**FARG‘ONA DAVLAT UNIVERSITETI HUZURIDAGI
ILMIY DARAJALAR BERUVCHI
DSc.03/30.12.2022.Fil.05.02 RAQAMLI ILMIY KENGASH**

BUXORO DAVLAT UNIVERSITETI

ATABOYEV NOZIMJON BOBOJON O‘G‘LI

**MEDIAMATNLAR DIAXRONIK KORPUSIDA TIL RIVOJINING EMPIRIK
TAHLIL TAMOYILLARI
(gazeta materiallari misolida)**

**10.00.11 – Til nazariyasi. Amaliy va
kompyuter lingvistikasi**

**FILOLOGIYA fanlari doktori (DSc) dissertatsiyasi
AVTOREFERATI**

Farg‘ona – 2024

Fan doktori (DSc) dissertatsiyasi avtoreferati mundarijasi

Contents of dissertation abstract of Doctor of Philology (DSc)

Содержание автореферата диссертации доктора филол. наук (DSc)

**FARG'ONA DAVLAT UNIVERSITETI HUZURIDAGI
ILMIY DARAJALAR BERUVCHI
DSc.03/30.12.2022.Fil.05.02 RAQAMLI ILMIY KENGASH**

BUXORO DAVLAT UNIVERSITETI

ATABOYEV NOZIMJON BOBOJON O'G'LI

**MEDIAMATNLAR DIAXRONIK KORPUSIDA TIL RIVOJINING EMPIRIK
TAHLIL TAMOYILLARI
(gazeta materiallari misolida)**

**10.00.11 – Til nazariyasi. Amaliy va
kompyuter lingvistikasi**

**FILOLOGIYA fanlari doktori (DSc) dissertatsiyasi
AVTOREFERATI**

Farg'ona – 2024

Fan doktori (DSc) dissertatsiyasi mavzusi O'zbekiston Respublikasi Oliy attestatsiya komissiyasida B2022.3.DSc/Fil465 raqam bilan ro'yxatga olingan.

Doktorlik dissertatsiyasi Buxoro davlat universitetida bajarilgan.

Dissertatsiya avtoreferati uch tilda (o'zbek, ingliz va rus (rezyume)) Farg'ona davlat universiteti veb-sahifasida (https://fdu.uz/) va "Ziyonet" Axborot ta'lim portalida (www.ziyonet.uz) joylashtirilgan.

Ilmiy maslahatchi: **Teshabayeva Dilfuza Mo'minovna**
filologiya fanlari doktori, professor

Rasmiy opponentlar: **Sultonova Shoxista Muxammadjanovna**
filologiya fanlari doktori (DSc), professor

Abduraxmonova Nilufar Zaynobiddin qizi
filologiya fanlari doktori (DSc), professor

Dr. Mario Bisiada
filologiya fanlari doktori (Ispaniya)

Yetakchi tashkilot: **Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti**

Dissertatsiya himoyasi Farg'ona davlat universiteti huzuridagi ilmiy darajalar beruvchi DSc.03/30.12.2019.Fil.05.02 raqamli Ilmiy kengashning "26" august 2024-yil soat 8:00dagi majlisida bo'lib o'tadi (Manzil: 100151, Farg'ona shahri, B.Marg'inoniy ko'chasi, 105-uy. Tel.: (99873) 244-57-82; faks: (99873) 244-44-01; e-mail: info@fdu.uz).

Dissertatsiya bilan Farg'ona davlat universitetining Axborot-resurs markazida tanishish mumkin (366 raqami bilan ro'yxatga olingan). Manzil: 100151, Farg'ona shahri, Murabbiylar ko'chasi, 19-uy. Tel.: (99873) 244-71-28.

Dissertatsiya avtoreferati 2024-yil "13" august kuni tarqatildi.
(2024-yil "13" august da 22 raqamli reyestr bayonnomasi).

S.M. Muminov
Ilmiy darajalar beruvchi Ilmiy kengash raisi, filol.f.d., professor

M.T. Zokirov
Ilmiy darajalar beruvchi Ilmiy kengash kotibi, filol.f.m, professor

A. Mamajonov
Ilmiy darajalar beruvchi Ilmiy kengash qoshidagi Ilmiy seminar raisi, filol.f.d., professor

<center>**KIRISH (fan doktori (DSc) dissertatsiyasi annotatsiyasi)**</center>

Tadqiqot mavzusining dolzarbligi va zarurati. Jahon tilshunosligida globallashuv tendensiyalarining jadallashayotgani bois madaniyatlararo hamda tillararo yaqinlashuv jarayoni yuzaga kelib, o'zaro yaqin aloqadagi tillar o'rtasidagi interferensiya kuchayishining sof leksik qatlam yaruslariga ta'sirini o'rganish zamonaviy tilshunoslikning dolzarb muammosi sifatida muhim ahamiyat kasb etadi. Shunga ko'ra, tilni jamiyatning uzviy qismi sifatida til sohibidan ajralmagan holda, ya'ni antropotsentrik paradigma nuqtayi nazaridan tadqiq etilayotgani tilshunoslar oldiga tilning tarixiy evolyutsiyasiga uyg'unlashgan bugungi holatini o'rganib, uning kelajagiga oid tadrijiy tahlillarga asoslangan ilmiy farazlarni ishlab chiqish talabini qo'ymoqda. Bu borada xronologik tartiblangan, aniq statistik me'yorlarga tayanib, asliyatdan olingan autentik materialar/matnlar asosida shakllantiriladigan diaxronik korpuslar tatbiqi o'ziga xos yondashuvni taqozo etadi.

Dunyo lisoniy tadqiqotlar doirasida keng tarqalib, zamonaviy ilm-fan talablariga mos ravishda kompyuter texnologiyalaridan foydalanib, taraqqiy etayotgan korpus lingvistikasi o'zining metodologiyasi asosida tilshunoslikning barcha muammolariga yechim izlashning yangicha qulay usullarini taklif etmoqda. Metodlar tizimining kompyuter lingvistikasi, matn lingvistikasi, matematik lingvistika va amaliy lingvistika kabi fan tarmoqlarining uyg'unligidan hosil bo'lganligi bu borada olib borilayotgan izlanishlarning mohiyatan dolzarbligini belgilaydi.

Mamlakatimizda davlat tili maqomini yanada mustahkamlash borasida keng qamrovli islohotlarning siyosiy irodaga tayanganligi tilshunoslar oldiga ona tilimizda bitilgan qadimiy qo'lyozmalardan tortib, bugungi kun lisoniy voqelik qamrovidagi matniy birliklargacha bo'lgan manbalarning elektron bazasini, ya'ni korpusini yaratib, yangicha yondashuv asosida tadqiq etish vazifasini yuklamoqda. "Ma'lumki, har qanday tilni o'rganish va rivojlantirishda, birinchi navbatda, uning nazariy-metodik asoslari katta o'rin tutadi. Shular qatorida o'zbek tilining elektron milliy korpusi va dasturiy mahsulotlarning o'zbekcha ilovalarini yaratish, o'zbek tilini Internet tarmog'ida ommalashtirish va munosib o'rin egallashini ta'minlash, davlat tilini o'rgatishga hamda matn tahririga mo'ljallangan dasturlarni ishlab chiqish va amaliyotga tatbiq qilish bo'yicha ham jiddiy ishlar olib borilayotganini qayd etish zarur"[1]. Shunday ekan, tadqiqot obyekti sifatida tanlangan gazeta materiallari misolida o'zbek tilining diaxronik korpusini ilk bor ishlab chiqish asnosida ona tilimizning leksik qatlami rivojini empirik tahlil etish orqali uning til yarusidagi o'zgarishlarni qayd etib borish kabi muhim vazifalar ilmiy ishning zaruriyatini belgilaydi.

O'zbekiston Respublikasi Prezidentining 2019-yil 21-oktabrdagi PF-5850-son "O'zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to'g'risida", 2020-yil 20-oktabrdagi PF-6084-son "Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida"gi farmonlari, 2021-yil 19-maydagi PQ-5117-son "O'zbekiston

---

[1] Prezident Shavkat Mirziyoyevning O'zbek tili bayrami munosabati bilan O'zbekiston xalqiga tabrigi. https://president.uz/oz/lists/view/6784 [murojaat sanasi 29.02.2024]

Respublikasida xorijiy tillarni o'rganishni ommalashtirish faoliyatini sifat jihatidan yangi bosqichga olib chiqish chora-tadbirlari to'g'risida", 2024-yil 12-yanvardagi PQ-19-son "Xorijiy davlatlarda O'zbekiston tarixi va madaniyati, o'zbek tili va adabiyotini o'rganishni keng targ'ib qilish orqali turistlar oqimini ko'paytirish hamda mamlakatimiz nufuzini yanada oshirishga doir qo'shimcha chora-tadbirlar to'g'risida"gi qarorlari, O'zbekiston Respublikasi Vazirlar Mahkamasining 2020-yil 11-martdagi 139-son "O'zbek tili va adabiyoti bo'yicha fundamental va amaliy tadqiqotlar samaradorligini yanada oshirish chora-tadbirlari to'g'risida", 2020-yil 3-martdagi 117-son "Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti faoliyatini yanada takomillashtirish chora-tadbirlari to'g'risida"gi qarori hamda mavzuga oid boshqa me'yoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda ushbu dissertatsiya tadqiqoti muayyan darajada xizmat qiladi.

Tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo'nalishlariga mosligi. Dissertatsiya respublika fan va texnologiyalar taraqqiyotining I. "Axborotlashgan jamiyat va demokratik davlatni ijtimoiy, huquqiy, iqtisodiy, madaniy, ma'naviy-ma'rifiy rivojlantirishda innovatsion g'oyalar tizimini shakllantirish va ularni amalga oshirish yo'llari" ustuvor yo'nalishiga muvofiq bajarilgan.

Dissertatsiya mavzusi bo'yicha xorijiy ilmiy-tadqiqotlar sharhi[2].

Lisoniy korpuslar yaratish amaliyoti hamda korpus lingvistikasi fan tarmog'ining bugungi yutuqlaridan foydalangan holda til tadqiqi masalasini o'rganish bo'yicha dunyoning yetakchi ilmiy markazlari va oliy ta'lim muassasalari, shu jumladan, Lancaster University, Lancaster (Birlashgan Qirollik), The University of Birmingham (Birlashgan Qirollik), Macquarie University, Sidney (Avstraliya), University of Portsmouth (Birlashgan Qirollik), Lancaster summer schools (Birlashgan Qirollik), University of Sydney, Sidney (Avstraliya), Buxoro davlat universiteti (O'zbekiston), Alisher Navoiy nomidagi O'zbek tili va adabiyoti universiteti (O'zbekiston) da olib borilmoqda.

Jahon tilshunosligida korpus lingvistikasi hamda lisoniy korpuslarni til tarixi, uning rivojlanish tendensiyalari hamda madaniyatlararo interferensiya muammolari tahlilida qo'llashga oid tadqiqotlar natijasida, jumladan, quyidagi ilmiy natijalar olingan: ingliz tilining tuzilishi va g'arb tillari tahlili uchun korpus lingvistik usullarning qo'llanilishi (Lancaster University, Lancaster (Birlashgan Qirollik), korpus lingvistikasi metodologiyasiga asoslangan ilk leksikografik tadqiqotlar hamda stilistika, nutq tahlili, psixolingvistika, kognitiv lingvistika va statistika kabi sohalar bilan korpus tahlil metodlari uyg'unligini aniqlash (The University of Birmingham (Birlashgan Qirollik), turli tadqiqot yo'nalishlari, jumladan, bola tilining rivoji, ikkinchi tilni o'rganish jarayoni, tarjima, sotsiolingvistika, ingliz tilining dunyo bo'ylab tarqalgan variantlari va matematik lingvistika kabilarni qamragan korpus lingvistik tadqiqotlarini amalga oshirish (Macquarie University, Sidney (Avstraliya)), milliardlab so'zlardan tarkib topgan korpuslar yordamida ayni bir tilning turli mintaqalarda, janrlarda va vaziyatlarda turlicha qo'llanilish sabablarini asoslash

---

[2] Sharh universitetlarning veb-saytlaridagi ma'lumotlar asosida taqdim etildi: https://www.lancaster.ac.uk/, https://www.birmingham.ac.uk/, https://www.mq.edu.au/, https://www.port.ac.uk/, https://wp.lancs.ac.uk/, https://www.sydney.edu.au/, https://buxdu.uz/

(University of Portsmouth (Birlashgan Qirollik), ingliz tili grammatikasi va lug'at qatlamini korpus lingvistikasi orqali o'qitish (Lancaster summer schools (Birlashgan Qirollik) ingliz tilini o'rganish va o'qitish jarayonida korpus tahlillaridan foydalanish (University of Sydney, Sidney (Avstraliya), inglizcha va o'zbekcha mediamatnlar diaxronik korpusini yaratish orqali tanlangan tillar rivojini kuzatish (Buxoro davlat universiteti (O'zbekiston), o'zbek tilining milliy, ta'limiy hamda elektron korpuslarini yaratish Alisher Navoiy nomidagi O'zbek tili va adabiyoti universiteti (O'zbekiston) kabi muammolar o'z yechimini topgan.

Dunyo tilshunosligida korpus tahlilariga asoslangan til rivoji taqdiqi va diaxronik korpuslar yaratish amaliyotini yaxlit tizimli o'rganish bo'yicha quyidagi ustuvor yo'nalishlarda tadqiqotlar olib borilmoqda: biologik til rivoji tadqiqida, xususan, yosh va jins kategoriyalari asnosida korpus tahlil tamoyillari, madaniy va mintaqaviy til rivojida hududiy, dialektik, sheva va lahjalarni o'z ichiga olgan til korpuslarini yaratish va boshqalar.

Muammoning o'rganilganlik darajasi. Jahon tilshunosligida G.Lich birinchilardan bo'lib, korpus lingvistikasining paydo bo'lishi, taraqqiyoti haqida ilmiy izlanish olib bordi[3]. Zamonaviy korpus lingvistikasi shakllanishiga doir tadqiqotlar S.Hyuston, D.Biber, S.Grays, M.Halliday va V.Teubert kabi ingliz olimlari tomonidan amalga oshirilgan[4]. Rus olimlari V. Baranov va V.Zaxarov ishlarida korpus lingvistikasi spetsifikasiga, N.Vladimov, S.Vadyaev, O.Gorkina, V.Mamontova va A.Magametova kabi olimlar "korpus lingvistikasi", "lisoniy korpus" tushunchasi va uni tasniflash muammolariga atroflicha yondashganlar[5]. Tojikistonlik olima D. Djafarova tadqiqotida korpus lingvistikasining kompyuter lingvistikasi va boshqa o'zaro o'xshash sohalar bilan qiyosiy tahlili amalga oshirilgan[6].

---

[3] Leech G. Corpora And Theories Of Linguistic Performance. In J.Svartvik (ed.). 1992. – 105–122 pp.

[4] Hunston S. Corpora in Applied Linguistics: Cambridge University Press, 2002. – 234 p.; Halliday M.A., TeubertW., Yallop C. and Cermakova A. Lexicology and Corpus Linguistics, – London, 2004. – 184 p.; Gries S.Th. Introduction to S. Gries and A. Stefanowitsch (eds.). Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis. – Berlin: Heidelberg, New York: Mouton de Gruyter, 2006. – 1-18 p.; Douglas B., UllaC. and Thomas A.U. Discourse On The Move. – Amsterdam, 2007. – 290 p.

[5] Баранов А.Н. Компьютерная лингвистика // Введение в прикладную лингвистику: учебное пособие. – М.: Едиториал УРСС, 2003. – 320 с.; Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие/ В.П. Захаров. – СПб, 2005. – 48 с. – Режим доступа:http://download.yandex.ru/class/zakharov/plan.pdf; Владимов Н.В. Корпусный подкод к решению переводческий проблем (на материале англиского и русского языках): Дисс....канд. филол. наук. – М.: РГБ, 2005. – С 26.; Вадяев С.Е.Лингвистические принципы построения и использование корпуса текстов для Исследования официально-делового стиля современного немецкого языка(на материале электронного корпуса «DER»): Дис ... канд. филол. наук. – Самара, 2005. – 160 с.; Горина О.Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке: Дисс. … канд. пед. наук. – М., 2014. – 321 с.; Мамонтова В.В. Особенности перевода сложносоставных слов с английского языка на русский (на материале корпуса публицистических текстов): Дисс. ... канд. филол.наук. – Ставрополь: 2008. – 139 с.; Магомедова А.Н.Корпусная лингвистика и контекстное разрешение лексической многозначности слов: Дисс. ... канд. филол. наук. – М.: РГБ, 2005. – 132 с.

[6] Джаъфарова Д.Ф. Модели лингвистического анализа текстов таджикского языка (на материале газелей Хафиза): Дисс. ... канд. филол. наук. – Душанбе, 2013. – 151 с.

Korpus tahlillaridan lug‘atshunoslikda foydalanish borasidagi ishlar A. Kilgarrif, K. Iztok, F. Karlson va I. Sojenin tadqiqotlarida kuzatilgan bo‘lsa[7], V. Rikov, A. Romanov va V. Plungyan kabi olimlar "korpus" tushunchasini matn tahlillarida qo‘llash metodini ishlab chiqishgan[8].

O‘zbekistonda ham bugungi kunda korpus lingvistikasi metodlarini qo‘llash, mavjud korpuslardan foydalanish borasida keng ko‘lamli ilmiy-tadqiqotlar olib borilmoqda. J. Jumaboeva ingliz tili korpuslaridan, xususan, BNC[9] tarkibidan semantik gradatsiyalarni topish, G. Ergasheva BNC korpusi asosida genderga oid konseptual metaforalarni tadqiq qilish, G. Sobirova esa BNC korpusini tillarni o‘qitishda qo‘llashga e’tibor qaratgan. Sh. Hamroeva mualliflik korpusini yaratish muammosiga yechim izlagan[10]. Shuningdek, o‘zbek tilining elektron, milliy, ta’limiy va parallel korpuslarini yaratish amaliyoti hamda ularning nazariy asoslarini ishlab chiqish borasida N.Z. Abduraxmonova, Sh.M. Hamroyeva, G.I. Toyirova, G.X. Begmatova, A.A. Raxmanova, O‘.M. Xoliyorov, D.Q. Elova va R.A. Karimov[11] kabi olimlarning salmoqli ilmiy-tadqiqotlarini misol qilib keltirish o‘rinlidir. Qolaversa, hozirda o‘zbek tilining lisoniy korpuslarini yaratish bo‘yicha amalga oshirilayotgan katta hajmli loyihalar o‘z natijasini bermoqda. Ularga misol qilib O‘zbek tilining elektron korpusi (https://uzbekcorpus.uz), O‘zbek tili milliy

[7] Kilgarriff A., Kosem I. Corpus Tools For Lexicographers. – Trojina: Institute for Applied Slovene Studies, 2012. – 37 p. УРЛ: https://www.researchgate.net/publication/264884809_Corpus_tools_for_lexicographers; Karlsson F. Lexicography And Corpus Linguistics. – Helsinki: Department of General Linguistics University of Helsinki, 1992. УРЛ: http://www.ling.helsinki.fi/~fkarlsso/INTJLLEX.pdf; Kilgarriff A. Using Corpora [And The Web] As Data Sources For Dictionaries. – Helsinki: 2012. – 21 p. https://www.sketchengine.eu/wp-content/uploads/Using_corpora_2012.pdf; Саженин И.И.Корпусные методы в лексикографии: опыт создания модели Словарного корпуса: Дисс. ... канд. филол. наук. – Новосибирск, 2013. 193 с.

[8] Рыков В.В. Корпус текстов новый тип словесного единства Текст. / В.В. Рыков. – Электрон, изд. – Режим доступа к изд.: http://rykov-cl.narod.ru/t31 .html.; Романов А.С. Методика и программный комплекс для идентификации автора неизвестного текста: Автореф. дисс....канд. тех. наук. - Томск, 2010. 26 с.; Плунгян В.А. Корпус как инструмент и как идеология: о некоторых урокахсовременной корпусной лингвистики// Русский язык в научном освещении. – М.: №2., 2008. – С. 7-20.

[9] BNC (BritishNationalCorpus) – Британия Миллий Корпуси: 100 миллион сўзли лисоний корпус.

[10] Сабирова Г. Умумтаълим мактаблари ўқувчиларига инглиз тили олмошларини ахборот технологиялари асосида ўргатиш методикаси: Филол. фалс. д-ри (PhD) ...дисс. автореф. – Тошкент: 2017. – 41 б. URL: http://uzswlu.uz/media/ba7be50d-84f1-c826-f935-6e7610cb6bf5.pdf; Эргашева Г. Турли тизимдаги тилларда гендерга оид терминологиянинг шаклланишида лингвистик ва экстралингвистик омиллар: Филол. фан. д-ри. .... дисс. автореф. – Тошкент: 2018. – 61 б. URL: http://demo.navoiy-uni.uz/sites/default/files/ziyonet.pdf; Джумабаева Ж. Ўзбек ва инглиз тилларида лексик ҳамда стилистик градуонимия: Филол. фан. д-ри. .... дисс. автореф. – Тошкент: 2016. – 82 б. URL: http://tashgiv.uz/wp-content/uploads/2017/01/ avtoref_jumabayeva.pdf; Хамроева Ш. Ўзбек тили мауллифлик корпусини тузишнинг лингвистик асослари: Филол.бўйича фалс. д-ри (PhD) ... дисс. автореф. – Қарши, 2018. – 52 б. URL: https://e-books.tsue.uz/pdf-ozbek-tili-mualliflik-korpusini-tuzishning-lingvistik-asoslari

[11] Abdurahmonova N.Z. O`zbek tili elektron korpusining kompyuter modellari: Filol. fan. d-ri. (DSc) diss. avtoreferati. – Toshkent, 2021. – 73 b.; Хамроева Ш.М. Ўзбек тили морфологик анализаторининг лингвистик таъминоти: Филол. фан. д-ри (DSc) дисс. автореферати – Фарғона, 2021. – 76 б.; Тоирова Г.И. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари: Филол. фан. д-ри (DSc) дисс. автореферати – Бухоро, 2021. – 72 б.; Бегматова Г.Х. Ўзбек тили миллий корпусида идиомалар базасини яратиш: Филол. фан. фалс. д-ри (PhD) дисс. – Термиз, 2021. – 143 б.; Рахманова А.А. Ўзбек тили миллий корпусини яратишдаги компьютер усуллари: Филол. фан. фалс. д-ри (PhD) дисс. – Тошкент, 2021. – 161 б.; Холиёров Ў.М. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари: Филол. фан. фалс. д-ри (PhD) дисс. – Термиз, 2021. – 181 б.; Каримов Р.А. Ўзбек-инглиз параллел корпусини тузишнинг лингвистик ва дастурий масалалари: Филол. фан. фалс. д-ри (PhD) дисс. – Бухоро, 2022. – 148 б.; Элова Д.Қ. Ўзбек тили корпуси бирликларининг услубий тегларини яратиш тамойиллари ва лингвистик таъминоти: Филол. фан. фалс. д-ри (PhD) дисс. автореф. – Тошкент, 2022. – 65 б.

korpusi (http://uzbekcorpora.uz), O'zbek tilining ta'limiy korpusi (https://uzschoolcorpara.uz) kabilarni salmoqli dasturiy mahsulotlar sifatida qayd etish mumkin.

Tadqiqotning dissertatsiya bajarilgan oliy ta'lim muassasasining ilmiy-tadqiqot ishlari rejalari bilan bog'liqligi. Dissertatsiya tadqiqoti Buxoro davlat universiteti ilmiy-tadqiqot ishlarining "Tilning leksik-semantik tizimi, qiyosiy-tipologik izlanishlar va adabiyotshunoslik muammolari" ilmiy mavzusining tarkibiy qismi bo'lib, Xorijiy tillar fakultetining ilmiy rejalariga kiritilgan.

Tadqiqotning maqsadi turli tizimli tillar mediamatnlari diaxronik korpusini yaratish orqali til rivoji tadiqiqining empirik tahlil tamoyillarini ochib berishdan iborat.

Tadqiqotning vazifalari:

diaxronik korpus yaratish va uni qo'llashga oid korpus lingvistikasi ilmiy-tadqiqot metodologiyasini takomillashtirish asnosida manba bilan ishlash bo'yicha diaxronik korpuslar bazasiga xos matnlar bilan ishlash talablarini ishlab chiqish;

mediamatnlarni tilning ijtimoiy ifodasi hamda til boyishining asosiy resursi sifatida o'ziga xos xususiyatlarini yoritish;

mediamatnlar diaxronik korpusini yaratishning nazariy va amaliy ahamiyatini anglagan holda tanlangan tillar doirasida inglizcha va o'zbekcha gazeta matnlaridan tarkib topgan diaxronik korpusni yaratish;

mavjud ingliz tili korpuslari, jumladan, xronologik tartiblangan hamda me'yorlashtirilgan lisoniy korpuslarning amaliy ahamiyatini ochib berish;

o'zbekcha va inglizcha gazeta materiallari misolida mediamatnlar diaxronik korpusining amaliy ahamiyati, jumladan, leksik qatlamning yillar kesimida o'zgarishiga statistik asos sifatida qo'llanishi mumkin bo'lgan baza ekanligini isbotlash;

mediamatnlar diaxronik koprusining tadqiq etilayotgan tillararo lingvomadaniy tushunchalar muqoyasasi uchun yetarli manba ekanligini tadqiq qilish;

mediamatnlar diaxronik koprusining tilning leksik qatlami rivojini kuzatishga doir yondashuvlarni nazariy asoslash.

Tadqiqotning obyekti sifatida "Oila va jamiyat" va "Anaheim Bulletin" gazetalarining 1991–2020-yillar oralig'ida chop etilgan nashrlari asosida ishlab chiqilgan mediamatnlar diaxronik korpusi tanlangan.

