



Received: November 14, 2025
Accepted: December 22, 2025
Available online: December 25, 2025

Nozimjon Ataboyev

Filologiya fanlari doktori (DSc), professor
Buxoro davlat universiteti
Buxoro, O'zbekiston

TIL EVOLYUTSIYASI TAHLILIDA KORPUS LINGVISTIK METODLARNING AHAMIYATI

ANNOTATSIYA

Mazkur maqolada korpus lingvistikasi fan tarmog'ining shakllanishi, rivojlanish bosqichlari hamda lisoniy korpuslar yaratish amaliyotining jahon tilshunosligidagi o'rni atroficha tahlil qilinadi. Tadqiqotda yetakchi xorijiy va mahalliy ilmiy markazlar tajribasi asosida korpus lingvistik metodlarning til tadqiqidagi samaradorligi yoritilib, ularning til taraqqiyoti, diaxronik o'zgarishlar, madaniyatlararo interferensiya va til ta'limidagi ahamiyati ochib beriladi. Lancaster University, University of Birmingham, Macquarie University, University of Sydney kabi nufuzli oliy ta'lim muassasalarida olib borilgan korpus lingvistik tadqiqotlar misolida korpus tahlillarining lug'atshunoslik, stilistika, psixolingvistika, kognitiv lingvistika, tarjima va til o'qitish metodikasidagi imkoniyatlari ko'rsatib beriladi.

Maqolada korpus lingvistikasining tarixiy taraqqiyot bosqichlari – dastlabki konkordanslarga asoslangan qo'lyozma korpuslardan boshlab mega va gigakorpuslarga bo'lgan rivojlanish jarayoni ilmiy manbalar asosida izohlanadi. Xususan, British National Corpus, International Corpus of English, Google Books Ngram, COCA kabi yirik korpuslar misolida katta hajmdagi lingvistik ma'lumotlarning til tadqiqidagi roli tahlil qilinadi. Shuningdek, maxsus va kichik korpuslarning representativlik tamoyillari hamda ularning muayyan nutq turlarini o'rganishdagi ustun jihatlari asoslab beriladi.

Tadqiqotda rus, arab, koreys, xitoy va hind tilshunosligida yaratilgan milliy va maxsus korpuslar tajribasi ham yoritilib, ko'p tilli

Nozimjon Ataboev

Professor, Doctor of Sciences in Philology (DSc)
Bukhara State University
Bukhara, Uzbekistan
E-mail: n.b.ataboyev@buxdu.uz
ORCID iD: 0000-0002-9756-6849

TIL EVOLYUTSIYASI TAHLILIDA KORPUS LINGVISTIK METODLARNING AHAMIYATI

ABSTRACT

This article provides a comprehensive analysis of the formation of the discipline of corpus linguistics, its stages of development, and the role of the practice of creating language corpora in world linguistics. The study, based on the experience of leading foreign and domestic scientific centers, highlights the effectiveness of corpus linguistic methods in language research, and reveals their significance in language development, diachronic changes, intercultural interference, and language education. The possibilities of corpus analysis in lexicography, stylistics, psycholinguistics, cognitive linguistics, translation, and language teaching methodologies are demonstrated on the example of corpus linguistic studies conducted at such prestigious higher education institutions as Lancaster University, University of Birmingham, Macquarie University, and University of Sydney.

The article explains the historical stages of corpus linguistics development - from manuscript corpora based on initial concordances to mega- and gigacorpora - based on scientific sources. In particular, the role of large-scale linguistic data in language research is analyzed using the example of large corpora such as the British National Corpus, International Corpus of English, Google Books Ngram, COCA. The principles of representativeness of special and small corpora and their advantages in studying certain types of speech are also substantiated.

The study also highlights the experience of national and special corpora created in Russian, Arabic, Korean, Chinese and Indian linguistics, and emphasizes the relevance of creating

jamiyatlarda korpus yaratishning dolzarbligi ta'kidlanadi. Ayniqsa, o'zbek tilining elektron, milliy va ta'limiy korpuslarini yaratish borasida olib borilayotgan ishlar tahlil qilinib, o'zbek tilining diaxronik korpusini shakllantirish zarurati ilmiy jihatdan asoslanadi. Maqola natijalari korpus lingvistikasi asosida til rivojini tizimli o'rganish uchun muhim nazariy va amaliy ahamiyatga ega.

Maqolada til taraqqiyotini tahlil qilishga oid nazariy muammo korpus lingvistikasi metodologiyasi nuqtayi nazaridan o'rganilgan. Tadqiqot jarayonida ingliz, rus, koreys va o'zbek tillarining mavjud korpuslarining ichki imkoniyatlari qidiruv natijalarini tahlil qilish asosida atroflicha yoritilib, tegishli ilmiy xulosalar chiqarilgan.

Kalit so'zlar: korpus lingvistikasi, lisoniy korpus, diaxronik korpus, milliy korpus, mega va gigakorpuslar, korpus tahlili, til taraqqiyoti, madaniyatlararo interferensiya, leksikografiya, til ta'limi, korpusga asoslangan tadqiqot, o'zbek tili korpuslari.

corpora in multilingual societies. In particular, the work on the creation of electronic, national and educational corpora of the Uzbek language is analyzed, and the need to form a diachronic corpus of the Uzbek language is scientifically substantiated. The results of the article are of significant theoretical and practical importance for the systematic study of language development based on corpus linguistics.

The article studies the theoretical problem of analyzing language development from the perspective of corpus linguistics methodology. In the course of the research, the internal capabilities of existing corpora of English, Russian, Korean and Uzbek languages were thoroughly investigated based on the analysis of search results, and relevant scientific conclusions were drawn.

Key words: corpus linguistics, corpus, diachronic corpus, national corpus, mega and gigacorpora, corpus analysis, language development, intercultural interference, lexicography, language education, corpus-based research, Uzbek language corpora.

KIRISH

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 turli-

cha qo'llanilish sabablarini asoslash (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 taqdiqida, 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.

TADQIQOT METODLARI

Muammoning o'rganilganlik darajasiga nazar tashlaydigan bo'lsak, dunyoda shu turdagi tadqiqotlarga kenng yo'l ochilayotganini ko'rishimiz mumkin. Xususan, jahon tilshunosligida G.Lich birinchilardan bo'lib, korpus lingvistikasining paydo bo'lishi, taraqqiyoti haqida ilmiy izlanish olib bordi [31]. Zamonaviy korpus lingvistikasi shakllanishiga doir tadqiqotlar S.Hyuston, D.Biber, S.Grays, M.Halliday va V.Teubert kabi ingliz olimlari tomonidan amalga oshirilgan [19; 5; 15; 16]. Rus olimlari V. Baranov va V.Zaxarov ishlarida korpus lingvistikasi spetsifikasiga [45; 52], N.Vladimov, S.Vadyaev, O.Gorkina, V.Mamontova va A.Magametova kabi olimlar "korpus lingvistikasi", "lisoniy korpus" tushunchasi va uni tasniflash muammolariga atroflicha yondashganlar [48; 47; 49; 55; 54]. Tojikistonlik olimi D. Djafarova tadqiqotida korpus lingvistikasining kompyuter lingvistikasi va boshqa o'zaro o'xshash sohalar bilan qiyosiy tahlili amalga oshirilgan [50].

Korpus tahlillaridan lug'atshunoslikda foydalanish borasidagi ishlar A. Kilgarriif, F. Karlson va I. Sojenin tadqiqotlarida kuzatilgan bo'lsa [24; 22; 61], A. Romanov va V. Plungyan kabi olimlar "korpus" tushunchasini matn tahlillarida qo'llash metodini ishlab chiqishgan [59; 57].

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 tarkibidan semantik gradatsiyalarni topish [51], G. Ergasheva BNC korpusi asosida genderga oid konseptual metaforalarni tadqiq qilish [67], G. Sobirova esa BNC korpusini tillarni o'qitishda qo'llashga e'tibor qaratgan [60]. Sh. Hamroeva mualliflik korpusini yaratish muammosiga yechim izlagan [64]. 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 kabi olimlarning salmoqli ilmiy-tadqiqotlarini misol qilib keltirish o'rinlidir [1; 63;

62; 46; 58; 65; 66; 53]. 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 korpusi (<http://uzbekcorpora.uz>), O‘zbek tilining ta’limiy korpusi (<https://uzschoolcorpara.uz>) kabilarni salmoqli dasturiy mahsulotlar sifatida qayd etish mumkin.

NATIJALAR VA MUNOZARA

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

Korpus lingvistikasi hozirgi taraqqiyotiga qadar bir qancha davrlarni bosib o‘tdi, mustaqil fan tarmog‘i sifatida o‘tgan asr 70-yillarining oxirida shakllandi hamda uning asosi hisoblangan tilshunoslikka oid tadqiq usul, metod va tahlil vositalari XII asrdan buyon ma’lum [7]. Korpuslarning rivojlanishida matn hajmi va ularni saralash tamoyillarining yuzaga kelishi bir qancha davrlarni qamraydi. Dastlab, elektron shakldagi korpuslarning vujudga kelishigacha bo‘lgan davr XIII asrdan 1960-yillarning boshlarigacha bo‘lgan vaqtni o‘z ichiga oladi. Texnologiyaning rivojlanishi korpus lingvistikasining ildam qadam qo‘yishiga turtki bo‘ldi, ya’ni tadqiqotchi va tilshunoslar juda ko‘p vaqtini so‘zlarni saralab chiqishga sarflash majburiyatidan xalos bo‘ldi. Bu davr 1970-yillar, ya’ni dastlabki shaxsiy kompyuterlar yaratilishi bilan bog‘lanadi. Korpus tilshunosligining tadqiq predmetini aniqlash tillarning real hayotda qo‘llanilish imkoniyatlari va til foydalanuvchilarining tilni qay yo‘sinda qo‘llashlarini doimiy ravishda o‘rganib, tadqiq qilish orqali amalga oshiriladi. Bu jarayon tilshunoslarning muammoga turli nuqtayi nazarlari bilan yanada mukammallashib boradi.

