

**Nozimjon ATABOYEV,**  
*BuxDU dotsenti, Filologiya fanlari bo'yicha falsafa doktori (PhD)*  
 anb929292@gmail.com  
 E-mail:n.b.ataboyev@buxdu.uz

*Buxoro davlat universiteti dotsenti, PhD M.Axmedova taqrizi asosida*

**MEDIA MATNLAR KORPUSINING MAVJUD INGLIZ TILI DIAXRONI KORPUSLARI PRIZMASIDA  
 TAHLILI**  
 Annotatsiya

Maqolada yaratilgan mediamatnlr diaxronik korpusi va hamda mavjud ingliz tilining shu turdagi korpuslar o'rtasidagi o'xshashlik va farqli jihatlar tahlilga totiladi. Keltirilgan o'zaro qiyoslash aniq chizmalar asosida amalga oshiriladi.

**Kalit so'zlar:** Korpus lingvistikasi, korpus, mediamatn, mediamatnlr diaxronik korpusi

**ANALYSIS OF THE MEDIA TEXTS CORPUS IN THE PRISM OF EXISTING ENGLISH DIACHRONIC  
 CORPORA**  
 Annotation

The article analyzes the similarities and differences between the diachronic corpus of media texts and the corpus of the existing English language. The given mutual comparison is carried out on the basis of clear drawings.

**Key words:** corpus linguistics, corpus, media text, diachronic corpus of media texts

**АНАЛИЗ КОРПУСА МЕДИАТЕКСТОВ В ПРИЗМЕ СУЩЕСТВУЮЩИХ АНГЛИЙСКИХ  
 ДИАХРОНИЧЕСКИХ КОРПОРА**  
 Аннотация

В статье анализируются сходства и различия диахронического корпуса медиатекстов и корпуса существующего английского языка. Данное взаимное сравнение осуществляется на основе наглядных рисунков.

**Ключевые слова:** корпусная лингвистика, корпус, медиатекст, диахронический корпус медиатекстов.

**Kirish.** Jahon tilshunosligida o'tgan asrning 60-yillarida lisoniy korpuslarni yaratish va ularning tatbiqi bilan shug'ullanuvchi – korpus lingvistikasining kompyuter texnologiyalari bilan o'zaro aloqasi jadallashdi. XIX asrning boshlarida tilshunoslik sohasi korpus lingvistikasining nazariy tamoyillari shakllanishi natijasida ilk namuna – Brovn korpusi yaratildi. Natijada, insoniyatning intellektual boyligini elektron korpus asosida avtomatik raqamlashtirish zamonaviy tilshunoslikning bugungi kundagi dolzarb vazifalari sirasiga kiritildi. Shu sababli til yoki tilning muayyan qismini o'zida ifoda etgan, yozma va og'zaki matnlarni umumlashtirgan elektron lisoniy baza – korpus tahlil metodlarining amaliy ifodasi bo'lgan milliy korpuslar til rivoji, leksik qatlam boyishi va uning bugungi kuni xususidagi statistik tahlilini aniq sonlarda takomillashtirish dolzarb ahamiyat kasb etadi.

**Mavzuga oid adabiyotlar tahlili.** Zamonaviy korpuslar taqdim etayotgan keng imkoniyatlarni aniqroq ko'rsatish uchun yuqorida sanab o'tilgan ingliz tili korpuslari konkordansida taqdim etiladigan ma'lumotlar shakli va formatimuhokamaga tortildi. Birinchi o'rinda Google Books n-gram korpusiko'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 tilisidan 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 o'z ichiga korpus konkordansda o'n yilliklar tartibida BChlarini jadval yoki grafika ko'rinishida taqdim etadi. Bunday ko'rinishdagi qidiruvlar asosan so'z qo'llashning rivoji yoki etimologiyasini o'rganishda qo'l keladi. Masalan, oddiy qidiruvda internet so'zi kiritildi va konkordans natijalari kiritilgan birlik yagona bo'lgani uchun bir chiziqli grafika orqali berildi (1-chizma). Grafikadan ko'rish mumkinki, ushbu tushuncha yigirmanchi asrning so'ngi ikki o'n yillikda keng qo'llanila boshlangan, lekin ungacha so'z chastotasi deyarli iste'molda bo'lmaganini ko'rsatyapti. Bu esa uning yaratilishi va ilm-fanda keng qo'llanila boshlangan davrlarga ishora qiladi.