Tadqiqotning predmetini "Oila va jamiyat" va "Anaheim Bulletin" gazetalarining 1991–2020-yillar oralig'ida chop etilgan nashrlari asosida yaratilgan turli tizimli tillar mediamatnlari diaxronik koprusidan olingan statistik tahlillar va qidiruv natijalarida leksik, semantik, kollokatsion va lingvomadaniy birliklarning tadrijiy tahlili tashkil etadi.

Tadqiqotning usullari. Tadqiqot jarayonida kontent tahlil, tasniflash, tavsiflash, qiyoslash, statistik, lingvokognitiv, korpus va empirik tahlil metodlaridan foydalanilgan.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

ingliz, rus, xitoy, arab, ispan va koreys tillarining mavjud diaxronik korpuslar manbasiga qo'yiladigan xronologik uzviylik, transliteratsiya, davriy aniq miqdor, vaqt oralig'ining qat'iyligi kabi me'yorlari orqali lisoniy korpuslar bazasi bilan

ishlash tamoyillarining autentiklik hamda reprezentativlik kabi funksional jihatlari ochib berilgan;

o‘zbek tili lisoniy korpuslarining sinxronik korpus tamoyillariga mosligi hamda dunyoning lisoniy manzarasidagi lingvomadaniy evrilishda tillarning leksik qatlam yaruslaridagi tadrijiy rivojini dalillash orqali o‘rganilayotgan tillar diaxronik korpusini yaratishning xronologik tartib, davriy qamrov, lingvistik va ekstralingvistik omillari ilmiy asoslangan;

“Oila va jamiyat” va “Anaheim Bulletin” gazeta materiallarining xronologik ketma-ketlikdagi o‘ttiz yillik bazasini qamrab, o‘rganilayotgan tillar tarkibida umumiy hamda yillik statistik tahlil tamoyillarini avtomatik bajaruvchi o‘zbek va ingliz mediamatnlari diaxronik korpusi (https://mediatextcorpus.com/) veb-dasturi yaratish orqali turli tizimli tillarda kollokatsiyalarning xronologik birikuvchanligiga oid istorizmlar tadrijiy aniqlangan;

tillararo lingvomadaniy tushunchalarning konseptual freymini ishlab chiqish orqali tadqiq etilayotgan tillar bazasida tahlilga tortilayotgan lingvokuluremaga oid milliy, siyosiy va ijtimoiy yondashuvning o‘zgarib borish tendensiyalarini tadrijiy kuzatish imkoniyatining mavjudligi isbotlangan;

diaxronik korpuslar vositasida o’rganilayotgan tillar lug‘at qatlami rivojining asosiy hodisalari: neologizmlar va arxaizmlarni qayd etib borish funksiyasi imkoniyatlarining mavjudligi empirik dalillangan.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

korpus lingvistikasi metodlar tizimi zamirida diaxronik korpuslar yaratishga yo‘naltirilgan ba zaviy, ya’ni manba bilan ishlash me’yorlari takomillashtirilgan;

“Oila va jamiyat” va “Anaheim Bulletin” gazetalari bazasida 1991-2020-yillar oralig‘idagi uch dekadada nashr etilgan maqolalar tarkibidan miqdoriy me’yorlashtirilgan o‘zbek va ingliz gazetalari mediamatnlari diaxronik korpusi (https://mediatextcorpus.com/) yaratilgan;

diaxronik lisoniy korpuslar yaratish va ularni takomillashtirishning amaliy va nazariy tamoyil hamda ko‘nikmalari ishlab chiqilgan.

Tadqiqot natijalarining ishonchliligi Tadqiqot yuzasidan chiqarilgan xulosalar tadqiqotchi tomonidan e’lon qilingan ilmiy ishlarda o‘z ifodasini topganligi hamda bu xulosalar ilmiy jamoatchilik tomonidan ijobiy baholanganligi, diaxron mediamatnlar korpusining xususiyatiga oid qo‘yilgan muammo kontent tahlil, tasniflash, tavsiflash, qiyoslash, statistik, kognitiv, korpus va empirik tahlil metodlari asosida aniqlab berilgani, taklif hamda tavsiyalarning amaliyotga joriy etilgani, olingan natijalarning vakolatli tashkilotlar tomonidan tasdiqlanganligi bilan izohlanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqot natijalarining ilmiy ahamiyati izlanish xulosalaridan o‘zbek tilining diaxronik korpuslarini yaratish va uni mukammallashtirishning nazariy asoslarini ishlab chiqishda, korpus lingvistikasi yo‘nalishida tadqiqotlar olib borishda ilmiy-nazariy manba sifatida foydalanish mumkinligi bilan belgilanadi.

Tadqiqot natijalarining amaliy ahamiyati bugungi kunda ilmiy izlanish metodologiyasi sifatida qaralayotgan bo‘lsa-da, yaqin kelajakda fan sifatida taraqqiy etishi kutilayotgan korpus lingvistikasini o‘qitish jarayonida dastur, rejalar tuzish

hamda mavzularni bayon etishda nazariy manba vazifasini o‘tashi, sinxronik hamda diaxronik korpuslarni yaratish va mavjud korpuslarni qo‘llash bo‘yicha amaliy namuna bo‘la olishi bilan izohlanadi.

Tadqiqot natijalarining joriy qilinishi. Korpus lingvistikasi rivojida diaxronik korpuslar tadqiqini nazariy anglagan holda ingliz va o‘zbek gazeta matnlari korpusini yaratish orqali til rivoji tadiqiqining empirik tahlil tamoyillarini ochib berish bo‘yicha ishlab chiqilgan amaliy taklif va tadqiqot natijalari asosida:

ingliz, rus, xitoy, arab, ispan va koreys tillarining mavjud diaxronik korpuslar manbasiga qo‘yiladigan xronologik uzviylik, transliteratsiya, davriy aniq miqdor, vaqt oralig‘ining qat’iyligi kabi me’yorlari orqali lisoniy korpuslar bazasi bilan ishlash tamoyillarining autentiklik hamda reprezentativlik kabi funksional jihatlari ochib berilishiga oid tahlillardan Mahalla va oila ilmiy-tadqiqot institutining 2022-2023-yillarda bajarilgan JHBL-94 – sonli “Chekka qishloq hududlardagi ayollar daftari”ga kiritilgan ijtimoiy himoyaga muhtoj qizlarga kompyuter grafikasi dasturlarini o‘qitish orqali ularni ish bilan ta’minlash” nomli amaliy loyihasida foydalanilgan (“Mahalla va oila” ilmiy-tadqiqot institutining 2023-yil 4-sentabrdagi 01-09/1607-son ma’lumotnomasi). Natijada loyiha ishtirokchilarida diaxronik korpus yaratish, ularni amalda qo‘llashning tizimli metodlari borasida ilmiy-amaliy ko‘nikmalar hosil qilingan;

o‘zbek tili lisoniy korpuslarining sinxronik korpus tamoyillariga mosligi hamda dunyoning lisoniy manzarasidagi lingvomadaniy evrilishda tillarning leksik qatlam yaruslaridagi tadrijiy rivojini dalillash orqali o‘rganilayotgan tillar diaxronik korpusini yaratishning xronologik tartib, davriy qamrov, lingvistik va ekstralingvistik omillari ilmiy asoslanganliga doir nazariy qarashlar hamda tahlillardan O‘zbekiston Respublikasi Madaniyat vazirligi jurnallarining birlashgan tahririyati faoliyatida hamda O‘zbekiston milliy teleradiokomaniyasining “Foreign languages” telekanalida namoyish etilgan “Good Morning Uzbekistan” ko‘rsatuvida unumli foydalanilgan (O‘zbekiston Respublikasi Madaniyat Vazirligi “Guliston” va “Teatr” jurnallari birlashgan tahririyatining 2023-yil 20-oktabrdagi 43-son ma’lumotnomasi hamda O‘zbekiston milliy teleradiokomaniyasi O‘zbekiston teleradiokanali davlat muassasasining 2023-yil 27-noyabrdagi 01-23-1774-son ma’lumotnomasi). Natijada tahririyat a’zolari hamda ko‘rsatuv mas’ullari yordamida o‘zbek tilining diaxronik korpusini yaratishga bo‘lgan ehtiyojning mavjudligi ommalashtirilgan;

“Oila va jamiyat” va “Anaheim Bulletin” gazeta materiallarining xronologik ketma-ketlikdagi o‘ttiz yillik bazasini qamrab, o‘rganilayotgan tillar tarkibida umumiy hamda yillik statistik tahlil tamoyillarini avtomatik bajaruvchi o‘zbek va ingliz mediamatnlari diaxronik korpusi (https://mediatextcorpus.com/) veb-dasturi yaratish orqali turli tizimli tillarda kollokatsiyalarning xronologik birikuvchanligiga oid istorizmlar tadrijiy aniqlanganligiga oid ilmiy farazlar hamda veb-dastur O‘zbekiston Respublikasi Adliya vazirligiga 2023-yil 16-yanvarda taqdim etilgan DGU 2023 0264-raqamli talabnoma tegishli tekshiruvlardan o‘tkazilgach, O‘zbekistion Respublikasining “Elektron hisoblash mashinalari uchun yaratilgan dasturlar va ma’lumotlar bazalarining huquqiy himoyasi to‘g‘risida”gi Qununiga asosan rasmiy ro‘yxatdan o‘tkazilgan (O‘zbekiston Respublikasi Adliya

vazirligining 2023-yil 3-fevraldagi DGU 21899-son guvohnomasi). Natijada amalga oshirilgan tadqiqot natijalari va yaratilgan diaxronik korpusga nisbatan muallif huquqimizning kafolatiga yuridik asos qo'lga kiritilgan;

tillararo lingvomadaniy tushunchalarning konseptual freymini ishlab chiqish orqali tadqiq etilayotgan tillar bazasida tahlilga tortilayotgan lingvokulturemaga oid milliy, siyosiy va ijtimoiy yondashuvning o'zgarib borish tendensiyalarini tadrijiy kuzatish imkoniyatining mavjudligi isbotlanganiga oid ilmiy xulosa va takliflardan 2020-2021-yillarda O'zbekiston davlat xoreografiya akademiyasida amalga oshirilgan "O'zbek milliy raqs san'atini targ'ib etishga bag'ishlangan veb-sayt va multimediamahsulotlari to'plamini yaratish" mavzusidagi FZ-2022081663 raqamli fundamental ilmiy-amaliy loyihada foydalanilgan (O'zbekiston Respublikasi madaniyat vazirligi huzuridagi O'zbekiston davlat xoreografiya akademiyasining 2023-yil 8-iyundagi 01/04-61-son ma'lumotnomasi). Natijada o'zbekcha va inglizcha gazeta materiallari bazasida tillarning lug'at qatlamida milliy raqs san'ati terminlarining boyish darajasini empirik kuzatish imkoniyatini chuqur o'rganishda dissertatsiya ilmiy manba sifatida xizmat qilgan;

diaxronik korpuslar vositasida o'rganilayotgan tillar lug'at qatlami rivojining asosiy hodisalari: neologizmlar va arxaizmlarni qayd etib borish funksiyasi imkoniyatlarining mavjudligi empirik dalillanganligiga doir amaliy taklif va tavsiyalardan Germaniya akademik almashinuv xizmati DAAD tomonidan moliyalashtirilgan ID-57663348-sonli "European Perspectives in Dialogue: Central Asia and Western Research – Muloqotda Yevropa istiqbollari: Markaziy Osiyo va G'arb tadqiqotlari" nomli loyihada foydalanilgan (Buxoro davlat universitetining 2023-yil 25-sentabrdagi 06-01-0425-son ma'lumotnomasi). Natijada o'zbek va ingliz tillari leksikografiyasi rivojida muhim hissa qo'shishi jihatidan ahamiyatli neologizmlar va arxaizmlarning lug'at qatlamida sodir bo'lish xronologiyasi to'grisidagi nazariy bilimlar Germaniyaning Gottingen universiteti magistrlari hamda professor-o'qituvchilariga seminar shaklida tushuntirilgan.

Tadqiqot natijalarning aprobatsiyasi. Tadqiqot natijalari bo'yicha 5 ta xalqaro va 3 ta respublika ilmiy-amaliy anjumanlarda qilingan ma'ruzalarda jamoatchilik muhokamasidan o'tkazilgan.

Tadqiqot natijalarining e'lon qilinganligi. Dissertatsiya mavzusi bo'yicha jami 26 ta ilmiy ish chop etilgan, jumladan, 1 ta chastotali-izohli terminologik lug'at, 2 ta monografiya, O'zbekiston Respublikasi Oliy attestatsiya komissiyasining doktorlik dissertatsiyalari asosiy ilmiy natijalari chop etish uchun tavsiya qilingan ilmiy nashrlarda 15 ta maqola, ulardan 10 tasi respublika va 5 tasi xorijiy jurnalda nashr etilgan.

Dissertatsiyaning tuzilishi va hajmi. Dissertatsiya tarkibi kirish, to'rt bob, xulosa, foydalanilgan adabiyotlar ro'yxati va ilovadan iborat bo'lib, umumiy hajmi 236 sahifani tashkil etadi.

DISSERTATSIYANING ASOSIY MAZMUNI

Kirish qismida tadqiqot mavzusining dolzarbligi va zarurati asoslangan, tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor

yo‘nalishlariga mosligi ko‘rsatilgan, dissertatsiya mavzusi bo‘yicha xorijiy ilmiy-tadqiqotlar sharhi, muammoning o‘rganilganlik darajasi, dissertatsiya tadqiqotining dissertatsiya bajarilgan oliy ta’lim ilmiy-tadqiqot ishlari rejalari bilan bog‘liqligi, tadqiqotning maqsadi va vazifalari, obyekti va predmeti tavsiflangan, usullari, ilmiy yangiligi, amaliy natijalari bayon qilingan, olingan natijalarning ishonchliligi, ilmiy va amaliy ahamiyati ochib berilgan, tadqiqot natijalarining amaliyotga joriy qilinishi, aprobatsiyasi, nashr etilgan ishlar va dissertatsiya tuzilishi bo‘yicha ma’lumotlar keltirilgan.

Dissertatsiyaning “Til rivoji tadqiqida korpus lingvistikasi metodologiyasining o‘rni” nomli birinchi bobida korpus lingvistikasi, korpus tahlil metodlari yordamida til leksik qatlamining boyish mexanizmlari tadqiqi, mediamatnlar tilning ijtimoiy ifodasi hamda til boyishining asosiy resursi ekanligi hamda mavjud diaxronik korpuslarning funksional xususiyatlarini ochib berishga qaratilgan nazariy tahlillar amalga oshirilgan dunyo va mamlakat miqyosidagi tadqiqotlar izohlangan.

Korpus lingvistikasi lisoniy korpuslar yaratish amaliyoti orqali tilni tadqiq etuvchi tilshunoslikning alohida sohasi hisoblanadi. Soha rivoji til korpuslarining lingvistik muammolarga yechim topishda keng qo‘llanilishi bilan chambarchas bog‘liq. Ma’lumki, korpus lingvistikasi asosan lug‘atshunoslikka oid tadqiqotlarga yo‘naltirilgan metodlar jamlanmasiga ega bo‘lsa-da, ularni stilistika, sintaksis, til o‘qitish metodikasi hamda til rivoji tadqiqiga oid boshqa fan tarmoqlariga mos ravishda ishlatish mumkin. Uning bugungi kunga qadar taraqqiyot tavsifi yuqoridagi fikrni yaqqol dalillaydi.

Korpuslar onlayn tizimga o‘tishi ortidan qidiruvga berilgan leksik birliklarni topish tezligi oshdi va foydalanuvchilar soni tobora kengayib bormoqda. To‘g‘ridan to‘g‘ri kirish veb-brauzer orqali onlayn qidiruvda amalga oshiriladi. M. Davies to‘rtinchi avlod konkordanserlarini gibrid korpuslar deb ataydi, chunki ularning interfeysi morfemik, leksik, sintaktik va frazeologik birlik darajalarida korpus yaratish va so‘zlarning faollik darajalariga oid tahlilini o‘tkazish uchun o‘ziga xos umumiy maydon hisoblanadi.

Tadqiqotlar shuni ko‘rsatdiki, ko‘p tilli aholi tarkibidan tashkil topgan Hindiston uchun milliy til korpusini yaratish amaliyoti bir qancha murakkabliklarni keltirib chiqaradi, zero, ushbu davlat konstitutsiyasida 2 ta milliy hamda 22 ta mahalliy til davlat tili sifatida qabul qilingan. Bizningcha, ushbu bir davlat hududidagi tillar o‘zaro munosabat, muloqot va yagona mediamakon ta’sirida bir-birlarini to‘ldirib, o‘z o‘rnida bir-birlarining lug‘at boyliklarining qisqarishiga sabab bo‘lmoqda. Nazarimizda, ushbu millat uchun ko‘p tilli korpusning yaratilishi tillarning ayni paytdagi rang-barangligini saqlab qolardi.

Bu haqida hind tilshunos olimi N. Dash ham aynan yuqoridagi fikrlarni ma’qullagan holda shunday deydi: “Ko‘p tilli davlat sifatida Hindiston til gigantidir. U turli etnik va lingvistik jamoalarning ko‘p sonli mavjud “tirik” tillarini saqlab kelmoqda. Korpuslarning yetishmasligi tufayli bu tillarning o‘z texnologik taraqqiyotiga erisha olmayotganligidan aziyat chekmoqda. Korpuslarni yaratish savodxonlik darajasini oshirish, yo‘qolib ketish xavfi ostida turgan hind tillarini saqlab qolish, ingliz tilining haddan tashqari tajovuzidan o‘z

ahamiyatini yo‘qotgan tillarni himoya qilish uchun til ta’limini jadallashtirishi mumkin. Korpuslar ularga lingvistik imperializm muhitida muomalada qolishga yordam berishi mumkin bo‘lgan yagona lisoniy bazalardir. Bundan tashqari, u yo‘qolgan mavqeyini tiklash uchun statistik jihatdan ishonchli ma’lumotlarni taqdim eta olishi bilan xarakterlidir”[12]. Olimning bu boradagi xavotiri o‘rinli, aynan globalashuv davrida tadqiqotlar to‘g‘ri yo‘naltirilmasa, tillarning bugungi rivojida boshqa tildagi ommalashgan mediamatn elementlari oqimining ortishi oqibatida kuchsiz tillar yo‘qolib ketishi yoki xitoy tilining mandarin shevasidek ko‘plab chet tillarini singdirishga mahkum bo‘lib qoladi. Olib borilgan tadqiqotda ham aynan mana shu masalaga alohida e’tibor qaratildi. Zero, tadqiqot obyektlari bo‘lgan materiallar tahlili asosida xulosa qiladigan bo‘lsak, rivojlanishda, ishlab chiqarishda takomillashgan tillar borki, ularning ta’siri mediamakon prizmasida kengayib borishi inobatga olinganda, o‘zbek tilining ham diaxronik korpusini yaratish amaliyotini yo‘lga qo‘yish zamonning ayni talabi demakdir.

Fikrimizcha, har qanday jamiyat borki, boshqa bir insonlar guruhi bilan muloqotga ehtiyoj sezadi. Bu muloqot yaqinlashgani sayin quyidagi bosqichda va ketma-ketlikda til rivojiga ta’sir etib boradi:

1. Ijtimoiy qarashlar o‘zaro uyg‘unlashadi.
2. Diniy e’tiqod va dunyoviy idiologik yaqinlashish sodir bo‘ladi.
3. O‘zaro ilmiy, texnik va maishiy yutuqlar hamda ixtirolar baham ko‘riladi.
4. Kiyinish, urf-odat, marosimlar, bir so‘z bilan aytganda, madaniyatlar interferensiyasi yuzaga keladi.
5. Yuqoridagilar oqibatida leksik qatlam tillar interferensiyasi ta’sirida o‘zgarib, boyib boradi.

Umuman olganda, ta’kidlab o‘tganimiz yuqoridagi besh bosqichli uzviylikda boradigan tillar interferensiyasi bilan tugovchi jarayon bugungi kecha-kunduzda yanada jadallashib bormoqda. Chunki muloqotning hozirgi jarayoni faktorlari oldingi odam-odam va jamiyat-jamiyat kategoriyasidan tubdan farq qilib, axborot almashinuvining yangi, zamonaviy shakllari asosida qurilmoqda. Bizningcha, bu mediamaydondagi muloqotning tezlashishi madaniyatlar va tillar sofligining hamda rang-barangligining ko‘lamini haminqadar qisqartirib bormoqda. Shuning uchun, tadqiqot obyekti bo‘lgan til rivoji tadqiqi korpuslar prizmasidan statistik asoslab o‘rganilmas ekan, tillarning bugungi holatini saqlab qolish, kelajakdagi tadqiqotlarga asos sifatida qoldirish amaliyoti yuzaga chiqmaydi.

Shuningdek, so‘nggi yillarda respublika miqyosida himoya qilingan dissertatsiyalarda ham mediamatnning lisoniy xususiyatlari alohida tahlilga tortilmoqda, ular turkumiga quyidagilarni kiritish mumkin: S.A. Atkamovaning turli tizimli tillar gazeta matnlari asosida frazeologizmni ifodalash va baholashning uslubiy vositalari[13], Sh.A. Usmonovaning mediadiskursda gazeta tili

---

[12] Dash N.S. Language Corpora: Present Indian Need. – Indian Statistical Institute, Kolkata. 2004. – P 1. Available at https://www.researchgate.net/publication/2938590_Language_Corpora_Present_Indian_Need

[13] Атқамова С. А. Фразеологизм медиаматнда ифодалаш ва баҳолашнинг услубий воситаси сифатида (турли тизимли тиллар газета матнлари асосида): Филол.фан.фалс.д-ри (PhD) дисс. автореф. – Тошкент, 2022. – 58 б.

tendensiyalari[14], N.S. Nasrulloyevaning oʻzbek va ingliz tillarida kompyuter va internet jargonlarining chogʻishtirma tadqiqiga oid izlanishlari[15], M.R. Artikovaning mediamatnda presedent fenomenining lingvomadaniy aspekti tahliliga oid tadqiqotlari aynan mediadiskursiga oid lisoniy muammolar yechimiga qaratilgan[16]. Aksariyat olimlar ommaviy axborot vositalarini isteʼmolchi sifatida tadqiq etishadi: faqat turli ishlab chiqarish shakllari orqali, ommaviy axborot vositalarini sanoatning yakuniy natijasi – mahsulot reklamasi haqida xabar beruvchi tizim sifatida oʻrganishadi. Yaʼni gazeta, jurnal va komikslarni oʻqiymiz, filmlar va telekoʻrsatuvlar tomosha qilamiz, radio va musiqa tinglaymiz, shuningdek, internetdan foydalanamiz va kompyuter oʻyinlarini oʻynaymiz va, albatta, turli joylarda turli mediamahsulotlarga duch kelamiz – qoʻshiq radioda, filmga saundtrek sifatida yoki fonda taqdim etilishi mumkin.

Dissertatsiyaning **"Diaxronik korpuslar yaratish nazariyasi va bosqichlari"** deb nomlangan ikkinchi bobida ingliz va oʻzbek mediamatnlar diaxronik korpusini yaratishning nazariy va amaliy ahamiyati, diaxronik korpus yaratishda baza va meʼyor tamoyillari va ularni yaratish bosqichlari izchilligi tavsifi tadqiq qilingan.

Diaxron va sinxron tadqiqotlar va yondashuvlarga toʻxtalinganda, *sinxron* va *diaxron* lingvistik tahlilga bir-birini toʻldiruvchi ikki nuqtai nazar sifatida qarash oʻrinlidir. *Sinxron* yondashuv (qadimgi yunonchadan: syan – "birga" va chrós – "vaqt") tilni uning tarixini hisobga olmagan holda vaqtning bir lahzasida koʻrib chiqadi. Demak, sinxron tilshunoslik tadqiq etilayotgan tilni muayyan vaqtning oʻzida, yaʼni ayni hozirgi zamon doirasida tasvirlash, tahlil qilish va oʻrganishga qaratilgan. Bundan farqli ravishda, diaxron (dia – "orqali" va chrós – "vaqt" soʻzlaridan) yondashuv, tarixiy tilshunoslikda boʻlgani kabi, tilning tarix orqali rivojlanishi va evolyutsiyasini koʻrib chiqadi[17].

Oʻzbekcha elektron Qomus lugʻatida diaxron tadqiqotlar quyidagicha izohlanadi: *Diaxroniya – til tizimining lingvistik tadqiqot manbai sifatidagi tarixiy taraqqiyoti, til taraqqiyotini vaqt nuqtayi nazaridan oʻrganish.* Tilshunoslikda diaxroniya tushunchasi sinxroniyaga qarama-qarshi qoʻyiladi. Diaxronik tilshunoslik baʼzan qiyosiy-tarixiy tilshunoslik hamda tarixiy fonetika bilan aynan bir fan tarmoqlari deb qaralgan. Keyinchalik uning tadqiqot obyekti sifatida tildagi oʻzgarishlarni, bu oʻzgarishlarning sabablarini, paydo boʻlish davrini oʻrganish nazarda tutildi. XX asrning 2-yarmida diaxronik yondashuv, til tarixini davrlarga boʻlib oʻrganish bilan bogʻliq sof tarixiy yoʻnalishdan farqli oʻlaroq, til tizimida sodir boʻladigan diaxronik oʻzgarishlarni, tilning tizim sifatidagi taraqqiyoti qonuniyatlarini oʻrganishga qaratildi. Diaxroniya til faoliyati xususiyatlarini

[14] Усмонова Ш.А. Медиадискурсда газета тили тенденциялари (ўзбек ва инглиз тиллари мисолида): Филол.фан.фалс.д-ри (PhD) дисс. автореф. – Тошкент, 2022. – 56 б.

[15] Насруллоева Н. С. Ўзбек ва инглиз тилларида компьютер ва интернет жаргонларининг чоғиштирма тадқиқи: Филол.фан.фалс.д-ри (PhD) дисс. автореф. – Самарқанд, 2021 – 57 б.