Korpus lingvistikasi xususida so‘z borar ekan, sohaning obykti sifatida bevosita o‘rganiladigan va matnlar bilan ishlashning zamonaviy yondashuvi bo‘lgan *korpus* termini alohida izoh talab etadi.

Korpus so‘zining lug‘aviy ma’nosi, dastavval, tor doirada qo‘llanilgan: diniy, adabiy janrdagi muayyan yozuvchi asarlarining tarkibi alifbo tartibida, ya’ni *concordance* – konkordans qatorlarida joylashtirilishi tushunilgan. Korpusning bu ma’noda ishlatilishi elektronlashtirilgan korpuslar vujudga kelgungacha bo‘lgan davrni qamrab oladi. Korpuslar bu davr mobaynida diniy, adabiy va leksikografik tadqiqotlar uchun shakllangan; bu jarayon uzoq muddat, ko‘p mehnatni talab qiladigan qo‘l mehnatiga ehtiyoj tug‘dirgan.

Korpuslar elektron shaklga kelgungunga qadar bo‘lgan davrda matn tarkibida so‘z va iboralar qaysi o‘rinlarda foydalanilganligini ko‘rsatuvchi o‘ziga xos belgilar-

dan foydalanilgan. Shunday konkordans qatorlarning dastlabkilaridan biri – Bibliya muqaddas kitobiga asoslanib yaratilgan bo‘lib, konkordans qatorlarida barcha so‘zlar alifbo tartibida joylashtirilgan; ular qo‘llangan joylar ko‘rsatib borilgan, bu esa, o‘z navbatida, bir so‘zning qanday ma‘nolarda qo‘llanganligini o‘rganish imkonini bergan. A. Kruden ushbu Bibliya konkordans qatorlariga *lug‘at* yoki *teglangan baza* sifatida qarashni o‘rtaga tashlaydi [8: P. 154]. Olimning fikriga qo‘shilgan holda, alifbo ketma-ketligida tartiblangan, so‘zlarning faollik darajasi ko‘rsatilgan har qanday bazani chastotali lug‘atlar turkumiga kiritishni ma‘qullaymiz. Bunday tajribalar insonning diniy va badiiy asarlarga bo‘lgan qiziqishini aks ettiribgina qolmay, asarni o‘qishdan oldin o‘quvchiga so‘zlarning ma‘nosini anglashga imkon berib, o‘qish jarayonida kuzatiladigan ehtimoliy qiyinchiliklarning oldini oladi.

XXI asrning boshlarida konkordans dasturlari yuqori funksionalligi bilan ajralib turadi: bitta dastur tezda so‘zlar ro‘yxatini, kollokatsion birliklarini tuzish, so‘zlarning faollik darajasini aniqlash kabi tahlillarni amalga oshirishga qodir [35: P. 35].

1990-yillarning boshidan texnik imkoniyatlar olimlarga yirik hajmli korpuslarni tuzish va rivojlantirishga imkon berdi. Bu davrda yaratilgan korpuslar o‘z oldiga tilning yozma va og‘zaki nutqda namoyon bo‘ladigan, shu bilan tilning barcha xilmaxilligini aks ettiradigan keng ko‘lamli shakllarini qamrab olishni maqsad sifatida belgilab olgan. Og‘zaki korpuslarni fonetik, morfologik, leksik, sintaktik va diskursiv darajada avtomatik ravishda teglash mumkin bo‘ldi. Bundan tashqari, konkordanslarni avtomatlashtirilgan tarzda qayta ishlash bo‘yicha bir qator dasturlar paydo bo‘ldi.

G. Kennedy, P. Baker, A. Hardie, T. McEnery kabi olimlar kompyuter qo‘llanilmagan [23; 4], inson resursini keng qamrovda jalb qilishga ehtiyoj mavjud bo‘lgan davrni birinchi avlod korpuslari deb atashsa, 80-yillarning oxiridan boshlab rivojlangan korpuslarni ikkinchi avlod korpuslari yoki mega-korpuslar deb nomlashadi, chunki ularning hajmi 100 millionga yaqin so‘zlarni qamragan. Ushbu korpuslarga *Longman Corpus Network*, *Longman Corpus Network (1991)*, *Bank of English*, *The Bank English*, *BoE (1993)*, *British National Corpus*, *BNC (1994)*, *The American National Corpus*, *ANC (2008)* kiradi. Keyinchalik, ingliz tilining og‘zaki korpusi ham Britaniya milliy korpusi (*British National Corpus*, *BNC*)ning so‘zlashuv qismiga kiritildi [6].

O‘z davrining yana bir mashhur loyihasi 1996-yilda S. Grinbaum boshchiligida London Universitet Kollejiida ishlab chiqilgan ingliz tilining milliy variantlari korpusi – *The International Corpus of English (ICE)* edi. Loyihaning maqsadi ingliz tilining mintaqalararo variantlari matnlarini to‘plash bo‘lgan. Subkorpuslarga ingliz tilining mintaqalararo variantlarining og‘zaki va yozma matnlari kiritilgan. Bu mintaqalar: Britaniya (ICE-GB), Sharqiy Afrika, Hindiston, Yangi Zelandiya, Singapur, Kanada, Gonkong, Yamayka, Filippin, AQSh, Kamerun, Fidji, Irlandiya, Keniya, Malta, Malayziya, Pokiston, Syerra-Leone, Shri-Lanka, Trinidad va Tobago. Respondentlar ingliz tilida so‘zlashadigan maktabda o‘rta maktab ma‘lumotiga ega 18 yoshdan katta shaxslar bo‘lgan. Barcha subkorpuslarda matnlarning 60 foizini yozma matnlar va 40 foizini og‘zaki transkriptlardan iboratligi ta‘minlangan. Dialogik nutq subkorpusi og‘zaki nutqning quyidagi janrlarini o‘z ichiga oladi: shaxsiy suhbatlar (shaxsiy

uchrashuv va telefon orqali suhbatlar) va ommaviy (dars, radio va televideniyaedagi suhbatlar, radio intervyu, parlament muhokamalari, biznes muzokaralari). Monolog nutq subkorpusi ikki qismga bo'linadi: birinchisi, spontan nutq manbalarini o'z ichiga oladi (sharh, namoyish va suddagi chiqishlar); ikkinchi qismda oldindan tayyorlangan nutqlar kiritilgan (televideniya va radio yangiliklari, televizion va radiosuhbatlar (tok-shoular). 2006-yilga kelib nutqning audioyozuvlari korpusga kiritila boshlandi [29]. ICE korpusi me'yorlashtirilgan diaxronik korpuslar turiga kirganligi bois kengroq tahlilga muhtoj.

Shunday qilib, ikkinchi avlod korpuslarining hajmi kamida yuz million token bo'lgan korpuslar bo'lib, ularning maqsadi yozma va og'zaki nutqning barcha turlarini namoyish etishdir. Korpus tuzuvchilar til sohiblarining turli qatlami uchun og'zaki va yozma nutqning iloji boricha ko'proq janr va uslublarini taqdim etishga intilashadi. Odatda, bu – korpus yaratish talablariga muvofiq onlayn tarzda to'plangan va teglangan korpuslarga xos tushunchadir. Milliy korpus tuzuvchidan uzoq vaqt, sinchkovlik talab etadi; matn tanlashning reprezentativlik tamoyillari va ikkinchi avlod korpuslariga xos bo'lgan qoidalar asosida tuziladi. 1990-yillarda British National Corpus yangi milliy korpus namunasi sifatida ishlatilgan. Biroq tadqiqot mavzusida nazarda tutilgan diaxronik korpuslar me'yorlariga javob bermasligi aniqlandi. Bizningcha, korpus bazasi xronologik uzluksizlikdagi muayyan vaqt oralig'ini aks ettirmas ekan, uni milliy til ifodachisi sifatida e'tirof etish o'rinli bo'lmaydi.

1987-yildan 2004-yilgacha bo'lgan davrda korpuslarni yig'ish, metamarkerlarni tuzish, shuningdek, matnlarni avtomatlashtirilgan tarzda teglash dasturlari ishlab chiqildi va sohaning yangi bosqichi – uchinchi davriylik yuzaga chiqdi. Buni *Uchinchi avlod korpuslari yoki gigakorpuslar* deya nomlashdi. 2010-yil boshlari katta texnik imkoniyatlarning paydo bo'lishi bilan ajralib turadi: to'rtinchi avlod konkordanslari – *BNCweb* (2009), *CQPweb* (2012), *SketchEngine* (2013), *Wmatrix* (2013) ishlab chiqilgan bo'lib, ular uchinchi avlod konkordanslariga o'xshashdir.

Korpuslar onlayn tizimga o'tishi ortidan qidiruvga berilgan leksik birliklarni topish tezligi oshdi, natijada, foydalanuvchilar soni kengayib bormoqda. To'g'ridan to'g'ri kirish veb-brauzer orqali onlayn qidiruvda amalga oshiriladi [35; 30]. M. Davies to'rtinchi avlod konkordanslarini 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 [10].

