoridagilardan uch avlod ikki jinsli namunalarni berishda farq qiladi, ma'lumot beruvchilar 19 yoshdan 70 yoshgacha. Fayllardan foydalanishga ruxsat olish uchun Anna-Liisa Vasko (anna-liisa.vasko@helsinki.fi) yoki Kirsti Peitsara (kirsti.peitsara@helsinki.fi) bilan bog'lanish kerak.

Ma'lum bir hududning shevasi asosida yaratilgan dialektal korpuslar ham bor. Masalan: Kuban dialektal korpusi shunday korpuslardan hisoblanadi. Kuban dialektal korpusi 18 yil davomida Kubanning g'arbiy qishloqlari an'anaviy madaniyatini o'rganish asosida yaratilgan. Bu loyiha Rossiya gumanitar jamg'armasi, Krasnodar o'lkasi ma'muriyati tomonidan qo'llab-quvvatlangan. 2014-yilda Rossiya gumanitar fondi va Krasnodar o'lkasi ma'muriyati tomonidan "Kuban dialekt madaniyatining elektron korpusini yaratish" loyihasi tasdiqlandi. 2016-yilda Rossiya gumanitar fondidan (2016-yildan buyon u Rossiya fundamental tadqiqotlar fondi tarkibiga kiradi) ekspeditsiya ishlari uchun grant olinadi. Korpus tarkibiga "marosim madaniyati", "ma'naviy madaniyat", "hunarmandchilik madaniyati", "kundalik madaniyat", "xalq hunarmandchiligi va san'ati", "oilaviy turmush tarzi" kabi kichik korpuslar kiradi. Hozirgi vaqtda "Ritual madaniyat" (to'y marosimi ma'ruzasi) va "ma'naviy madaniyat" ("mifologiya", "xalq pravoslavligi" ma'ruzalari) kabi korpuslar mazmunini shakllantirish bo'yicha ishlar olib borilmoqda. 2015-yilda loyi-

hani moliyalashtirish davom etdi. Korpusga turli subkorporalarning nutqlari materiallari bo‘yicha lingvokulturologik tahlil o‘tkazildi. Korpus uchun yig‘ilgan materiallarning keyinchalik lenta va video kassetalar, elektron vositalar, yozilgan dialekt nutqining raqamli versiyalari yaratildi. Turli tematik nutqlarni tahlil qilish jarayonida ularning tarkibiy va semantik birligini tartibga soluvchi mikro-mavzular, tushunchalar to‘plamini aniqlash mumkin bo‘ldi.

Xulosa

Xulosa qilib aytadigan bo‘lsak, dialektal korpuslarni o‘rganish va tahlil qilish turli dialektlarning lingvistik xususiyatlari haqida qimmatli ma’lumotlarni beradi. Dialekt matnlarining katta to‘plamlarini to‘plash va tekshirish orqali tadqiqotchilar bir dialektni boshqasidan ajratib turuvchi lingvistik xususiyatlarni chuqurroq tushunishlari mumkin. Dialektal korpus bilan ishlash tajribasi turli qiyinchiliklar va mulohazalarni o‘z ichiga oladi. Birinchidan, dialektal matnlarni to‘plash va tuzish ko‘p vaqt va mehnat talab qiladigan jarayon. Tadqiqotchilar turli xil manbalardan matnlarni aniqlashlari va to‘plashlari kerak, ular dialektal o‘zgarishlarning keng doirasini ifodalaydi. Korpus qurilgandan so‘ng, tadqiqotchilar turli lingvistik vositalar va usullardan foydalangan holda ma’lumotlarni tahlil qilishadi. Bu tahlil muayyan shevaga xos bo‘lgan fonetik, fonologik, morfologik, sintaktik va leksik xususiyatlarni aniqlashni o‘z ichiga olishi mumkin. Bundan tashqari, tadqiqotchilar lahjani shakllantirgan tildan foydalanish qonuniyatlarini, sotsiolingvistik omillarni va tarixiy ta’sirlarni o‘rganadi. Bunday tadqiqot natijalari tilshunoslik, sotsiolingvistika, antropologiya va tilni saqlash kabi turli sohalarga ta’sir qiladi.