[16] Артикова М.Р. Медиаматнда прецедент феноменининг лингвомаданий аспекти (ўзбек ва испан газеталари мисолида) : Филол.фан.фалс.д-ри (PhD) дисс. автореф. – Тошкент, 2022 – 54 б

[17] Giacalone Ramat, Anna; Mauri, Caterina; Molinelli, Piera, eds. (2013). Synchrony and Diachrony: A dynamic interface. Philadelphia, Pennsylvania: John Benjamins North America. pp. 17-18. ISBN 978-9027272072. [Retrieved on January 3, 2024]

zamonda, jumladan, sinxroniyada bilishga imkon beruvchi vosita boʻlib, umuman, til nazariyasining muhim qismidir[18].

Diaxronik tilshunoslik tushunchasi nazariy manbalarda tarixiy / qiyosiy tilshunoslik bilan bir xil tushuncha sifatida qaralishini kuzatish mumkin. Maʼlumki, qiyosiy tilshunoslik – tilshunoslikning qarindosh tillarni, yaʼni genetik jihatdan oʻzaro bogʻliq tillarni oʻrganuvchi, ular oʻrtasidagi munosabatlarni aniqlovchi hamda ularning zamon va makon boʻyicha tadrijiy taraqqiyotini tavsiflovchi sohasi; til oilalarining, shu tizimlardagi ayrim tillar va elementlarning kelib chiqishini aniqlash, jumladan, tillar oʻrtasidagi genetik qarindoshlikni – ularning yagona bir manba (bobo til)dan kelib chiqqanligini aniqlash (tillarning genealogik tasnifi) ham qiyosiy / tarixiy tilshunoslikning maqsadlari doirasiga kiradi[19].

Aynan tarixiy tilshunoslikning asosiy vazifalari turiga kiruvchi genetik tasniflashga doir ilmiy qarashlarga tayangan holda, qiyoslanayotgan tillarning ushbu tasnifning qaysi turiga mansubligini koʻrib chiqish oʻrinli. Ingliz tili *hind-ovrupo tillari oilasi*ga kiruvchi *german tillar guruhi*ga kiradi. Oʻzbek tili esa *turkiy tillar oilasi*ga, ikkinchi qarashga koʻra, ushbu oila *oltoy tillar katta oilasiga* mansubligi aytiladi. Yana ham aniqroq aytganda, oʻzbek tili turkiy tillar oilasiga kiruvchi *qarluq guruhiga* kiritiladi va bu borada oʻzbek tiliga eng yaqin til *yangi uygʻur tillari* hisoblanadi[20].

Shunday ekan, *diaxron* terminining korpus lingvistikasiga ham kiritilishi uning mazmuniy va vazifaviy xususiyatlariga soya solmaydi. Yaʼni diaxronik korpus turli davrlardagi matnlarni oʻz ichiga olgan korpus boʻlib, tilning rivojlanishi yoki oʻzgarishini tadqiq qilish uchun ishlatiladi. Bu borada M. Hilpert va S.Th. Gries kabi olimlarning talqinicha, diaxronik korpus deganda, hozirgi ogʻzaki ingliz tilining diaxronik korpusida boʻlgani kabi, ketma-ket vaqt oraligʻida tildan foydalanishning oʻxshash turlarini ifodalovchi matn resurslari tushuniladi, ushbu xususiyat mohiyatiga koʻra korpus kamida ikki vaqt oraligʻini oʻz ichiga qamrab olishi talab etiladi[21].

Demak, diaxronik korpus tushunchasi korpus lingvistikasi fan tarmogʻida muhim ahamiyatga ega boʻlib, korpus tahlil metodlaridan foydalanishning yangi bosqichini yuzaga chiqargan. Bizningcha, diaxronik korpuslardan tilshunoslikning quyidagi sohalarida qoʻllash orqali quyidagi ilmiy-nazariy savollarga maqbul yechim izlash maqsadida yangicha yondashish joiz:

birinchidan, diaxronik korpuslar lugʻatshunoslik sohasida qoʻllanilib, unda izohtalab soʻzlarning nafaqat qoʻllanishiga oid xronologik maʼlumot, qolaversa, korpus tarkibidagi manbaning tizimli tanlanganligiga qarab, uning etimologiyasiga ham nazar tashlay oladi;

---

[18] https://qomus.info/oz/encyclopedia/d/diaxroniya/ [murojaat sanasi 31.01.2024]

[19] Tomsen V., Istoriya yazmkoznaniya do konsa XIX veka, M., 1938; Meye A., Vvedeniye v sravnitelnoye izucheniye indoyevropeyskix yaznkov, 3 izd., per. s frann., M. — L., 1938 available at https://openlibrary.org/books/OL16367943M/Istoriya_yazykovedeniya_do_kontsa_XIX_veka [retrieved on December 5, 2023]

[20] Irisqulov A.T. Tilshunoslikka kirish. / Darslik. – Toshkent: "Yoshlar matbuoti" bosmaxonasi. 2009. – B. 185-187

[21] Hilpert M. and Gries S.Th. Quantitative approachesto diachronic corpuslinguistics / The Cambridge Handbook of English Historical Linguistics // - Publisher: Cambridge University Press, 2016. – P. 36 Available from https://www.researchgate.net/publication/325893284_Quantitative_approaches_to_diachronic_corpus_linguistics

ikkinchidan, diaxronik korpuslar bir vaqtning o'zida tarixiy korpus tushunchasi bilan bir xil maqomda qo'llaniladi hamda til rivoji tadqiqini o'rganishga alohida e'tibor qaratadi. Jumladan, leksikologiyaning eskirmas muammolari tarkibiga kiruvchi neologizmlar, arxaizmlar, istorizmlar kabi ilmiy muammolarga aniq statistik yechimlarni taklif etadi;

uchinchidan, diaxronik korpuslar gender tilshunosligi sohasida qo'llanilishi jahon tilshunosligida mavjud amaliyotlar sirasiga kiradi. Bu o'rinda, tarixiy korpuslar qarama-qarshi jins vakillarining ijod namunalari ayri holda jamlangan, davrlashtirilgan korpus bazalarini shakllantirish orqali ayol yoki erkak yozuvchilarning hayot falsafasi, dunyoqarashi, his-tuyg'ularni aks ettirish uslublarini ular yozgan badiiy asarlar, yozishmalar, gazeta yoki jurnal maqolalari hamda intervyu / suhbatlar namunalari asosida lingvostatistik tahlil qilib aniqlash imkoni mavjud;

to'rtinchidan, diaxronik korpuslar yaratish amaliyoti manbashunoslar, tarjimonlar va tarixiy asarlar yozishni maqsad qilgan shoir hamda yozuvchilar uchun muhim ahamiyat kasb etadi. Bu orqali, yuqorida sanab o'tilgan til ixlosmandlari diaxronik korpus qamrab olgan davrda qo'llanilgan faol leksik qatlamdan foydalangan holda tadqiq qilinayotgan vaqt oralig'idagi madaniyat, millat va tilning haqiqiy ruhiyatini his qiladi hamda uni qo'llash orqali shu ruhni saqlab qoladi;

beshinchidan, bizningcha eng muhimi, bugungi kunda rivojlanib borayotgan tezkor zamonda, iqtisodiy rivojlangan xalqlar va ularning tillari yashab qoladi. Bundan keskin fikrni bildira turib, o'ylaymizki, diaxronik korpuslar yaratish orqali tillararo interferensiya chiziqlarini statistik tahlil qilish orqali, til sohiblarini ushbu oqimdan saqlashga qaratilgan davlat dasturlarini ishlab chiqish muhim ahamiyat kasb etadi.

Bu kabi diaxronik korpuslarni qo'llashga doir ilmiy farazlarni ko'plab sanab o'tish mumkin. Fikrimizcha, agar diaxronik korpuslardan oqilona foydalanilsa, ularning tarkibi belgilangan me'yorlarga ko'ra to'g'ri shakllantirilsa, bu turdagi korpus tahlillari asosida dunyo tillarini o'zaro tarixiy aloqadorlik, siyosiy va hududiy yaqinlik, iqtisodiy qaramlik va hokazo madaniy hamda ijtimoiy jihatdan qayta tasniflash imkoniyati yuzaga chiqadi.

Bu o'rinda, diaxronik korpuslarni yaratish amaliyotida ingliz tilining ancha ildamlab borayotganini inobatga olganda, ushbu tilda yaratilgan shu turdagi korpuslarni sanab o'tish, ularning hajmi va xususiyatlarini inobatga olganda, o'zbek tilining tarixiy/ diaxronik korpusini yaratishga bo'lgan ehtiyoj aniq planda ko'zga tashlanadi. Bugungi kunda martinweisser.org saytiga ko'ra, diaxronik korpuslar hajmi soni, korpus yaratilgan davriga mos ravishda me'yorlashtirilgan holda yaratilishi ahamiyatlidir. Saytda sanab o'tilgan inglizcha diaxronik korpuslarning ba'zilari quyida izohlandi[22]. "1-jadvalga qarang".

---

[22] *Historical & Diachronic Corpora* Available from
http://www.martinweisser.org/corpora_site/historical_corpora.html [Retrieved on 04.02.2024]

Inglizcha *tarixiy / diaxronik* korpuslar

ARCHER Corpus (A Representative Corpus of Historical English Registers)	1650–1990-yillar oralig'idagi tarixiy davrlardan olingan ilk zamonaviy ingliz tiliga mansub yozma va "nutqqa asoslangan" janrlardan shakllangan jami 955 ta matn, ya'ni 1 789 309 ta so'zni o'z ichiga oladi.
Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English	Ingliz matnlarining Xelsinki korpusining qadimiy ingliz bo'limi 5 000 dan 10 000 so'zgacha bo'lgan qadimiy ingliz tili tarkibidan olingan manbalar asosida 106 210 so'zni o'z ichiga oladi.
Century of Prose Corpus	1680–1780-yillar oralig'ida yashab ijod qilgan jami 120 muallif ijodi asosida nashr etilgan adabiy va adabiy bo'lmagan ingliz tilidagi yarim million so'zdan tarkib topgan.
Complete Corpus of Old English	Toronto universitetida tuzilgan barcha qadimiy ingliz tili davrida yaratilgan 3022 ta matnlarni ifodalovchi korpus.
Corpus of English Dialogues, 1560-1760 (CED)	Erta zamonaviy ingliz tilidagi nutq shakllarini ifoda etuvchi 177 ta matndan tashkil topgan korpus 1 million 200 ming so'zni o'z ichiga oladi.
Corpus of Newsbooks	1653-yil dekabrdan 1654-yil maygacha *Tomason* nashriyotida chop etilgan barcha gazetalardan olingan taxminan 800 000 so'zdan tarkib topgan matn.
Corpus of Early English Correspondence (CEEC) & the Parsed Corpus of Early English Correspondence (PCEEC),	Korpus umumiy hisobda jami 2 million 700 ming so'zdan tashkil topgan bo'lib, u o'z ichiga 1410-yildan 1681-yilgacha 450 ming so'z; 1402-1663 yillardan "Erta yozishmalar korpusi" 50 ming atrofida so'z; 1681–1800-yillarni qamrab olgan oraliqdan jami 2 million 200 ming so'z kiritilgan.
Corpus of Early English Medical Writing & Corpus of Middle English Medical Texts (MEMT)	1375–1800- yillardagi tibbiy nashrlar va qo'lyozmalar korpusi bo'lib, 10-12 ming so'zli risolalar bilan to'ldirilgan hamda o'rta asrlar bo'limi 500 minga yaqin so'zlarni o'z ichiga oladi.
The Málaga Corpus of Early English Scientific Prose	1350–1900-yillarda ishlab chiqilgan tibbiy materiallarni o'z ichiga olgan tarixiy diaxronik

	korpus. Korpus deyarli olti yuz yillik davrni o'z ichiga olgan uchta subkorporaga bo'lingan taxminan olti million so'zni o'z ichiga oladi.
Newdigate Newsletters	1674–1692-yillarga tegishli qo'lyozma nashrlardan jami 750 ming so'zdan iborat.
TIME Magazine Corpus	TIME jurnalining 1923–2006-yillar oralig'idagi nashrlari negizida shakllantirilgan 100 million so'zdan iborat korpus. Vaqt o'tishi bilan so'z va iboralarning qo'llanishi yoki kamayishi, ma'nosi o'zgarganligini ko'rish imkonini beruvchi ma'lumotlar bilan ta'minlaydi.
Zürich Corpus of English Newspapers (ZEN)	1660-yillardan XX asr boshlarigacha bo'lgan vaqt oralig'idagi London gazeta nashrlaridan tashkil topgan.

Yuqorida sanab o'tilgan lisoniy korpuslarni ingliz korpus lingvistikasining yutuqlari sifatida e'tirof etgan holda, o'zbek tili doirasida ham bu borada bir qator ilmiy izlanishlar amalga oshirilayotganini ta'kidlash joiz. Ular sirasiga quyidagilarni kiritish mumkin:

1. O'zbek tilining ilk til korpusi sifatida qaralayotgan *uzbekcorpus.uz* elektron korpusi filologiya fanlari doktori (DSc), professor N.Z.Abdurahmonova boshchiligidagi loyiha a'zolari tomonidan bir necha yillik olib borilgan tadqiqotlar, ilmiy hamkorlik va loyihalardan olingan natijalar mahsuli hisoblanadi. O'zbek tili korpusi o'zbek tilida yaratilgan lingvistik lug'atlar, web sahifalar, o'zbek tilining morfologik ma'lumotlar bazasi, o'quv adabiyotlar hamda turli janrdagi ilmiy, rasmiy va badiiy matnlar majmuasidan tashkil topgan. Korpus – og'zaki va yozma matnlar yig'indisi sifatida, ma'lumotlarni boshqarish va foydalanish imkoniyatiga ega dasturiy ta'minot hisoblanadi. Ushbu korpusning tarkibida ichki – subkorpuslar ham mavjud: parallel korpus, mualliflik korpusi, ta'limiy korpus[23].

2. O'zbek tilining ta'limiy korpusi https://uzschoolcorpara.uz/ havolasida mavjud. Ta'limiy korpus ona tili hamda xorijiy tilni o'rganishda birdek ahamiyatli. Tilni o'rgatishda lug'at boyligining kattaligini ko'rsata olish, so'zning qo'llanish imkoniyatini u yoki bu grammatik konstruksiya orqali tushuntirish uchun misollar massivini ko'rsatishda korpus juda qo'l keladi. Til ta'limi uchun muhim bo'lgan misolning doimiy yangilanib borishi, buni aks ettirib turish xususiyati hamda imkoniyati faqat korpusda mavjud. O'qituvchi yangi, ishonarli, cheksiz hamda xilma-xil misollarni shu yerdan topa oladi, topshiriq, mashqlarni belgilashda qiynalmaydi, bir necha daqiqada mavzu bo'yicha yangi-yangi misollardan iborat topshiriqlarni tayyorlay oladi[24]. Tizim mualliflari qatorida t.f.f.d.(PhD), dots.

---

[23] https://uzbekcorpus.uz/aboutCorpus [murajaat sanasi 02.02.2024]
[24] https://uzschoolcorpara.uz/#home [murajaat sanasi 02.02.2024]

B.B. Elov, f.f.d., prof. B.R.Mengliyev, f.f.d., dots. Sh.M. Hamroyeva va boshqalar kiradi.

3. *Uzbek community corpus based on material from 2017* (2017-yildan to'plangan manbalar asosida yaratilgan o'zbek jamiyati korpusi) Leipzig universiteti, Leipzigdagi Saksoniya fanlar va gumanitar fanlar akademiyasi va Amaliy informatika instituti loyihasi asosida tuzilgan bo'lib, o'z ichiga jami 663 ming 119 ta gapdan iborat 9 million 256 ming 1 ta so'zni qamrab oladi[25]. Ushbu korpus avtomatik internet materiallari tarkibidan jamlanganligi tufayli o'zbek tilining bugungi kundagi muammolaridan biri bo'lgan alfavit, ya'ni, internet sahifalarida yuklanadigan matnlar ikki xil lotin va kirill grafikasida kiritilishi oqibatida qidiruv jarayonida so'zlar ikki turli grafikada alohida izlanishi murakkablik tug'diradi.

4. O'zbek tili milliy korpusi dasturiy mahsulot sifatida O'zbekiston Respublikasi innovatsiya vazirligining 2021 yilgi 30-turdagi "O'zbek tilining milliy korpusini loyihalash va dasturiy majmua ishlab chiqish" (2020-2022) mavzusidagi amaliy loyiha doirasida bajarilgan. O'rnatilgan tartibda o'tkazilgan tanlov natijalari bo'yicha ushbu loyihani bajarish uchun O'zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi O'zbek tilini rivojlantirish jamg'armasi bilan shartnoma tuzilgan[26]. Ushbu korpus http://uzbekcorpora.uz/ havolasida mavjud.

Yuqorida sanab o'tilgan o'zbek tili korpuslarining hech biri diaxron korpuslar sirasiga kirmaydi. Aksincha, korpuslar xronologik tartiblanmaganligi, tarixiy ketma-ketlikda matnlarning kiritilmaganligi bois ularni sinxron korpuslar rukniga daxldor deyish mumkin.

O'zbek tili diaxronik korpusining yaratilishiga bo'lgan ehtiyoj ayni mana shu hozirgi kunda mavjud o'zbek tili korpuslarining yuqoridagi tavsifiga uzviy bog'liq. Ya'ni bugungi kunda o'zbek tilining boyib borishini kuzatish uchun mo'ljallangan *monitor* korpuslarning mavjud emasligi bu boradagi ilmiy-nazariy xulosalarning ishonchliligini shubha ostida qoldiradi. Zero, tilning diaxronik korpusi tilning davriy o'zgarishlariga mos ravishda manbalarni jamlagan holda lug'at boyligining rivoji xususida aniq statistik ma'lumotni taqdim etadi.

Shu o'rinda mavjud tajribani qo'llagan holda, o'zbek tilining ham diaxronik korpusini yaratishga nazariy asos sifatida mazkur tadqiqot ishi amalga oshirilmoqda. Bunda, ikki tilli, ya'ni o'zbek va ingliz tillarining diaxronik korpusini ishlab chiqish ishning maqsadi etib belgilangan. Yozma va og'zaki matnlarni keng ko'lamda topish mumkin bo'lgan manba – mediamuloqot vositalari hisoblanar ekan, bizningcha, yaratilajak korpusimizning o'zagini aynan mediamatnlar asosida shakllantirish joiz. Demak, nazarda tutilgan korpusning jamlanishi olimlar va til o'rganuvchilarga quyidagi imkoniyatlarni yaratishi mumkin:

---

[25] *Leipzig Corpora Collection:* Uzbek community corpus based on material from 2017. Leipzig Corpora Collection. Dataset. https://corpora.uni-leipzig.de?corpusId=uzb_community_2017 [retrieved on 02.02.2024]

[26] http://uzbekcorpora.uz/buyurtmachi [murajaat sanasi 02.02.2024]
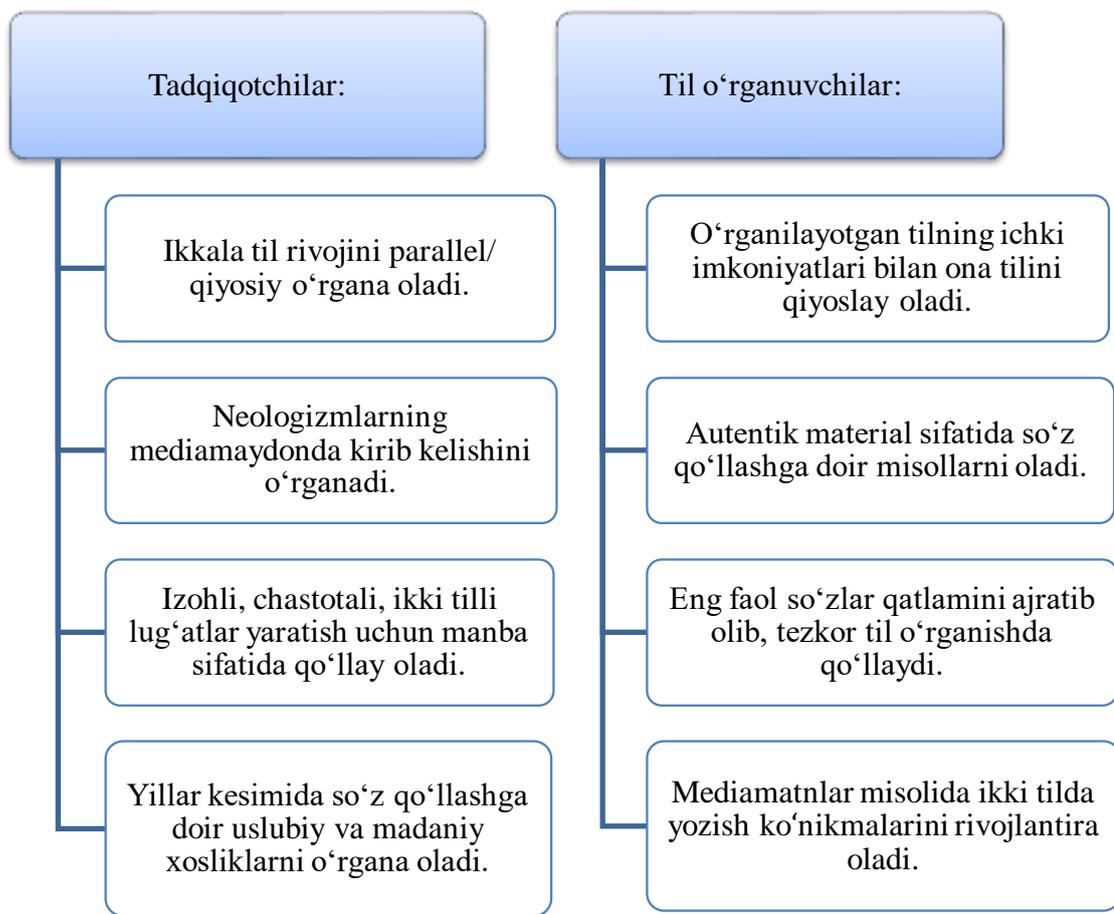
1-rasm. Ingliz va o'zbek mediamatnlar diaxronik korpusining ilmiy amaliy sohadagi dolzarbligi

Diaxronik / tarixiy korpuslar soni va hajmi bo'yicha o'sishda davom etmoqda, bugungi kunda ularni tahlil qilish usullari ham samaraliroq bo'lib, yanada takomillashmoqda. Shu bilan birga, murakkab diaxronik korpuslarni hisobga olmaganda, diaxronik ma'lumotlar tadqiqotchiga doimo muhim muammolarni yuzaga chiqarib beradi.

Diaxronik korpuslar foydalanuvchilar uchun kerakli manba sifatida e'tirof etilishi uchun uni, albatta, to'g'ri shakllantirish zarur. Fikrimizcha, yuqorida keltirib o'tilgan manbaga qo'yilgan talablarni diaxron xarakterdagi korpus bazalarini jamlashda va saralashda joriy qilish mumkin. Biroq diaxronik korpuslar uchun manba yig'ishning hamda uni korpus tarkibiga kiritishning ba'zi farqli tomonlarini inobatga olish o'rinli. Ular quyidagicha tasniflandi:

1. Diaxronik korpuslarda boshqa turdagi korpuslardan farqli ravishda xronologik uzviylikda, tarixiylikdan zamonaviylikka qarab yo'naltirilgan tartibda matnlarni yig'ish maqsadga muvofiq;

2. Dunyo tillari orasida ko'plab tillar tarixiy rivojida siyosiy noturg'unlik oqibatida yoki biror tarixiy voqelik ortidan yozuv grafikasida alfavit o'zgarishlarini boshidan o'tkazgan. Bunday tillarning qadimiy yoki tarixiy bitiklarini korpusga jamlaganda, matnlarni to'g'ri va aniq transliteratsiya qilish zarur. O'zbek tili ham yaqin tarixida bir emas, uch marta o'z yozuv grafikasini o'zgartirilganligi tufayli

bugun yig'iladigan diaxronik korpuslarga manba kiritishda ayni shu holat kuzatiladi hamda yechim yuqorida ta'kidlangandek bo'lsa, maqsadga muvofiq hisoblanadi;

3. Diaxronik korpuslar tarixiylik nuqtai nazaridan barcha davrlarda ham birdek manbaning mavjud emasligi sababli teng hajmda matnlar kiritish borasidagi me'yorga har doim ham amal qila olmaydi. Biroq yaratilayotgan korpus yaqin tarix, ya'ni 20-30 yil oldingi davrni qamrab olishi nazarda tutilsa, unda manbaning cheklanmaganligi tufayli miqdoriy chegaraga qo'yiladigan talab, ya'ni teng hajmda kiritish majburiyatiga amal qilish shart;

4. Har qanday diaxronik korpus yaratilishidan oldin tadqiqotchi tomonidan belgilab olingan davriylik oz muddatni emas, aksincha, kamida bir dekadali vaqt oralig'ini qamrab olishi kerak. Aksi holda, til rivoji tadqiqiga qaratilgan izlanishlar samarasi yuqori darajada bo'lmasligi ehtimoli yuzaga keladi.

"Gazeta materiallari asosida mediamatnlar diaxronik korpusini yaratish amaliyoti" nomli uchinchi bob diaxronik korpusni yaratishda matniy manbalar bilan ishlash mexanizmlari, asoslari va saralash xronologiyasi, korpus bazasini avtomatlashtirish va dizayn muammolari hamda yaratilgan korpusni baholash va chastotali qidiruv-tahlil natijalari tavsifini ochib berishga qaratilgan.