Korpuslar hajmining o'sish tendensiyasi 2000-yillardan keyin ham davom etdi. A. Mauranen [33], S. Kubler va X. Zinsmayster ushbu avlodni “korpus qanchalik katta bo'lsa, shuncha yaxshi” shiori bilan tavsiflaydi [27: P. 10] va L. Flowerdrew birinchi bo'lib bu davrni gigakorpuslar avlodi davri deb atadi [12: P. 11]. Ayni paytda, bir nechta yangi korpuslar (COCA, Google Books Ngram) paydo bo'ldi va ularning hajmi bir necha milliard tokenni tashkil etdi. Korpuslarning katta hajmi chastotali tadqiqotlarni keng ko'lamda o'tkazish; uch, to'rt va undan ortiq so'zlardan iborat kollokatsiyalarni o'rganishga imkon berdi. Bunday turkumlarni D. Biber va K. Hyland leksik to'plamlar (lexical bundles) deb atashadi [5; 20], bunda bitta so'z

o'zgaruvchan bo'lishi mumkin. Masalan, beshtalik so'z birikmasida: *in the beginning of the, in the end of the, in the form of the* – o'zgaruvchi uchinchi so'zdir. Keyinchalik bu kollokatsiyalar n-gramm deb nomlandi, bu yerda bigramlar ikki so'zdan, uch gramm uchta so'zdan va n-gramlar n so'zlardan iborat bo'lgan birikmalardir [38: P. 33]. Hozirgi vaqtda bunday kollokatsiyalarni aniqlash, ko'pincha internetning o'zida topish imkonini beradi, albatta, bu imkoniyat yirik – gigakorpuslarning yaratilishi orqali vujudga keldi. Bundan tashqari, bunday korpuslar n-gramm chastotasini 1800-yildan 2010-yilgacha turli vaqt oralig'idagi grafigini yaratish imkoniyatini taklif etadi. Bu esa, diaxronik korpuslar yaratishning eng muhim talablaridan biridir.

Mega va gigakorpuslarning paydo bo'lishi shuni ko'rsatdiki, katta ma'lumotni qamrab oluvchi korpuslar muayyan kasb egalarining nutqini o'rganish uchun yaroqsiz, chunki katta korpuslar hajmiga qaramay, asosan, og'zaki va yozma nutqning eng keng tarqalgan janrlari matnlarini o'z ichiga oladi. 1990-yillar oxiri 2000-yillarning boshlarida maxsus korpuslarning reprezentativlik prinsiplari sezilarli darajada kichikroq hajmlarda kuzatilayotganligi isbotlandi, chunki ikkala atama va neytral so'zlarning chastotasi barqaror hamda bir xil bo'lib qolmoqda. Shu nuqtayi nazardan, L. Flowerdewning ta'kidlashicha, 250000 belgidan kam bo'lgan yozma korpuslar kichik hisoblanadi [12: P. 24].

Shunday qilib, bu davr korpus tilshunosligi metodlarining butunjahon tarmog'i bilan birlashishi bilan tavsiflanadi: *NOW* va *GloWbe* korpuslarida bo'lgani kabi, Butunjahon Internet tarmog'iga korpus sifatida qarash (*Google Books* korpusi) va Butunjahon Internet tarmog'iga o'z vositalari bilan kirish (*SketchEngine*, *BNCweb*) hamda matnlarni avtomatik ravishda yuklab olish uchun dasturlar yaratildi. Bundan tashqari, muayyan so'zning katta ma'lumot to'plamlarida ishlatilishini, masalan, vaqt o'tishi bilan yozma (*Google Books*) yoki og'zaki nutqda so'zning shakli va ma'nosining o'zgarishini kuzatib borish imkoni vujudga keldi.

Qolaversa, ushbu lisoniy tadqiqotlarga yo'naltirilgan korpuslarni yaratish amaliyoti dunyoning barcha ilmiy maktablarining diqqat markazida bo'lib kelmoqda. Masalan, koreyalik olimlar o'z tadqiqotlarida koreys tilidagi hissiyot turlarini tasniflashda korpus tahlil metodlarini, shu asosda Koreys ruhiyati tahlil korpusi (Korean Emotion Analysis Corpus)ni ishlab chiqishgan [21]. Olimlar yaratilgan korpusning tuzilishiga asoslanib, his-tuyg'ularni tasniflashda hissiyotlar sinflari orasidagi farqlar haqida chuqurroq tushunchaga ega bo'libgina qolmay, hissiyotlarni tahlil qilish yondashuvlarini baholash imkonini beruvchi yangi standart ma'lumotlar to'plamini yaratishgan. Bu esa, o'z o'rnida, korpus bazalarining asosli yig'indisiga qarab, tadqiqotchiga kuzatish va tasniflash imkonini yarata olishini ko'rsatadi.

Rus tilshunoslarining korpus lingvistikasi rivojiga qo'shayotgan ulkan hissasini rus tilining milliy korpusi *Национальный корпус русского языка* misolida ko'rsatish mumkin [56]. Rus tilining milliy korpusi 2004-yil 29-apreldan beri onlayn qidiruv interfeysi orqali qisman kirish mumkin bo'lgan rus tilini aks ettiruvchi lingvistik korpusdir. Mazkur korpus ustida Rossiya Fanlar Akademiyasi Rus tili instituti tomonidan doimiy izlanishlar olib borilmoqda. Korpus hozirda 1 milliarddan ortiq so'z shaklini o'z ichiga oladi, ular avtomatik ravishda lemmatizatsiya qilinadi va

POS-/grammema-teglar, ya'ni har bir orfografik shakl uchun barcha mumkin bo'lgan morfologik tahlillarni qamrab oladi. Lemmalar, POS, grammatik elementlar va ularning kombinatsiyalarini qidirish mumkin. Bundan tashqari, subkorpuslarida 6 million so'z shakl qo'lda amalga oshirilgan omonimiya bilan mavjud. Yuqorida ta'kidlangan morfologik omonimiyaga oid qo'shimcha metadataga ega bo'lgan subkorpus ham avtomatik ravishda teglanadi. Butun korpusda leksik semantika (LS) bo'yicha qidirish mumkin bo'lgan teglar mavjud bo'lib, ular morfosemantik POS kichik sinflari (ot, refleksiv olmosh va boshqalar), tegishli LS xususiyatlari (mavzuli sinf, sabab, baho), hosila (kichiklashtiruvchi, qo'shimchalardan tuzilgan qo'shimchalar sifatdosh va hokazo)lardan iborat.

Arab tilshunoslari ham bu borada keng qamrovli tadqiqotlarni amalga oshirib kelmoqdalar. Masalan, tilni qayta ishlashning yangi vositasi *Arabic Corpus Processing Tools ACPTs 4.6 Version* ishlab chiqilgan bo'lib, u arabcha va inglizcha katta hajmdagi matnlarni tahlil qilish uchun mustaqil ochiq/bepul manba hisoblanadi. 8 gigabaytdan ortiq shaxsiy kompyuterlarga moslangan 50 million so'zdan ortiq bazani saqlaydi. Yangi *ACPT* korpus lingvistikasi va korpus lingvistik tahlil adabiyotlari, ayniqsa, statistik paketlarda qo'llaniladigan murakkab zamonaviy funksiyalarga ega. Korpus lingvistik tahlil uchun mos keladigan boshqa vositalar bilan taqqoslaganda, *ACPT*larni korpusni tahlil qilish uchun eng samarali vosita sifatida qayd etilishiga olib keladi. Til o'rgatish masalalariga e'tibor qaratilganda, til o'rgatish/o'rganish jarayonini yaxshilashda foydalanish mumkin bo'lgan mos vositalarning eng muhim vazifalari ushbu korpusga birlashtirilgan [3]. Shunday ishlarning davomi sifatida birinchi marta yaratilgan Xalqaro Arab Korpusi, ya'ni *International Corpus of Arabic (ICA)*ni yaratish loyihasini keltirish mumkin. Foydalanuvchilarga korpus ma'lumotlari bilan bir necha usullar orqali o'zaro aloqada bo'lishga imkon beruvchi interfeysga ega 100 million tahlil qilingan tokenlarni kiritish nazarda tutilgan. *ICA* 2006-yilda yaratilgan arab tilini aks ettiruvchi representativ korpus bo'lib, u butun arab dunyosida qo'llaniladigan Zamonaviy standart arab tilini qamrab olishga mo'ljallangan. *ICA Bibliotheca Alexandrina Morphological Analysis Enhancer (BAM-AE)* tomonidan tahlil qilingan. *BAMAE Buckwalter Arabic Morfologik Analizatoriga (BAMA)*ga asoslangan [37].

Yuqorida ko'rsatilgan Arab tilining xalqaro korpusini yaratishga asos bo'lgan elementlarga alohida to'xtalish o'rinli, nazarimizda. *Alexandrina Bibliotheca* nomi ostidagi veb-sahifa Arabcha xalqaro korpusning o'zagi demakdir. *Bibliotheca Alexandrina* (BA) Misrdagi yetakchi xalqaro tashkilotlardan biri bo'lib, madaniyat va bilimlarni tarqatishda, shuningdek, ilmiy-tadqiqotlarni qo'llab-quvvatlashda muhim ahamiyat kasb etadi. U butun arab dunyosida qo'llaniladigan arab tilining representativ korpusini yaratishga katta urinish sifatida Xalqaro Arab Korpusini (*ICA*) yaratish bo'yicha ulkan loyihani boshlagan, chunki bu til bo'yicha tadqiqotlarni qo'llab-quvvatlashning eng keng tarqalgan yo'li sifatida qaraladi. Korpus yaratilgandan so'ng tahlil qilingan shakl tadqiqotchilar uchun lingvistik manba sifatida mavjud bo'lgan, birinchi tahlil qilingan arab korpusi bo'lgan. Bu, shuningdek, arab tilida so'zlashuvchi jamoada milliy izlanishlarning birinchi tizimli tahlilidir. O'z nazariya va til tavsiflarini asossiz faktlarga emas, balki haqiqiy ma'lumotlarga asoslangan bo'lishi kerak

deb hisoblaydigan tilshunoslar uchun juda foydali manba bo'la oladi [2].

