



Received: July 6, 2023
Accepted: September 20, 2023
Available online: September 25, 2023

Nozimjon Ataboev

Filologiya fanlari bo'yicha falsafa
doktori (PhD), dotsent
Buxoro davlat universiteti
Buxoro, O'zbekiston

Nozimjon Ataboev

Associate Professor, Doctor of Philosophy (PhD)
in Philological Sciences
Bukhara State University
Bukhara, Uzbekistan
E-mail: n.b.ataboev@buxdu.uz

DIAXRONIK KORPUSLAR YARATISHDA MANBALAR BILAN ISHLASH

ANNOTATSIYA

Ushbu maqola o'z oldiga ingliz tilining mavjud diaxronik korpuslari yaratilish tartibi, manbalarni yig'ish, ular ustida ishlash va saralash amaliyotini tadqiq qilgan holda, o'zbek tilining ingliz tili bilan hamohang diaxronik korpusini yaratishning ahamiyatini tahlil qilishni maqsad qilib belgilagan. Qolaversa, har qanday lisoniy korpus yaratish amaliyoti muhim faktorlar qatorida turuvchi kerakli, yetarli va asosli manba, ya'ni, matniy birliklar bilan ishlashning aynan diaxronik korpuslarga xos xususiyatlari va ahamiyatli tomonlari nazariy jihatdan muhokamaga tortilgan. Maqolada berilgan nazariy qarashlar, ilmiy faktlar hamda tanqidiy fikrlarni yoritishda kontent tahlil, tasniflash, tavsiflash, qiyoslash va empirik tahlil metodlaridan foydalanilgan.

Ushbu ishda Sh.M. Hamroyeva, Y.O. Knyazeva, T.MakEneri, R.Syao, Y.Tono, R.A. Karimov, A.A. Rahmanova, U.M. Xoliyorov, G.H. Begmatova va boshqa ilmiy nazariyalarga asoslanib, til korpuslarining manbalariga qo'yiladigan umumiy hamda diaxronik korpus bazasiga kiritiladigan matniy birliklarga nisbatan talab etiladigan xususiy tamoyillar ishlab chiqilgan. Bunda umumiy tamoyillarga tarkibiga manbaning autentik material bo'lishi; o'zaro teng hajmda to'planishi; hech qanday o'zgartirish yoki to'g'rilash kabi tahririy jarayonlarga tortilmasligi; umumiy leksik miqdoriy hajmi bir milliondan kam bo'lmasligi; reprezentativlik tamoyiliga mos kelishi; zamon va makondagi mutanosibli kabilarni kiritish o'rinli ekanligi aniqlandi. Aynan diaxronik korpuslarga xos bo'lgan xususiy tamoyillar sifatida quyidagi to'rt talab alohida sanab o'tildi: 1) xronologik uzviylikda matnlarni yig'ish; 2) tarixiy voqelik

WORKING WITH SOURCES IN CREATING DIACHRONIC CORPUS

ABSTRACT

The article discusses the practice of collecting and working on existing diachronic corpora of the English language, as well as sorting them. It also covers the important factors involved in creating any linguistic corpus, such as the necessary, sufficient and reasonable sources. The article goes into detail about the specific features and important aspects of working with text units that are specific to diachronic corpora. Theoretical views, scientific facts, and critical opinions are illuminated by using content analysis, classification, description, comparison, and empirical analysis methods.

Based on the theoretical viewpoints of Sh.M. Hamroyeva, Y.O. Knyazeva, T.McEnery, R.Xiao, Y.Tono, R.A. Karimov, A.A. Rakhmanova, U.M. Kholiyorov, G.H. Begmatova, and other scientific theories, specific principles have been developed for textual units to be included in the general and diachronic corpus base. The general principles include the following: the material should be authentic and unedited; the total lexical volume should not be less than one million and should comply with the principle of representativeness; it was determined that proportionality in time and space is appropriate. The following four requirements have been listed separately as specific principles for diachronic corpora: 1) texts should be collected in chronological order; 2) transliteration of ancient or historical inscriptions of languages that have undergone alphabetic changes in writing graphics according to historical facts; 3) a relative approach to the standard requirements for entering texts in an equal volume; 4) the corpus should cover at least one decade.

ortidan yozuv grafikasida alifbo o'zgarishlarini boshidan o'tkazgan tillarning qadimiy yoki tarixiy bitiklarini transliteratsiya qilish; 3) teng hajmda matnlar kiritish borasidagi me'yor talablariga nisbiy yondashish; 4) kamida bir dekadali vaqt oralig'ini qamrab olish.

Maqola diaxronik korpus tushunchasi korpus lingvistikasi fan tarmog'ida katta ahamiyatga ega bo'lib, korpus tahlil metodlaridan foydalanishning yangi bosqichini yuzaga chiqarayotgani holda, unda sodir bo'layotgan tarixiy-tadrijiy rivojlanish bosqichlarini o'rganish borasidagi muhim qadamlar sirasiga kirishi xususidagi xulosalar bilan yakunlangan.

Kalit so'zlar: korpus lingvistikasi, matn, korpus, diaxronik korpus, baza, konkordans, autentik material, reprezentativlik.

KIRISH

Korpus lingvistikasi tadqiqot metodologiyasi sifatida qo'llanilar ekan, o'zining ko'lami jihatidan tilshunoslikning barcha sohalarida ijobiy imkoniyatlarni taqdim etayotgani sir emas. Korpus tahlillari tilshunoslik tadqiqotlarining ko'plab sohalarida qo'llanilishi mumkin: 1) *leksikografiya*: bu muayyan so'zlarning lingvistik va tilga bog'liq bo'lmagan jihatlarini o'rganishga imkon beradi va lug'at tuzuvchilarga yordam beradi; 2) *sotsiolingvistik*: korpusga asoslangan usullar ilgari amalga oshirib bo'lmaydigan dialektlar va registrnlarni o'rganishga imkon beradi; 3) *diskurs tahlili*: korpora ko'plab diskurs namunalari va polisemantik so'zlar doirasida tadqiqotlar o'tkazishi mumkin; 4) *morfologiya*: so'z variantlarining chastotasi, tarqalishi va funksiyalari tahlil qilinishi mumkin; 5) *fonologiya*: fonetik tarqalishning turli xil usullarini o'rganish mumkin; 6) *sintaksis*: tilning tuzilishini va tilda jumalarning shakl va o'zaro grammatik munosabatini izohlovchi empirik dalillar keltirishi mumkin; 7) *qiyosiy tipologiya*: tillar o'rtasidagi o'xshashlik va farqlarni o'quvchilar parallel korpora orqali tahlil qilishlari mumkin; 8) *chet tillarini o'qitish*: bunday keng ko'lamli tadqiqotlar samarali o'quv materiallari va til o'qitishni yangicha loyihalashda yordam beradi [Dobrik, 2009].