O'zbek tilshunosligi, xususan, leksikografiyada ham korpus lingvistikasi metodlaridan foydalanilganligini kuzatish mumkin, biroq ular aynan korpus tahlillariga asoslanganligi qayd etilmaydi. Masalan, *"O'zbek xalq dostonlarining chastotali izohli frazeologik lug'ati"*ni yaratish borasida qilingan ishlar sirasiga matniy birliklarning yig'ilganligi, ularni kompyuter xotirasiga kiritib, alfavit-chastotali lug'atlar olinganligi aytiladi: "… lug'atni tuzish jarayonida dastlab "Alpomish", "Ravshan", "Rustam" dostonlarining matnlari kompyuter xotirasiga kiritildi. Kompyuter xotirasida to'plangan ana shu matnlar asosida kompyuter yordamida 28499 so'z shaklidan iborat, uchta chastotali va uchta alfavit-chastotali lug'atlar olindi: "Alpomish" dostonining umumiy matn hajmi – 14029, ya'ni – 49,2 % ni; "Ravshan" dostoni – 7059, ya'ni 24,7 %ni, "Rustam" dostoni 7411 so'z shakli 26,2% ini tashkil qiladi. Matnlarni qayta tekshirish jarayonida mazkur matnlarda jami 1041 FB qo'llanganligi ma'lum bo'ldi, ular xalq dostonlari matnida 1641 marta qaytarilgani aniqlandi"[27]. Ko'rinib turganidek, biror o'rinda qaysi elektron yoki kompyuter dasturi qo'llanilganligi eslatib o'tilmaydi. Vaholanki, bu jarayon korpus lingvistikasining asosiy maqsadi bo'lgan *korpuslar yaratish* orqali lisoniy bazaga oid statistik axborotni tez fursatda olish imkonini yaratishga xos birlamchi vazifalar turkumiga kiradi.

Bizningcha, yuqoridagi holatni inobatga olganda, o'zbek lingvistikasi korpusga asoslangan nazariy tadqiqotlarning amaliy jabhalarda qo'llashga oid yangi yondashuvlarni talab qilmoqda. Demak, yaratilishi kerak bo'lgan korpuslar qatorida turuvchi *O'zbek va ingliz gazetalari mediamatnlari diaxronik korpusini yaratish* aynan shu turdagi natijaviy ilmiy izlanishlar sirasiga kiradi.

---

[27] O'zbek xalq dostonlarining chastotali izohli frazeologik lug'ati /B.Yo'ldoshev, D.O'rinboyeva; Mas'ul muharrir B.O'rinboyev; O'zbekiston Respublikasi oliy va o'rta maxsus ta'lim vazirligi, Alisher Navoiy nomidagi Samarqand Davlat universiteti. – Samarqand, 2013. – B. 4 https://elib.buxdupi.uz/books/34891.pdf

J. Sinclair[28] va F. Douglas[29] kabi olimlarning fikricha, korpusga kiritiladigan matnlarning ba'zilari uzun bo'lgani tufayli, ularning to'liq hajmini jalb etmasdan, balki matn bo'laklaridan foydalanish ba'zan matnning faqatgina bir turi ustun bo'lmasligi uchun korpusda qo'llaniladi. Biroq to'liq matnlardan foydalanishning ustunligiga doir ko'plab yondashuvlar mavjud, deya xulosa qilishadi. Bizning nazarimizda, olimlarning fikrlari o'rinli va matnlar hajman teng bo'lishi amaliyotda kamdan-kam uchragani bois, korpusga jamlanadigan matnning to'liq shaklini kiritish uning mazmunan va mantiqan tog'ri tahlil qilinishiga asos bo'ladi.

Bugungi kunda yaratilayotgan ko'plab korpuslar, xususan, o'zbek tilining zamonaviy korpuslari asosan onlayn, ya'ni internet materiallaridan shakllantirilayotganligini kuzatish mumkin. Bizningcha, korpus yaratish maqsad qilinganda, undan kutiladigan natijani inobatga olgan holda, imkon doirasida manbaning ishonchliligi, autentik ekanligi hamda eng muhimi uning ertangi kunga saqlanishiga e'tibor qaratilishi joiz.

Yuqoridagi fikrlardan kelib chiqib, mediamatnlar diaxronik korpusini yaratishda material sifatida "Oila va jamiyat" va "Anaheim Bulletin" gazetalarining 1991-2020-yillar oralig'ida chop etilgan nashr sonlari tanlandi. Bu orqali ishonchlilik, reprezentativlik va haqqoniylik tamoyillariga asos bo'la oladigan materiallar sifatida e'tirof etilishi mumkin bo'lgan baza shakllantirildi. Bunda tanlab olingan o'zbek tilidagi "Oila va jamiyat" va ingliz tilidagi "Anaheim Bulletin" gazetalarining 30 yillik nashrlaridan olingan matnlar manba sifatida tanlanganligiga asos sifatida quyidagilarni sanab o'tish mumkin:

Birinchidan, yuqorida aytilgan 30 yillik muddat, ya'ni 1991-2020-yillar oralig'i o'zbek tili uchun 1989-yil 21-oktabrdagi "O'zbekiston Respublikasining davlat tili to'g'risida"gi qonuniga muvofiq unga davlat tili maqomi berilganidan keyingi taraqqiyot, ya'ni tillarning samarali rivojiga mustaqillikning uchta o'n yilligi dolzarblik kasb etadi. Ingliz tilining ham aynan shu davrlarda chop etilgan gazeta nashrlaridan yig'ilgan matnlar bilan ishlash maqsad qilindi.

Ikkinchidan, "Oila va jamiyat" va "Anaheim Bulletin" gazetalari nisbatan kam o'rganilgan, haftalik chiqadigan, nashrlar chop ettirish boshlanish davri bir-biriga yaqin bo'lgan hamda eng muhimi ijtimoiy gazetalar sirasiga kirishi inobatga olinib, tanlangan. Yanada aniqroq aytganda, "Anaheim Bulletin" gazetasi 1966-yildan mustaqil gazeta sifatida Amerika Qo'shma Shtatlari Kaliforniya shtatida joylashgan eng ko'p aholi istiqomat qiluvchi Anaheim shahrida chiqarib kelinmoqda. Gazeta o'zining tarixida 1992-yilning 6-avgust sanasidan boshlab haftada bir marta chop etiladigan nashrlar sirasiga o'tgan, bu esa 1991-yil 5-sentabr sanasidan "Oila va jamiyat" gazetasining har haftaning payshanba kuni chop etilishi yo'lga qo'yilgan. Davriy yaqinlik mavjudligi hamda ular ko'taradigan ilm-fan, madaniyat, san'at, ishlab chiqarish, tadbirkorlik, qishloq xo'jaligi va yana bir nechta sohalarga bag'ishlangan ruknlari o'zaro qiyoslashga mos gazetalar ekanligini ko'rsatdi.

---

[28] Sinclair J. Corpus and text – basic principles. In M. Wynne (Ed.) // Developing linguistic corpora: A guide to good practice. Oxford: Oxbow Books. 2005. – pp. 1-16

[29] Douglas F. M. The Scottish corpus of texts and speech: Problems of corpus design // Literary and Linguistic Computing, 18(1). 2003. – pp. 23-37.

Uchinchidan, "Oila va jamiyat" va "Anaheim Bulletin" gazetalarining tanlangan davr oralig'idagi nashrlari arxivining mavjudligi, garchi har ikkala gazetalarning eski nashrlarida kompyuter o'qiy oladigan imkoniyatning mavjud emasligi kabi muammoning uchraginini inobatga olmaganda, aynan shu manbalar tanlanganiga asos bo'ladi.

Har ikkala baza o'zida gazetalarning skanerlangan holatini saqlovchi https://newspaperarchive.com/ hamda http://press.natlib.uz/ru/publications rasmiy saytlardan olindi. Ushbu saytlarda barcha ma'lumotlar foydalanuvchilar uchun ochiq ekanligi va cheklovlarning mavjud emasligi tadqiqotda tug'ilishi mumkin bo'lgan mualliflik huquqi (copyright) tamoyillariga zid harakat bajarilmagini ko'rsatadi. Matnlar bilan ishlash me'yorlari quyidagi tartibda uch bosqichda amalga oshirildi:

1-bosqich: https://newspaperarchive.com va http://press.natlib.uz rasmiy saytlaridan mavjud skanerlangan gazeta nashrlari to'plandi. Bunda, ularning katta hajmli jildlarda yillar kesimida, mavjud nashrlar ixtiyoriy tarzda, matnlarni o'qiy olinishidagi sifat ko'rsatkichiga tayangan holda saralab tartiblandi;

2-bosqich: jamlangan va xronologik tartiblangan baza Microsoft Word dasturi orqali qo'l mehnati yordamida birma-bir qayta terildi va *.doc kengaytmali hujjatlar shaklida saqlandi. Bu turdagi hujjatlarda har ikkala tildan olingan ma'lumotlar maxsus ishlab chiqilgan jadvallarga joylashtirildi hamda ushbu jadvallarni umumlashtirgan alohida jildlar yaratilib saqlandi. Xronologik yig'ilgan manbalarga yillar kesimida +/- 5000 so'zli bazalar shakllantirish talabini qo'yish lozim topildi. Zero, miqdoriy me'yorning bo'lishi qidirilayotgan so'zning qo'llanishiga oid yillar kesimida faollik darajasini ko'rsatuvchi statistik ma'lumotning ishonchliligini isbotlaydi;

3-bosqich: yuqorida sanab o'tilgan bosqichlardan o'tgan, yig'ilgan manba Microsoft Excel dasturida ma'lumotlar o'rganilayotgan tillar kesimida alohida ikkita *.xls faylda vertikal ketma-ketlikda jamlandi va kompyuter xotirasi yoki internet sahifalariga yuklashga mos holatga keltirildi. Ushbu dastur yordamida ko'rib chiqilganda, umumiy jamlangan matnlar hajmi umumiy hisobda ikkala bazadan 316 ming 602 ta so'zli manba yig'ilgani aniqlandi. Hozirgi holatda, *Anaheim Bulletin* gazetasining 1991-yilning 2-avgustdagi sonidan 2020-yilning 8-dekabr sanasida chiqqan sonigacha bo'lgan muddatda yig'ilgan 30 yillik materiallar har bir yil hisobiga 5000 so'z atrofidagi va undan ortiq bo'lgan (bunga maqolani to'liq olingani sabab bo'ldi) hajmdagi matnlarni jamlagan umumiy 158 ming 819 ta so'z to'plandi.

Jamlangan materiallar ustida qayta ishlandi, saralandi, statistik birlamchi axborot olindi hamda ular mos ravishda 158 ming 819 va 157 ming 783 ta so'zni tashkil etayotganligi ikki qiyoslanayotgan tillardan yig'ilgan manbaning o'zaro katta hajmda farqlanmasligini ko'rsatmoqda. Bu degani korpus bazasidagi qiyosiy plandagi statistik tahlillar ishonarli va haqqoniy axborot sifatida qabul qilinishi mumkin.

Biz tomonimizdan yaratilgan *"Oila va jamiyat"* hamda *"Anaheim Bulletin"* gazetalarining 30 yillik bazasidan yig'ilgan gazeta matnlari asosida shakllantirilgan

ingliz va o‘zbek mediamatnlar diaxronik korpusi hozirda https://mediatextcorpus.com havolasida mavjud. Quyida zamonaviy elektron gadjetlarga osonlik bilan skanerlash orqali yuqorida aytib o‘tilgan havolani yuklab olish uchun qulay QR kod formasi ishlab chiqildi. "2-rasmga qarang".

2-rasm. Ingliz va o‘zbek mediamatnlar diaxronik korpusi havolasiga o‘tish

Umumiy uch fasldan iborat bo‘lgan dissertatsiyaning to‘rtinchi bobi **"Mediamatnlar diaxronik korpusining amaliy va funksional ahamiyati"** deb nomlanib, o‘z ichiga diaxronik korpusda faol leksik qatlamning statistik asoslari, jumladan, korpusning yillar kesimida o‘zgaruvchan yuqori chastotali so‘zlarga oid tahliliy imkoniyatlari hamda so‘zlarning shakliy xilma-xilligi, o‘zak va morfologik birlik munosabatiga, ingliz va o‘zbek mediamatn diaxronik korpuslarida lingvomadaniy tushunchalar muqoyasasiga, tillararo o‘zlashgan uslub, neologizmlar va ijtimoiy tushunchalar tahliliga doir amaliy izlanishlar natijasini umumlashtiradi.

Yaratilgan korpus avtomatik taqdim etadigan jadvallardan ko‘rinib turibdiki, *the* aniq artikli ingliz tilidan olingan 30 yillik bazaning barcha yillarida birdek eng faol qo‘llanilgan. Umumiy hisobda, 9434 marta (BCh) qo‘llanilgan bo‘lib, jami so‘zlar miqdoriga nisbatan olinganda, taqriban 4,69 foiz (MCh)ni tashkil etganini ko‘rish mumkin. Artiklning boshqa turlari, xususan, *a* noaniqlikni bildiruvchi shakli jami 3612 marta, *an* esa 595 marta qo‘llangan. Ularning mos ravishda unli yoki undosh tovush bilan boshlanganligiga ko‘ra birlikdagi sanoqli ot oldida kelishini inobatga oladigan bo‘lsak, bu ingliz tilidagi shunday turdagi otlarning ko‘prog‘i undosh tovush bilan boshlanishi to‘g‘risida farazni keltirib chiqaradi. O‘zbek tili bazasidan avtomatik chiqarilgan statistikada esa yillar davomida eng faol qo‘llanilgan so‘zlar farqlanib turganini kuzatish mumkin. Ular qatorida *ham* yuklamasi 12 yil mobaynida, *va* bog‘lovchisi 12 yil davomida, *bilan* ko‘makchisi 5 yil va bir soni umumiy hisobda faqatgina bir yil oralig‘ida qo‘llangan eng faol so‘z sifatida qayd etiladi. Umumiy hisobda esa, 2054 marta qo‘llanilgan *va* bog‘lovchisi aynan eng yuqori chastotali ekanligi ma’lum bo‘ldi hamda uning o‘zbekcha baza miqdorining umumiy hajmiga nisbatlangani 1,01 foizni tashkil etishi aniqlandi. Demak har ikkala tildan olingan so‘zlar jami bazaviy miqdorga nisbatan katta ulushga egaligi ortidan chog‘ishtirilayotgan tillarda keng qo‘llanilishi xususida xulosaga kelish mumkin.

Keyingi e’tiborli holat shuki, *to, of, in, for, on, with* kabi bir qator predloglar qolgan shu turdagi so‘zlarga qaraganda, yuqoriroq chastotada qo‘llanilganini

jadvallar aks ettirib turibdi. Ular tarkibidan eng faol ishlatilgani esa *to* predlogi bo'lib, umumiy hisobda 4290 marotaba ishlatilgan bo'lib, 2,13 foizli nisbat bilan ingliz tili bazasida ikkinchi eng faol so'z hisoblanadi. Nazarimizda, leksik birlikning funksional vazifalarining ko'pligi uning faollik darajasini oshiradi. *Macmillan English dictionary*[30] ga ko'ra *to* so'zining ikki, ya'ni predloglik va ravishlik holati bo'lib, jami 21 ta turli ma'no va vazifalarda keladi. Jumladan, yo'nalishni ko'rsatuvchi predlog, fe'lning *infinitiv* shakli tarkibida, *going to, ought to, used to* kabi turli fe'l birikmalari tarkibida, soatlarni aytishda va boshqa o'rinlarda qo'llanilishi keltirilgan. Korpus yuqoridagi fikrlarni to'liq o'rinli ekanligini isbotlay oladigan baza ekanligi ma'lum bo'ldi.

Ingliz tiliga oid korpus tahlillariga e'tibor berganimizda, *said, was, had* kabi o'tgan zamon (past simple) formasidagi hamda *is, have, has, are* hozirgi zamon (present simple) fe'l shakllarining ham eng faol so'zlar qatorida qo'llanilganini ko'rish mumkin. Bu o'zbek tili holatida ham kuzatilib, *bo'lgan, deydi, deb, bo'lib, olib, bor, bo'ldi* kabi o'tgan zamon, ravishdosh, hozirgi zamon formalaridagi fe'l shakllari misolida ko'rinadi. Har ikkala tildan olingan matnlarning gazeta maqolalari ekanligini inobatga olganimizda, ularning barchasi aynan axborotni ikki asosiy zamon kategoriyasi shakllari orqali yetkazib berishini anglash o'rinli bo'ladi.

Ingliz va o'zbek tillari tarkibidan eng faollari qatorida olmosh so'z turkumiga mansub so'zlarni ko'rish mumkin. Misol uchun, *I, it, her, he, they, their, this, we* kabi kishilik va *demonstrative* / ko'rsatish olmoshlari inglizcha matnlar bazasida yuqori chastotali 20 ta eng faol so'zlar qatorida turishini ko'rish mumkin. O'zbek tili misolida olib qaralganda, bunday so'z turkumi tarkibidan *shu, bu, u, uning, men, ular, biz, siz, meni, ularning* kabilarni keltirish mumkin. Ular tarkibidan men / I olmoshining shakliy o'zgarishlarining tahlili natijasida quyidagilar aniqlandi: *I (885), me (127), my (183), mine (2), myself (10)*; *men (466), meni (162), menga (148), menda (5), mendan (32), o'zim (28), mening (117), meniki (1)*. Har ikkala holatda bosh kelishikda kelgan olmosh shakllari faolroq, -niki qo'shimchasi orqali o'zaro alternativ bo'lgan egalikni otsiz kelib, oldin eslatilgan otning ma'nosini ifodalovchi olmosh turlari nibatan pastroq chastotaga ega ekanligi kelib chiqdi.

Ingliz tili bazasidan *Anaheim, school, community, council, center, students, scored, three, abdominal, Bayless, high, city, orange, Ryans* kabi o'zbek tili bazasidan esa *sog'lom, muhabbat, paranji, onam, Abbos, bolalar, Xudoyberdi, Sobir, momo, respublikasi, oila, oilalar, O'zbekiston, nos, Ozoda* kabi umum iste'moldagi umumiy va atoqli ot so'z turkumiga tegishli so'zlarning ushbu ro'yxatda kelishi korpusga kiritilgan bazaning miqdori nisbatan kam ekanligi hamda bu bazadagi cheklangan sondagi maqolalar kontentining xususiyatlariga bog'liq bo'lib qolganligi bilan xarakterlanadi. Agarda ushbu baza miqdori hozirgi holatidan bir necha marotaba katta bo'lganida, bizningcha, bu so'zlar eng faol qo'llaniluvchi so'zlar qatoriga kira olmas edi. Bunga asos sifatida ularning uch dekadali bazada faqatgina bir yoki ikki yillik ma'lumotlar bazasida yuqori chastotali so'zlar qatorida ko'zga tashlanishini ham keltirish mumkin.

---

[30] Macmillan English Dictionary for Advanced Learners. edited by Rundell M. and others. – Macmillan Publishers Limeted, 2007. – P. 1573-1574.

Korpusda *Navro'z*[31] so'zi qidirilganda, jami 27 ta misol mavjudligini ko'rish mumkin bo'ldi. Ular tarkibidan kollokatsion birliklar ajratib olinganda, quyidagi faol birikuvchi jumlalar aniqlandi: *Navro'z arafasi; Navro'z ayyomi; Navro'z bayrami; Navro'zi olam; Navro'z kunlari; Navro'zi Ajam.*

Yuqoridagi misollarning barchasi tushunarli bo'lib, bugungi kunda ham til sohiblari tomonidan faol qo'llanilishi ma'lum. Biroq so'ngda berilgan va korpusda jami 3 o'rinda birikib qo'llanilgan *Navro'zi Ajam* jumlasi e'tiborni tortdi. Kengroq konteksda olib qaralganda, quyidagi gaplarda qo'llanilgani ma'lum bo'ldi: 1) S.Saidovning 1992-yilda chiqarilgan "Jimlik" nomli maqolasida *"To'g'ri, har bir san'atkordan "Navruzi Ajam"ni talab qilib bo'lmaydi, lekin... San'atda shaldir-shuldir ko'payib ketdi..."*[32] degan jumla uchraydi; 2) Sirojiddin Sayyidning 1992-yildagi "Sholikor" qishloq sho'rosi binosi nomli maqolasida ham ikki o'rinda qo'llanilgani ma'lum bo'ladi. *Birinchisi, "Mirzo Boburning "Bahor ayyomidurdagi yigitlikning ayvonidur" satri bilan boshlanguvchi nozik g'azalini "Navro'zi Ajami" satriga solib aytganida, umrning, tiriklikning go'zal salomi yanglig' bitilgan bu g'azalning qadimiy kuy pardalari ichra vido yanglig' zoru nigoron yangraganida qanday hikmat bor? "* mazmunida berilgan bo'lsa, *ikkinchi o'rinda, "Navro'zi Ajam" ulug' kuy, zor ko'ngilning mangu, tuganmas navhasidir u. Undagi yig'ini so'zga ko'chirish qiyin"* deyiladi. Ushbu kengroq konteksdan olingan misollar orqali korpus foydalanuvchisi *Navro'zi Ajam*ning qadimiy o'zbek xalq kuyi ekanligini anglay oladi.

Keyingi o'rinda ingliz madaniyatiga xos bo'lgan *Christmas* bayramining shu til sohiblari ongida ifodalanishiga oid misollar tahlilini berishni joiz topdik. Mediamatnlar diaxronik korpusida *Christmas*[33] so'zi qidiruvga berilganida jami 24 marta qo'llanilgani aniq bo'ldi. Ushbu so'z bilan birikib keluvchi kollokatsion birliklar ajratib olinganda, quyidagilar kelib chiqdi: *Christmas project; Christmas cards; Christmas services; Christmas Eve; Christmas tree; Christmas Party; Christmas time; Christmas Day.*

Yuqoridagilarni ko'rib chiqib, barchasi umumiy madaniyatga xos tushunchalarni anglatishini anglash qiyin emas. Ular orasidan *Christmas services* birikmasi qo'shimcha izoh talab etayotgandek, nazarimizda. Kengroq konteksda ko'rilganda, korpusda maqola 681-raqam bilan belgilangan bo'lib, 2011-yilda Sh. Nlapcich tomonidan yozilgan *"Toy drive planned tonight for babies"* sarlavha ostidagi matn tarkibidan olingani ma'lum bo'ldi. *"Any donations are appreciated. CANYON HILLS PRESBYTERIAN PLANNING CHRISTMAS SERVICES Canyon Hills Presbyterian Church will offer two Christmas Eve worship services, one at 5 p.m. and a candlelight service at 11 p.m."* – mazmunidagi matn qismi olib qaralganda, xristianlik e'tiqodiga sobit jamiyat vakillari uchun diniy bayram qismi hisoblanishi va bayramga tashrif buyuruvchi, xor tarkibida diniy qo'shiqlarni kuylovchi bolalarga sovg'a sifatida beriluvchi ehsonlarning qabul qilinishi aytiladi. Shuningdek, xizmat jarayoni ikkiga bo'linishi nazarda tutiladi. Ulardan birinchisi *Christmas Eve worship services,* ya'ni *ibodat xizmati* bo'lsa, ikkinchisi *a candlelight*

---

[31] O'sha manba https://mediatextcorpus.com/search?q=navr&language=2&match_type=0&year=

[32] O'sha manba https://mediatextcorpus.com/article/46

[33] O'sha manba https://mediatextcorpus.com/search?q=christmas&language=1&match_type=1&year=

service, ya'ni *sham yoqish xizmati* hisoblanadi. Bu orqali ommalashgan bayram an'analariga muvofiq tashkiliy masalalarni muvofiqlashtiruvchi xizmatlarni taklif etuvchi jamoat birlashmalari shakllanganligi aniqlandi. Bizningcha, bu turdagi xizmatlar kelajakda o'z faoliyatini tugatishi yoki rivojlanishi mumkin, biroq uning tildagi aksi gazeta matni arxivlarida saqlanishi korpusning mediamatnlarga asoslanganini izohlaydi.

XULOSA

1. O'tgan asr ikkinchi yarmining boshlanishida yuzaga kelgan tilshunoslik ilmining yangi tarmog'i bo'lgan korpus lingivsitikasi o'zining tadqiqot metodologiyasiga egaligi bilan xarakterlanib, zamonaviy ilm-fan, xususan, antropotsentrik tadqiqotlarga oid muammolarni yechishda keng qo'llanilmoqda. Uning obyekti hisoblangan lisoniy korpusga o'zida tilning asliyatdan olingan autentik materiallar asosida shakllanuvchi, reprezentativlik tamoyiliga javob beruvchi, til xususidagi ilmiy farazlarga amaliy yechim beruvchi, me'yoriy hajm talablariga bo'ysunuvchi elektron lisoniy baza sifatida yondashish o'rinli.

2. Ijtimoiy hodisa hisoblnadigan til doimiy rivojlanishdan to'xtamasligi hamda hozirgi zamon globallashuv davridagi xalqlarning keskin yaqinlashuvi oqibatida uning bugungi holati oldingiga va kelajakdagi holatidan tubdan farqlanishiga doir farazlarni keltirib chiqarmoqda. Bunday tahlikali vaziyatni to'gri baholagan holda, lingvistik tadqiqotlarning maqsad va vazifalari mavjud tillarning hali global internet hamda axborot xurujidan ta'sirlanmagan qatlam vakillaridan milliy sinxron korpuslarni yozib olib shakllantirishda dolzarblik kasb etadi. Zero, ushbu jarayon uzluksiz davom ettirib boriladigan bo'lsa, vaqt o'tishi bilan yaratib borilgan lingvistik baza tarixiy qiymatga ega bo'lib, u keyingi bosqichda diaxronik korpus maqomiga erishadi.

3. Diaxronik korpuslar o'zi ifoda etayotgan lison tarkibidan aniq belgilab olingan tarixiy davrda yaratilgan asosan yozma matnlar tarkibidan yaratilishini ko'zda tutadi, chunki og'zaki nutqqa oid materiallarni jamlash, ular ustida ishlash nisbatan murakkablik kasb etadi. Ya'ni tarixiy voqelik radio va televideniye tarmog'ining rivojidan oldingi insoniyatning taraqqiyot bosqichlariga xos og'zaki manbani topish qiyin. Biroq tillararo interferensiyaga moyil til qatlami, dialekt yoki sheva, xalq og'zaki ijodi namunalarining turli shakllarini qayd etish orqali tilning rivojlanish tendensiyalarini ochib beruvchi ishonchli manbaga erishish mumkin. Bunda nutqning ushbu shaklini transkripsiya qoidalariga moslab, diaxronik korpus tarkibiga kiritishga ehtiyoj yuzaga keladi.