Xitoy tilshunosligi ham korpus tilshunosligiga oid tadqiqotlarda ildamlab borayotgan sohalar qatoriga kiradi. Chunki kompyuterga asoslangan tarjimonni yaratishda olingan tajribalarga tayanib, Guangdong chet tillari universitetining tarjimonlik fakultetida o'qitish tizimi, ta'limga yo'naltirilgan tarjima va parallel korpusni ishlab chiqish tamoyil va tartiblarini umumlashtiradi. Universitet olimlari tomonidan tarjimonlarni tayyorlash uchun onlayn avtonom o'quv platformasiga tarjima korpusini integratsiyalash bo'yicha keyingi tadqiqotlar olib borilgan. 2003- yildan boshlab PACCEL-S-dagi ba'zi matnlarni tanlab, ParaConc dasturi yordamida talaffuz xatosi, grammatik xato va pauza holatlarini miqdoriy baholovchi jarayon o'rganib chiqilgan. Ularning fikricha, xitoylik o'quvchilarning tarjimonlik faoliyatida eng ko'p uchraydigan xato talaffuz, asosan, unli tovushlarning talaffuzi bilan bog'liq. 183 nafar imtihon oluvchi tomonidan talqin qilingan 732 ta jumla orasida o'rtacha har bir jumlada bitta orfoepik xato kuzatiladi; har bir imtihon oluvchi 3,64 ta shunday xatoga yo'l qo'ygan. Yaratilgan korpus – ParaConc esa 90 ta grammatik xatoni ko'rsatadi, ulardan tez-tez uchraydigan holatlar birlik va ko'plik shakllarini noto'g'ri ishlatish, nutq birliklarini noo'rin qo'llashdan iboratligini ko'rsatgan [18: P. 65]. Xitoy tilshunoslari tadqiqotlari misolida aytish mumkinki, korpus tahlillari asosida talaffuzda uchraydigan xatoliklarni aniqlash va bartaraf etish imkoniyati ham mavjud.

Bundan tashqari, *a corpus of spoken Chinese* (og'zaki xitoy tili korpusi) ham yaratilgan, ochiq korpus bo'lishi mo'ljallangan, til sohibi bo'lmaganlar va ona tilida so'zlashuvchilar tomonidan ishlab chiqarilgan xitoy tilining og'zaki nutqi transkriptlari to'plamidir [44: P. 283]. Bu korpus o'zi maqsad qilgan mandarin xitoy tilining shu paytgacha yaratilmagan og'zaki lisoniy korpusining ilk varianti bo'lib, xitoy tilida uchraydigan, ingliz tilidan o'zlashgan ko'plab so'zlarni o'rganish uchun manba sifatida xizmat qila oladi.

Shuningdek, Y. Liu, M. Xiaohui Qin, L. Wang va Ch. Huang kabi xitoylik olimlar *CCAEC: A Corpus of Chinese-based Asian Englishes* (xitoychaga asoslangan Osiyo ingliz tillari korpusi)ni yaratishgan [32]. Bu ham o'z o'rnida korpuslarning tillararo interferensiya hamda hukmron tilning boshqa tillarga ta'sir doirasini o'rganish uchun keng yo'l ochib bera oladi.

Tadqiqotlar shuni ko'rsatadiki, ko'p tilli aholidan 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 davlat hududidagi tillar o'zaro munosabat, muloqot va yagona mediamakon ta'sirida bir-birlarini to'ldirib, o'z o'rnida, bir-birining lug'at boyliklarining qisqarishiga sabab bo'lmoqda. Nazarimizda, ushbu millat uchun ko'p tilli korpusning yaratilishi tillarning ayni paytdagi rang-barangligini saqlab qoladi.

Bu haqda hind tilshunos olimi N. Dash ham 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 imperiazizm jangida omon qolishga yordam berishi joiz bo‘lgan yagona lisoniy bazadir. Bundan tashqari, u yo‘qolgan mavqeyini tiklash uchun statistik jihatdan ishonchli ma’lumotlarni taqdim eta olishi bilan xarakterli” [9: P. 1]. Olimning kuyunchaklik bilan aytgan fikriga ko‘ra, globallashuv 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. Taqdiqotimizda ham aynan shu masala ko‘tarilmoqda. Zero, tadqiqot obyekti bo‘lgan materiallar asosida xulosa qiladigan bo‘lsak, rivojlanishda, ishlab chiqarishda bizdan oldinda bo‘lgan tillar borki, ularning ta’siri mediamakon prizmasida kengayib borishini inobatga olganda, o‘zbek tilining ham diaxronik korpusini yaratish amaliyotini yo‘lga qo‘yish zamon talabidir.

Yuqoridagi fikrlarni rad etmagan holda, til rivoji tadqiqi deganda so‘z, iboralarning, umuman olganda, leksik qatlamga kirib kelgan har qanday lisoniy birlikning etimologiyasini tahlil etish nazarda tutiladi. Bu sohada til korpuslarini qo‘llash amaliyoti K. Harringtonning ishlarida ko‘zga tashlanadi [17]. Olim yopiq jamiyat etnografiyasi tadqiqida korpus tilshunosligining o‘rni umumiy til yoki madaniyat mavjud bo‘lmaganda muloqotga kirishish muammosiga duch kelgan odamlarning instinktiv *topqir*ligini aniqlash uchun Irlandiyadagi qochoqlarni qabul qilish markazida muhojirlarning o‘zaro munosabatlarini tahlil qiladi. Uch yillik etnografik kuzatish, amaliy metodologiya, korpus tilshunosligi va suhbatlarning innovatsion gibril tahlilidan foydalangan holda quyidagilarga erishilgan:

- 98 000 so‘zdan iborat korpusga asoslangan;
- rezidentlar va markazning ingliz tilida so‘zlashuvchi xodimlari bilan o‘zaro munosabatlari ingliz tilidan foydalanishning jarayonda o‘rganilishiga asoslangan;
- nutq hamjamiyati, kommunikativ kompetensiya va tillararo interferensiya kabi konstruksiyalar qiyinlashib borishi hamda bir tilning sofligiga putur yetishi aniqlangan.

Bizningcha, 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 etadi:

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

Umuman olganda, ta’kidlab o‘tganimiz yuqoridagi besh bosqichli uzviylikda sodir bo‘ladigan tillar interferensiyasi bilan yakunlovchi jarayon bugungi kecha- kun-

duzda 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 soffligi, 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. Buni esa hech bir bahona bilan oqlab bo‘lmaydi.

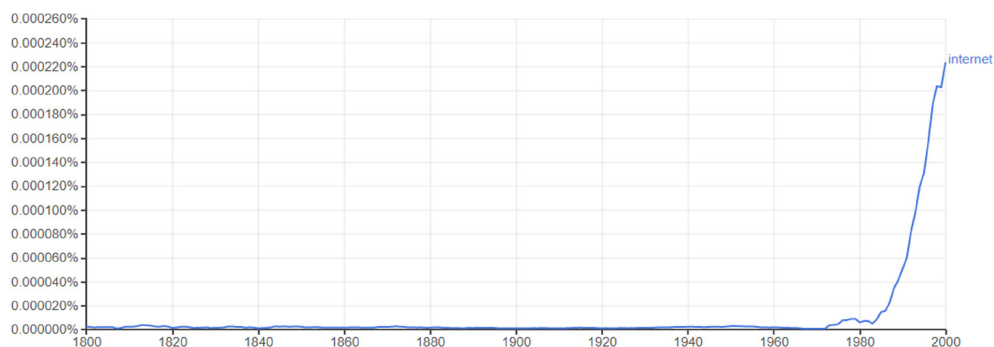
Shvetsiyalik olim M. Kytöning qarashlariga ko‘ra, elektron tarixiy korpus va korpus metodologiyasi tillarning hozirgi holatini o‘rganish va baholashga qaratilgan, kelajakda yuzaga kelishi mumkin bo‘lgan lisoniy o‘zgarishlarni ko‘rib chiqishga imkon beruvchi tadqiqot usulidir. Bu bilan olim korpus lingvistik metodologiyasining keng doirasi ichida tarixiy korpus tilshunosligi til tarixi va o‘zgarishini o‘rganish uchun jozibadorlikni sezilarli darajada oshirgan jonli soha sifatida paydo bo‘lganligini ta’kidlaydi. Darhaqiqat, agar korpus lingvistikasi metodologiyasi turtki bo‘lmaganida, dalillarga asoslangan tarixiy tilshunoslikni tadqiq qilish jarayoni tugallanmagan bo‘lar edi. Tez sur‘atlarda rivojlanib borayotgan hayot va tadqiqotlar, akademik martaba yo‘lidagi raqobat kuchayayotgan va yosh olimlarga taqdim etilayotgan imkoniyat bir davrda, uslubiy jihatdan oson bo‘lgan sohaning kelajagi yorqin ekanligi shubhasiz. Tarixiy korpus va boshqa elektron resurslar ham til tarixini o‘rganishni jozibador qildi: ular vositasida ishlash o‘quvchilarni individual va interaktiv tarzda jalb qiladi [28: P. 417].

Dunyo tadqiqotlari orasida zamonaviy izlanishlar alohida o‘rin egallaydi. Bunga misol qilib, yuqorida ta’kidlangan tarixiy korpuslarni misol keltirish mumkin. Tarixiy tushunchasi korpusning eskiligiga ishora qilmaydi, aksincha, korpus qamrab olgan tilning oldingi namunalari, arxiv hujjatlari, tarixiy qo‘lyozmalar, oldin nashr etilgan materaillarni izohlaydi. Bu jarayonning jadal rivojini, ayniqsa, ingliz tilidagi COHA korpusi misolida ko‘rish mumkin.