Ushbu jarayonning yaqqol isboti sifatida bugungi kunda yaratilgan va yaratilayotgan milliardlab so'zlarni o'z ichiga olgan lisoniy korpuslarni sanash kifoya qiladi. Shular ichidan aynan ingliz tili tadqiqi uchun yaratilgan korpuslarga to'xtalib, ularning qisqacha maqsad va vazifalarini ko'rsatish hamda ulardan olingan ma'lumotlarni tahlillarda qo'llash orqali korpus lingvistikasi metodologiyasining amaliy tatbiqini yanada aniqroq planda ko'rsatish imkoni paydo bo'ladi. T.MakEneri ta'biri bilan aytganda, 1960-yillardan beri davom etib kelayotgan ingliz tili korpusi ustida olib borilayotgan ilmiy-amaliy ishlar aynan korpus lingvistikasi rivojida ko'p

ahamiyat kasb etgan [McEnery & Hardie, 2012; 71]. Ilk elektron korpus (Brown corpus) va bugungi kundagi eng katta hajmli (Google books n-gram) 200 milliard soʻzdan tashkil topgan korpus ham aynan ingliz tilida yaratilgan.

Diaxron va sinxron tadqiqotlar va yondashuvlarga toʻxtalganda *sinxron* va *diaxron* lingvistik tahlilga bir-birini toʻldiruvchi ikkita nuqtayi 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 maʼlum bir 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.

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 qaramaqarshi 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 zamonda, jumladan, sinxroniyada bilishga imkon beruvchi vosita boʻlib, umuman til nazariyasining muhim qismidir.

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 [Tomsen, 1938].

Aynan tarixiy tilshunoslikning asosiy vazifalari turiga kiruvchi genetik tasniflashga doir ilmiy qarashlarga tayangan holda qiyoslanayotgan tillarning ushbu tasnifning qaysi turiga tushishini koʻrib chiqish oʻrinli. Ingliz tili *Hind-ovrupo tillari oilasiga* kiruvchi *german tillar guruhiga* kiradi. Oʻzbek tili esa, *Turkiy tillar oilasiga*, ikkinchi qarashga koʻra, ushbu oila *Oltoy tillar katta oilasiga* mansubligi aytiladi. Yana ham aniqroq qilib aytganda, oʻzbek tili turkiy tillar oilasiga kiruvchi *Qarluq guruhiga* kiritiladi va bu borada oʻzbek tiliga eng yaqin til *yangi uygʻur tillari* hisoblanadi [Irisqulov, 2009; 185–187].

Shunday ekan, *diaxron* terminini korpus lingvistikasiga ham kiritilishi uning mazmuniy va vazifaviy xususiyatlariga soya solmaydi. Yaʼni, diaxronik korpus

turli davrlardagi matnlarni o'z ichiga olgan korpus bo'lib, tilning rivojlanishi yoki o'zgarishini tadqiq qilish uchun ishlatiladi. Bu borada M.Hilpert va S.Th. Gris kabi olimlarning fikriga quloq solinsa, diaxronik korpus deganda, hozirgi og'zaki ingliz tilining diaxronik korpusida bo'lgani kabi ketma-ket vaqt oralig'ida tildan foydalanishning o'xshash turlarini ifodalovchi matn resurslari tushuniladi, ushbu xususiyati pirovardida korpus kamida ikkita vaqt oralig'ini o'z ichiga qamrab olishi talab etiladi [Hilpert & Gries, 2016; 36].

Bizningcha, diaxronik kopuslar va, umuman olganda, barcha turdagi lisoniy korpuslarni yaratish amaliyotida matniy birliklarni to'plash, saralash va bazani shakllantirishning ahamiyati kattadir. Chunki, ma'lum maqsadda, ya'ni tilning biror qatlami, dialekti, hudud yoki vaqt oralig'iga doir ilmiy tadqiqot olib borishni maqsad qilgan lingvist borki, u o'zi shakllantirayotgan korpusning tarkibini oldindan belgilashi maqsadiga erishini kafolatidir.

ASOSIY QISM

Diaxronik korpus yaratishda baza va me'yor tamoyillari

Tilshunoslikda mavjud sohalarining bugungi kundagi eng zamonaviy sohalarini sirasiga korpus lingvistikasini kiritish mumkin. Korpus lingvistikasi dunyo tillarining elektronlashtirilishi, ya'ni til korpuslariga jamlanishi hamda bu orqali tillarni o'rganishga oid ham amaliy, ham ilmiy-nazariy izlanishlarni osonlashtirishni o'z oldiga maqsad qilib olgan. Korpus lingvistikasiga asoslangan yondashuv ikki turli bo'lib, ularni quyidagicha izohlash mumkin:

– birinchidan, tilning hozirgi zamonidan jamlangan manbadan tuzilgan korpuslar yaratish bilan shug'ullanadi. Bunday turdagi korpuslar ayni holat to'g'risidagi qat'iy xulosa va ayni tilning kelajagi xususidagi ilmiy farazlarni o'rta tashlaydi;

– ikkinchidan, tilning o'tmishidan, tarixiy biror bosqichidan yig'ilgan korpuslarni shakllantiradi. Bu turdagi korpuslar tahlil natijalaridan tilning rivojlanish bosqichlarini o'rgangan holda uning bugungi kundagi maqomini to'g'ri belgilash mumkin bo'ladi.

S.Gris korpus lingvistikasi amaliyoti taqdim etadigan imkoniyatlarga to'xtalib, korpusga asoslangan miqdoriy tahlillarni quyidagicha izohlaydi: 1) muhim nazariy qarashlarni ochish uchun obyektiv usuldir; 2) eng ishonchli tekshirishga imkon beradi; 3) ko'pincha yakka tartibdagi mulohazalarga asoslangan tadqiqotlarga qaraganda empirik jihatdan ko'p qirrali izlanishlarni yuzaga keltiradi; 4) tabiiy til tarkibidan saralab olingan og'zaki va yozma matnlardan tashkil topgan korpuslarda olib boriladi [Gries, 2006].

Korpusga kiritiladigan bazaning xususiyatlari, uning tarkibiy, miqdoriy va ifodaviylikka oid talablarini ko'rib chiqish uchun, quyida lisoniy korpus haqidagi qarashlar va fikrlarning bir nechtasini keltirib o'tamiz:

S.Hanston lisoniy korpusni quyidagicha izohlaydi: "Korpus, bu – tabiiy ravishda uchraydigan til namunalarining elektron saqlanadigan to'plami. Ko'pgina zamonaviy korpuslar hajmi kamida bir million so'zdan iborat bo'lib, qisqa to'liq matnlardan yoki uzun matnlarning parchalaridan jamlanadi. Odatda, matnlar muloqot turini yoki turli tillarni ifodalash uchun tanlanadi; masalan, tarix darsliklarida qo'llaniladigan

ingliz tilini yoki Kanadada qoʻllaniladigan fransuz tilini yoki genetik modifikatsiyaga oid Internet muhokamalarini korpusda ifodalash uchun aynan shu sohalaridan baza shakllantirish mumkin. Korpuslar maxsus dasturiy taʼminotdan foydalanish orqali tekshiriladi” [Hunston, 2006; 234]. Olimaning fikriga tayanadigan boʻlsak, til korpuslariga manba kiritish uchun, albatta, tilning tadqiqot prizmasi ostiga olingan zamon va makondagi chegarasini belgilab olish zarurati yuzaga chiqadi. Bunda, oʻrganilayotgan hudud shevasi yoki ijtimoiy qatlamning xarakter shakllarini inobatga olish kerak.