1-chizma.

Google Books n-gram korpusida internet leksemasining qo'llanilish diaxronikasi

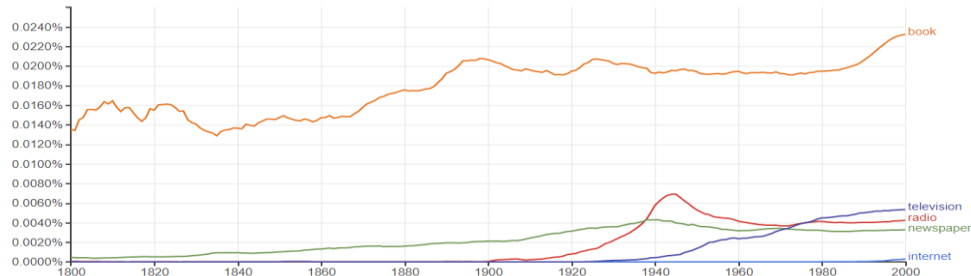


Google Books n-gram korpusida murakkab qidiruv amalga oshirilganda kiritmalar soni beshtani tashkil etdi. Ular internet, book, television, radio va newspaper kabi axborot almashishning boshqa turlari ham kiritmalar sifatida vergul orqali oraliq masofa tashlamasdan (probilsiz) konkordans qidiruviga berildi. Natijalar shuni ko'rsatdiki, har bir so'z o'z rivojlanish va shu asnoda tilga kirib kelishi mutanosibligida yillar oralig'ida mos ravishda kam va ko'p chastotalarni taqdim etdi (2-

chizma). 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 qoʻllanila boshlangan boʻlsa, kitob – book har ikkala asrda ham qoʻllanilish chastotasi jihatidan qolgan axborot vositalaridan eng koʻp qoʻllanilganini aniq koʻrish mumkin. Bu tahlillar yuqoridagilarning jamiyatga kirib kelish tarixi haqidagi bilimlarga ega boʻlmay turib ham, lisoniy korpus tahlillarini oʻrganish orqali kerakli va eng muhimi asosli xulosalarga kelish qiyin kechmaydi.

2-chizma.

Google Books n-gram korpusida internet, book, television, radio va newspaper leksemalarining qoʻllanilish diaxronikasi



Sketch engine formatining eng muhim va foydali tomonlaridan biri unda soʻzlar roʻyxati (word list)ning avtomatik shakllantirib berilishidir. Har qanday korpus kiritilgandan soʻng uning tarkibida qoʻllanilgan matnlarning soʻz-tur shaklda alohida tartiblanadi va har bir soʻz-turning chastotasiga koʻra jadvalda oʻrin egallaydi. Eng birinchi oʻrindagi soʻz-tur shu korpusda eng koʻp qoʻllangan soʻz-soʻzlarimbirlashtiradi. Shunday kuzatuvlarning biri Sketch engine internet korpuslar yaratish va ular bilan ishlash saytida mavjud BNC tarkibida qoʻllanilgan soʻz-turlarning chastotali tartibdagi jadvali shakllantirildi. BNC Milliy lisoniy korpus maqomiga ega boʻlganligi uchun undan olingan natijalar asosida chiqarilgan xulosalar oʻz-oʻzidan britaniyacha ingliz tiliga aloqador deb qarash oʻrinli boʻladi. Shunga asoslanib, eʼtirof etish mumkinki, ushbu tilda eng koʻp qoʻllaniladigan soʻzlar qaysilar ekani tez va aniq sonlarda koʻrsatib berildi. Unga koʻra, BNC tarkibida soʻz-soʻz (token)larning soni 95 548 323ni koʻrsatdi, bu chastota 155 434 soʻz-turning takroriy qoʻllanishidan kelib chiqqani maʼlum boʻldi. Ularning eng koʻp qoʻllanilgan yigirmatasi ajratib olindi (3-jadval).