Tarixiy Amerika ingliz tili korpusi (The Corpus of Historical American English) tarixiy ingliz tilining eng katta tuzilgan korpusidir. COHA 1820-2010- yillardagi 475 million so‘zdan ortiq matnni o‘z ichiga oladi. Bu ingliz tilining boshqa taqqoslanadigan tarixiy korpuslaridan 50-100 baravar kattaroqdir. Qolaversa, korpus o‘n yilliklar bo‘yicha janrlar kesimida muvozanatlangan, ya’ni miqdoriy tenglashtirilgan. Korpusning yaratilishi 2008-2010-yillardagi Milliy gumanitar fanlar jamg‘armasining (NEH) granti bilan bog‘liq [40]. Shu turdagi korpuslarni ingliz tili misolida ham yaratish ahamiyat kasb etadi, nazarimizda. Zero, materiallar elektronlashtirilayotgan bir zamonda qo‘lyozma va nashr etilgan adabiyotlarning qurbi pasayib borsa, elektron shaklda korpuslarga umumlashtirib qo‘yilsa, har qanday lingvistik izlanishlarni amalga oshirayotgan tadqiqotchi tilning tarixiga qarab buguni xususida asosli xulosaga kela oladi.

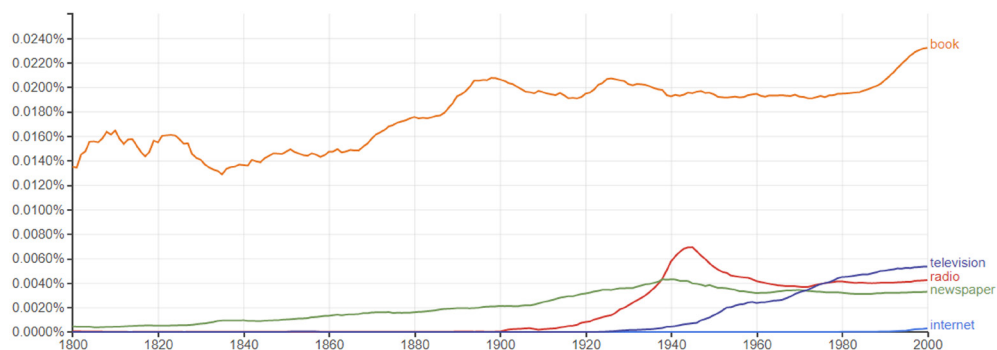
Zamonaviy korpuslar taqdim etayotgan keng imkoniyatlarni aniqroq ko‘rsatish uchun ingliz tili korpuslari konkordansida taqdim etiladigan ma’lumotlar shakli va formati muhokamaga tortildi. Birinchi o‘rinda *Google Books n-gram* korpusi ko‘rib

chiqildi, chunki hajm jihatidan jami 200 milliarddan ko‘proq so‘zli matnlarni qamrab oladi, bundan 155 milliard amerikacha ingliz tiliga to‘g‘ri kelsa, 34 milliard britaniya ingliz tilidan olingan matnlardan tashkil topgan. Ushbu korpusning o‘ziga xos tomoni shundaki, u oddiy va murakkab qidiruv natijalarini 1800-yildan 2000-yilning birinchi o‘n yilligigacha bo‘lgan ikki asr oralig‘ida yozilgan kitoblarni; konkordansda o‘n yilliklar tartibida birlamchi chastotalarini jadval yoki grafika ko‘rinishida taqdim etadi. Bunday ko‘rinishdagi qidiruv, asosan, so‘z qo‘llash rivoji yoki etimologiyasini o‘rganishda qo‘l keladi. Masalan, oddiy qidiruvda *internet* so‘zi kiritildi, konkordans natijalari kiritilgan birlik yagona bo‘lgani uchun bir chiziqli grafika orqali berildi. “1- rasmga qarang”. Grafikadan ko‘rish mumkinki, ushbu tushuncha yigirmanchi asrning so‘nggi ikki o‘n yilligida keng qo‘llanila boshlangan, lekin ungacha so‘zning deyarli iste‘molda bo‘lmaganini ko‘rsatyapti. Bu esa uning kirib kelishi, ilm-fanda keng qo‘llanila boshlangan davrlarga ishora qiladi.



1-rasm. Google Books n-gram korpusida internet leksemasining qo‘llanilish diaxronikasi

Google Books n-gram korpusida murakkab qidiruv amalga oshirilganda, kiritmalar soni beshtani tashkil etdi. Ular *internet*, *book*, *television*, *radio* va *newspaper* kabi axborot almashishning boshqa turlari ham kiritmalar sifatida vergul orqali oraliq masofa tashlamasdan (probilsiz) konkordans qidiruviga berildi. Natijalar shuni ko‘rsatdiki, har bir so‘z o‘z rivojlanish, shu asnoda tilga kirib kelishi mutanosibligida yillar oralig‘ida mos ravishda kam va ko‘p chastotalarni taqdim etdi “2-rasmga qarang”. Ushbu korpus natijalaridan *internet-internet* XX asrning oxirida; televizor – *television* XX asrning 40-yillaridan; radio – *radio* XX asrning boshidan va gazeta – *newspaper* XIX asrning boshlaridan beri qo‘llanila boshlangan bo‘lsa, kitob – *book* har ikkala asrda ham qo‘llanilish chastotasi jihatidan qolgan axborot vositalaridan eng ko‘p qo‘llanilganini aniq ko‘rish mumkin. Bu tahlillar yuqoridagilarning jamiyatga kirib kelish tarixi haqidagi bilimlarga ega bo‘lmay turib ham, lingvistik korpus tahlillarini o‘rganish orqali kerakli, eng muhimi, asosli xulosaga kelishga yordam beradi.



2-rasm. Google Books n-gram korpusida *internet, book, television, radio* va *newspaper* leksemalarining qo'llanilish diaxronikasi

Keyingi qidiruv *English Web 2015 (enTenTen15)* diaxronik korpusida o'tkazildi. Unda mavjud bo'lgan lemmalash imkoniyati amalda qo'llandi. Bunda korpusda murakkab qidiruv tugmasi bosiladi va lemma (so'z) kiritiladi. Misol tariqasida, *money* – pul so'zi lemmalandi va quyidagi ma'lumotlar tahlil qilindi:

1) *money* – pul so'ziga birikib keladigan aniqlovchi sifatlar ajratib berildi. Masalan, *much money, easy money, big money, soft money, dirty money, private money, dark money* va boshqalar;

2) *money* – pul so'zi aniqlovchi sifatida kelib, boshqa bir ot bilan birikib kelishi holatlari ko'rsatib berildi. Masalan, *money pit, money market, money worries, money maker, money loserva* boshqalar;

3) *money* – pul to'ldiruvchi sifatida kelib, qaysi fe'llar bilan birikishi ko'rsatildi. Masalan, *earn money, spend money, donate money, transfer money, receive money, funnel money, refund money* va boshqalar;

4) *money* – pul ega sifatida kelib, o'zidan keyin sifat bilan yasalgan ot-kesimlar tarkibi ko'rsatildi. Masalan, *money is tight, money is scarce, money is worthless, money is important, money is hard, money is short, money is neutral, money is bad* va boshqalar;

5) *money* – pul ega sifatida kelib o'zidan keyin ot yoki olmosh bilan yasalgan ot-kesimlar tarkibi ko'rsatildi. Masalan, *money is not speech, money isn't everything, money is not motivator, money is power, money is the answer, money is a problem, money is a big consideration, money is no excuse, money is a human invention, money is a way, money is a skill, love of money is the root of all evil* va boshqalar.

Bundan tashqari, lemmalash uchun qidiruvga berilgan so'zning boshqa birikish xususiyatlari, ya'ni olmoshlar bilan kelishi, idiomalar tarkibida kelishi, aniqlanish bilan birikishi, boshqa turdagi vosita hamda holatiga ko'ra farqlanuvchi birikmalarni aniq ko'rsatib beradi. Keltirilgan misollarni tushunishda yoki qo'llanilishida qiyinchilik bo'lsa, uni konkordansda va kengroq kontekstda ko'rish imkoniyatining mavjudligi foydalanuvchi uchun juda muhim.

English Web 2015 (enTenTen15), *BNC* va *Google Books n-gram* kabi ingliz tili korpuslari taqdim etishi mumkin bo'lgan keng qidiruv va konkordans natijalari

ko'rib chiqildi. Bunda har bir korpusning imkoniyatlarini ko'rsatib beruvchi oddiy va murakkab qidiruvlar amalga oshirildi. *English Web 2015 (enTenTen15)* korpusidan lemma olish amaliyoti ko'rsatib o'tildi, u taqdim etgan natijalar ishda o'z ifodasini topdi. *BNC*da so'z-turlarning birlamchi chastota asosida tartiblangan jadvalini hosil qilish orqali ko'rib chiqildi, ulardan eng yuqori chastotali yigirmatasi ajratib olinib, lingvistik xususiyatlarini tahlil etish orqali ingliz tilidagi eng yuqori chastotali so'zlar funksional so'zlar ekani aniqlandi. *Google Books n-gram* korpusida axborot vositalari murakkab va oddiy qidiruvga berildi. Qidiruv natijalarining konkordansda grafik chizmalar shaklida berilishi kuzatildi. Bu chizmalar ishda keltirilib, qidiruvga berilgan kiritmalarning etimologik rivojini o'rganish imkoniyati mavjudligi xususida muayyan xulosaga kelindi.

Lisoniy korpuslarni qo'llash tabiiy til muhitida tadqiqotchilarga qiziqarli lingvistik birliklarning belgilarini (so'zlar, jumlar, grammatik va sintaktik konstruksiyalar) kuzatish imkonini beradi. Sun'iy ravishda qurilgan kontekstlar emas, balki haqiqiy reallikdan olingan manbalar qo'llaniladi. Ya'ni korpus yaratishda u qamrab olayotgan matnlar muayyan tilni o'zida to'liq aks ettirishni ta'minlash zarurati qo'yiladi.