T.MakEneri va A.Uilsonning fikricha: “korpus tushunchasi dastavval birdan ortiq boʻlgan matnlar toʻplamiga nisbatan ishlatilgan va bundan ortiq hech qanday maʼnoni anglatmagan. Ammo hozirgi rivojlanib borayotgan tilshunoslikda “korpus” kengroq maʼnoda ishlatilmoqda: barcha matn turlarini oʻz ichiga oladigan va ulardan namunalar keltirilgan; aniq oʻlchamga ega boʻlgan; kompyuter texnologiyalari oʻqiy oladigan tarzda shakllantirilgan; barcha uchun standart tarzda maʼlumot topa olish imkoniyatini beradigan matnlar jamlanmasi” [McEnery & Wilson, 2001]. Soha olimlari lisoniy korpuslarga barcha turdagi uslubga xos boʻlgan matniy birliklarni qamrab olish shartini oʻrtaga tashlaydi. Qolaversa, korpusga kiritiladigan matnlarga miqdoriy meʼyorni belgilash talabini ham aynan shu tilshunoslar izohida uchratish mumkin.

J.Sinklarning korpus borasidagi fikrlarini qayd fikrlari shunday: “Korpus lingvistik maʼlumotlarning toʻplami boʻlib, yozma matnlar yoki yozib olingan nutqning transkripsiyasidan tarkib topadi va bu lingvistik tavsifning boshlangʻich nuqtasi yoki til haqidagi farazlarni tekshirish vositasi sifatida qoʻllanilishi mumkin” [Sinkler, 1991]. J.Sinkler oʻz oʻrnida lisoniy korpuslarni shakllantirishda matnning xususiyatiga qarab yondashish kerakligini maʼqullaydi. Unga koʻra, bazaga kiritilayotgan korpus maʼlumotlari ogʻzaki yoki yozma nutq shaklida boʻlishiga asoslanib, uni bir xil qidiruvda muammo tugʻdirmaydigan holatda kiritilishi borasidagi talabni oʻrtaga tashlagan.

G.Begmatovaning fikricha, til korpusining ilmiy-amaliy ahamiyati quyidagilardan iborat: soha vakillari uchun elektron qidiruv manba; til yashab qolishining zamonaviy shakli; qidiruvning ixtisoslashuvi; vaqtning tejallishi; lugʻat tuzish uchun qimmatli manba [Begmatova, 2021; 62]. G.Begmatovaning taʼrifida keltirilgan korpus tilning mavjudlik prinsiplariga boʻysungan holda, uning yoʻqolib ketishini oldini olishga qaratilgan nazariy manba ekanligi toʻgʻrisidagi fikrni maʼqullaymiz. Bizningcha, til korpuslarini yaratish ehtiyoji bugungi kunda davlat tili maqomida boʻlmagan, siyosiy qaramlikdagi millat tillari, xorijiy kuchli davlat taʼsiriga tushib qolgan davlat hududida istiqomat qiluvchi xalqning milliy tili va rivojlanish markazi sifatida chet tillarni, xususan, ingliz tilini yosh avlodga oʻqitish, singdirish yoʻlini tanlagan hukumatlar tomonidan boshqarilayotgan millat tili uchun juda muhim boʻlib, bunday tillarning bugungi variantini bir avloq almashmay turib – ertaga topib boʻlmay qoladi. Bu esa tilshunoslik nazdida “haqiqiy fojea”dir. Bu borada, korpuslar tarkibiga kiritiladigan manba haqqoniy, autentik va reprezentativlik xususiyatiga ega boʻlgan matniy birlikdan tashkil topishi shart.

Korpus tushunchasi empirik lingvistika uchun asos boʻlib, maʼlum matnlarni

manual kuzatish, tahlil qilishdan farq qiladi. Prinsipial jihatdan bir nechta matnlardan iborat har qanday to'plamni korpus deb atash mumkin. Zero, "korpus" atamasi shunchaki lug'aviy "tana" ma'nosini berishi ma'lum va shuning uchun korpusga istalgan matn "qismi, bo'lagi" sifatida qarashlar uchrab turadi. Ammo "korpus" atamasi zamonaviy tilshunoslik kontekstida ishlatilganda, ko'pincha bu oddiy ta'riflar nazarda tutilganidan ko'ra aniqroq izohlanishga muhtoj. T.MakEneri va boshqalar bu borada "...korpusni, qat'iy aytganda, mavjud til me'yorlaridagi tabiiy sodir bo'ladigan jarayondan yig'ilgan lisoniy baza sifatida ta'riflash mumkin" – degan fikrni bildirishadi. Qolaversa, olimlarning nazdida korpusni turlicha izohlash mumkin, ammo korpusda quyidagi xususiyatlar bo'lishi shart:

- 1) kompyuter o'qiy oladigan;
- 2) haqiqiy/autentik matnlar (shu jumladan, og'zaki ma'lumotlarning transkriptlari) yig'indisi bo'lgan;
- 3) har bir sohadan misollar olingan;
- 4) muayyan til yoki til qatlamini ifoda etadigan.

Yuqorida sanab o'tilgan olimlarning fikrlariga qo'shilgan holda, ularning kompyuter o'qiy oladigan (machine readable) shakldagi matnlarni korpuslarga kiritish me'yorini korpus manbalari bilan ishlash, saralash va jamlash borasida nazarda tutish o'rinli.

O'zbek korpus lingvistikasi ham bugungi kunda rivojlanib bormoqda. Buning misoli sifatida o'zbek olimlarining korpus yaratish borasidagi ildam qadamlarini sanab o'tish joiz bo'lib, ular tomonidan bildirilgan hamda o'zlari yaratgan lisoniy korpuslar amaliyotida inobatga olingan manbaga qo'yilgan quyidagi talablarni ma'qullayamiz:

Tadqiqotchi A.A. Raxmanova o'z tadqiqotlarida o'zbek tili milliy korpusini yaratish bo'yicha uning matniy birliklariga nisbatan qo'yiladigan talablarni hamda quyidagi tamoyillarini ishlab chiqqan: 1. Korpus matnlarini tanlash va tizimlashtirish tamoyili; 2. Matnlarni grafik jihatdan bir xillastirish tamoyili; 3. Matnlarni morfologik jihatdan qayta ishlash tamoyili; 4. Matnlarni sintaktik jihatdan qayta ishlash tamoyili; 5. Matnlarni semantik jihatdan qayta ishlash tamoyili; 6. Matnlarni razmetkalash va qidiruv tizimiga moslash tamoyili; 7. Korpus dizaynini amalga oshirish [Raxmanova, 2021; 65].

O'zbek tilining ta'limiy korpusini yaratish bo'yicha olib borgan tadqiqotida O'.M. Xoliyorov "til korpusida foydalanuvchiga taqdim etilgan ma'lumot xotik tarzda bo'lganligi sababli material tanlashda o'qituvchi mas'uliyatni o'z zimmasiga olishi talab etiladi. Shunda ta'limiy korpus til ta'limida foydali elektron ta'limiy resurs bo'lib xizmat qiladi. Ushbu vositani qo'llash orqali ta'lim oluvchida til materialini (korpus asosida) mustaqil tahlil qilish, tanqidiy va analitik fikrlash ko'nikmasini hosil qiladi. Murakkabligiga qaramasdan, o'quv materialini ishlab chiqishda ta'limiy korpus effektiv manba bo'la oladi" [Xoliyorov, 2021; 36] – deya ta'kidlaydi.