4. Har qanday diaxronik korpus yaratish amaliyotida, undan kutiladigan maqsad, ya'ni tillarning tadrijiy taraqqiyotini o'rganishga moslangan korpus interfeysi o'z ichiga, albatta, qidiruv tizimi, ro'yxatdan o'tish burchagi, manbadagi faol so'zlar ro'yxati va statistikasining ochiq ma'lumoti, manbalar bilan kengroq tanishish imkoniyati, yillar ketma-ketligiga alohida solnoma ma'lumotlariga bo'lgan havolalar va har bir kiritilgan matn xususidagi statistik hamda xronologik axborot olish imkoniyatini qamrab olishi shart.

5. Diaxronik korpuslar yaratish koʻp yillik, kamida bir dekadali oraliqdagi manbani yigʻishga tayangan holda uzoq muddatli ishlanishlar namunasi ekan, uning interfeysi va dizaynini foydalanuvchiga qulay shakllantirish korpus tuzuvchilarning pirovard maqsadi boʻlishi shart. Dizayndagi kichkina detalning ham ahamiyatdan chetda qolishi ortidan uzoq yillik mehnat mahsulini yuzaga chiqmay qolishiga sabab boʻlishi mumkin. Shuning uchun, diaxronik korpuslar yaratishda ularning konkordans natijalarini takomillashtirish, davriy farqlanuvchi, tizimli davomiylikka ega xronologik tartibda berilishi yoki kerakli davr hisobidan cheklanishiga doir funksiyalar kiritilishi lozim.

6. Korpus materiallariga qoʻyiladigan umumiy tamoyillar tarkibiga manbaning autentik material boʻlishi; oʻzaro teng hajmda toʻplanishi; hech qanday oʻzgartirish yoki toʻgʻrilash kabi tahririy jarayonlarga tortilmasligi; umumiy leksik miqdoriy hajmi bir milliondan kam boʻlmasligi; reprezentativlik tamoyiliga mos kelishi; zamon va makondagi mutanosibligi kabilarni kiritish oʻrinli ekanligi aniqlandi. Aynan diaxronik korpuslarga xos boʻlgan xususiy tamoyillar sifatida quyidagilar kiradi: 1) xronologik uzviylikda matnlarni yigʻish; 2) tarixiy voqelik ortidan yozuv grafikasida alfavit oʻzgarishlarini boshidan oʻtkazgan tillarning qadimiy yoki tarixiy bitiklarini transliteratsiya qilish; 3) teng hajmda matnlar kiritish borasidagi meʼyor talablariga nisbiy yondashish; 4) kamida 1 dekadali vaqt oraligʻini qamrab olish.

7. Diaxronik korpuslarning monitor korpuslar turiga kirishini inobatga olganda, uning tarkibiga kiritiladigan bazaviy miqdor va talab tilning boyishiga oid barcha voqelikni qayd etib borish barobarida zamon bilan hamnafas shakllanib borishga moyil boʻlgan matniy birliklar boʻlishiga oid qatʼiy talab joriy qilish zarur. Bu borada bugungi kunda ommalashgan internet tarmoqlarida kundalik yuklanayotgan materiallardan foydalanishni ularning oʻzgaruvchanlik, yaʼni matnni doimiy tahrir qilish mumkinligiga boʻlgan xususiyati sababli maʼlumotning kiritilgan vaqtiga xos ekanligini shubha ostiga qoldirgani tufayli amaliyotga joriy etib boʻlmaydi. Shunday ekan, qogʻozga chop etiladigan yoki online nashr etiladigan hamda barcha til meʼyorlariga rioya qilgan holda yaratiladigan va eng muhimi mediamuloqotning ajralmas qismi sifatida til sohiblarining ijtimoiy, milliy va madaniy hayot tarzini muttasil yoritib boruvchi gazeta matnlaridan diaxronik korpuslar uchun manba sifatida foydalanish oʻzini oqlaydi.

8. Takomillashtirilgan korpus lingvistikasi metodologiyasini olti bosqichga boʻlingan quyidagi tizimlilikka asoslangan metodlar tashkil etadi: yoʻnaltiruvchi metodlar, manba bilan ishlash yoki bazaviy metodlar, metadata qoʻshish metodlari, avtomatlashtirish metodlari, baholash metodlari, chastotali qidiruv-analiz metodlari. Ushbu har bir bosqich uzviylik bilan bir-birini toʻldiruvchi metodlar sistematik uzviylikka asoslangan holda ketma-ket amalga oshiriladi hamda barcha turdagi, xususan, diaxronik korpuslar yaratishda ham qoʻllash oʻrinli.

9. Mediamatnlar diaxronik korpusini shakllantirishda aniq belgilab olingan tamoyillarga rioya qilgan holda, tilning muayyan tarixiy bosqichiga oid matniy birliklarni yigʻish, saralash va avtomatlashtirish jarayoniga tortilishi qatʼiy tamoyillarga asoslanmasa, korpus oʻz oldiga qoʻygan maqsadga erisha olmaydi. Shuni inobatga olib, diaxronik korpus takibiga badiiy adabiyot namunalarini kiritish

maqsadga muvofiq emas. Zero, badiiy matnda muallif tomonidan asar ifodalayotgan zamon va makon ruhiyatini berish uchun maqsadli ravishda tarixiy, arxaik yoki neologizmlardan foydalanilishi oqibatida diaxronik davrlashtirishga nomutanosib natijalarni keltirib chiqarishi mumkin.

10. Mediamatnlar diaxronik korpusini shakllantirishda tanlab olingan o'rganilayotgan tillar doirasida "Anaheim Bulletin" va "Oila va jamiyat" gazetalari tanlab olindi. Gazeta materiallaridan aniq me'yorga asoslangan holda matnlar yig'ish bosqichi yakuniga ko'ra "Anaheim Bulletin" gazetasidan 1991-yilning 2-avgustdagi sonidan 2020-yilning 20-dekabr sonigacha hamda Oila va jamiyat" gazetasidan 1991-yilning 26-sentabrdagi 4-sonidan 2020-yilning aprel oyida chiqqan 16-sonigacha yig'ilgan 30 yillik materiallar har bir yil hisobiga 5000 so'z atrofidagi va undan ortiq bo'lgan (bunda maqola matni to'liq olingani sabab bo'ldi) hajmdagi matnlarni jamlagan, mos ravishda, umumiy 158 ming 819 ta va 157 ming 783 ta so'zni tashkil etuvchi baza shakllantirildi. Bunda, bazadagi yillar kesimidagi xronologik bazaviy-miqdoriy bir xillik yaratilgan korpusning me'yorlashtirilgan (balanced) turga kiritilishini asoslaydi.

11. Mediamatnlar diaxronik korpusi taqdim etadigan statistik axborotga tayangan holda aytish mumkinki, *the* aniq artikli ingliz tilidan olingan 30 yillik bazaning barcha yillarida birdek eng faol so'z sifatida qo'llanilgan hamda bu o'zbek tili bazasidan avtomatik chiqarilgan statistikada esa yillar davomida eng faol qo'llanilgan so'zlar farqlanib turganini kuzatish mumkin. Ular qatorida *ham* so'z-turi 12 yil mobaynida, *va* so'z-turi 12 yil davomida, *bilan* so'z-turi 5 yil va *bir* so'z-turi umumiy hisobda faqatgina bir yil oralig'ida eng faol so'z sifatida qayd etiladi. Tadqiq etilgan tillar bazasida xronologik kuzatish natijalariga ko'ra eng faol o'nlikni, asosan, yordamchi, ko'makchi, artikl, olmosh va vazifadosh so'zlar tashkil etganligini kuzatish mumkin. Ayni natijalar mavjud til korpuslari tahlillariga qiyoslanganda ham yaqin o'xshashlikni aks ettiradi. Bu esa mediamatnlar diaxronik korpusi va shu turdagi yaratilishi mumkin bo'lgan diaxronik korpuslarning tilning lug'at qatlamining faollik darajasiga ko'ra tartiblash imkoniyatini yaratishiga oid ilmiy farazni asoslaydi.

12. Ingliz va o'zbek tillari mediamatnlar diaxronik korpusi o'zi ifoda etuvchi lisoniy bazaga tayanib, ayni tillarning madaniyatini chuqurroq anglashga qaratilgan tahlillar, jumladan, lingvokulturologik birliklarning gazeta matnlarida berilishiga oid nazariy xulosalarga kelishda xizmat qiluvchi statistik hamda so'z qo'llashga doir tezkor, aniq, ishonchli misollarni taqdim etadi. Bunda qidiruvga beriladigan tushuncha yoki konseptning mohiyatini anglash va qiyoslanayotgan tillar doirasida muqoyasa qilish bilan bir qatorda ularning konseptual freymini yaratish imkoniyati yuzaga keladi.

13. Ingliz va o'zbek tillari mediamatnlar diaxronik korpusi mohiyatan o'zi qamrab olgan 1991–2020-yillar oralig'idagi davrga xos tanlangan gazeta matnlari tarkibidan yillar ketma-ketligida so'zlarning faollik darajasidan kamayib borishiga qadar avtomatik hisoblab beradi. Ayni vaziyat tadqiqotchiga tilning lug'at tarkibiga kirib kelgan yangi, tarixiy yoki arxaik leksik birliklarni empirik tahlil tamoyillariga mos ravishda aniqlashga imkon beradi. Ushbu korpus bazasining hozirgi hajmi qiyoslanayotgan tillarga oid yakuniy xulosalarni chiqarishga imkon bermasa-da,

ishlab chiqilgan loyiha hajmini kengaytirib borish oqibatida bunga asosli ravishda erishish mumkin, deb hisoblaymiz.

14. Mediamatnlar diaxronik korpusi so'zlarning o'zaro birikuvchanligiga oid aniq misollar, statistik axborot va kollokatsion birliklarning kontekstda qo'llanilishini anglash uchun matn bo'laklarini taqdim etadi. Ular asosida turli tizimli tillarda o'rganiladigan tushunchaning semantik maydonini hosil qilishda nafaqat so'zning ma'noviy guruhlarini hosil qilish mumkin, qolaversa, so'z birikmalari va kollokatsion birliklarning xronologik birikuvchanligiga oid istorizmlarni tadrijiy aniqlash imkoni yuzaga keladi. Bunda konkordans qidiruviga kiritilgan leksemaga birikuvchanlikka moyil lisoniy birliklar, jumladan morfemalarning ham o'zak+qo'shimcha munosabatiga oid faollik darajasini aniq xronologik kuzatish mumkinligi so'zning shakl va mazmun o'zgarishiga moyilligini tadrijiy o'rganish imkoniyatini beradi.

15. Internet tarmog'ida yuklanayotgan elektron til resurslari, jumladan, lisoniy korpuslarning yaratilishi tillarning yo'qolib ketishiga oid farazlarni inkor etuvchi omillar sirasiga kirishi borasidagi e'tirof o'rinli. Zero, tilshunoslikka oid izlanishlar bazasida yaratilayotgan ilmiy tadqiqotlar mavjud tilning unsurlariga daxl qilishga emas, balki uning xususiyatlarni tahlil qilishga qaratilgan zamon va makonni aks ettiruvchi lingvistik bazalarni jamlab borishga yo'naltirilishi zarur. Dissertatsiyaning amaliy natijasi bo'lmish ingliz va o'zbek tillarining mediamatnlar diaxronik korpusining yaratilishi tillarning ayni paytdagi holatiga baho berish emas, ko'zda tutilgan davr mobaynidagi uzluksiz rivojining kvantitativ aksini ko'rsatishi bilan xarakterlidir.

SCIENTIFIC COUNCIL ON AWARDING SCIENTIFIC DEGREES DSc.03/30.12.2022.Fil.05.02 AT FERGANA STATE UNIVERSITY

BUKHARA STATE UNIVERSITY

ATABOEV NOZIMJON BOBOJON UGLI

PRINCIPLES OF EMPIRICAL ANALYSIS OF LANGUAGE DEVELOPMENT IN THE DIACHRONIC CORPUS OF MEDIA TEXTS (in the example of Newspaper materials)

10.00.11 – Theory of language. Applied and computational linguistics

DISSERTATION ABSTRACT for the doctor of philological sciences (DSc)

Fergana – 2024

The theme of the dissertation for doctor of philological sciences (DSc) was registered at the Supreme Attestation Commission of the Republic of Uzbekistan under the number B2022.3.DSc/Fil465.

The doctoral dissertation was carried out at the Bukhara State University.

The abstract of the dissertation is posted in three languages (Uzbek, English, Russian (abstract)) on the website of Fergana state university www.fardu.uz and on the website ZiyoNet information and educational portal www.ziyonet.uz.

|	
Scientific supervisor:	**Teshabaeva Dilfuza Muminovna,** Doctor of Philological Sciences, Professor
Official opponents:	**Sultonova Shokhista Mukhammadjanovna** Doctor of Philological Sciences (DSc), Professor
	Abdurakhmonova Nilufar Zaynobiddin kizi Doctor of Philological Sciences (DSc), Professor
	Dr. Mario Bisiada Doctor of Philological Sciences (Spain)
Leading organization:	**Tashkent State University of Uzbek Language and Literature named after Alisher Navoi**

Defense of the Dissertation will take place on "26th" of August 2024, at 8:00 a.m. at a meeting of Scientific Council DSc.03/30.12.2019.Fil.05.02 under Fergana State University (Address: 105, Marginoni Street, Fergana, 100151. Tel: (99873) 244-57-82; fax: (99873) 244-44-01; e-mail: info@fdu.uz.

The doctoral dissertation can be reviewed at the Information Recourse Center at Fergana State University (registered under the number 366). The address: 150100, 19, Murabbiylar street, Fergana city Tel: (+99873) 244-44-02, fax: (+99873) 244-44-93, e-mail: fardu_info@umail.uz

The abstract of the dissertation sent out on "13th" of August 2024
Mailing report №22 on "13th" of August 2024
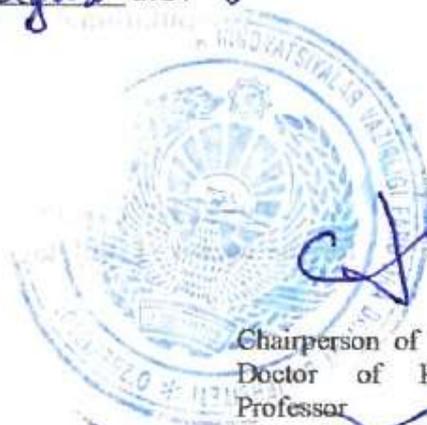
S.M. Muminov
Chairperson of the Scientific Council, Doctor of Philological Sciences, Professor

M.T. Zokirov
Scientific Secretary of the Scientific Council, Candidate of Philological Sciences, Professor
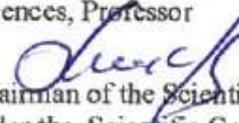
A. Mamajonov
Chairman of the Scientific seminar under the Scientific Council, Doctor of Philological Sciences, Professor

INTRODUCTION (abstract of DSc dissertation)

Topicality and necessity of the thesis. In world linguistics, due to the acceleration of globalization trends, intercultural and interlinguistic rapprochement is occurring and the interference between closely related languages is strengthened, the question of studying the effects on the levels of the pure lexical layer is emerging. Accordingly, the fact that language is being studied as an integral part of society without separating it from the owner of the language, that is, from the perspective of the anthropocentric paradigm, presents to linguists the requirement to study the current state of the language in harmony with its historical evolution, and to develop scientific hypotheses based on analyzes of its future. In this regard, the implementation of diachronic corpora, which are chronologically ordered and formed on the basis of authentic materials/texts taken from the original based on precise statistical standards, is of great importance.

Corpus linguistics, which is widely spread in the framework of linguistic research in the world and is progressing using computer technologies in accordance with the requirements of modern science, offers new convenient ways to find solutions to all problems of linguistics based on its methodology. The fact that the system of methods is formed by the combination of disciplines such as computer linguistics, text linguistics, mathematical linguistics and applied linguistics guarantees that the research conducted in this regard is essentially relevant.

The fact that comprehensive reforms to further strengthen the status of the state language in our country are based on political will imposes on linguists the task of creating an electronic database, i.e. a corpus, of sources from ancient manuscripts written in our native language to textual units covering today's linguistic reality, and researching them based on a new approach. As the Prezident of our Republic, Sh.M. Mirziyoyev noted, "It is known that in the study and development of any language, first of all, its theoretical and methodological foundations play an important role. Among these, it is necessary to note that serious work is being carried out on creating the electronic national corpus of the Uzbek language and Uzbek applications of software products, popularizing the Uzbek language on the Internet, ensuring that it occupies a worthy place, developing and implementing the programs for teaching the state language and text editing."[34] Therefore, in the course of the first development of the diachronic corpus of the Uzbek language, on the example of the newspaper materials selected as the research object, important tasks such as recording the changes in the linguistic layer by empirically analyzing the development of the lexical layer of our native language were determined.

The present scientific study serves to a certain degree to achieve the objectives set out in the following legal documents: Decree PF-6084 of the President of the Republic of Uzbekistan "On measures to further develop the Uzbek language in our country and improve the language policy" as of October 20, 2020, Decree PF-5850 "On measures to fundamentally increase the prestige and position of the Uzbek language as a state language" as of October 21, 2019, Resolution PQ-5117 "On measures to bring the

---

[34] Prezident Shavkat Mirziyoyevning Oʻzbek tili bayrami munosabati bilan Oʻzbekiston xalqiga tabrigi. https://president.uz/oz/lists/view/6784 [murojaat sanasi 29.02.2024]

activities of popularization of foreign language learning to a qualitatively new level in the Republic of Uzbekistan" as of May 19, 2021, Resolution PQ-19 "On Additional measures to increase the flow of tourists and further increase the prestige of our country by widely promoting the study of the history and culture of Uzbekistan, the Uzbek language and literature in foreign countries" as of January 12, 2024, Resolution of the Cabinet of Ministers of the Republic of Uzbekistan No. 139 "On measures to further increase the effectiveness of fundamental and applied research on the Uzbek language and literature" as of March 11, 2020, Resolution No. 117 "On measures to further improve the activities of the Tashkent State University of Uzbek Language and Literature named after Alisher Navoi" as of March 3, 2020.

Relevance of the research to the priority areas of science and developing technology of the Republic. The research was carried out under the priority directions of science and developing technology of the Republic, "The Formation of the Innovative System of Ideas in Social, Legal, Economic, Cultural, Spiritual and Educational Development of the Democratic State and their Ways of Implementation in the Information Society."

A review of foreign research on the topic of dissertation[35].

The practice of creating language corpora and the use of today's achievements of the scientific network of corpus linguistics have achieved important scientific results in theoretical and practical researches conducted in leading scientific centers and higher education institutions of the world including the following: Lancaster University, Lancaster (United Kingdom), The University of Birmingham (United Kingdom), Macquarie University, Sydney (Australia), University of Portsmouth (United Kingdom)), Lancaster summer schools (United Kingdom), University of Sydney, Sydney (Australia), Bukhara State University (Uzbekistan), Tashkent State University of Uzbek Language and Literature named after Alisher Navoi (Uzbekistan).

As a result of research on corpus linguistics and the use of linguistic corpora in the analysis of the history of language, its development trends and problems of intercultural interference in world linguistics, the following scientific results were obtained, among others: the problems including the application of corpus linguistic methods to the structure of English and the analysis of Western languages (Lancaster University, Lancaster (United Kingdom), early lexicographic research based on corpus linguistics methodology and fields such as stylistics, discourse analysis, psycholinguistics, cognitive linguistics and statistics (The University of Birmingham (United Kingdom) to determine the coherence of corpus analysis methods, various research areas, including child language development, second language learning, translation, sociolinguistics, global variants of English and (Macquarie University, Sydney (Australia), using corpora consisting of billions of words to justify the reasons for the different use of the same language in different regions, genres and situations (University of Portsmouth, (United Kingdom)), teaching English grammar and vocabulary through corpus linguistics (Lancaster summer schools (United Kingdom)

---

[35] Sharh universitetlarning veb-saytlaridagi ma'lumotlar asosida taqdim etildi: https://www.lancaster.ac.uk/, https://www.birmingham.ac.uk/, https://www.mq.edu.au/, https://www.port.ac.uk/, https://wp.lancs.ac.uk/, https://www.sydney.edu.au/, https://buxdu.uz/

using corpus analysis in the process of learning and teaching English (University of Sydney (Australia), English and monitoring the development of selected languages by creating a diachronic corpus of Uzbek media texts (Bukhara State University (Uzbekistan), creation of national, educational and electronic corpora of the Uzbek language (Tashkent State University of Uzbek Language and Literature named after Alisher Navoi (Uzbekistan) have been solved.

In world linguistics, the study of language development based on corpus analysis and the integrated systematic study of the practice of creating diachronic corpora are conducted in the following priority directions: principles of corpus analysis in the study of biological language development, in particular, according to age and gender categories, cultural and regional language, creation of language corpora including regional dialects and language variants in development, etc.

Problem development status. In world linguistics, G. Leach was one of the first to conduct scientific research on the emergence and development of corpus linguistics[36]. Studies on the formation of modern corpus linguistics were carried out by English scholars such as S. Houston, D. Biber, S. Gries, M. Halliday and W.Teubert[37]. In the works of Russian scientists V. Baranov and V. Zakharov, corpus linguistics was specified. Scientists such as N. Vladimov, S. Vadyaev, O. Gorkina, V.Mamontova and A. Magametova studied the concept of "corpus linguistics," "linguistic corpus" and the problems of its classification in detail[38]. A comparative analysis of corpus linguistics with computer linguistics and other similar fields was carried out in the research of Tajik scientist D. Djafarova[39].

A. Kilgarriff, K. Iztok, F. Karlsson, and I. Sozhenin's studies on the use of corpus analysis in lexicology have been observed[40], Scientists such as V. Rikov, A. Romanov

[36] Leech G. Corpora And Theories Of Linguistic Performance. In J.Svartvik (ed.). 1992. – 105–122 pp.

[37] Hunston S. Corpora in Applied Linguistics: Cambridge University Press, 2002. – 234 p.; Halliday M.A., TeubertW., Yallop C. and Cermakova A. Lexicology and Corpus Linguistics, – London, 2004. – 184 p.; Gries S.Th. Introduction to S. Gries and A. Stefanowitsch (eds.). Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis. – Berlin: Heidelberg, New York: Mouton de Gruyter, 2006. – 1-18 p.; Douglas B., UllaC. and ThomasA.U. Discourse On The Move. – Amsterdam, 2007. – 290 p.

[38] Баранов А.Н. Компьютерная лингвистика // Введение в прикладную лингвистику: учебное пособие. – М.: Едиториал УРСС, 2003. – 320 с.; Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие/ В.П. Захаров. – СПб, 2005. – 48с. – Режим доступа:http://download.yandex.ru/class/zakharov/plan.pdf; ВладимовН.В. Корпусный подход к решению переводческий проблем (на материале англиского и русского языках). Дис....канд. филол. наук. – М.: РГБ, 2005. – С 26.; Вадяев С.Е.Лингвистические принципы построения и использование корпуса текстов для Исследования официально-делового стиля современного немецкого языка(на материале электронного корпуса «DER») диссер... для кандидата филол. наук. – Самара, 2005. – 160 с.; Горина О.Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке. Дисс. … канд. пед. наук. – М., 2014. – 321с.; Мамонтова В.В. Особенности перевода сложносоставных слов с английского языка на русский (на материале корпуса публицистических текстов).Дис. ... канд. фил.наук. – Ставрополь: 2008. – 139 с.; Магомедова А.Н.Корпусная лингвистика и контекстное разрешение лексической многозначности слов. Дис. ... канд. филол. наук – М.: РГБ, 2005. – 132 с.

[39] Джаъфарова Д.Ф. Модели лингвистического анализа текстов таджикского языка (на материале газелей Хафиза) дисс. ... канд. фил. наук. – Душанбе, 2013. – 151 с.

[40] Kilgarriff A. and Kosem I. Corpus Tools For Lexicographers. – Trojina: Institute for Applied Slovene Studies, 2012. – 37 p. УРЛ: https://www.researchgate.net/publication/264884809_Corpus_tools_for_lexicographers; Karlsson F. Lexicography And Corpus Linguistics. – Helsinki: Department of General Linguistics University of Helsinki, 1992. УРЛ: http://www.ling.helsinki.fi/~fkarlsso/INTJLLEX.pdf; Kilgarriff A. Using Corpora [And The Web] As Data Sources For Dictionaries. – Helsinki: 2012. – 21 p. https://www.sketchengine.eu/wp-content/uploads/Using_corpora_2012.pdf; Саженин И.И.Корпусные методы в лексикографии: опыт создания

and V. Plungyan developed the method of using the concept of "corpus" in text analysis[41].

In Uzbekistan today, large-scale scientific research is being conducted on the application of corpus linguistics methods and the use of existing corpora. J. Jumaboeva focused on finding semantic gradations from English language corpora, in particular from BNC content, G. Ergasheva researched gender conceptual metaphors based on BNC corpus, and G. Sobirova focused on using BNC corpus in language teaching. Sh.M. Hamroeva was looking for a solution to the problem of creating an author corpus[42]. Also, regarding the practice of creating electronic, national, educational and parallel corpora of the Uzbek language and the development of their theoretical foundations, It is appropriate to cite important scientific research by scientists like N.Z. Abdurakhmanova, Sh.M. Hamroyeva, G.I.Toyirova, G.Kh. Begmatova, A.A. Rakhmanova, O.M. Kholyirov, D.Q. Elova and R.A. Karimov as an example[43]. In addition, large-scale projects on the creation of linguistic corpora of the Uzbek language are giving good results. For example, the electronic corpus of the Uzbek language (https://uzbekcorpus.uz), the national corpus of the Uzbek language (http://uzbekcorpora.uz), and the educational corpus of the Uzbek language (https://uzschoolcorpara.uz) can be noted as important software products.