Korpus lingvistikasi so'nggi o'n yillikda, xususan, tilshunoslikda amaliy ishlarni qo'llash, lug'atlar va grammatik topshiriqlar tayyorlash nuqtai nazaridan tobora ko'proq faollashmoqda. Shu bilan birga, yangi yo'nalishning nazariy asoslarini tushunish turli yondashuvlar asosida amalga oshirilayotgan maxsus tadqiqotlarni qamrab olmoqda. Korpus lingvistikasining asosiy tushunchalarini aniqlash, lingvistik bilimlar tuzilishidagi o'rni va korpus lingvistikasi metodikasini tadqiq etish muhim ahamiyat kasb etadi. Korpus lingvistikasini ayni tadqiqot vositasi sifatida ishlatish faqat tilni ifoda etuvchi predmet bo'lib qoladi.

Korpus lingvistikasi korpusning umumiy prinsiplari va til xususiyatlarini ishlab chiqish bilan shug'ullanar ekan, lisoniy korpus muayyan me'yorlarga muvofiq to'plangan va muayyan standartlarga muvofiq yaratilgan va maxsus qidiruv tizimi bilan ta'minlangan matnlar to'plamidir. Ba'zan korpus – to'plangan, til, janr, muallif, matnlarni yaratish davrini birlashtirgan matnlar to'plami ham deyiladi hamda tilning rivojlanishi kuzatishga imkon beruvchi manba sifatida qaraladi.

XULOSA

Tillar shakllanib, kengayib borgani sari ularning ta'sir doirasi ham kuchayib boradi yoki doira nisbatan torayishi mumkin. Bunda til sohibining intellekt va iqtisodiy barqarorligining trayektoriyasiga mos ravishda ikkala holat rivojlanib boradi. Zero, har qanday til o'z ichida rivojlanishi uchun jamiyatda va hayot tarzida zarur bo'lgan barcha qurilma, texnika va jihozlar shu jamiyatda yaratilgan bo'lishi kerak. Buning uchun jamiyat a'zolari o'z intellektini to'g'ri yo'naltirishi talab etiladi. Aks holda, yuqorida sanalgan maishiy ashyolar chet eldan kirib keladi, chunki unga ehtiyoj mavjud, ular o'zi bilan jamiyat tilining lug'at tarkibiga yangi nomlarni ham olib keladi.

O'tgan asr ikkinchi yarmining boshlanishida yuzaga kelgan tilshunoslik ilmin-

ing yangi tarmog'i bo'lgan korpus lingvistikasi 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 aslyatdan 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.

Ijtimoiy hodisa hisoblanadigan 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'g'ri 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.

Diaxronik korpuslar o'zi ifoda etayotgan til 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.

FOYDALANILGAN ADABIYOTLAR

1. Abdurahmonova N.Z. O'zbek tili elektron korpusining kompyuter modellari: Filol. fan. d-ri. (DSc) diss. avtoreferati. – Toshkent, 2021. – 73 b.;
2. Alexandrina Bibliotheca available from <https://www.bibalex.org/ica/en/About.aspx> [Retrieved on September 23, 2023]
3. Almujaivel S. and Al-thubaity A. Arabic Corpus Processing Tools for Corpus Linguistics and Language Teaching // The Globalization of Second Language Acquisition and Teacher Education At: Fukuoka Volume: 2. – 2016. – 4 p. available at https://www.researchgate.net/publication/309351881_Arabic_Corpus_Processing_Tools_for_Corpus_Linguistics_and_Language_Teaching
4. Baker P., Hardie A., McEnery T. Glossary of Corpus Linguistics. –Edinburg: Edinburgh University Press, 2006. – 192 p.
5. Biber D. University Language: A Corpus-based Study of Spoken and Written Registers. - Amsterdam: John Benjamins, 2006. – 261 p.
6. BNC (BritishNationalCorpus) available from: <https://www.english-corpora.org/bnc/> [Retrieved on 03.08.2024]
7. Corpus linguistics // available from: https://en.wikipedia.org/wiki/Corpus_linguistics

- [Retrieved on 03.08.2023]
8. Cruden A. A Complete Concordance to the Holy Scriptures of Old and New Testament. – London: Fleming H. Revell Company, 1737. – 756 p.
 9. Dash N.S. Language Corpora: Present Indian Need. – Kolkata: Indian Statistical Institute, 2004. – P 1. Available from https://www.researchgate.net/publication/2938590_Language_Corpora_Present_Indian_Need
 10. Davies M. Corpora: an introduction // *The Cambridge handbook of Corpus Linguistics* / ed. by D. Biber, R. Reppen. – Cambridge University Press, 2015. – P. 11-31.
 11. Douglas B., UllaC. and Thomas A.U. *Discourse On The Move*. – Amsterdam, 2007. – 290 p.
 12. Flowerdew L. The argument for using English specialized corpora to understand academic and professional language // *Discourse in professions: perspectives from Corpus Linguistics* / ed. by U. Connor, T. Upton, 2004. – P. 11-33.
 13. Flowerdew L. The argument for using English specialized corpora to understand academic and professional language // *Discourse in professions: perspectives from Corpus Linguistics* / ed. by U. Connor, T. Upton, 2004. – P. 11-33.;
 14. Flowerdew L. The argument for using English specialized corpora to understand academic and professional language // *Discourse in professions: perspectives from Corpus Linguistics* / ed. by U. Connor, T. Upton, 2004. – P. 11-33.;
 15. 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.;
 16. Halliday M.A., TeubertW., Yallop C. and Cermakova A. *Lexicology and Corpus Linguistics*, – London, 2004. – 184 p.;
 17. Harrington, K. The Role of Corpus Linguistics in the Ethnography of a Closed Community: Survival Communication. // *International Journal of Corpus Linguistics*, 24 (4), 2018. – pp. 541-547. ISSN 1384-6655
 18. Hu K. and Kim K. H. (eds.), *Corpus-based Translation and Interpreting. Studies in Chinese Contexts*, Palgrave Studies in Translating and Interpreting, https://doi.org/10.1007/978-3-030-21440-1_3 // written by B. Wang (*) – China: Guangdong University of Foreign Studies, Guangzhou. 2009. – pp. 61-87.
 19. Hunston S. *Corpora in Applied Linguistics*: Cambridge University Press, 2002. – 234 p.;
 20. Hyland K. As it can be seen: Lexical bundles and disciplinary variation // *English for Specific Purposes*. 2008. Vol. 27. – P. 4-21.
 21. Jung Y. and others. A corpus-based approach to classifying emotions using Korean linguistic features // *Cluster Comput* (2017) 20: – pp. 583-595. DOI 10.1007/s10586-017-0777-8
 22. Karlsson F. *Lexicography And Corpus Linguistics*. – Helsinki: Department of General Linguistics University of Helsinki, 1992. YPJI: <http://www.ling.helsinki.fi/~fkarlssso/INTJLLEX.pdf>;
 23. Kennedy G. *An Introduction to Corpus linguistics*. – London and New York: Addison Wesley Longman limited, 1998. – 315 p.
 24. Kilgarrieff 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;
 25. Kilgarrieff A., Kosem I. *Corpus Tools For Lexicographers*. – Trojina: Institute for Applied Slovene Studies, 2012. – 37 p. YPJI: https://www.researchgate.net/publication/264884809_Corpus_tools_for_lexicographers;
 26. Koester A. *Building small specialized corpora* // *The Routledge handbook of corpus linguistics*. 2010. – P. 66-80;
 27. Kuebler S., Zinsmeister H. *Corpus Linguistics and Linguistically Annotated Corpora*. – London: Bloomsbury Publishing, 2015. – P. 10.
 28. Kytö M. *Corpora and historical linguistics* // *RBLA, Belo Horizonte*, v. 11, n. 2. – Sweden, Uppsala: Uppsala University, 2011 – P. 417.