Yanada aniqroq qilib aytganda, o'zbek tili ta'limiy korpusi materiali sifatida hozirgi o'zbek adabiy tilidagi rasmiy, ilmiy-ommabop, badiiy hamda publisistik uslubdagi matnlar tanlangan. Korpus reprezentativlik xususiyatini to'la namoyon etib, tarkibidagi xilma-xil matnlar janr jihatdan quyidagi proporsiyaga ega: 1) matnlarning

badiiy uslubga tegishli qismi Abdulla Qahhor, Said Ahmad, Tog‘ay Murod, Asqad Muxtor qalamiga mansub qissa, hikoya, roman, felyeton, esdalik kabi janrlarga oid asarlari qamrab olingan; 2) publitsistik uslubdagi matnlar tarkibiga kun.uz, daryo.uz saytlaridagi maqolalar kiritilgan; 3) rasmiy uslubdagi materiallarini lex.uz saytidan yuridik hujjat matnlari tashkil etadi; 4) ilmiy-ommabop uslubdagi materiallar, asosan, maktab darsliklari matnlaridan iborat [Xoliyorov, 2021; 37]. Mantlarni fragmentga bo‘lish, nomatniy belgi (jadval, rasm, chizma) larni olib tashlash sublimetext dasturi vositasida amalga oshirilgan.

R.A. Karimov o‘zbek-ingliz tillari parallel korpusini yaratishning nazariy asoslari ustida izlanish olib borgan hamda ularning manbaga oid xususiyatlarini quyidagicha sharhlagan: “...ko‘p tilli korpuslardan farqli o‘laroq, parallel korpus faqat tarjima matnlaridan tashkil topadi. Ular nisbatan kichik hajmli; muvofiqlashtirilgan bo‘ladi. Masalan, *Rus tili milliy korpusi* tarkibidagi Rus-ingliz parallel korpusi hajmi asosiy korpus hajmining 10 foizini tashkil qiladi. Buning sababi juda oddiy: til egalarining turli sohalaridagi nutqini topish oson: ta‘lim, madaniyat, ijtimoiy hayot va h.k. Tarjima matnlar esa ijtimoiy hayotning ma‘lum sohasiga tegishli matnlar orasidan tanlab olinadi. Tarjima matni chegaralangan masshtabda – xorijiy tilni o‘qitish, turli chaqiriq, yig‘ilish nutqi hamda konferensiya materiallaridan to‘planadi” [Karimov, 2021; 53].

Sh.Hamroyeva korpus tahlili natijalarini quyidagicha sharhlaydi [Hamroyeva, 2018; 18]:

- belgilangan so‘zning turli kontekstdagi barcha shakldagi ko‘rinishi;
- tilning lug‘at tarkibidagi o‘rni va variantlari;
- belgilangan so‘z bilan birikish imkoniyatiga ega so‘zlar ro‘yxati;
- u yoki bu yozuvchining ayni so‘zdan foydalanish chastotasi yoki statistikasi;
- so‘zning denotativ, konnotativ ma‘nosi;
- leksik qo‘llanishning yashirin modeli (imkoniyati);
- til taraqqiyotining turli davrida qo‘llanish holati.

Bizningcha, korpus lingvistikasi taqdim etadigan eng katta va muhim ijobiy imkoniyat – tabiiy tilni zamon, makon va soha nuqtayi nazaridan saralangan reprezentativlik xususiyatlarini o‘zida aks ettirgan matnlarni jamlab, keyingi avlodga yetkazishidir. Tilning mavjudlik va rivojlanish tamoyillari – davlat tili maqomida bo‘lishi (makon), uning foydalanuvchilari, ya‘ni til sohiblarining mavjudligi (nisbiy ravishda, zamon va soha) kabi xususiyatlarni o‘zida aks ettira oladigan milliy korpuslarning yaratilishi tilning buguni va ertasini solishtirish imkonini beradi. Globallashuv natijasida kirib kelgan turli yot tushunchalar til tarkibiga salbiy ta‘sir o‘tkazadi. Agar tilning milliy korpusi yaratilsa, qamrab olingan vaqt oralig‘idagi til tarkibi aniq saqlab qolinadi. Bunga misol tarzida Britaniya Milliy Korpusini (BNC) keltirish mumkin. O‘zida 100 million so‘zli lisoniy bazani saqlovchi BNC 1993–1997-yillar oralig‘idagi tilni o‘zida ifoda etadi va bu baza bir necha asrdan so‘ng ham tarkiban o‘zgarmasdan saqlanadi.

Korpus lingvistikasining asosiy maqsadi korpus yaratish bo‘lsa, uni yaratish prinsiplarini ishlab chiqish esa uning muammosi hisoblanadi. Bu borada Sh.Hamroyeva korpus tarkibiga kiradigan matn turlarini ishlab chiqib ularni to‘rt guruhga bo‘lgan

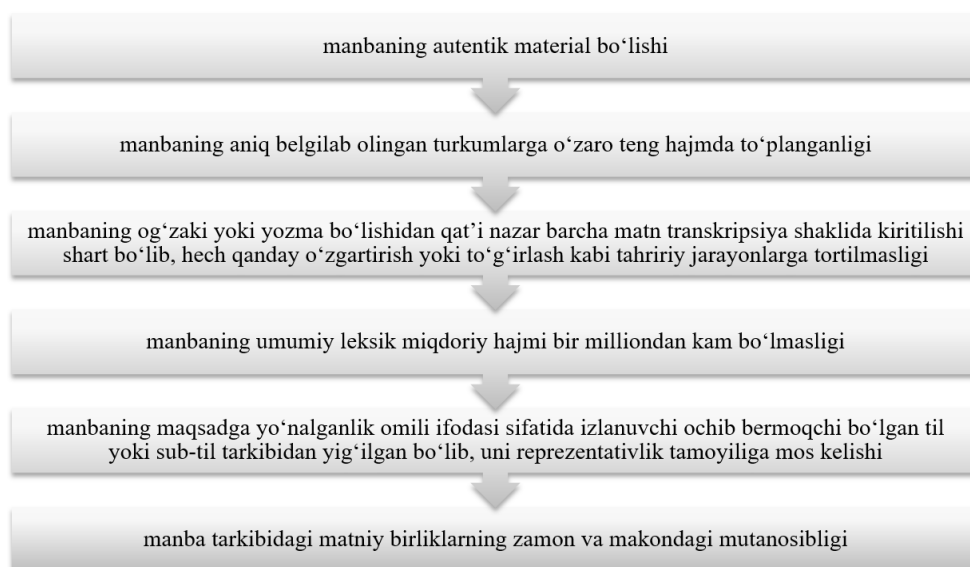
[Hamroyeva, 2018; 18-19]:

- 1) alohida bir muallif yoki bir necha yozuvchi asaridan olingan matnlar;
- 2) ma'lum davr (bir necha o'n yillik yoki yuz yillik)ni qamrab olgan matnlar;
- 3) belgilangan mavzudagi zamonaviy matnlar;
- 4) til va jamiyatning bugungi holatini aks ettiruvchi zamonaviy matnlar.

Korpus matnlarini yig'ish va saralash mezonlari xususida yetarlicha izlanishlar olib borilmaganligini inobatga olib, korpus tarkibiga kiritiladigan matniy lisoniy ma'lumotlarga nisbatan mezon talablarini taklif etish dolzarbdir. Korpus matnlariga qo'yiladigan mezon talablarining asosiylari quyidagilar (1-rasmga qarang):

1-rasm.