Relevance of the dissertation research with the plans of the scientific research works of the higher educational institution where the dissertation has been conducted. The theme of the dissertation is an integral part of the plan of scientific research at Bukhara State University on "Lexical-semantic systems of language,

модели Словарного корпуса Диссертация на соискание ученой степени кандидата филологических наук Новосибирск — 2013. 193 с.

[41] Рыков В.В. Корпус текстов новый тип словесного единства Текст. / В.В. Рыков. – Электрон, изд. – Режим доступа к изд.: http://rykov-cl.narod.ru/t31 .html.; Романов А.С. Методика и программный комплекс для идентификации автора неизвестного текста: Автореф. дисс....канд. тех. наук. - Томск, 2010. 26 с.; Плунгян В.А. Корпус как инструмент и как идеология: о некоторых урокахсовременной корпусной лингвистики// Русский язык в научном освещении.–М.: №2., 2008. – С. 7-20.

[42] Сабирова Г. Умумтаълим мактаблари ўкувчиларига инглиз тили олмошларини ахборот технологиялари асосида ўргатиш методикаси: Филол. фалс. д-ри (PhD) ...дисс. автореф. – Тошкент: 2017. – 41 б. URL: http://uzswlu.uz/media/ba7be50d-84f1-c826-f935-6e7610cb6bf5.pdf; Эргашева Г. Турли тизимдаги тилларда гендерга оид терминологиянинг шаклланишида лингвистик ва экстралингвистик омиллар: Фил. фан. д-ри. .... дисс. автореф. – Тошкент: 2018. – 61 б. URL: http://demo.navoiy-uni.uz/sites/default/files/ziyonet.pdf; Джумабаева Ж. Ўзбек ва инглиз тилларида лексик ҳамда стилистик градуонимия. докторлик диссертацияси автореферати. – Тошкент: 2016. – 82 б. URL: http://tashgiv.uz/wp-content/uploads/2017/01/ avtoref_jumabayeva.pdf; Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Ф.б. фалс. д-ри (PhD) ... дисс. автореф. – Қарши: 2018. – 52 б. URL: https://e-books.tsue.uz/pdf-ozbek-tili-mualliflik-korpusini-tuzishning-lingvistik-asoslari

[43] Abdurahmonova N.Z. O`zbek tili elektron korpusining kompyuter modellari: Filol. fan. d-ri. (DSc) diss. avtoreferati. – Toshkent, 2021. – 73 b.; Хамроева Ш.М. Ўзбек тили морфологик анализаторининг лингвистик таъминоти: Филол. фан. д-ри (DSc) дисс. автореферати – Фарғона, 2021. – 76 б.; Тоирова Г.И. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари: : Филол. фан. д-ри (DSc) дисс. автореферати – Бухоро, 2021. – 72 б.; Бегматова Г.Х. Ўзбек тили миллий корпусида идиомалар базасини яратиш: Филол. фан. фалс. д-ри (PhD) дисс. – Термиз, 2021. – 143 б.; Рахманова А.А. Ўзбек тили миллий корпусини яратишдаги компьютер усуллари: Филол. фан. фалс. д-ри (PhD) дисс. – Тошкент, 2021. – 161 б.; Холиёров Ў.М. Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари: Филол. фан. фалс. д-ри (PhD) дисс. – Термиз, 2021. – 181 б.; Каримов Р.А. Ўзбек-инглиз параллел корпусини тузишнинг лингвистик ва дастурий масалалари: Филол. фан. фалс. д-ри (PhD) дисс. – Бухоро, 2022. – 148 б.; Элова Д.Қ. Ўзбек тили корпуси бирликларининг услубий тегларини яратиш тамойиллари ва лингвистик таъминоти: Филол. фан. фалс. д-ри (PhD) дисс. автореф. – Тошкент, 2022. – 65 б.

comparative typological research and problems of literary criticism." The topic of the dissertation was included in the scientific development plan of the Foreign Languages faculty of the university.

The purpose of the research is to reveal the principles of empirical analysis of language development research by creating a diachronic corpus of English and Uzbek media texts with a theoretical understanding of the application of corpora in the development of corpus linguistics.

The tasks of the research work are as follows:

to interpret corpus linguistic approaches to the enrichment of the lexical layer of the language using corpus analysis methods;

to develop the requirements for working with texts specific to the database of diachronic corpora on working with the source in the course of improving the research methodology of corpus linguistics related to the creation of a diachronic corpus and its application;

to elucidate the specific features of media texts as a social expression of language and the main resource for language enrichment;

to realize the theoretical and practical importance of creating a diachronic corpus of English and Uzbek media texts, creating a diachronic corpus consisting of English and Uzbek newspaper texts within the selected languages;

to reveal the practical importance of existing English language corpora, including chronologically ordered and standardized linguistic corpora;

to prove that the diachronic corpus of media texts in Uzbek and English newspapers is a base that can be used as a statistical basis for practical significance, including the change of the lexical layer over the years;

to conduct research on whether the diachronic corpus of English and Uzbek media texts is a sufficient source for the comparison of interlinguistic and linguistic concepts;

to justify the approaches to monitoring the development of the lexical layer of the language of the diachronic bridge of English and Uzbek media texts theoretically.

The objects of the research work are "Oila va jamiyat" and "Anaheim Bulletin" newspapers' volumes published between 1991 and 2020.

The subject of the research work is the gradual comparison of lexical units in statistical analyses and search results obtained from the diachronic corpus of English and Uzbek media texts created on the basis of publications of "Oila va jamiyat" and "Anaheim Bulletin" newspapers between 1991-2020.

Methods of the research. Content analysis, classification, descriptive, comparison, statistical, linguo-cognitive, corpus-based and empirical analysis methods were used in the process of the research.

Scientific novelty of the research work includes the following:

based on the achievements of world corpus linguistics, in particular, the theoretical and practical foundations of the creation of existing diachronic corpora of English, Russian, Chinese, Arabic, Spanish, and Korean languages, requirements such as chronological coherence, transliteration, periodic quantitative norm, time

interval accuracy have been developed, and specific aspects of the principles of working with the base of linguistic corpora have been revealed;

it has been determined that the Uzbek language acquires the character of a synconic corpus from the point of view of the periodical coverage of the currently developed linguistic corpora, and the need to create a diachronic corpus of the Uzbek language has been scientifically justified due to the fact that the lexical layers of the languages are damaged along with world cultures in the process of globalization;

a diachronic corpus of Uzbek and English mediatexts, covering a thirty-year base of Uzbek "Oila va jamiyat" and English "Anaheim Bulletin" newspaper materials in chronological order, automatically performing the principles of general and annual statistical analysis in the compared languages (https://mediatextcorpus.com/) web application has been created for the first time;

based on the analysis of the diachronic corpus of Uzbek and English media texts, along with the comparative development of the conceptual frame of interlinguistic and cultural concepts, it has been practically proven that there is an opportunity to gradually observe the changing tendencies of the national, political and social approach to the linguoculturemes that are being analyzed on the basis of the compared languages;

on the example of the diachronic corpus of Uzbek and English mediatexts, taking into account the main standards of textual elements such as authenticity, representativeness and chronological ordering, the main phenomena of the development of the vocabulary layer of the languages that are being compared by means of diachronic corpuses that are perfectly processed: the possibilities of the function of recording neologisms and archaisms existence principles have been proved using practical examples.

Practical results of the research work are as follows:

based on the system of methods of corpus linguistics, the basic standards for working with sources aimed at creating diachronic corpora have been improved;

a diachronic corpus of media texts from Uzbek and English newspapers, quantitatively normalized from the content of articles published in three decades between 1991 and 2020 in the database of "Oila va jamiyat" and "Anaheim Bulletin" newspapers (https://mediatextcorpus.com/), has been created and it is available to use;

practical and theoretical principles, as well as skills for creating and perfecting diachronic linguistic corpora, have been developed.

Authenticity of the research results. The conclusions drawn from the research were expressed in the scientific works published by the researcher and these conclusions were positively evaluated by the scientific community. It is explained by the fact that it was determined on the basis of cognitive, corpus and empirical analysis methods, the suggestions and recommendations were put into practice, and the obtained results were confirmed by competent organizations.

Scientific and practical value of the research results. The scientific significance of the research is determined by the fact that the research results can be used as a scientific-theoretical source for creating diachronic corpora of the Uzbek

language, developing theoretical foundations for its improvement, and conducting research in the field of corpus linguistics.

Although the practical value of research is considered today as a methodology of scientific research, it is expected to develop as a science in the near future. It is explained by the fact that it can be a practical example of the application of existing corpuses.

Implementation of the research results. In the development of corpus linguistics, with a theoretical understanding of the study of diachronic corpora, by creating a corpus of English and Uzbek newspaper texts, based on the practical proposal and research results developed to reveal the principles of empirical analysis of language development research:

the results of the thesis as development of the requirements such as chronological coherence, transliteration, periodic quantitative norm, time interval accuracy based on the achievements of world corpus linguistics, in particular, the theoretical and practical foundations of the creation of existing diachronic corpora of English, Russian, Chinese, Arabic, Spanish, and Korean languages, and revealing the analysis related to the principles of working with the base of linguistic corpora were used in the practical project JHBL-94 of "Mahalla va oila" scientific research institute entitled "Employment of graphic designers by training them" carried out in 2022-2023 (Reference of "Mahalla va oila" scientific research institute registered with № 01-09/1607 as of September 4, 2023). As a result, the participants in the project developed scientific and practical skills in creating a diachronic corpus and systematic methods of their practical application;

theoretical views and analyzes related to the scientifical justifaction of the need to create a diachronic corpus of the Uzbek language on the basis that the lexical layers of the languages along with the cultures of the world are being damaged in the process of globalization and the Uzbek language's currently developed linguistic corpora were effectively used in the activities of the joint editorial office of the magazines of the Ministry of Culture of the Republic of Uzbekistan and in the program "Good Morning Uzbekistan" TV program on the TV channel "Foreign languages" of the National Television and Radio Company of Uzbekistan (Reference of Ministry of Culture of the Republic of Uzbekistan "Guliston" and "Teatr" magazine's joint editorial board registered with № 43 as of October 20, 2023 and Reference of the state institution of Uzbek TV and Radio Channel under the National Television and Radio Company of Uzbekistan registered with № 01-23-1774 as of November 27, 2023). As a result, the existence of the need to create a diachronic corpus of the Uzbek language was popularized with the help of editorial staff and program managers;

the created diachronic corpus of Uzbek and English media texts, covering a thirty-year base of Uzbek "Oila va jamiyat" and English "Anaheim Bulletin" newspaper materials in chronological order, automatically performing the principles of general and annual statistical analysis in the compared languages (https://mediatextcorpus.com/) web program was officially registered by the Ministry of Justice of the Republic of Uzbekistan as one of "Programs and data created for electronic computing machines" according to the Law of the Republic of

Uzbekistan "On Legal Protection of Bases" after the digital application of DGU 2023 0264 submitted on January 16, 2023 (Reference of the Ministry of Justice of the Republic of Uzbekistan registered with № DGU 21899 as of February 3, 2023). As a result, a legal basis was obtained for the guarantee of our author's rights regarding the results of the conducted research and the created diachronic corpus;

based on the analysis of the diachronic corpus of Uzbek and English media texts, as well as the comparative development of the conceptual framework of interlinguistic and cultural concepts, scientific conclusions and proposals regarding the practical proof of the possibility of gradually observing the changing tendencies of the national, political and social approach to the linguoculturemes compared were implemented within the framework of the fundamental scientific-practical project FZ-2022081663 of the State Academy of Choreography of Uzbekistan on the topic "Creation of a website and a collection of multimedia products dedicated to the promotion of the Uzbek national dance art" carried out in 2020-2021 (Reference of the State Choreography Academy of Uzbekistan under the Ministry of Culture of the Republic of Uzbekistan registered with № 01/04-61 as of June 8, 2023). The dissertation served as a scientific source for the in-depth study of the possibility of empirical observation of the enrichment level of national dance art terms in the vocabulary layer of languages on the basis of Uzbek and English newspaper texts;

the practical suggestions and recommendations on the proof of the principles of existence using practical examples on the example of the diachronic corpus of Uzbek and English media texts, taking into account the main standards of textual elements such as authenticity, representativeness and chronological ordering, and the main phenomena of the development of the vocabulary layer of the languages compared by means of diachronic corpuses that are perfectly processed: the possibilities of the function of recording neologisms and archaisms were applied within the framework of the project ID-57663348 entitled "European Perspectives in Dialogue: Central Asia and Western Research," funded by the German Academic Exchange Service DAAD (Reference of Bukhara State University registered with № 06-01-0425 as of September 25, 2023). As a result, the theoretical knowledge about the chronology of the occurrence of neologisms and archaisms in the vocabulary layer, which is important in terms of making a great contribution to the development of the lexicography of the Uzbek and English languages, was explained in the form of a seminar to the MA students and professors of the University of Gottingen, Germany.

Approbation of the research results. The results of this study were discussed in public at 5 international and 3 national scientific conferences.

Publication of the research results. 26 scientific papers on the topic of the dissertation have been published in total. 2 monographs, 1 dictionary, and 15 articles in the scientific publications recommended for publication of the main scientific results of doctoral dissertations of the Higher Attestation Commission of the Republic of Uzbekistan: 10 were published in the national and 5 in foreign journals.

The outline of the thesis. The dissertation consists of an introduction, four chapters, a conclusion and a list of references. The total volume of the work is 236 pages.

THE MAIN CONTENT OF THE DISSERTATION

The introduction of the dissertation contains such issues as topicality and necessity of the thesis, relevant research priority areas of science and developing technology of the Republic, a review of foreign research on the topic of dissertation, problem development status, relevance of the dissertation research with the plans of the scientific- research works of the higher educational institution where the dissertation has been conducted, the aim and the tasks of the research work, the objects and the subject of the research work, methods of the research, scientific novelty of the research work, practical results of the research work, authenticity of the research results, scientific and practical value of the research results, implementation of the research results, approbation of the research results, publication of the research results, the outline of the thesis.

The first chapter of the dissertation entitled "**The Role of Corpus Linguistics Methodology in the Study of Language Development**" contains analyses on the basis of world and country-level studies on corpus linguistics, the study of the enrichment of the lexical layer of the language using corpus analysis methods, media texts as the main resource of language social expression and language enrichment, and theoretical analyses aimed at revealing the functional characteristics of existing diachronic corpora.

Corpus linguistics is a branch of linguistics that studies language through the practice of creating linguistic corpora. The development of the field is closely related to the widespread use of language corpora in finding solutions to linguistic problems. It is known that corpus linguistics has a collection of methods focused mainly on lexicography, that is, lexicographic studies, but they are compatible with other branches of science related to stylistics, syntax and sentence construction, language teaching methodology, and language development research. can be used. His description of progress to date clearly demonstrates the above point.

After the transition of the corpora to the online system, the speed of finding the lexical units in search has increased and the number of users is expanding. Direct access is done through an online search through a web browser. M. Davies calls the fourth-generation concordancers hybrid corpora because their interface is a unique common area for creating a corpus at the levels of morphemic, lexical, syntactic and phraseological units and conducting analysis of words on the activity levels.

Studies have shown that the practice of creating a national language corpus for India, which consists of a multilingual population, causes a number of complications, because the constitution of this country recognizes 2 national and 22 local languages as state languages. In our opinion, the languages in the territory of this country complement each other under the influence of interaction, communication and a single media space, causing a reduction in each other's vocabulary. In our opinion, the creation of a multilingual corpus for this nation could preserve the current diversity of languages.

About this, the Indian linguist N. Dash, approving the above ideas, states: "As a multilingual country, India is a language giant. It maintains a large number of existing "living" languages of various ethnic and linguistic communities. Due to the

lack of corpora, these languages suffer from the fact that they cannot achieve their technological development. Building corpora can boost language education to improve literacy, save endangered Indian languages, and protect languages that have lost their importance from the encroachment of English. Corpora are the only linguistic bases that can help them survive the battle of linguistic imperialism. In addition, it is characterized by the fact that it can provide statistically reliable data for the recovery of the lost position"[44]. The deep consideration of the scientist is that in the era of globalization, if research is not properly oriented, in the current development of languages, due to the increase in the flow of popularized media text elements in other languages, weak languages will disappear or will be condemned to assimilate many foreign languages, such as the Mandarin dialect of the Chinese language. This issue was raised in our research. After all, based on the materials that are the objects of research, if we conclude that there are languages that are ahead of us in development and production, and taking into account the fact that their influence is expanding in the prism of media space, the practice of creating a diachronic corpus of the Uzbek language as well It is the demand of the time to set it up.

We believe that any society has a need for communication with another group of people. As this communication approaches, it affects language development in the following stages and sequences:

1. Social views harmonize with each other;

2. Religious faith and secular ideological rapprochement occur;

3. Mutual sharing of scientific, technical and household achievements and inventions;

4. Dress, customs, rituals – in a word, the interference of cultures;

5. As a result of the above, the lexical layer changes and enriches under the influence of language interference.

In general, the process of languages interference in the five-stage sequence mentioned above is accelerating today. Because the factors of the current process of communication are fundamentally different from the previous person-person and society-society categories and are being built on the basis of new, modern forms of information exchange. In our opinion, the acceleration of communication in this media field is reducing the scope of the purity and diversity of cultures and languages. Therefore, as long as the study of language development, which is the object of research, is not studied statistically based on the prism of corpora, the practice of preserving the current state of languages and leaving them as a basis for future research will not emerge. And this cannot be improved by any excuse.

Also, the linguistic features of the media text are being analyzed separately in the national dissertations defended in recent years. S.Atkamova's dissertation on phraseologism as a methical means of expression and evaluation in media text

---

[44] Dash N.S. Language Corpora: Present Indian Need. – Indian Statistical Institute, Kolkata. 2004. – P 1. Available at https://www.researchgate.net/publication/2938590_Language_Corpora_Present_Indian_Need

(based on newspaper texts of various systematic languages) [45], Sh.A.Usmonova's scientific work titled "Tendencies of newspaper language in media discourse (as an example of Uzbek and English)" [46], N.S. Nasrulloyeva's study of computer and Internet slang in Uzbek and English research on crossbreeding[47], and M.R. Artikova's research on the linguistic and cultural aspect of the phenomenon of precedent in media text (in the case of Uzbek and Spanish newspapers)[48] is aimed at the solution of linguistic problems related to media discourse. Most scholars study the media as consumers—only through various forms of production, the media as a system that communicates the end result of industry. That is, we read newspapers, magazines and comic books, watch movies and TV shows, listen to radio and music, use the Internet and play computer games. Of course, we encounter different media products in different places. come on - the song can be played on the radio, as a soundtrack to a movie, or in the background.

In the second chapter of the dissertation, entitled "**Theory and stages of creating diachronic corpora**", the theoretical and practical importance of creating a diachronic corpus of English and Uzbek media texts, the description of the consistency of the basic and normative principles and stages of their creation in the creation of a diachronic corpus were studied.

When discussing diachronic and synchronic studies and approaches, it is appropriate to consider synchronic and diachronic linguistic analysis as two complementary perspectives. The synchronic approach (from the ancient Greek: syan- "together" and chrós- "time") examines the language at a single moment in time without taking into account its history. Therefore, synchronic linguistics aims to describe, analyze and study the language under study at a certain time, that is, within the present time. In contrast, the diachronic (from dia-"through" and chrós-"time") approach looks at the development and evolution of language through history, as in historical linguistics[49].

In the Uzbek electronic dictionary, diachronic studies are explained as follows: Diachrony is the historical development of the language system as a source of linguistic research – the study of language development from the perspective of time. In linguistics, the concept of diachrony is opposed to synchrony. Diachronic linguistics is sometimes considered to be the same branch of science as comparative-historical linguistics and historical phonetics. Later, the study of changes in the language, the reasons for these changes, and the period of their appearance were considered as the object of its research. In the second half of the 20th century, the

---

[45] Атқамова С. А. фразеологизм медиаматнда ифодалаш ва баҳолашнинг услубий воситаси сифатида (турли тизимли тиллар газета матнлари асосида) // Филология фанлари фалсафа доктори (PhD) дис. автореферати. – Тошкент, 2022. – 58 б.

[46] Усмонова Ш.А. медиадискурсда газета тили тенденциялари (ўзбек ва инглиз тиллари мисолида) // Филология фанлари фалсафа доктори (PhD) дис. автореферати. – Тошкент, 2022. – 56 б.

[47] Насруллоева Н. С. Ўзбек ва инглиз тилларида компьютер ва интернет жаргонларининг чоғиштирма тадқиқи // филология фан. фалсафа доктори (PhD) дис. Автореферати. – Самарқанд, 2021 – 57 б.

[48] Артикова М.Р. Медиаматнда прецедент феноменининг лингвомаданий аспекти (ўзбек ва испан газеталари мисолида) // филология фан. фалсафа доктори (PhD) дис. автореферати. – Тошкент, 2022 – 54 б

[49] Giacalone Ramat, Anna; Mauri, Caterina; Molinelli, Piera, eds. (2013). Synchrony and Diachrony: A dynamic interface. Philadelphia, Pennsylvania: John Benjamins North America. pp. 17-18. ISBN 978-9027272072. [Retrieved on January 3, 2024]

diachronic approach focused on the study of the diachronic changes occurring in the language system, the laws of the development of the language as a system, in contrast to the purely historical approach related to the study of the history of the language divided into periods. Diachrony is a tool that allows us to know the features of language activity in time, including synchrony, and is an important part of language theory in general[50].

It can be observed that the concept of diachronic linguistics is treated as the same concept as historical/comparative linguistics in theoretical sources. It is known that comparative linguistics is a branch of linguistics that studies related languages, that is, genetically related languages, determines the relationship between them and describes their gradual development in time and space; Determining the origin of language families, certain languages and elements in these systems, including genetic kinship between languages and determining whether they originate from a single source (ancestral language) (genealogical classification of languages), is also part of the goals of comparative / historical linguistics[51].

Based on the scientific views on genetic classification, which is one of the main tasks of historical linguistics, it is appropriate to consider the type of classification of the languages being compared. English belongs to the group of Germanic languages belonging to the Indo-European language family. The Uzbek language belongs to the Turkic language family, and according to the second view, this family belongs to the Altai language family. More precisely, the Uzbek language is included in the Qarluq group of the Turkic language family, and the closest language to the Uzbek language in this regard is the new Uighur languages[52].

Therefore, the inclusion of the diachronic term in corpus linguistics does not overshadow its substantive and functional characteristics. That is, a diachronic corpus is a corpus containing texts from different periods and is used to study language development or change. In this regard, According to scholars such as M. Hilpert and S.Th. Gries, a diachronic corpus refers to textual resources representing similar types of language use at successive time intervals, as in the case of the diachronic corpus of modern spoken English, the characteristic of which is that the corpus takes at least two time periods[53].

Therefore, the concept of diachronic corpus is of great importance in the field of corpus linguistics, and it has created a new stage in the use of corpus analysis methods. In our opinion, by using diachronic corpora in the following areas of linguistics, it is possible to approach the scientific-theoretical questions in a new way:

---

[50] https://qomus.info/oz/encyclopedia/d/diaxroniya/ [murojaat sanasi 31.01.2024]

[51] Tomsen V., Istoriya yazmkoznaniya do konsa XIX veka, M., 1938; Meye A., Vvedeniye v sravnitelnoye izucheniye indoyevropeyskix yaznkov, 3 izd., per. s frann., M. — L., 1938 available at https://openlibrary.org/books/OL16367943M/Istoriya_yazykovedeniya_do_kontsa_XIX_veka [retrieved on December 5, 2023]

[52] Irisqulov A.T. Tilshunoslikka kirish. / Darslik. – Toshkent: "Yoshlar matbuoti" bosmaxonasi. 2009. – B. 185-187

[53] Hilpert M. and Gries S.Th. Quantitative approachesto diachronic corpuslinguistics / The Cambridge Handbook of English Historical Linguistics // - Publisher: Cambridge University Press, 2016. – P. 36 Available from https://www.researchgate.net/publication/325893284_Quantitative_approaches_to_diachronic_corpus_linguistics

- firstly, diachronic corpora are used in the field of lexicology, in which they can provide not only chronological information about the use of words, but also look at their etymology, depending on the systematic selection of the source in the corpus;

- secondly, diachronic corpora are used at the same time as the concept of historical corpus and pay special attention to the study of language development. In particular, it offers accurate statistical solutions to scientific problems such as neologisms, archaisms, historicisms, which are part of the timeless problems of lexicology;

- thirdly, the use of diachronic corpora in the field of gender linguistics is one of the existing practices in world linguistics. In this place, historical corpora are collected separately from the works of representatives of the opposite sex, and by forming periodized corpus bases, the life philosophy, worldview, and emotional feelings of female or male writers are reflected in their artistic works, correspondence, newspaper or magazine articles. It is possible to determine by linguostatistical analysis based on samples of interviews and conversations;

- fourthly, the practice of creating diachronic corpora is important for source scholars, translators, poets and writers who want to write historical works. In this way, the above-mentioned language lovers can feel the true spirit of the culture, nation and language of the period under study by using the active lexical layer used in the period covered by the diachronic corpus, and preserve this spirit by applying it to publishing work remains;

- fifthly, and most importantly, in today's rapidly developing world, "strong" peoples and their languages will survive. Expressing a more extreme opinion, we think that by creating diachronic corpora through the statistical analysis of lines of interlinguistic interference, it is possible to understand the importance of developing state programs aimed at keeping the speakers of the language as far away from this flow as possible.

There are many scientific hypotheses for the use of such diachronic corpora. In our opinion, if diachronic corpora are used rationally and if their composition is correctly formed according to the established norms, based on the analysis of this type of corpus, it is possible to analyze the languages of the world based on their historical connection, political and territorial proximity, economic dependence, etc., the possibility of reclassification arises culturally and socially.