Foydalanilgan adabiyotlar

Холова М. ўзбек миллий шевалари корпусини тузишнинг лингвистик асослари (Бойсун тумани “ж”ловчи шевалари мисолида) Фил. фан. бўйича фалсафа доктори (PhD)...дисс. Автореф. – Термиз, 2022.

http://rusling.narod.ru/qqq_corp_nonslav_other.htm

http://www.corpusrural.es

https://freidok.uni-freiburg.de/proj/

Nordic Dialect Corpus (uio.no)

https://www.helsinki.fi/varieng/CoRD/corpora/Dialects/fieldwork

Региональная этнолингвистика (ethnolex.ru)

GENERAL DESCRIPTION OF DIALECTAL CORPSES: EXPERIMENT AND ANALYSIS

Ruxsora Muftillayeva[1]

Abstract. Dialectal variations in language are a fascinating area of study that sheds light on the diversity of human communication. This abstract presents a general description of dialectal corpses, focusing on the experimental methods and analytical approaches used to explore these linguistic phenomena. By examining dialectal corpses, researchers can uncover valuable insights into the historical evolution and geographical distribution of dialects. This abstract provides an overview of the experimental techniques, data collection methods, and statistical analyses commonly employed in studying dialectal corpses. Furthermore, it highlights the importance of understanding dialectal variation in linguistics and its implications for language preservation and cultural heritage.

Key words: *dialectal corpus, language culture logic, interview, syntax, morphology.*

References

Xolova M. o'zbek milliy shevalari korpusini tuzishning lingvistik asoslari (Boysun tumani "j"lovchi shevalari misolida): Fil. fan. bo'yicha falsafa doktori (PhD)...diss. Avtoref. – Termiz, 2022. http://rusling.narod.ru/qqq_corp_nonslav_other.htm

http://www.corpusrural.es

https://freidok.uni-freiburg.de/proj/

Nordic Dialect Corpus (uio.no)

https://www.helsinki.fi/varieng/CoRD/corpora/Dialects/fieldwork

Региональная этнолингвистика (ethnolex.ru)

[1]*Muftillayeva Ruxsora Toshmuhammad qizi* –Master of degree, Alisher Navo'i Tashkent State University of Uzbek Language and Literature.
E-mail: *ruxsoramuftillayeva851@gmail.com*
ORCID: 0009-0007-3046-3953

Jurnal 2017-yil 26-oktyabrda O‘zbekiston Respublikasi Matbuot
va axborot agentligi tomonidan 0936-raqam bilan ro‘yxatdan o‘tgan.

Jurnal O'zbekiston Respublikasi Oliy Attestatsiya Komissiyasi tomonidan
filologiya fanlari bo'yicha falsafa doktori (PhD) va fan doktori (DSc)
dissertatsiyalari asosiy ilmiy natijalari chop etilishi lozim bo'lgan ro'yxatga
kiritilgan (30.10.2021. № 308/6).

Tahririyatga kelgan maqolalar mualliflarga qaytarilmaydi.

Manzil: Toshkent shahri, Yakkasaroy tumani, Yusuf Xos
Hojib ko‘chasi 103-uy.
Telefonlar: +99871 281-45-11, +99871 281-41-93.
Website: compling.tsuull.uz
E-mail: kompling@navoiy-uni.uz

Bosishga 25.12.2023-yilda ruxsat etildi.
Bichimi 70x100 1/16, Ofset bosma. “Cambria” garniturasi.
Shartli b.t. 7,51. Nashr b.t. 7,62.

“O‘zbekiston: til va madaniyat” jurnali tahririyatida
tayyorlandi va sahifalandi.
“YASHNOBOD NASHR” bosmaxonasida chop etildi.
Adadi 300 nusxa. Buyurtma №2.
Bosmaxona manzili: Toshkent shahar Yashnobod tumani,
58-a harbiy shaharcha.