Umumiy til korpuslari manbasiga qo'yiladigan talablar



Bizningcha, korpus maqsadga yo'nalganligi bilan bir qatorda hajm jihatidan ham izlanuvchi ehtiyojlarini qondirishi shartligini ko'zda tutish kerak. Zero, hajm qanchalik kam bo'lsa, undan olingan misollar soni konkordansda tor qamrovli empirik analiz qilish imkoniyatini taqdim etadi. Masalan, Brovn korpusi hajm jihatidan bir million so'zni tashkil etganligi sababli, undan olingan qidiruv natijalari soni butun til xususida xulosa chiqarishga yetarli emas. Hajmning ko'pligi esa o'z-o'zidan tahlillar ko'lamini oshiradi.

Y.Knyazeva korpus tarkibiga kiruvchi har qanday matn birligi "qayta ishlov" jarayonidan o'tishi kerakligini ta'kidlaydi. Masalan, og'zaki nutq shakllaridan to'planadigan bo'lsa, yozib olingan materiallar transkripsiyanishi talab etiladi [Knyazeva, 2009; 97]. Agar material yozma bo'lsa, uni talab etilgan programmada matniy elektron shaklga o'tkazish shart. Bu birlamchi ish bo'lib, ushbu jarayonlar bilan bir qatorda matnlarni "belgilash (mark-up)" bosqichi ham mavjud. MakEneri va boshqalar korpuslarni teglash terminiga quyidagi izohni beradi [McEnery et al., 2006; 22]: "Korpuslarni teglash – matnning o'zi to'g'risida ma'lumot

beradigan va formatlash, chop etish kabi boshqa jarayonlarni tashkil etadigan elektron holatda saqlanadigan hujjatlarga kiritilgan standard kodlar sistemasidir”. Shuningdek, T.MakEneri teglash jarayonining muhimligini uch omil orqali asoslaydi. Birinchidan, korpus ma'lumotlari asosan mavjud til tarkibidan yig'iladi. Ya'ni lingvistik qo'llanishlarni ifoda etuvchi misollar real kontekstdan olingan bo'lsa-da, kontekstual ma'lumotni bera olmaydi. Vaziyat, holat va manbaga oid ma'lumotlar o'z-o'zidan matnda keltirilmaydi. Ikkinchidan, matnlarning nomlanishi yoki shunga o'xshash xususiyatlari asosida guruhlash ham juda kam ekstra-matniy (extra-textual) ma'lumotlarni bera oladi. Belgilash jarayoni esa korpuslarning sifat darajasini va unda matnlarning strukturasi hamda manbasiga oid ma'lumotlarni beradi. Uchinchidan, og'zaki nutq namunalaridan yig'ilgan matnlar transkripsiyasi o'zida belgilashlarni talab etadi. Masalan, paralingvistik birliklar (kulgi va h.k.) va nutqdagi to'xtalishlar belgilash asosida ko'rsatilishi kerak.

Diaxronik korpuslarning manba va matniy elementlariga qo'yiladigan talablarni ishlab chiqish uchun bu borada yutuqqa erishgan, tajribalarga tayanib yaratilgan mavjud shu turdagi korpuslarni tahlilga tortish maqsadga muvofiq bo'lar edi.

M.Hilpert va S.Th. Gris diaxronik korpus yaratish tamoyillari xususida to'xtalib, ularning eng yaxshi namunalari qatorida english.corpora.org saytida joylashtirilgan korpusni misol qilib keltiradi: "... 2010-yilda M.Devis tomonidan yig'ilgan *Tarixiy Amerika ingliz tili korpusi* (COHA), kuzatuvchi korpus bo'lib, uni jamlash paytida yigirma bir ketma-ket o'n yillik, ya'ni, ikki asrlik vaqt oralig'idagi ingliz tilidan foydalanishga oid namunalarini yig'adi. Zamonaviy ingliz diaxronik korpuslari turli xil matn turlarini ifodalaydi va ularni tegishli vaqt muddati prizmasida farqlaydi, lekin bu diaxronik korpuslarning dizayn xususiyati bo'lib, matn turini doimiy ravishda saqlab turadi, shuning uchun ma'lum bir matn turidagi diaxronik til o'zgarishini kuzatish imkoniyati mavjud bo'ladi" [Hilpert & Gries, 2016; 36].

Diaxron korpuslarga muqobil termin sifatida tillar rivojini kuzatib borish maqsadida monitor korpus tushunchasi ham mavjud. Monitor korpus – bu vaqt o'tishi bilan hajmi katta bo'lgan va turli xil materiallarni o'z ichiga olgan ma'lumotlar to'plami. Har xil turdagi materiallarning taqribiy nisbati vaqt o'tishi bilan farq qilishi mumkin. Birmingem universitetida ishlab chiqilgan Bank of English (BoE) monitor korpusining eng mashhur namunasi. BoE loyihasi 1980-yillarda boshlangan va o'shandan beri hozirga qadar yarim milliarddan ortiq so'zgacha kengaydi.

Helsinki korpusi mavjud bo'lib, uning tarkibiy qismi sifatida diaxronik korpus ham yaratilgan va u qadimiy, o'rta va ilk zamonaviy (Britaniya) ingliz tili davrlarida yozilgan matnlarning asosiy tanlovini va mintaqaviy turkumlarga ajratilgan qo'shimcha qismni o'z ichiga oladi. Qisqacha matnlardan tashqari, qo'lyozmalardan olingan nusxalarning uzunligi 2000 dan 10 000 so'zgacha yetadi. Hozirgi vaqtda korpusning qadimiy ingliz tili bo'limida 413 ming 300 so'z, o'rta ingliz bo'limida 608 ming 600 so'z va zamonaviy ingliz bo'limida 551 ming so'z, jami olib qaraganda, 1 million 572 ming 800 so'z mavjud (1-jadvalga qarang). Ushbu jadval <http://icame.uib.no/hc/> saytidan olindi, tarjima esa muallifdan.

1-jadval.

Helsinki korpusi tarkibidagi diaxronik korpusning tarkibiy/bazaviy tuzulishi

Davriy bo'linish	So'zlar soni	%
<i>Qadimiy ingliz tili</i>		
I 850	2 190	0.5
II 850-950	92 050	22.3
III 950-1050	251 630	60.9
IV 1050-1150	67 380	16.3
Umumiy	413 250	100.0
<i>O'rta ingliz tili</i>		
I 1150-1250	113 010	18.6
II 1250-1350	97 480	16.0
III 1350-1420	184 230	30.3
IV 1420-1500	213 850	35.1
Umumiy	608 570	100.0
<i>Ilk zamonaviy ingliz tili</i>		
I 1500-1570	190 160	34.5
II 1570-1640	189 800	34.5
III 1640-1710	171 040	31.0
Umumiy	551 000	100.0

Yana bir mashhur *bugungi og'zaki ingliz tilining diaxronik korpusi* (The Diachronic Corpus of Present-Day Spoken English) to'g'risida to'xtab o'tishni joiz deb topdik. Ushbu korpus o'zi mansub bo'lgan turdagi 1990-yillarning boshida to'plangan *ICE-GB* korpusidan 400 mingdan ortiq va 1960-yillarning oxiridan toki 1980-yillarning boshigacha bo'lgan oraliqqa tegishli London-Lund Korpusidan 400 ming so'zlarini o'z ichiga oladi. Orfografik transkripsiyalar bir xil mezonlarga muvofiq me'yorlashtirilgan va izohlangan. *Bugungi og'zaki ingliz tilining diaxronik korpusi* (The Diachronic Corpus of Present-Day Spoken English) og'zaki ingliz tili grammatikasida so'nggi o'zgarishlarni o'rganish uchun beqiyos manbadir.