At this point, taking into account that the English language is leading the way in the practice of creating diachronic corpora, listing the corpora of this type created in this language, taking into account their size and characteristics, the need to create a historical or diachronic corpus of the Uzbek language is clearly in the plan. Today, according to martinweisser.org, it is important to create diachronic corpora with the size number normalized to the period of creation of the corpora. Some of the English diachronic corpora listed on the site are listed below[54] (see table 1).

---

[54] *Historical & Diachronic Corpora* Available from http://www.martinweisser.org/corpora_site/historical_corpora.html [Retrieved on 04.02.2024]

Table 1. English historical/diachronic corpora

ARCHER Corpus (A Representative Corpus of Historical English Registers)	It contains a total of 955 texts, i.e. 1,789,309 words, formed from written and "speech-based" genres from 7 historical periods between 1650 and 1990, covering examples from Early Modern English.
Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English	The Old English section of the Helsinki Corpus of English texts contains 106,210 words from Old English sources, a sample of longer texts between 5,000 and 10,000 words in length.
Century of Prose Corpus	It consists of half a million words of literary and non-literary English published by 120 authors who lived and worked between 1680 and 1780.
Complete Corpus of Old English	A corpus representing 3,022 texts from all Old English periods compiled at the University of Toronto.
Corpus of English Dialogues, 1560-1760 (CED)	The corpus consists of 177 texts representing early modern English speech forms and contains 1,200,000 words.
Corpus of Newsbooks	A text of approximately 800,000 words taken from all of Thomason's papers published between December 1653 and May 1654.
Corpus of Early English Correspondence (CEEC) & the Parsed Corpus of Early English Correspondence (PCEEC),	The corpus consists of a total of 2,700,000 words, including 450,000 words from 1410 to 1681; "Corpus of Early Correspondence" from 1402-1663 contains about 50,000 words; A total of 2,200,000 words were included from the period 1681-1800.
Corpus of Early English Medical Writing & Corpus of Middle English Medical Texts (MEMT)	It is a corpus of medical publications and manuscripts of the years 1375-1800, filled with treatises of 10-12 thousand words, and the medieval section contains about 500 thousand words.
The Málaga Corpus of Early English Scientific Prose	Historical diachronic corpus containing medical materials developed in 1350-1900. The corpus contains approximately six million words divided into three subcorpora spanning a period of almost six hundred years.
Newdigate Newsletters	It contains a total of 750,000 words from manuscript publications from 1674-1692.
TIME Magazine Corpus	A corpus of 100 million words from TIME magazine publications from 1923-2006. The use or decline of words and phrases over time provides information that allows you to see whether their meaning has changed.
Zürich Corpus of English Newspapers (ZEN)	It consists of London newspaper publications from the 1660s to the beginning of the 20th century.

Recognizing the above-mentioned linguistic corpora as achievements of English corpus linguistics, it should be noted that a number of scientific studies are being carried out in this regard within the framework of the Uzbek language. Among them can be included:

1. The uzbekcorpus.uz electronic corpus, which is considered as the first linguistic corpus of the Uzbek language, was created by Doctor of Philology (DSc), Associate Professor N.Z. It is the result of several years of research, scientific cooperation and project results by members of the project led by Abdurahmonova. The current state of the Uzbek language corpus consists of a complex of linguistic dictionaries, web pages, a morphological database of the Uzbek language, educational literature and scientific, official and artistic texts of various genres. A corpus is a collection of spoken and written texts, software with the ability to manage and use data. This corpus also includes an internal subcorpus: the parallel corpus, the author corpus, the educational corpus[55].

2. The educational corpus of the Uzbek language is available at the link https://uzschoolcorpara.uz/. Educational corpus is equally important in learning a mother tongue and a foreign language. In language teaching, the corpus is very useful for showing the size of the vocabulary and for showing the array of examples for explaining the possibility of using a word through one or another grammatical construction. Only a corpus has the ability to reflect this constant updating of examples, which is important for language learning. Here, the teacher can find new, convincing, unlimited and diverse examples, it is not difficult to set assignments and exercises, he can prepare assignments with new examples on the topic in a few minutes[56]. The authors of the system include a PhD, Assoc. Prof. B.B. Elov, DSc, Prof. B.R. Mengliyev, PhD, Assoc. Prof. Sh.M. Hamroyeva and others.

3. *The Uzbek community corpus, created on the basis of sources collected in 2017* was compiled on the basis of the project of the University of Leipzig, the Saxon Academy of Sciences and Humanities and the Institute of Applied Informatics in Leipzig. It includes a total of 9 million 256 thousand words in 663 thousand 119 sentences[57]. Due to the fact that this corpus is compiled from the contents of automatic internet materials, the alphabet, which is one of the problems of the Uzbek language today, that is, the texts uploaded on the Internet pages are entered in two different Latin and Cyrillic graphics.

4. The national corpus of the Uzbek language as a software product was implemented within the framework of the practical project of the 30th round of 2021, "Designing the national corpus of the Uzbek language and developing a software package" (2020-2022) of the Ministry of Innovation of the Republic of Uzbekistan. According to the results of the competition held in the established order, a contract was concluded with the Fund for the Development of the Uzbek Language under the

---

[55] https://uzbekcorpus.uz/aboutCorpus [murajaat sanasi 02.02.2024]
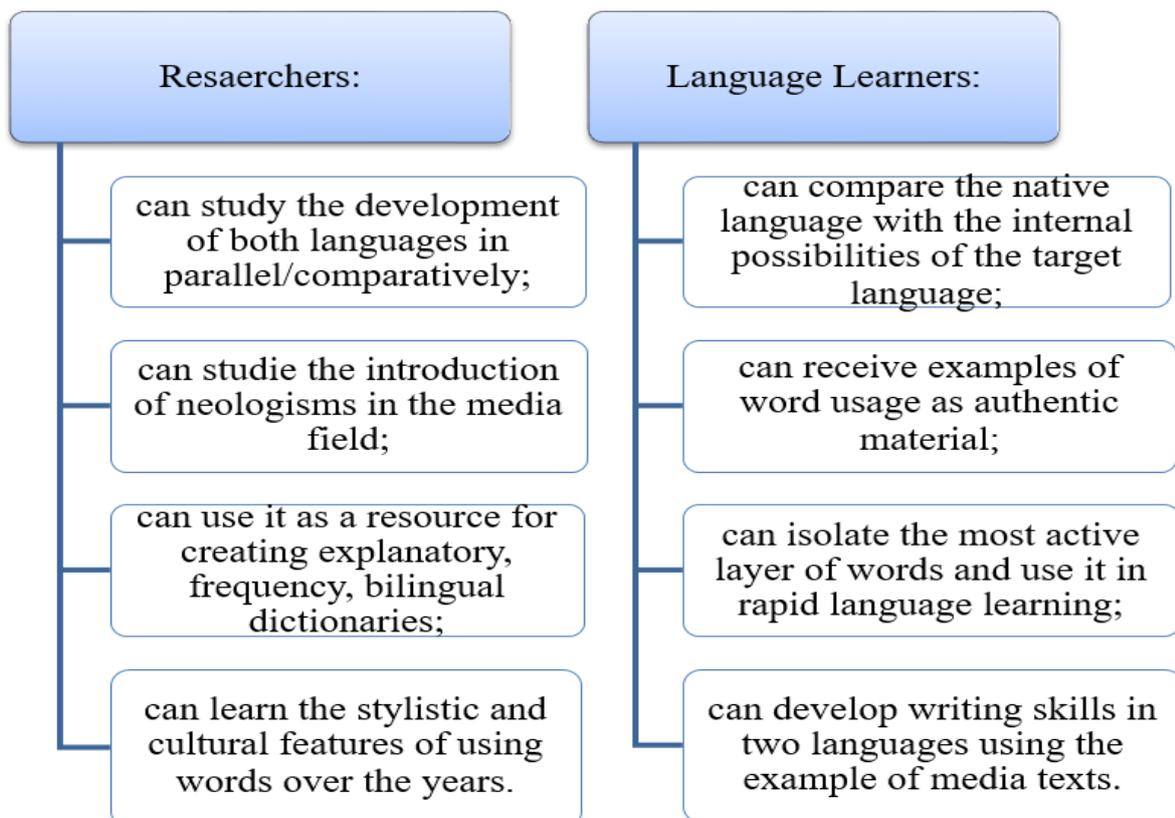[56] https://uzschoolcorpara.uz/#home [murajaat sanasi 02.02.2024]
[57] *Leipzig Corpora Collection:* Uzbek community corpus based on material from 2017. Leipzig Corpora Collection. Dataset. https://corpora.uni-leipzig.de?corpusId=uzb_community_2017 [retrieved on 02.02.2024]

Cabinet of Ministers of the Republic of Uzbekistan for the implementation of this project[58]. This corpus is available at the link http://uzbekcorpora.uz/.

All the corpora of the Uzbek language listed above are not among diachronic corpora. On the contrary, since the corpora are not chronologically arranged, and the texts are not included in the historical sequence, they can be said to be in the column of synchronic corpora.

The need for the creation of a diachronic corpus of the Uzbek language is inextricably linked to the above description of the corpus of the Uzbek language that exists today. That is, the absence of monitor corpora designed to monitor the enrichment of the Uzbek language today makes researchers question the reliability of their conclusions in this regard. After all, the diachronic corpus of the language provides accurate statistical information on the development of the vocabulary, aggregating sources in accordance with the periodic changes of the language.

Here, using the existing experience, research work is being carried out as a theoretical basis for the creation of a diachronic corpus of the Uzbek language. In this case, the goal of our work is to develop the diachronic corpus of bilinguals, that is, Uzbek and English. As the source of written texts can be found on a large scale, media communication means are considered, in our opinion, it is permissible to form the core of our corpus to be created on the basis of media texts. Therefore, the accumulation of the intended corpus can create the following opportunities for language scholars and language learners:

Resaerchers:
- can study the development of both languages in parallel/comparatively;
- can studie the introduction of neologisms in the media field;
- can use it as a resource for creating explanatory, frequency, bilingual dictionaries;
- can learn the stylistic and cultural features of using words over the years.

Language Learners:
- can compare the native language with the internal possibilities of the target language;
- can receive examples of word usage as authentic material;
- can isolate the most active layer of words and use it in rapid language learning;
- can develop writing skills in two languages using the example of media texts.

Picture 1. The relevance of the diachronic corpus of English and Uzbek media texts in the scientific and practical field

---

[58] http://uzbekcorpora.uz/buyurtmachi [murajaat sanasi 02.02.2024]

Diachronic/historical corpora continue to grow in number and size, and the methods of their analysis are becoming more efficient and sophisticated. At the same time, except for complex diachronic corpora, diachronic data always present important problems to the researcher.

In order for diachronic corpora to be recognized as a useful resource for users, it is necessary to form them correctly. In our opinion, the above-mentioned requirements for the source can be introduced in the collection and sorting of corpus databases of a diachronic nature. However, for diachronic corpora, it is appropriate to take into account some different aspects of source collection and its inclusion in the corpus. We consider them as follows:

1. In diachronic corpora, unlike other types of corpora, it is appropriate to collect texts in a chronological order, from historical to modernity;

2. Among the languages of the world, many languages in their historical development have undergone alphabetic changes in their writing graphics as a result of political instability or some historical reality. When compiling the ancient or historical writings of such languages into a corpus, it is necessary to correctly and precisely transliterate the texts. Due to the fact that the Uzbek language has changed its writing graphics not once but three times in its recent history, the same situation is observed in the introduction of sources in the diachronic corpora collected today, and the solution will be appropriate if it is as stated above;

3. Diachronic corpora cannot always follow the norm of including texts in equal volume due to the fact that the same source does not exist in all periods from the point of view of history. However, if the corpus being created is intended to cover recent history, that is, the period of 20-30 years ago, then due to the unlimitedness of the source, it is necessary to comply with the requirement of the quantitative limit, that is, the obligation to include equal amounts;

4. Before the creation of any diachronic corpus, the periodicity determined by the researcher should cover at least a decade, rather than a short period. Otherwise, there is a possibility that the results of research aimed at language development will not be high.

The third chapter, entitled **"The practice of creating a diachronic corpus of media texts of Uzbek and English newspapers"** is aimed at revealing the mechanisms of working with textual sources in the creation of a diachronic corpus, the basics and sorting chronology, automation and design problems of the corpus base, as well as the description of the created corpus evaluation and frequency search-analysis results.

It can be observed that the methods of corpus linguistics are used in Uzbek linguistics, especially in lexicography, but it is not noted that they are based on corpus analysis. For example, it is said that during the work on the creation of a frequency-annotated phraseological dictionary of Uzbek folk epics, textual units were collected, entered into the computer memory, and alphabetic-frequency dictionaries were obtained: "... in the process of creating the dictionary, first Alpomish, The texts of the epics "Ravshan" and "Rustam" were entered into the computer memory. On the basis of these texts collected in the computer memory, dictionaries of 28,499 word forms, three frequency and three alphabet-frequency

dictionaries were obtained: the total text volume of the epic "Alpomish" is 14,029, i.e. – 49.2%; the "Ravshan" epic – 7059, that is, 24.7%, "Rustam" epic – 7411 word forms – 26.2%. During the re-examination of the texts, it was found that a total of 1041 FBs were used in these texts, and it was found that they were returned 1641 times in the text of folk epics[59]. As it can be seen, there is no mention of which electronic or computer software was used. However, this process belongs to the series of primary tasks specific to the creation of corpora, which is the main goal of corpus linguistics, to quickly obtain statistical information about the linguistic base.

In our opinion, taking into account the above situation, the science of Uzbek linguistics requires new approaches to the application of corpus-based theoretical research in practical aspects. Therefore, creating a diachronic corpus of media texts from Uzbek and English newspapers, which is among the corpora to be created, is one of the results of this type of scientific research.

According to scholars such as J. Sinclair[60] and F. Douglas[61], because some of the texts included in the corpus are long, rather than involving their full size, the use of text fragments is sometimes used so that only one type of text is not dominant in the corpus. However, they conclude that there are many approaches to the superiority of using full texts. In our opinion, since it is rare in practice for the opinions of scientists to be reasonable and the texts to be equal in size, the inclusion of the full form of the text included in the corpus is the basis for its correct analysis in terms of content and logic.

It can be observed that many corpora being created today, in particular, modern corpora of the Uzbek language, are formed mainly online, that is, from Internet materials. In our opinion, when creating a corpus, taking into account the expected result, it is permissible to pay attention to the reliability, authenticity of the source and, most importantly, its preservation for tomorrow.

Based on the above considerations, the issues of "Oila va jamiyat" and "Anaheim Bulletin" newspapers published between 1991 and 2020 were selected as the object of our research. Through this, a basis was formed that could be recognized as materials that could be based on the principles of reliability, representativeness and authenticity. In this case, the texts taken from the 30-year editions of the Uzbek "Oila va jamiyat" and the English "Anaheim Bulletin" newspapers can be listed as the basis for choosing them as sources:

First, the above-mentioned 30-year period, i.e., the period 1991-2020, for the Uzbek language, according to the law of October 21, 1989, "On the State Language of the Republic of Uzbekistan," has the status of a state language. The revolution after the grant, that is, the first three decades of the independence of the state, which is considered to lead to the most effective development of languages, is relevant. It

---

[59] O'zbek xalq dostonlarining chastotali izohli frazeologik lug'ati /B.Yo'ldoshev, D.O'rinboyeva; Mas'ul muharrir B.O'rinboyev; O'zbekiston Respublikasi oily va o'rta maxsus ta'lim vazirligi, Alisher Navoiy nomidagi Samarqand Davlat universiteti. – Samarqand, 2013. – B. 4 https://elib.buxdupi.uz/books/34891.pdf

[60] Sinclair J. Corpus and text - basic principles. In M. Wynne (Ed.) // Developing linguistic corpora: A guide to good practice. Oxford: Oxbow Books. 2005. – pp. 1-16

[61] Douglas F. M. The Scottish corpus of texts and speech: Problems of corpus design // Literary and Linguistic Computing, 18(1). 2003. – pp. 23-37.

was aimed at working with the texts collected from the newspaper publications of the same period of the English language.

Second, Family and Society and Anaheim Bulletin were selected because they are relatively little-studied weekly newspapers, have similar publication dates, and most importantly, are social newspapers. More specifically, the Anaheim Bulletin newspaper has been published since 1966 as an independent newspaper in Anaheim, California, the most populous city in the United States. In its history, the newspaper has been published once a week since August 6, 1992, and since September 5, 1991, the newspaper "Family and Society" has been published every Thursday. The existence of periodical closeness and the columns devoted to science, culture, art, production, entrepreneurship, agriculture and several other fields that they raise have shown that they are newspapers suitable for mutual comparison.

Third, the availability of archives of Family and Community and Anaheim Bulletin newspapers from the selected time period, despite the fact that older editions of both newspapers are not computer readable, justified the selection of these sources.

Both databases were obtained from the official sites https://newspaperarchive.com/ and http://press.natlib.uz/ru/publications, which contain scanned copies of newspapers. The fact that all information on these sites is open to users and the absence of restrictions indicates that research is not in violation of the principles of copyright. The norms for working with texts were implemented in three stages in the following order:

Step 1: Available scanned newspaper publications were collected from the official websites, https://newspaperarchive.com and http://press.natlib.uz. In this case, the existing editions in their large volumes were randomly sorted based on the readability of the texts;

Step 2: The compiled and chronologically arranged database was retyped one by one by hand using Microsoft Word and saved as *.doc extension documents. In these types of documents, information from both languages was placed in specially developed tables, and separate folders summarizing these tables were created and kept. We found it necessary to impose a requirement to create databases of +/- 5000 words for chronologically collected sources. After all, the presence of a quantitative norm proves the reliability of statistical information showing the level of activity in the cross-section of years related to the use of the searched word.

Step 3: The collected source, which has passed the above-mentioned steps, was compiled in Microsoft Excel in two separate *.xls files in the section of the languages being compared, in a vertical sequence and made suitable for uploading to computer memory or web pages. When viewed with the help of this program, it was found that the total volume of aggregated texts was 316 thousand 602 word resources collected from both databases. In the current case, the 30 years of material collected from the Anaheim Bulletin from the August 2, 1991 issue to the December 20, 2020 issue totals around 5,000 words and more each year. (due to the fact that we received the article in its entirety) a total of 158819 words were collected.

As it can be seen, the collected materials were processed, sorted, statistical primary information was obtained, and the fact that they constitute 158819 and

157 thousand 783 words, respectively, shows that the sources collected from the two compared languages are not significantly different from each other. . This means that comparative statistical analyses based on the corpus base can be accepted as reliable and truthful information.

Diachronic Corpus of English and Uzbek Media Texts created by us on the basis of newspaper texts collected from the 30-year database of Family and Society and Anaheim Bulletin newspapers is now available at the link https://mediatextcorpus.com. Below is a handy QR code form designed for modern electronic gadgets to grab the aforementioned link by scanning easily (see Picture 2).

Picture 2. QR code for the link to the diachronic corpus of English and Uzbek media texts

The fourth chapter of the dissertation, which consists of three general chapters, is titled as "Practical significance of the diachronic corpus of media texts of Uzbek and English newspapers" and includes the statistical foundations of the active lexical layer in the diachronic corpus, including the high frequency of the corpus changing over the years. It summarizes the results of practical research on the analytical possibilities of words and the variety of forms of words, the relationship between the root and morphological unity, the comparison of linguistic and cultural concepts in the diachronic corpora of English and Uzbek media texts, the analysis of cross-linguistic styles, neologisms and social concepts.

The tables automatically presented by the generated corpus show that the definite article was used as the most active word in all years of the 30-year base in English. In total, it was used 9,434 times, which is about 4.69% of the total number of words. Other types of the article, in particular, the ambiguous form a, were used a total of 3612 times, and an was used 595 times. Considering that they come before singular nouns depending on whether they start with a vowel or a consonant, respectively, this suggests that most of these types of nouns in English start with a consonant. In the statistics automatically extracted from the Uzbek language database, it can be observed that the most actively used words differ over the years. Among them, the word type is the most active word for 12 years, and the word type is the most active word for 12 years, and the word type is the most active word for 5 years, and the word type is the most active word in a total of only one year is recorded as in the general calculation, it was found that the word was used 2054 times and it was the word with the highest frequency, and it was determined that its proportion to the total volume of the Uzbek base is 1.01 percent. Therefore, it can

be concluded that the words taken from both languages have a large share of the total base amount, and that they are widely used in the languages being confused.

The table shows that a number of prepositions such as to, of, in, for, on, with are used with a higher frequency than other words of this type. The most actively used part of them is the type of word to, which is used 4290 times in total, and is the second most active word in the English language base with a ratio of 2.13%. In our opinion, the large number of functional tasks of a lexical unit increases its level of activity. According to the Macmillan English dictionary[62], the word TO has two cases, i.e., prepositional and adverbial, and there are a total of 21 different meanings and functions. In particular, it is used as a preposition indicating the direction, in the infinitive form of the verb, in various verb combinations such as going to, ought to, used to, telling the time and in other places. It turned out that the corpus is a base that can prove the above points to be completely reasonable.

When we pay attention to the analysis of the corpus of the English language, it is among the most active words in the past tense (past simple) and is, have, has, are (present simple) verbs. We can see that it is done. This is also observed in the case of the Uzbek language, it can be seen in the example of the verb forms in the past tense, past participle, and present tense forms such as say, say, be, take, have, and have. Considering that the texts from both languages are newspaper articles, it is reasonable to realize that they all convey information through the forms of two main time categories.

Among the most active of the English and Uzbek languages, you can see word types belonging to the pronoun group. For example, we can see that personal and demonstrative pronouns such as *I, it, her, he, they, their, this, we* are among the 20 most active words with high frequency in the English text base. Taking the example of the Uzbek language, it is possible to cite such words as *shu, bu, u, uning, men, ular, biz, siz, meni, ularning*. If we analyze the form changes of the pronoun мен / I separately from their composition, we have the following result*: I (885), me (127), my (183), mine (2), myself (10); men (466), meni (162), menga (148), menda (5), mendan (32), o'zim (28), mening (117), meniki (1)*. In both cases, it was found that the forms of the pronouns in the main agreement are more active and have a lower frequency than the types of pronouns that express the meaning of the mentioned noun without a noun.

Anaheim, school, community, council, center, students, scored, three, abdominal, Bayless, high, city, orange, Ryans from the English language base, and *sog'lom, muhabbat, paranji, onam, Abbos, bolalar, Xudoyberdi, Sobir, momo, respublikasi, oila, oilalar, O'zbekiston, nos, Ozoda* from the Uzbek language base are included in this list of words belonging to common and popular nouns, and the amount of the base included in the corpus is relatively small, and it is characterized by the fact that a limited number of articles in the database depend on the characteristics of their content. If the amount of this base was many times larger than it is now, we think that these words would not be among the most actively used

---

[62] Macmillan English Dictionary for Advanced Learners. edited by Rundell M. and others. – Macmillan Publishers Limeted, 2007. – P. 1573-1574.

words. As a basis for this, it is possible to cite the fact that they stand out among words with high frequency in a database of only one or two years in a three-decade database.

Searching for the word *Navro‘z*[63] in the corpus revealed a total of 27 examples. When collocational units were separated from their composition, the following active compound sentences were identified: *Navro‘z arafasi; Navro‘z ayyomi; Navro‘z bayrami; Navro‘zi olam; Navro‘z kunlari; Navro‘zi Ajam.* All this is reflected in examples of the correct use of sentences related to Nowruz. All of the above examples are understandable and are known to be actively used by native speakers today. However, the phrase *Navro‘zi Ajam*, mentioned last and used together in 3 places in the corpus, attracted attention. In a broader context, it was found to be used in the following sentences: 1) In S. Saidov's article entitled "*Jimlik*" published in 1992, it can be seen the following example: "*To‘g‘ri, har bir san'atkordan "Navruzi Ajam"ni talab qilib bo‘lmaydi, lekin... San'atda shaldir-shuldir ko‘payib ketdi...*"[64]; 2) Sirojiddin Sayyid's article in 1992 titled "Shalikor" village council building shows that it is used in two places. First is given as "*Mirzo Boburning "Bahor ayyomidurdagi yigitlikning ayvonidur" satri bilan boshlanguvchi nozik g‘azalini "Navro‘zi Ajami" satriga solib aytganida, umrning, tiriklikning go‘zal salomi yanglig‘ bitilgan bu g‘azalning qadimiy kuy pardalari ichra vido yanglig‘ zoru nigoron yangraganida qanday hikmat bor?* "; while the second is given as the following: "*Navro‘zi Ajam" ulug‘ kuy, zor ko‘ngilning mangu, tuganmas navhasidir u. Undagi yig‘ini so‘zga ko‘chirish qiyin*". Through examples taken from this broader context, the corpus user can understand that *Navrozi Ajam* is an ancient Uzbek folk music.

In the next place, we found it permissible to give an analysis of the examples of the expression of the *Christmas holiday*, which is characteristic of the English culture, in the minds of the speakers of this language. When the word *Christmas*[65] was searched in the diachronic corpus of media texts, it was found that it was used a total of 24 times. When the collocational units combined with this word type were separated, the following results were obtained: *Christmas project; Christmas cards; Christmas services; Christmas Eve; Christmas tree; Christmas Party; Christmas time; Christmas Day.*

Looking at the above, it is not difficult to understand that all of them refer to common cultural concepts. Among them, the combination of *Christmas services* seems to require additional explanation. When viewed in a broader context, the article is marked with the number 681 in the corpus, and in 2011. It turns out that it was taken from the text under the title *Toy drive planned tonight for babies* written by Sh. Nlapcic. "*Any donations are appreciated. CANYON HILLS PRESBYTERIAN PLANNING CHRISTMAS SERVICES Canyon Hills Presbyterian Church will offer two Christmas Eve worship services, one at 5 p.m. and a candlelight service at 11 p.m.*" – when looking at the content of the text, it is said that it is considered a part of a religious holiday for representatives of a fixed society in the Christian faith,

---

[63] O‘sha manba https://mediatextcorpus.com/search?q=navr&language=2&match_type=0&year=

[64] O‘sha manba https://mediatextcorpus.com/article/46

[65] O‘sha manba https://mediatextcorpus.com/search?q=christmas&language=1&match_type=1&year=

and donations given as gifts to children who sing religious songs in the choir visiting the holiday are accepted. It is also envisaged that the service process will be divided into two. The first of them is *a Christmas Eve worship service, and the second is a candlelight service.* Through this, it was determined that public associations were formed that offer services for coordinating organizational matters in accordance with popular holiday traditions. In our opinion, this type of services may cease to exist or develop in the future, but the fact that its linguistic reflection is stored in newspaper text archives explains the fact that the corpus is based on media texts.