29. Lamel L., Cole R. Spoken Language Corpora // Survey of the State of the Art in Human Language Technology. – New York, 1997. – P. 338-391.
30. Laurence A. A Critical Look at Software Tools in Corpus Linguistics // Linguistic Research. – Tokyo, № 30 (2) // 2013. – P. 141-161.
31. Leech G. Corpora And Theories Of Linguistic Performance. In J.Svartvik (ed.). 1992. – 105–122 pp.
32. Liu Y. and others CCAE: A Corpus of Chinese-based Asian Englishes. Available from <https://arxiv.org/abs/2310.05381v1> [Retrieved on 30.10.2023]
33. Mauranen A. Speaking professionally in L2 // Variation and change in spoken and written discourse: Perspectives from Corpus Linguistics / ed. by J. Bamford, S. Cavalereri, G. Diani, 2013. – P. 5-31.
34. Mauranen A. Speaking professionally in L2 // Variation and change in spoken and written discourse: Perspectives from Corpus Linguistics / ed. by J. Bamford, S. Cavalereri, G. Diani, 2013. – P. 5-31;
35. McEnery T., Hardie A. Corpus Linguistics: Method, theory and practice. – New York: Cambridge university press, 2012. – P. 35.
36. McEnery T., Hardie A. Corpus Linguistics: Method, theory and practice. – New York: Cambridge university press, 2012. – P. 35.
37. Nagi M. The International Corpus of Arabic: Compilation, Analysis and Evaluation. // Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP). – January 2014. – pp. 8-17. available at https://www.researchgate.net/publication/301404178_The_International_Corpus_of_Arabic_Compilation_Analysis_and_Evaluation
38. Rayson P. Computational tools and methods for corpus compilation and analysis // The Cambridge handbook of English corpus linguistics / ed. by D. Biber, R. Reppen. – Cambridge univer. press, 2015. – P. 32-49.
39. The British National Corpus. Available from: <http://www.natcorp.ox.ac.uk> The Corpus of Historical American English: Available from <https://www.english-corpora.org/coha/> [Retrieved on 31.10.2023]
40. The International Corpus of English. Available from:: <http://www.ucl.ac.uk/englishusage/projects/ice.htm>
41. The Longman Corpus Network. Available from: <http://www.longmandictionariesusa.com/longman/corpus> The International Corpus of English. Available from: <http://www.ucl.ac.uk/englishusage/projects/ice.htm>;
42. Tognini-Bonelli E. Corpus linguistics at work. – Amsterdam: John Benjamins, 2001.; The Longman Corpus Network. Available from: <http://www.longmandictionariesusa.com/longman/corpus> [Retrieved on 20.06.2021].
43. Wang J. Recent Progress in Corpus Linguistics in China // International Journal of Corpus Linguistics // 6(2). DOI: 10.1075/ijcl.6.2.05wan – China, 2002. – pp. 281-304.
44. Baranov A.N. Kompyuternaya lingvistika // Vvedenie v prikladnuyu lingvistiku: uchebnoe posobie. – M.: Yeditorial URSS, 2003. – 320 s.;
45. Begmatova G.X. O'zbek tili milliy korpusida idiomalar bazasini yaratish: Filol. fan. fals. d-ri (PhD) diss. – Termiz, 2021. – 143 b.;
46. Vadyaev S.E. Lingvisticheskie principy postroeniya i ispolzovanie korpusa tekstov dlya Issledovaniya ofitsialno-delovogo stilya sovremennogo nemeskogo yazyka (na materiale elektronno korpusa «DER»): Dis ... kand. filol. nauk. – Samara, 2005. – 160 s.;
47. Vladimov N.V. Korpusnyy podkod k resheniyu perevodcheskiy problem (na meteriale angliskogo i russkogo yazykax): Disc....kand. filol. nauk. – M.: RGB, 2005. – S 26.;
48. Gorina O.G. Ispolzovanie texnologiy korpusnoy lingvistiki dlya razvitiya leksicheskix navykov studentov-regionovedov v professionalno-orientirovannom obshchenii na angliyskom yazyke: Diss. ... kand. ped. nauk. – M., 2014. – 321 s.;
49. Dja'farova D.F. Modeli lingvisticheskogo analiza tekstov tadjikskego yazyka (na materiale

- gazeley Xafiza): Diss. ... kand. filol. nauk. – Dushanbe, 2013. – 151 s.
50. Djumabaeva J. O'zbek va ingliz tillarida leksik hamda stilistik graduonimiya: Filol. fan. d-ri. diss. avtoref. – Toshkent: 2016. – 82 b. URL: http://tashgiv.uz/wp-content/uploads/2017/01/avtoref_jumabayeveva.pdf;
 51. Zaxarov V.P. Korpusnaya lingvistika: Uchebno-metod. posobie/ V.P. Zaxarov. – SPb, 2005. – 48 s. – Rejim dostupa: <http://download.yandex.ru/class/zaxarov/plan.pdf>;
 52. Karimov R.A. O'zbek-ingliz parallel korpusini tuzishning lingvistik va dasturiy masalalari: Filol. fan. fals. d-ri (PhD) diss. – Buxoro, 2022. – 148 b.;
 53. Magomedova A.N. Korpusnaya lingvistika i kontekstnoe razreshenie leksicheskoy mnogoznachnosti slov: Disc. ... kand. filol. nauk. – M.: RGB, 2005. – 132 c.
 54. Mamontova V.V. Osobennosti perevoda slojnosostavnyx slov s angliyskogo yazyka na russkiy (na materiale korpusa publitsisticheskix tekstov): Disc. ... kand. filol. nauk. – Stavropol: 2008. – 139 c.;
 55. Natsionalnyy korpus russkogo yazyka. Dostupno ot <https://ruscorpora.ru/>
 56. Plungyan V.A. Korpus kak instrument i kak ideologiya: o nekotoryx urokax sovremennoy korpusnoy lingvistiki // Russkiy yazyk v nauchnom osvещenii. – M.: №2., 2008. – S. 7-20.
 57. Raxmanova A.A. O'zbek tili milliy korpusini yaratishdagi kompyuter usullari: Filol. fan. fals. d-ri (PhD) diss. – Toshkent, 2021. – 161 b.;
 58. Romanov A.S. Metodika i programmnyy kompleks dlya identifikatsii avtora neizvestnogo teksta: Avtoref. diss. kand. tex. nauk. – Tomsk, 2010. 26 s.;
 59. Sabirova G. Umumta'lim maktablari o'quvchilariga ingliz tili olmoshlarini axborot texnologiyalari asosida o'rgatish metodikasi: Filol. fals. d-ri (PhD) ...diss. avtoref. – Toshkent: 2017. – 41 b. URL: <http://uzswlu.uz/media/ba7be50d-84f1-c826-f935-6e7610cb6bf5.pdf>;
 60. Sajenin I.I. Korpusnye metody v leksikografii: opyt sozdaniya modeli Slovarnogo korpusa: Disc. ... kand. filol. nauk. – Novosibirsk, 2013. 193 s.
 61. Toirova G.I. O'zbek tili milliy korpusini yaratishning nazariy va amaliy masalalari: Filol. fan. d-ri (DSc) diss. avtoreferati – Buxoro, 2021. – 72 b.;
 62. Xamroeva Sh.M. O'zbek tili morfologik analizatorining lingvistik ta'minoti: Filol. fan. d-ri (DSc) diss. avtoreferati – Farg'ona, 2021. – 76 b.;
 63. Xamroeva Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Filol. bo'yicha fals. d-ri (PhD) ... diss. avtoref. – Qarshi, 2018. – 52 b. URL: <https://e-books.tsue.uz/pdf-ozbek-tili-mualliflik-korpusini-tuzishning-lingvistik-asoslari>
 64. Xoliyorov O'.M. O'zbek tili ta'limiy korpusini tuzishning lingvistik asoslari: Filol. fan. fals. d-ri (PhD) diss. – Termiz, 2021. – 181 b.;
 65. Elova D.Q. O'zbek tili korpusi birliklarining uslubiy teglarini yaratish tamoyillari va lingvistik ta'minoti: Filol. fan. fals. d-ri (PhD) diss. avtoref. – Toshkent, 2022. – 65 b.
 66. Ergasheva G. Turli tizimdagi tillarda genderga oid terminologiyaning shakllanishida lingvistik va ekstralingvistik omillar: Filol. fan. d-ri. diss. avtoref. – Toshkent: 2018. – 61 b. URL: <http://demo.navoiy-uni.uz/sites/default/files/ziyonet.pdf>;

REFERENCES

1. Abdurahmonova N.Z. O'zbek tili elektron korpusining kompyuter modellari: Filol. fan. d-ri. (DSc) diss. avtoreferati. – Toshkent, 2021. – 73 b.;
2. Alexandrina Bibliotheca available from <https://www.bibalex.org/ica/en/About.aspx> [Retrieved on September 23, 2023]
3. Almujaivel S. and Al-thubaity A. Arabic Corpus Processing Tools for Corpus Linguistics and Language Teaching // The Globalization of Second Language Acquisition and Teacher Education At: Fukuoka Volume: 2. – 2016. – 4 p. available at https://www.researchgate.net/publication/309351881_Arabic_Corpus_Processing_Tools_for_Corpus_Linguistics_and_Language_Teaching

4. Baker P., Hardie A., McEnery T. *Glossary of Corpus Linguistics*. –Edinburg: Edinburgh University Press, 2006. – 192 p.
5. Biber D. *University Language: A Corpus-based Study of Spoken and Written Registers*. - Amsterdam: John Benjamins, 2006. – 261 p.
6. BNC (BritishNationalCorpus) – Британия Миллий Корпуси: 100 миллион сўзли лисоний корпус.
7. Corpus linguistics // available from: https://en.wikipedia.org/wiki/Corpus_linguistics [Retrieved on 03.08.2023]
8. Cruden A. *A Complete Concordance to the Holy Scriptures of Old and New Testament*. – London: Fleming H. Revell Company, 1737. – 756 p.
9. Dash N.S. *Language Corpora: Present Indian Need*. – Kolkata: Indian Statistical Institute, 2004. – P 1. Available from https://www.researchgate.net/publication/2938590_Language_Corpora_Present_Indian_Need
10. Davies M. *Corpora: an introduction* // *The Cambridge handbook of Corpus Linguistics* / ed. by D. Biber, R. Reppen. – Cambridge University Press, 2015. – P. 11-31.
11. Douglas B., UllaC. and Thomas A.U. *Discourse On The Move*. – Amsterdam, 2007. – 290 p.
12. Flowerdew L. The argument for using English specialized corpora to understand academic and professional language // *Discourse in professions: perspectives from Corpus Linguistics* / ed. by U. Connor, T. Upton, 2004. – P. 11-33.
13. Flowerdew L. The argument for using English specialized corpora to understand academic and professional language // *Discourse in professions: perspectives from Corpus Linguistics* / ed. by U. Connor, T. Upton, 2004. – P. 11-33.;
14. Flowerdew L. The argument for using English specialized corpora to understand academic and professional language // *Discourse in professions: perspectives from Corpus Linguistics* / ed. by U. Connor, T. Upton, 2004. – P. 11-33.;
15. 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.;
16. Halliday M.A., Teubert W., Yallop C. and Cermakova A. *Lexicology and Corpus Linguistics*, – London, 2004. – 184 p.;
17. Harrington, K. *The Role of Corpus Linguistics in the Ethnography of a Closed Community: Survival Communication*. // *International Journal of Corpus Linguistics*, 24 (4), 2018. – pp. 541-547. ISSN 1384-6655
18. Hu K. and Kim K. H. (eds.), *Corpus-based Translation and Interpreting. Studies in Chinese Contexts*, Palgrave Studies in Translating and Interpreting, https://doi.org/10.1007/978-3-030-21440-1_3 // written by B. Wang (*) – China: Guangdong University of Foreign Studies, Guangzhou. 2009. – pp. 61-87.
19. Hunston S. *Corpora in Applied Linguistics*: Cambridge University Press, 2002. – 234 p.;
20. Hyland K. As it can be seen: Lexical bundles and disciplinary variation // *English for Specific Purposes*. 2008. Vol. 27. – P. 4-21.
21. Jung Y. and others. A corpus-based approach to classifying emotions using Korean linguistic features // *Cluster Comput* (2017) 20: – pp. 583-595. DOI 10.1007/s10586-017-0777-8
22. Karlsson F. *Lexicography And Corpus Linguistics*. – Helsinki: Department of General Linguistics University of Helsinki, 1992. YPJ: <http://www.ling.helsinki.fi/~fkarlss/INTJLEX.pdf>;
23. Kennedy G. *An Introduction to Corpus linguistics*. – London and New York: Addison Wesley Longman limited, 1998. – 315 p.
24. 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;
25. Kilgarriff A., Kosem I. *Corpus Tools For Lexicographers*. – Trojina: Institute for Applied Slovene Studies, 2012. – 37 p. YPJ: https://www.researchgate.net/publication/264884809_