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 tamonlarini inobatga olish o'rinli. Ularni quyidagicha, deb hisoblaymiz:

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

2. Dunyo tillari orasida ko'plab tillar tarixiy rivojida siyosiy noturg'unlik oqibatida yoki biror tarixiy voqelik ortidan yozuv grafikasida alifbo 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 korpuslarda manba kiritishda ayni shu holat kuzatiladi hamda yechim yuqorida ta'kidlangandek bo'lsa maqsadga muvofiq bo'ladi.

3. Diaxronik korpuslar tarixiylik nuqtayi nazaridan barcha davrlarda ham birdek manbaning mavjud emasligi sababli teng hajmda matnlar kiritish borasidagi me'yorga har doim ham amal qila olmaydi. Biroq, agarda 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 belgilan olingan davriylik oz muddatni emas, aksincha, kamida 1 dekadali vaqt oralig'ini qamrab olishi kerak. Aks holda, til rivoji tadqiqiga qaratilgan izlanishlar samarasi yuqori darajada bo'lmasligi ehtimoli yuzaga keladi.

Demak, diaxronik korpus tushunchasi korpus lingvistikasi fan tarmog'ida katta ahamiyatga ega bo'lib, korpus tahlil metodlaridan foydalanishning yangi bosqichini yuzaga chiqargan. Bizningcha, diaxronik korpuslardan tilshunoslikning quyidagi sohalarida qo'llash orqali, javob talab ilmiy-nazariy savollarga yangicha yondashish mumkin:

– Birinchidan, diaxronik korpuslar lug'atshunoslik sohasida qo'llanilib, unda izohtalab so'zlarning nafaqat qo'llanishiga oid xronologik ma'lumot, qolaversa, korpus tarkibidagi manbaning tizimli tanlanganligiga qarab, uning etimologiyasiga ham nazar tashlay oladi.

– Ikkinchidan, diaxronik korpuslar bir vaqtning o'zida tarixiy korpus tushunchasi bilan bir xil maqomda qo'llanadi hamda til rivoji tadqiqini o'rganishga alohida e'tibor qaratadi. Jumladan, leksikologiyaning eskirmas muammolari tarkibiga kiruvchi neologizmlar, axraizmlar, 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, hissiy tuyg'ularning aks etish holatlarini 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 istovchi 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 qila oladi hamda uni noshirlik ishlarida qo'llash orqali shu ruhni saqlab qoladi.

– Beshinchidan, bizningcha eng muhimi, bugungi kundagi rivojlanib borayotgan tezkor zamonda, "kuchli" xalqlar va ularning tillari yashab qoladi. Bundan keskin fikrni bildirar turib, o'ylaymizki, diaxronik korpuslar yaratish orqali tillararo interferensiya chiziqlarini statistik tahlil qilish orqali til sohiblarini ushbu oqimdan haminqadar saqlashga qaratilgan davlat dasturlarini ishlab chiqish ahamiyatini anglash mumkin.

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 shu kabi ko'plab 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'zilarini quyida sanab o'tish mumkin (2-jadvalga qarang).

2-jadval.

Inglizcha tarixiy/diaxronik korpuslar

<i>ARCHER Corpus (A Representative Corpus of Historical English Registers)</i>	Ilk zamonaviy ingliz tilidan misollarni qamrab olgan 1650–1990 yillar oralig'idagi 7ta tarixiy davrdan olingan yozma va “nutqqa asoslangan” janrlardan shakllangan jami 955ta matn, ya'ni 1 789 309ta so'zni o'z ichiga oladi. Uppsala, Xelsinki, Frayburg, Haydelberg, Lankaster, Manchester va Michigan universitetlari bilan korpus qamrovini kengaytirish bo'yicha doimiy hamkorlikdagi tadqiqot ishlari olib borilmoqda.
<i>Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English</i>	Ingliz matnlarining Xelsinki korpusining qadimiy ingliz bo'limi uzunroq matnlardan namunalar uzunligi 5000dan 10000 so'zgacha bo'lgan qadimiy ingliz tili tarkibidan olingan man-balari 106210 so'zni o'z ichiga oladi. Korpus matnlari kompo-zitsiya, mualliflar va janrlarning bir qator sanalarini ifodalaydi. Matnlar sintaktik va morfologik jihatdan teglangan va har bir so'z izohlanadi.
<i>Century of Prose Corpus</i>	1680–1780-yillar oralig'ida yashab ijod qilgan jami 120 muallif ijodidan nashr etilgan adabiy va adabiy bo'lmagan ingliz tilidagi yarim million so'zdan tarkib topgan.
<i>Complete Corpus of Old English</i>	Toronto universitetida tuzilgan barcha qadimiy ingliz tili davrida yaratilgan 3022ta matnlarni ifodalovchi korpus.
<i>Corpus of English Dialogues, 1560-1760 (CED)</i>	Ilk zamonaviy ingliz tilidagi nutq shakllarini ifoda etuvchi 177ta matndan tashkil topgan korpus 1 million 200 ming so'zni o'z ichiga oladi. CED manbalari matn sifatida dialog, sud jarayoni, jumladan, guvohlarning so'zlari, drama, komediya, didaktik asarlar va badiiy adabiyot namunalaridan iborat.
<i>Corpus of Newsbooks</i>	1653-yil dekabrda 1654-yil maygacha nashr etilgan <i>Tomason</i> nashrlari barcha gazetalardan olingan taxminan 800 000 so'zdan tarkib topgan matn.
<i>Corpus of Early English Correspondence (CEEC) & the Parsed Corpus of Early English Correspondence (PCEEC)</i>	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 “Ilk yozishmalar korpusi” 50 ming atrofida so'z; 1681–1800-yillarni qamrab olgan oraliqdan jami 2 million 200 ming so'z kiritilgan.

<i>Corpus of Early English Medical Writing & Corpus of Middle English Medical Texts (MEMT)</i>	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 mingga yaqin so‘zlarni o‘z ichiga oladi.
<i>The Málaga Corpus of Early English Scientific Prose</i>	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. U tibbiyotga oid qo‘lyozmalar, ya‘ni nazariy risolalar, jarrohlikka oid yozma ma‘lumot va retseptlari (dori qog‘ozi) to‘plamining taksonomiyasi nuqtayi nazaridan me‘yorlashtirilgan va u diaxronik nuqtayi nazardan lingvistik o‘zgarishlarni o‘rganish uchun ayniqsa foydali vosita sifatida ishlab chiqilgan. Korpus uch xil holatda taqdim etiladi: oddiy matn, me‘yorlashtirilgan va morfologik teglangan.
<i>Newdigate Newsletters</i>	1674–1692-yillarga tegishli qo‘lyozma nashrlardan jami 750 ming so‘zdan iborat.
<i>TIME Magazine Corpus</i>	TIME jurnalining 1923–2006-yillar oralig‘idagi nashrlaridan 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.
<i>Zürich Corpus of English Newspapers (ZEN)</i>	1660-yillardan XX asr boshlarigacha bo‘lgan vaqt oralig‘idagi London gazeta nashrlaridan tashkil topgan [https://martinweisser.org/corpora_site/historical_corpora.html].