CONCLUSION

1. Corpus linguistics, a new branch of linguistic science that emerged at the beginning of the second half of the last century, is characterized by its own research methodology and is widely used in solving problems related to modern science, in particular, anthropocentric studies. Its object is to approach the linguistic corpus as an electronic linguistic database that is formed on the basis of authentic materials taken from the original language, responds to the principle of representativeness, provides a practical solution to scientific hypotheses related to language, and obeys the requirements of normative volume.

2. The language, which is considered a social phenomenon, does not stop its continuous development, and as a result of the rapid convergence of nations in the era of globalization, its current state is fundamentally different from the previous and future states. With a correct assessment of such a dangerous situation, it becomes urgent to record and form national synchronous corpora from the representatives of the layers of languages that have not yet been affected by the global Internet and information attack. After all, if this process continues continuously, the linguistic base created over time will have historical value, and it will reach the status of a diachronic corpus at the next stage.

3. Diachronic corpora envisages the creation of mainly written texts created in the historical period clearly defined by the content of the language it represents, because collecting materials related to oral speech and working on them becomes relatively complicated. That is, it is difficult to find an oral source specific to the stages of human development before the development of the historical reality radio and television network. However, it is possible to reach a reliable source that reveals the development trends of the language by recording various forms of the language layer, slang or dialect, samples of folk oral creativity prone to interlinguistic interaction. In this case, there is a need to include this form of speech in the diachronic corpus in accordance with the rules of transcription.

4. In the practice of creating any diachronic corpus, the corpus interface adapted to the expected goal, i.e., the study of the gradual development of languages, necessarily includes a search system, a registration corner, a list of active words in the source and statistics of It must include open information, the possibility of getting to know the sources more widely, the links to the data of individual chronicles for a series of years, and the possibility of obtaining statistical and chronological information about each included text.

5. Since the creation of diachronic corpora is an example of long-term development, relying on the collection of sources with an interval of at least a decade, the creation of its user-friendly interface and design must be the ultimate goal of corpus creators. Even a small detail in the design may be overlooked, and the product of many years of work may not come to the surface. Therefore, when creating diachronic corpora, it is necessary to introduce functions to improve their concordance results, to provide them in a chronological order with periodic differentiation, with systematic continuity, or to limit them due to the required period.

6. The general principles applied to corpus materials include authentic material of the source; equal accumulation; not subject to editorial processes such as any changes or corrections; total lexical volume should not be less than one million; compliance with the principle of representativeness; it was determined that it is appropriate to include such things as proportionality in time and space. The following four requirements are the specific principles specific to diachronic corpora: 1) collecting texts in chronological order; 2) transliteration of ancient or historical inscriptions of languages that have undergone alphabetic changes in writing graphics following historical reality; 3) a relative approach to the standard requirements for entering texts in an equal volume; 4) to cover a period of at least 1 decade.

7. Taking into account the inclusion of diachronic corpora into the type of monitor corpora, it is necessary to introduce a strict requirement for the basic quantity included in it and the requirement to have textual units that tend to be formed in sync with the times while recording all the facts related to the enrichment of the language. In this regard, it is impossible to put into practice the use of materials that are read daily on the popular Internet networks today, due to their variability, that is, the fact that the text can be constantly edited, which calls into question the specificity of the information at the time of its entry. Therefore, the use of newspaper texts, which are printed on paper or published online and are created in compliance with all language norms, and most importantly, as an integral part of media communication, which continuously illuminate the social, national and cultural lifestyle of the speakers of the language, as a source for diachronic corpora. justifies it.

8. The methodology of improved corpus linguistics consists of methods based on the following system, divided into six stages: reference methods, working with the source or basic methods, methods of adding metadata, automation methods, evaluation methods, frequency search-analysis methods. Each of these stages is carried out sequentially, based on systematic coherence, and it is appropriate to use it in the creation of all types, in particular, diachronic corpora.

9. If it is not based on the strict principles of collecting, sorting and automating the process of collecting, sorting and automating textual units related to a certain historical stage of the language, observing the principles clearly defined in the formation of the diachronic corpus of media texts, the corpus cannot achieve its goal. Taking this into account, it is not appropriate to include samples of fiction in the diachronic corpus. After all, the purposeful use of historical, archaic or neologisms

by the author to give the spirit of the time and space represented by the work in the artistic text can cause results that are disproportionate to diachronic periodization.

10. Based on the linguistic base represented by the diachronic corpus of media texts of the English and Uzbek languages, analyses aimed at a deeper understanding of the culture of the same languages, including statistical and vocabulary prompts that serve to reach theoretical conclusions about the presentation of linguistic and cultural units in newspaper texts, provide clear, convincing examples. In this case, the researcher is given the opportunity to understand the essence of the concept or concept to be searched for and to make a comparison within the framework of the compared languages, as well as to create their conceptual frame.

11. Diachronic corpus in media texts of English and Uzbek languages is automatically calculated from the content of selected newspaper texts specific to the period between 1991-2020, covered by it, in order of decreasing words according to the level of activity. This situation allows the researcher to identify obsolete historical and/or archaic lexical units that have entered the vocabulary of the language in accordance with the principles of empirical analysis. Although the current size of this corpus base does not allow us to make final conclusions about the languages being compared, we believe that it can be reasonably achieved due to the expansion of the developed project.

12. Relying on the linguistic base represented by the diachronic corpus of English and Uzbek media texts, analyses aimed at a deeper understanding of the culture of these languages, including statistical and lexical skills that serve to come to theoretical conclusions about the presentation of linguistic and cultural units in newspaper texts, provide clear, convincing examples. In this case, there is an opportunity to understand the essence of the concept or concepts to be searched for and to create a conceptual frame for the compared languages.

13. The dichronic corpus of media texts in English and Uzbek languages is automatically calculated from the content of selected newspaper texts specific to the period between 1991 and 2020, which it covers, until the level of activity of words decreases in a series of years. This situation allows the researcher to identify new, historical, or archaic lexical units that have entered the vocabulary of the language in accordance with the principles of empirical analysis. Although the current size of this corpus base does not allow us to draw final conclusions about the languages being compared, we believe that this can be reasonably achieved as a result of expanding the size of the developed project.

14. A diachronic corpus of media texts provides concrete examples of word combinations, statistical information, and text fragments to understand the contextual use of collocational units. Based on them, when creating the semantic field of the concept studied in different systematic languages, it is possible to create not only the semantic groups of the word but also to gradually determine the historicisms related to the chronological contiguity of word combinations and collocational units. In this case, it is possible to precisely chronologically observe the level of activity related to the stem-suffix relationship of linguistic units, including morphemes, which are prone to conjugation to the lexeme included in the

concordance search, which provides an opportunity to gradually study the tendency of the word to change its form and content.

15. It is appropriate to acknowledge that electronic language resources uploaded on the Internet, including the creation of linguistic corpora, are among the factors that negate the hypothesis of the disappearance of languages. After all, the scientific research created on the basis of research related to linguistics should be directed not to touch the elements of the existing language but to collect the linguistic bases reflecting the time and space aimed at analyzing its features. The creation of a diachronic corpus of media texts of the English and Uzbek languages, which is a practical result of the dissertation, is characterized by the fact that it shows a quantitative reflection of the continuous development of the languages during the envisaged period rather than an assessment of the current state of the languages.

БУХАРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

АТАБОЕВ НОЗИМЖОН БОБОЖОН УГЛИ

ПРИНЦИПЫ ЭМПИРИЧЕСКОГО АНАЛИЗА РАЗВИТИЯ ЯЗЫКА В ДИАХРОНИЧЕСКОМ КОРПУСЕ МЕДИАТЕКСТОВ
(на примере газетных материалов)

10.00.11- Теория языка. Практическая и компьютерная лингвистика

АВТОРЕФЕРАТ
диссертации доктора философии (PhD) по ФИЛОЛОГИЧЕСКИМ наукам

Фергана – 2024

Тема диссертации доктора философии (PhD) по филологическим наукам зарегистрирована под номером B2022.3.DSc/Fil465 в Высшей Аттестационной комиссии Республики Узбекистан.

Диссертация выполнена в Бухарском государственном университете.

Автореферат диссертации на трех языках (узбекском, английском, русском, (резюме)) выложен на веб-странице Ферганского государственного университета по адресу: https://fdu.uz/ и в Информационном портале "Ziyonet (www.ziyonet.uz).

Научный консультант:	**Тешабаева Дилфуза Муминовна** доктор филологических наук, профессор
Официальные оппоненты:	**Султонова Шохиста Мухаммаджановна** доктор филологических наук (DSc), профессор
	Абдурахмонова Нилуфар Зайнобиддин кизи доктор филологических наук (DSc), профессор
	Марио Биснада Доктор филологических наук (Испания)
Ведущая организация:	**Ташкентский государственный университет узбекского языка и литературы имени Алишера Навои**

Защита диссертации состоится « 26 » авгует 2024 года в « 8 ;00 » часов на заседании Научного совета DSc.03/30.12.2029 Fil.05.02 при Ферганском государственном университете по адресу: 100151, г. Фергана, улица Маргиноний, 105. Тел.: (99873) 244-57-82; факс: 99873) 244-44-01; e-mail: info@fdu.uz

Диссертацией можно ознакомиться в Информационно-ресурсном центре Ферганского государственного университета (зарегистрирована под номером 366). (Адрес: 100138, 150100, г.Фергана, ул. Мураббийлар, 19.Тел: (+99873) 244-44-02. Факс: (+99873) 244-44-93. Эл.почта: fardu_info@umail.uz).

Автореферат диссертации разослан « 13 » авгует 2024 года
Протокол реестра № 22 от « 13 » авгует 2024 года

С.М.Муминов
Председатель Научного совета по присуждению ученых степеней, д.ф.н., профессор

М.Т.Зокиров
Секретарь Научного совета по присуждению ученых степеней, к.ф.н., профессор
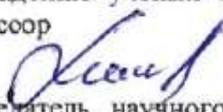
А.Мамажонов
Председатель научного семинара при Научном совете по присуждению ученых степеней, д.ф.н., профессор

ВВЕДЕНИЕ (аннотация диссертации доктора наук (DSc))

Целью исследования является раскрытие принципов эмпирического анализа исследования развития языка путем создания диахронического корпуса медиатекстов английских и узбекских языков.

Задачи исследования:

разработка требований к работе с текстами, относящимися для базы диахронических корпусов, по работе с источником на основе совершенствования научно-исследовательской методологии корпусной лингвистики относительно создания и применения диахронического корпуса;

освещение своеобразия медиатекстов как социального выражения языка и основного ресурса обогащения языка;

создание диахронического корпуса, состоящего из газетных текстов на английском и узбекском языках в рамках выбранных языков, с осознанием теоретической и практической значимости создания диахронического корпуса медиатекстов;

раскрытие практической важности существующих корпусов английского языка, в частности хронологически упорядоченных и нормированных языковых корпусов;

доказательство практической значимости диахронического корпуса медиатекстов на примере узбекских и английских газетных материалов, в том числе того, что она является базой, которая может быть применена в качестве статистической основы к изменениям лексического слоя в разрезе лет;

исследование того, является ли диахронический корпус медиатекстов достаточным источником для сравнения исследуемых межъязыковых лингвокультурных понятий;

теоретическое обоснование подходов к наблюдению за развитием лексического слоя языка диахронического корпуса медиатекстов.

Научная новизна исследовательской работы заключается в следующем:

раскрыты такие функциональные аспекты принципов работы с базой языковых корпусов, как аутентичность и репрезентативность через такие нормы как хронологическая непрерывность, транслитерация, периодическая точная величина, строгость временных интервалов, которые устанавливаются к источнику существующих диахронических корпусов английского, русского, китайского, арабского, испанского и корейского языков;

научно обоснованы хронологический порядок, периодический охват, лингвистические и экстралингвистические факторы формирования диахронического корпуса изучаемых языков через доказательство соответствия языковых корпусов узбекского языка принципам синхронного корпуса, а также эволюционного развития языков в ярусах лексического слоя лингвокультурному изменению языковой картины мира;

эволюционно определены историзмы, относящиеся к хронологической связности словосочетаний в различных системных языках посредством охвата в хронологической последовательности тридцатилетней базы газетных

материалов "Оила ва Жамият" и "Anaheim Bulletin", а также путем создания веб-программы (https://mediatextcorpus.com/) диахронического корпуса узбекских и английских медиатекстов, автоматически реализующую принципы общего и годового статистического анализа в составе изучаемых языков;

доказано, что в базе изучаемых языков имеется возможность эволюционного наблюдения за изменяющимися тенденциями национального, политического и социального подхода к анализируемой лингвокультуре путем разработки концептуального фрейма межъязыковых лингвокультурных понятий;

эмпирически доказано существование возможностей функции фиксации основных явлений развития лексического слоя изучаемых языков с помощью диахронических корпусов: неологизмов и архаизмов.

Внедренность результатов исследований. На основе разработанных практических предложений и результатов исследований по раскрытию принципов эмпирического анализа исследований развития языка развития путем создания корпуса газетных текстов узбекского и английского языков с теоретическим пониманием исследования диахронических корпусов в развитии корпусной лингвистики:

анализ относительно раскрытия таких функциональных аспектов принципов работы с базой языковых корпусов, как аутентичность и репрезентативность через такие нормы как хронологическая непрерывность, транслитерация, периодическая точная величина, строгость временных интервалов, которые устанавливаются к источнику существующих диахронических корпусов английского, русского, китайского, арабского, испанского и корейского языков использованы в практическом проекте № JHBL-94 "Трудоустройство девушек, нуждающихся в социальной защите, путем обучения их программам компьютерной графики, включенным в "Женскую тетрадь в отдаленных сельских районах" выполненного в 2022-2023 годах в научно – исследовательском Институте "Махалля и семья" (Справка № 01-09/1607 Научно – исследовательского института "Махалля и семья" от 4 сентября 2023 года). В результате у участников проекта сформировались научно-практические навыки создания диахронического корпуса, системных методов их применения на практике;

теоретические взгляды и анализ по научному обоснованию хронологического порядкв, периодического охвата, лингвистических и экстралингвистических факторов формирования диахронического корпуса изучаемых языков через доказательство соответствия языковых корпусов узбекского языка принципам синхронного корпуса, а также эволюционного развития языков в ярусах лексического слоя лингвокультурному изменению языковой картины мира были эффективно использованы в деятельности объединенной редакции журналов Министерства культуры Республики Узбекистан, а также в передаче "Good Morning Uzbekistan", транслировавшейся на телеканале "Foreign Languages" Национальной телерадиокомпании Узбекистана (Справка № 43 Объединенной редакции

журналов "Гулистан" и "Театр" Министерства культуры Республики Узбекистан от 20 октября 2023 года и справка № 01-23-1774 Государственного учреждения телерадиоканал Национальной телерадиокомпании Узбекистана от 27 ноября 2023 года). В результате была популяризирована существования необходимости создания диахронического корпуса узбекского языка с помощью членов редакции и ответственных за передачу;

научная гипотеза относительно эволюционного определения историзмов, относящихся к хронологической связности словосочетаний в различных системных языках посредством охвата в хронологической последовательности тридцатилетней базы газетных материалов "Оила ва Жамият" и "Anaheim Bulletin", а также путем создания веб-программы (https://mediatextcorpus.com/) диахронического корпуса узбекских и английских медиатекстов, автоматически реализующую принципы общего и годового статистического анализа в составе изучаемых языков после проведения соответствующих проверок заявки № DGU 2023 0264, поданной в Министерство юстиции Республики Узбекистан 16 января 2023 официально зарегистрирована в соответствии с Законом Республики Узбекистан "О правовой защите программ и баз данных, созданных для электронных вычислительных машин" (Свидетельство № DGU 21899 Министерства юстиции Республики Узбекистан от 3 февраля 2023 года). В результате получена юридическая основа для гарантии наших авторских прав на результаты проведенных исследований и на созданный диахронический корпус;

научные выводы и предложения относительно доказанности того, что в базе изучаемых языков имеется возможность эволюционного наблюдения за изменяющимися тенденциями национального, политического и социального подхода к анализируемой лингвокультуре путем разработки концептуального фрейма межъязыковых лингвокультурных понятий использованы в фундаментальном научно-практическом проекте №FZ-2022081663 "Создание веб-сайта и коллекции мультимедийных продуктов, посвященных продвижению узбекского национального танцевального искусства" реализованного в 2020-2021 годах в Узбекской Государственной академии хореографии (Справка № 01/04-61 Узбекской Государственной академии хореографии при Министерстве культуры Республики Узбекистан от 8 июня 2023 года). Диссертация послужила научным ресурсом при углубленном изучении возможности эмпирического наблюдения степени обогащения терминов национального танцевального искусства в словарном слое языков на базе узбекских и английских газетных текстов;

практические предложения и рекомендации относительно эмпирической доказанности существования возможностей функции фиксации основных явлений развития лексического слоя изучаемых языков с помощью диахронических корпусов: неологизмов и архаизмов использованы в проекте № ID-57663348 "European Perspectives in Dialogue: Central Asia and Western Research – Европейские перспективы в диалоге: исследования Центральной Азии и Запада" финансированного Службой академических обменов DAAD

Германии (Справка № 06-01-0425 Бухарского государственного университета от 25 сентября 2023 года). В результате теоретические знания о хронологии возникновения в словарном слое неологизмов и архаизмов, имеющих важное значение в развитии лексикографии узбекского и английского языков, были разъяснены в форме семинара магистрам и профессорам –преподавателям университета Gottingen Германии.

Структура диссертации. Содержание диссертации состоит из введения, четырех глав, заключения, списка использованной литературы и приложений, общий объем которого составляет 236 страниц.

E'LON QILINGAN ILMIY ISHLAR RO'YXATI
СПИСОК ОПУБЛИКОВАННЫХ НАУЧНЫХ РАБОТ
LIST OF PUBLISHED SCIENTIFIC WORKS

I bo'lim (I часть; I part)

1. Атабоев Н.Б. Инглиз тилининг диахроник корпусларининг функционал хусусиятлари (COCA мисолида). Монография. – Бухоро: Мухаммад полиграф, 2023. – 115 б.

2. Ataboyev N.B. Mediamatnlar diaxronik korpusida til rivojining empirik tahlil tamoyillari. Monografiya. – Buxoro: Muxammad poligraf, 2024. – 145 b.

3. Атабоев Н.Б. Корпус лингвистикасига оид терминларнинг частотали-изоҳли луғати. Луғат. – Бухоро: Мухаммад полиграф, 2023. – 44 б.

4. Ataboyev N.B. Mavjud ingliz tili diaxronik korpuslarining funksional xususiyatlari // Buxoro davlat universiteti ilmiy axboroti. – Buxoro, 2023 / № 11. – B. 144-151. [10.00.00 №1]

5. Ataboyev N.B. Til korpuslari – chet tilini o'qitishdagi so'z qo'llashga oid statistik aniq baza // Pedagogik mahorat ilmiy-nazariy va metodik jurnal. – Buxoro, 2024 / №1. – B. 269-274. [13.00.00 №23]

6. Ataboev N.B. The Theoritical Procedure Of Compiling A Diachronic Corpus // Pindus Journal Of Culture, Literature, and ELT. – Portugal, 2022 / V. 2 № 3. – pp. 121-124. [SJIF 2022: 7.598]

7. Ataboev N.B. Diachronic Corpus Creation Practices In World Science: On The Examples Of Korean, Chinese, Russian And Arabic Languages // Central Asian Journal Of Literature, Philosophy And Culture. – Spain, 2022 / V. 3. – pp. 58-61. [SJIF 2022: 7.62]

8. Ataboev N.B. Corpus Linguistics As A Methodology: A Set Of Methods Based On The Corpus Analyses // Electronic Journal Of Actual Problems Of Modern Science, Education And Training. – Khorazm, 2023 / V. 12. – pp. 23-27. [10.00.00 №22]

9. Ataboyev N.B. Mediamatnlar tilning ijtimoiy ifodasi hamda til boyishining asosiy resursi sifatida // O'zbekistonda xorijiy tillar: ilmiy-metodik elektron jurnal. – Toshkent, 2024 / № 1 (54). – B. 60-75. [10.00.00 №17]

10. Ataboyev N.B. Korpus lingvistikasi: korpus tahlil metodlari yordamida til leksik qatlamining boyishi tadqiqi // Buxoro davlat universiteti ilmiy axboroti. – Buxoro, 2024/ № 1. – B. 244-251. [10.00.00 №1]

11. Ataboyev N.B. Diaxronik korpuslar yaratishda manba bilan ishlash: ingliz va o'zbek mediamatnlar diaxronik korpusiga ehtiyoj // Filologiya masalalari jurnali. – Toshkent, 2023 / № 3. – B. 3-20. [10.00.00 №18]

12. Ataboyev N.B. Mediamatnlar korpusining mavjud ingliz tili diaxroni korpuslari prizmasida tahlili // O'zMU xabarlari. – Toshkent, 2024 / № 1/2/1. – B. 288-291. [10.00.00 №15]

13. Ataboyev N.B. Mediamatnlar diaxronik korpusini yaratishda matniy manbalar bilan ishlash asoslari va saralash xronologiyasi // Xorijiy lingvistika va lingvodidaktika jurnali. – Samarqand, 2024 / №1. – B. 151-155. [10.00.00 https://oak.uz/pages/4802]

14. Ataboyev N.B. O'zbek va ingliz tili mediamatnlar diaxronik korpus bazasini avtomatlashtirish va dizayn muammolari // Ta'lim va innovatsion tadqiqotlar xalqaro ilmiy-metodik jurnali. – Buxoro, 2024 / №1. – B. 9-14. [10.00.00 https://oak.uz/pages/4802]

15. Атабоев Н.Б. Инглиз тили корпусларида коллокацион бирликларнинг лингвомаданий тадқиқи // ЎзМУ хабарлари. – Тошкент, 2021 (1/4). – Б. 251-254. [10.00.00 №15]

16. Ataboev N.B. Language Evolution Studies Through Media Texts // American journal of Social and humanitarian research. – USA: Houston, 2023 / V.4, № 4, – pp. 204-209. [SJIF 2023: 7.754]

17. Атабоев Н.Б. Лисоний корпуслар таснифида диахроник корпусларнинг ўрни // Таълим ва инновацион тадқиқотлар халқаро илмий-методик журнали. – Бухоро, 2023 / №12. – Б. 67-71. [10.00.00 https://oak.uz/pages/4802]

18. Ataboev N.B. Corpus Based Analysis and Language Development: Theory and Practice // Best Journal of Innovation in Science, Research and Development. – USA, 2023 / V. 2, № 9, 2023. – pp. 449-456. [SJIF 2023: 6.952]

II bo'lim (II часть; II part)

19. Ataboev N.B. Data Collection And Normative Principles In Creating A Diachronic Corpus // Modern Aspects Of Economic And Social Sustainable Development: International E-Conference. Vol. 12: Special Issue. September 15. – Krakow, 2023. – pp. 83-85.

20. Ataboev N.B. Problems And Solutions: Indian Diachronic Corpora // International scientific review of the problems and prospects of modern science and economy: International E-Conference. Vol. 17: Special Issue. – Tashkent, September 18, 2023. – pp. 85-87.

21. Ataboyev N.B. O'zbek tilining mavjud sinxronik korpuslari va diaxronik korpus yaratish ehtiyoji // Yangi O'zbekistonda tabiiy va ijtimoiy-gumanitar fanlar: Respublika ilmiy amaliy konferensiyasi. – Tashkent, 2023 / 1(6). – B. 92-94.

22. Ataboyev N.B. Mediamatnlar diaxronik korpusi tushunchasi va zarurati // Conference on Universal Science Research 2023, 1(12). – B. 328-330. URL: https://zenodo.org/records/10637952

23. Ataboyev N.B. Inglizcha zamonaviy diaxronik va sinxronik korpuslari tahlili // XXI asrda innovatsion texnologiyalar, fan va ta'lim taraqqiyotidagi dolzarb muammolar: Respublika ilmiy-amaliy konferensiyasi. – Termiz, 2023 / 1(12). – B. 300-302.

24. Ataboev N.B. Practical Application of Corpus Analyses in Translation And its Advances // Translation and intercultural communication through the English literature and linguistics: International scientific-practical conference. – Bukhara, May 13, 2023. – pp. 10-14.

25. Атабоев Н.Б. COCA манбаларида морфологик бирликлар тадқиқи // Тил, адабиёт, таржима, адабий танқидчилик: замонавий ёндашувлар ва истиқболлар: халқаро илмий-амалий анжуман. – Бухоро, 2022. – Б. 15-16.

26. Ataboev N.B. Discourse Analysis of Media Texts about the Education System: Possibilities of Corpus and Cognitive Research // YuNESKOning "Nomoddiy madaniy merosni muhofaza qilish to'g'risida"gi konventsiyaning 20 yilligi, "Butunjahon madaniy va tabiiy merosini muhofaza qilish konventsiyasining 50 yilligi va Buxoro davlat universitetining 93 yilligiga bag'ishlangan "Moddiy va nomoddiy madaniy meros uyg'unligi": xalqaro ilmiy-amaliy anjuman. – Buxoro: Durdona nashriyoti, 2023. – B. 282-283.

Avtoreferat Farg'ona davlat universiteti
"Lingvistik tahrir va tarjimashunoslik"
Markazida tahrirdan o'tkazildi