- Corpus_tools_for_lexicographers;
26. Koester A. Building small specialized corpora // *The Routledge handbook of corpus linguistics*. 2010. – P. 66-80;
 27. Kuebler S., Zinsmeister H. *Corpus Linguistics and Linguistically Annotated Corpora*. – London: Bloomsbury Publishing, 2015. – P. 10.
 28. Kytö M. Corpora and historical linguistics // *RBLA, Belo Horizonte*, v. 11, n. 2. – Sweden, Uppsala: Uppsala University, 2011 – P. 417.
 29. Lamel L., Cole R. *Spoken Language Corpora // Survey of the State of the Art in Human Language Technology*. – New York, 1997. – P. 338-391.
 30. Laurence A. A Critical Look at Software Tools in Corpus Linguistics // *Linguistic Research*. – Tokyo, № 30 (2)// 2013. – P. 141-161.
 31. Leech G. Corpora And Theories Of Linguistic Performance. In J.Svartvik (ed.). 1992. – 105–122 pp.
 32. Liu Y. and others CCAE: A Corpus of Chinese-based Asian Englishes. Available from <https://arxiv.org/abs/2310.05381v1> [Retrieved on 30.10.2023]
 33. Mauranen A. Speaking professionally in L2 // *Variation and change in spoken and written discourse: Perspectives from Corpus Linguistics* / ed. by J. Bamford, S. Cavalereri, G. Diani, 2013. – P. 5-31.
 34. Mauranen A. Speaking professionally in L2 // *Variation and change in spoken and written discourse: Perspectives from Corpus Linguistics* / ed. by J. Bamford, S. Cavalereri, G. Diani, 2013. – P. 5-31;
 35. McEnery T., Hardie A. *Corpus Linguistics: Method, theory and practice*. – New York: Cambridge university press, 2012. – P. 35.
 36. McEnery T., Hardie A. *Corpus Linguistics: Method, theory and practice*. – New York: Cambridge university press, 2012. – P. 35.
 37. Nagi M. The International Corpus of Arabic: Compilation, Analysis and Evaluation. // *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*. – January 2014. – pp. 8-17. available at https://www.researchgate.net/publication/301404178_The_International_Corpus_of_Arabic_Compilation_Analysis_and_Evaluation
 38. Rayson P. Computational tools and methods for corpus compilation and analysis // *The Cambridge handbook of English corpus linguistics* / ed. by D. Biber, R. Reppen. – Cambridge univer. press, 2015. – P. 32-49.
 39. The British National Corpus. Available from: <http://www.natcorp.ox.ac.uk>
 40. The Corpus of Historical American English: Available from <https://www.english-corpora.org/coha/> [Retrieved on 31.10.2023]
 41. The International Corpus of English. Available from:: <http://www.ucl.ac.uk/englishusage/projects/ice.htm>
 42. The Longman Corpus Network. Available from: <http://www.longmandictionariesusa.com/longman/corpus.>; The International Corpus of English. Available from: <http://www.ucl.ac.uk/englishusage/projects/ice.htm>;
 43. Tognini-Bonelli E. *Corpus linguistics at work*. – Amsterdam: John Benjamins, 2001.; The Longman Corpus Network. Available from: <http://www.longmandictionariesusa.com/longman/corpus> [Retrieved on 20.06.2021].
 44. Wang J. Recent Progress in Corpus Linguistics in China // *International Journal of Corpus Linguistics* // 6(2). DOI: 10.1075/ijcl.6.2.05wan – China, 2002. – pp. 281-304.
 45. Baranov A.N. *Kompyuternaya lingvistika // Vvedenie v prikladnuyu lingvistiku: uchebnoe posobie*. – M.: Yeditorial URSS, 2003. – 320 s.;
 46. Begmatova G.X. *O'zbek tili milliy korpusida idiomalar bazasini yaratish: Filol. fan. false. d-ri (PhD) diss.* – Termiz, 2021. – 143 b.;
 47. Vadyaev, S.E. *Linguistic Principles of Constructing and Using a Corpus of Texts to Study the Official Business Style of Modern German (Based on the DER Electronic Corpus):*

- Diss. Cand. Philological Sciences – Samara, 2005, 160 p.;
48. Vladimov, N.V. Corpus Subcode for Solving Translation Problems (Based on English and Russian): Diss. Cand. Philological Sciences – Moscow: RGB, 2005, pp. 26–26;
 49. Gorina, O.G. Using Corpus Linguistics Technologies to Develop the Lexical Skills of Students Majoring in Regional Studies in Professionally Oriented Communication in English: Diss. ... Cand. Ped. Sci. – Moscow, 2014. – 321 p.;
 50. Djafarova D.F. Models of Linguistic Analysis of Tadjik-Language Texts (Based on the Gazeleys of Khafiz): Diss. ... Cand. Philological Sci. – Dushanbe, 2013. – 151 p.
 51. Djumabaeva J. O'zbek va ingliz tillarida leksik hamda stylistik graduonimiya: Philological fan. d-ri. diss. autoref. – Toshkent: 2016. – 82 b. URL: http://tashgiv.uz/wp-content/uploads/2017/01/avtoref_jumabayeva.pdf;
 52. Zaxarov V.P. Korpusnaya lingvistika: Educational method. posobie/
 53. V.P. Zaxarov. – SPb, 2005. – 48 s. – Rejim dostupa: <http://download.yandex.ru/class/zakharov/plan.pdf>;
 54. Karimov R.A. O'zbek-ingliz parallel korpusini tuzishning lingvistik va dasturiy masalalari: Filol. fan. false. d-ri (PhD) diss. – Buxoro, 2022. – 148 b.;
 55. Magomedova, A.N. Corpus Linguistics and Contextual Resolution of Lexical Polysemy: Discussion (Ph.D. in Philology). Moscow: RGB, 2005, 132 p.
 56. Mamontova, V.V. Specifics of Translating Compound Words from English into Russian (Based on a Corpus of Published Texts): Discussion (Ph.D. in Philology). Stavropol: 2008, 139 p.;
 57. National Corpus of the Russian Language. Available from <https://ruscorpora.ru/>
 58. Plungyan, V.A. Corpus as a tool and as an ideology: some lessons of modern corpus linguistics// Russian language in scientific coverage. – M.: No. 2., 2008. – P. 7-20.
 59. Rakhmanova A.A. Computer methods for creating the national corpus of the Uzbek language: Philol. science. fake Dr. (PhD) diss. - Tashkent, 2021. - 161 p.;
 60. Romanov A.S. Methodology and programming complex for the identification of the author of unknown text: Autoref. diss....kand. tekhn. nauk. - Tomsk, 2010. 26 p.;
 61. Sabirova G. Methodology of teaching English pronouns to students of secondary schools based on information technologies: Philol. fals. d-ri (PhD) ...diss. author's ref. – Tashkent: 2017. – 41 p. URL: <http://uzswlu.uz/media/ba7be50d-84f1-c826-f935-6e7610cb6bf5.pdf>;
 62. Sazhenin I.I. Corpus methods in lexicography: an experiment in creating a model of a dictionary corpus: Disc. ... kand. philol. nauk. – Novosibirsk, 2013. 193 p.
 63. Toirova G.I. Theoretical and practical issues of creating a national corpus of the Uzbek language: Philol. fan. d-ri (DSc) diss. abstract – Bukhara, 2021. – 72 p.;
 64. Khamroeva Sh.M. Linguistic support of the morphological analyzer of the Uzbek language: Philol. fan. dr. (DSc) diss. abstract – Fergana, 2021. – 76 p.;
 65. Khamroeva Sh. Linguistic foundations of creating an author's corpus of the Uzbek language: Philol.bo'icha fals. dr. (PhD) ... diss. abstract – Karshi, 2018. – 52 p. URL: <https://e-books.tsue.uz/pdf-ozbek-tili-mualliflik-korpusini-tuzishning-lingvistik-ososlari>
 66. Kholiyrov O'M. Linguistic foundations of creating an educational corpus of the Uzbek language: Philol. fan. dr. (PhD) diss. – Termez, 2021. – 181 p.;
 67. Elova D.K. Principles of creating stylistic tags of units of the Uzbek language corpus and their linguistic support: Philol. fan. fals. dr. (PhD) diss. author's ref. – Tashkent, 2022. – 65 p.
 68. Ergasheva G. Linguistic and extralinguistic factors in the formation of gender terminology in languages of different systems: Philol. fan. diss. author's ref. – Tashkent: 2018. – 61 p. URL: <http://demo.navoiy-uni.uz/sites/default/files/ziyonet.pdf>;