Yuqorida sanab o‘tilgan lisoniy korpuslar ingliz korpus lingvistikasining yutuqlari sifatida e‘tirof etiladi.

Ingliz va o‘zbek mediamatnlar diaxronik korpusini yaratishga bo‘lgan ehtiyoj tavsifi

O‘zbek tili doirasida ham bu borada bir qator ilmiy izlanishlar amalga oshirilmoqda. Ular sirasiga quyidagilarni kiritish mumkin:

1. O‘zbek tilining ilk til korpusi sifatida qaralayotgan *uzbekcorpus.uz* elektron korpusi filologiya fanlari doktori (DSc), dotsent 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 korpusining joriy holatida 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.

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 konstruktsiya 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. Tizim mualliflari qatoriga t.f.f.d. (PhD), dots. 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, Leipzigidagi Saksoniya fanlar va gumanitar fanlar akademiyasi va Amaliy informatika instituti loyihasi asosida tuzilgan bo'lib, o'z ichiga jami 663 ming 119ta gapdan iborat 9 million 256 ming 1ta so'zni qamrab oladi. Ushbu korpus avtomatik internet materiallari tarkibidan jamlanganligi tufayli o'zbek tilining bugungi kundagi muammolaridan biri bo'lgan alifbo, 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.

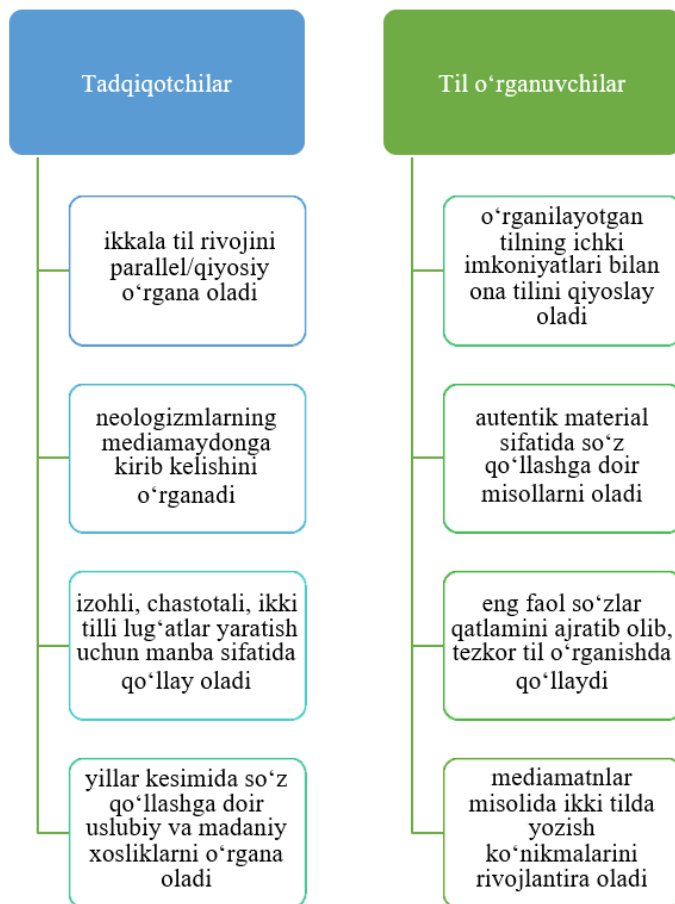
4. O'zbek tili milliy korpusi dasturiy mahsulot sifatida O'zbekiston Respublikasi Innovatsion vazirligining 2021-yilgi 30-turdagi "O'zbek tilining milliy korpusini loyihalash va dasturiy majmua ishlab chiqish" (2020-2022) mavzusidagi amaliy loyiha doirasida bajarilgan. O'rnatilgan tartibda o'tkazilgan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

Yuqorida sanab o'tilgan o'zbek tilining korpuslari barchasi diaxron korpuslar sirasiga kirmaydi. Aksincha, korpuslar xronologik tartiblanmaganligi, tarixiy ketma-ketlikda matnlarning kiritilmaganligi bois ularni sinxron korpuslari ruknida deyish mumkin.

O'zbek tilining diaxronik korpusi 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* korpuslarnig mavjud emasligi tadqiqotchilarni bu boradagi xulosalarining 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.

Ingliz tilining diaxronik korpuslari va ularning imkoniyatlari to'g'risida, umuman olganda, sohadagi yutuqlar borasida yuqorida sanab o'tildi. Shu o'rinda mavjud tajribani qo'llagan holda o'zbek tilining ham diaxronik korpusni yaratishga nazariy asos sifatida tadqiqot ishi amalga oshirilmoqda. Bunda ikki tilli, ya'ni o'zbek va ingliz tillarining diaxronik korpusini ishlab chiqish ishimizning maqsadi etib belgilangan. Yozma matnlarning keng ko'lamda topilish mumkin bo'lgan manba bu mediamuloqot vositalari hisoblanar ekan, bizningcha, yaratilajak korpusimizning o'zagini aynan mediamatnlar asosida shakllantirish joiz. Demak, nazarda tutilgan korpusning to'planishi nazariyotchi olimlar va til o'rganuvchilarga quyidagi imkoniyatlarni yaratishi mumkin (2-rasmga qarang):

2-rasm.

Ingliz va o'zbek mediamatnlar diaxronik korpusining ilmiy amaliy sohadagi dolzarbligi

Diaxronik/tarixiy korpuslar soni va hajmi bo'yicha o'sishda davom etmoqda va ularni tahlil qilish usullari ham samaraliroq va yanada murakkablashmoqda. Shu bilan birga, murakkab diaxronik korpuslarni hisobga olmaganda, diaxronik ma'lumotlar tadqiqotchiga doimo muhim muammolarni yuzaga chiqarib beradi. Diaxronik korpus tilshunosligining so'nggi yutuqlari borasidagi haqli ravishda ortib borayotgan qiziqishlarga qaramay, korpus dizayni va o'tmishdagi til tizimlarining kodlanishini va o'zgarishlarni tushuntirish uchun yangi turdagi noqardosh va qardosh tillar diaxronik korpuslarini yaratish kerak. Umuman olganda, diaxronik manbalardan foydalanganib olib boriladigan qiyosiy plandagi tadqiqotlar doimo o'zini oqlaydi. Zero, korpus uchun yig'iladigan lisoniy baza ham hajman, ham tarixiy davriga ko'ra barcha me'yorlarga javob beradi.

XULOSA

Ko'rinib turibdiki, ingliz va o'zbek mediamatnlar diaxronik korpusining yaratilishi ortidan ikkita bir-biriga qarindosh bo'lmagan, areal o'zaro ayri hududlarda

qo‘llaniluvchi ingliz va o‘zbek tillari o‘rtasidagi farqli va o‘xshash xarakterdagi ilmiy farazlarni asoslash imkoniyati yuzaga chiqadi. Korpusda mediamatnlar vositasida ma’lum bir yillar oralig‘idan olingan tarixiy nashrlarni qamrab oluvchi manba jamlanar ekan, o‘zining birorta diaxronik korpusiga ega bo‘lmagan o‘zbek tilining shu turdagi til korpusi yaratilibgina qolmay, bu borada katta tajribaga ega bo‘lgan ingliz tilining jalb qilinishi globallashuv davrida tillararo interferensiya, til leksik qatlamining boyishi va tillar rivojini tadqiq qilishda asosli manba sifatida xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR

1. Begmatova, G.H. (2021). O‘zbek tili milliy korpusida idiomalar bazasini yaratish. *Filol. fan. fals. dok. diss.* Termiz.
2. Dobrik, N. (2009). Extracting information from large digital corpora – A case study in quantitative methods in linguistics. *Primenjena Lingvistika*, 10, 103–113.
3. Gries, St. (2006). Introduction. In St.Gries & A.Stefanowitsch (Eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis* (pp. 1–17). Berlin/New York: Mouton de Gruyter.
4. Hamroyeva, Sh.M. (2018). *O‘zbek tili mualliflik korpusi tuzishning lingvistik asoslari. Filol. fan. fals. dok. diss.* Buxoro.
5. Hilpert, M., & Gries, S.Th. (2016). Quantitative approaches to diachronic corpus linguistics. In M.Kytö & P.Pahta (Eds.), *The Cambridge Handbook of English Historical Linguistics* (pp. 36–53). Cambridge: Cambridge University Press.
6. Hunston, S. (2006). Corpus Linguistics. In K.Brown (Ed.), *The Encyclopedia of Language and Linguistics* (2nd ed.) (pp. 234–248). Amsterdam: Elsevier Science. <https://doi.org/10.1016/B0-08-044854-2/00944-5>.
7. Irisqulov, A.T. (2009). *Tilshunoslikka kirish*. Toshkent: “Yoshlar matbuoti” bosmaxonasi.
8. Karimov, R.A. (2022). O‘zbek-ingliz parallel korpusini tuzishning lingvistik va dasturiy masalalari. *Filol. fan. fals. dok. diss.* Buxoro.
9. Knyazeva, Y.O. (2009). Strategii sinonimizacii v russkoj i britanskoj lingvokul'turah. *Diss. kand. fil. nauk.* Kursk.
10. Kytö, M. (1996). *Manual to the diachronic part of the Helsinki corpus of English texts: Coding conventions and lists of source texts* (3rd ed.). Helsinki: University of Helsinki.
11. McEnery, T. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
12. McEnery, T., & Hardie, A. (2012). *Corpus linguistics*. UK: Cambridge University press.
13. McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburg: Edinburgh University Press.
14. McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
15. Raxmanova, A.A. (2021). O‘zbek tili milliy korpusini yaratishdagi kompyuter usullari. *Filol. fan. fals. dok. diss.* Toshkent.
16. Ramat, A.G., Mauri, C., Molinelli, P. (2013). *Synchrony and diachrony: A dynamic interface*. Amsterdam: John Benjamins Publishing Company.
17. Sinkler, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
18. Tomsen, V. (1938). *Istoriya yazykovedeniya do kontsa XIX veka: Kratkiy obzor osnovnykh momentov* (3 izd.). Moskva: URSS.
19. Xoliyorov, U.M. (2021). O‘zbek tili ta’limiy korpusini tuzishning lingvistik asoslari. *Filol. fan. fals. dok. diss.* Termiz.
20. <http://uzbekcorpora.uz/buyurtmachi>.
21. https://corpora.uni-leipzig.de?corpusId=uzb_community_2017.
22. https://martinweisser.org/corpora_site/historical_corpora.html.

23. <https://qomus.info/oz/encyclopedia/d/diaxroniya/>.
24. <https://ruscorpora.ru/>.
25. <https://uzbekcorpus.uz/aboutCorpus>.
26. <https://uzschoolcorpara.uz/#home>.
27. <https://www.bibalex.org/ica/en/About.aspx>
28. <https://www.english-corpora.org/coha/>.
29. <https://www.ucl.ac.uk/english-usage/projects/dcpse/index.htm>.
30. <https://www.ucl.ac.uk/english-usage/projects/ice.htm>.

REFERENCES

1. Begmatova, G.H. (2021). Creating a database of idioms in the national corpus of the Uzbek language. *PhD thes. philol. scien.* Termez.
2. Dobrik, N. (2009). Extracting information from large digital corpora – A case study in quantitative methods in linguistics. *Primenjena Lingvistika*, 10, 103–113.
3. Gries, St. (2006). Introduction. In St.Gries & A.Stefanowitsch (Eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis* (pp. 1–17). Berlin/New York: Mouton de Gruyter.
4. Hamroyeva, Sh.M. (2018). Linguistic foundations of creating an author corpus of the Uzbek language. *PhD thes. philol. scien.* Bukhara.
5. Hilpert, M., & Gries, S.Th. (2016). Quantitative approaches to diachronic corpus linguistics. In M.Kytö & P.Pahta (Eds.), *The Cambridge Handbook of English Historical Linguistics* (pp. 36–53). Cambridge: Cambridge University Press.
6. Hunston, S. (2006). Corpus Linguistics. In K.Brown (Ed.), *The Encyclopedia of Language and Linguistics* (2nd ed.) (pp. 234–248). Amsterdam: Elsevier Science. <https://doi.org/10.1016/B0-08-044854-2/00944-5>.
7. Iriskulov, A.T. (2009). *Introduction to linguistics*. Tashkent: ‘Youth Press’ Printing House.
8. Karimov, R.A. (2022). Linguistic and programming issues of creating a Uzbek-English parallel corpus. *PhD thes. philol. scien.* Bukhara.
9. Kholiyorov, U.M. (2021). Linguistic foundations of creating the educational corpus of the Uzbek language. *PhD thes. philol. scien.* Termez.
10. Knyazeva, I.O. (2009). Synonymization strategies in Russian and British linguistic cultures. *Diss. cand. philol. scien.* Kursk.
11. Kytö, M. (1996). *Manual to the diachronic part of the Helsinki corpus of English texts: Coding conventions and lists of source texts* (3rd ed.). Helsinki: University of Helsinki.
12. McEnery, T. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
13. McEnery, T., & Hardie, A. (2012). *Corpus linguistics*. UK: Cambridge University press.
14. McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh: Edinburgh University Press.
15. McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
16. Rakhmanova, A.A. (2021). Computer methods in creating the national corpus of the Uzbek language. *PhD thes. philol. scien.* Tashkent.
17. Ramat, A.G., Mauri, C., Molinelli, P. (2013). *Synchrony and diachrony: A dynamic interface*. Amsterdam: John Benjamins Publishing Company.
18. Sinkler, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
19. Thomsen, V. (1938). *History of linguistics until the end of the twentieth century: A brief overview of the main points* (3rd ed.). Moscow: URSS.
20. <http://uzbekcorpora.uz/buyurtmachi>.
21. https://corpora.uni-leipzig.de?corpusId=uzb_community_2017.
22. https://martinweisser.org/corpora_site/historical_corpora.html.
23. <https://qomus.info/oz/encyclopedia/d/diaxroniya/>.

24. <https://ruscorpora.ru/>.
25. <https://uzbekcorpus.uz/aboutCorpus>.
26. <https://uzschoolcorpara.uz/#home>.
27. <https://www.bibalex.org/ica/en/About.aspx>
28. <https://www.english-corpora.org/coha/>.
29. <https://www.ucl.ac.uk/english-usage/projects/dcpse/index.htm>.
30. <https://www.ucl.ac.uk/english-usage/projects/ice.htm>.