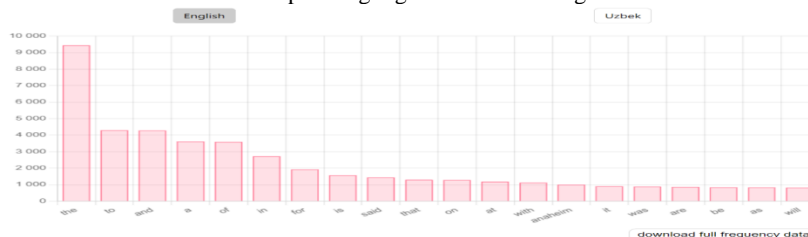
3-jadval

BNC tarkibida eng yuqori BCHli soʻzlar roʻyxati

r/d	Soʻz-tur	Bʻi	r/d	Soʻz-tur	Bʻi
1	The	6 054 939	11	For	880 805
2	Of	3 049 448	12	I	872 236
3	And	2 624 147	13	On	731 234
4	To	2 599 451	14	You	668 407
5	A	2 175 967	15	With	659 976
6	In	1 945 533	16	As	655 175
7	That	1 120 750	17	Be	651 542
8	It	1 054 366	18	He	641 241
9	Is	991 771	19	At	524 061
10	Was	883 547	20	By	513 428

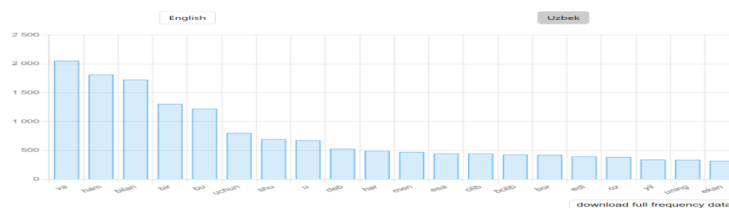
4-rasm

Media matnlari diaxronik korpusining ingliz tili bandida eng faol 20 ta leksik birlik



5-rasm

Media matnlari diaxronik korpusining oʻzbek tili bandida eng faol 20 ta leksik birlik



Yaratilgan media matnlari diaxronik korpusi oʻz ichiga jami 403500 ta soʻz birligidan tashkil topgan matniy elementlarni oladi. Ushbu matnlari “Oila va jamiyat” hamda “Anaheim Bulletin” kabi gazetalar tarkibidan jami 30 yillik

vaqt oralig'i, ya'ni 1990-yildan 2020-yilgacha bo'lgan muddatda chop etilgan nashrlardan yig'ilgan bo'lib, "Oila va jamiyat" gazetasidan 405 ta maqola hisobidan umumiy 202500 ta so'z birliklarini o'z ichiga olgan bo'lsa, bu miqdor "Annaheim Bulletin" gazetasidan 402 ta maqola va 201000 ta so'z sifatida aks etadi. Yuqorida keltirilgan ingliz tilining mavjud korpuslari tarkibidan olingan inglizcha so'zlarning faollik darajasi ro'yxati bilan solishtirilganda shu ma'lum bo'ladiki, biz yaratgan korpusda ham aynan shu funksional ko'makchi so'zlar tarkibiga kiruvchi artikl, predlog va ko'makchi fellar eng faol 20 talikda turadi. Farqli jihati shundaki, eng faol qo'llaniluvchi 20 talik so'zlar qatoriga 9-o'rinda said so'zi, 10-o'rinda unga ko'pincha ergashib keluvchi that bog'lovchisi kelgan bo'lsa, 14-o'rinda Annaheim so'zi qo'llanilganini ko'rishimiz mumkin. Bu esa, biz tomonimizdan yaratilgan korpus aynan media matnlar tarkibidan olinganini isbotlaydi va korpus matnlari aynan axborot yetkazish vazifasini bajarishiga ishora qilibgina qolmay, bu orqali tadqiqot natijalarining ishonchli ekanligini ham ko'rsatadi.

**Xulosa va takliflar:** Media matnlar diaxronik korpusining funksional xususiyatlari lisoniy birliklarning o'zaro semantik farqlanishi korpus tahlil natijalari asosida aniqlanadi. Ular sirasiga semantik birikmalarni o'rganishga qaratilgan izlanishlar uchun korpusning tutgan o'rni kiritilishi mumkin. Sintaktik tadqiqotlarda Media matnlar diaxronik korpusidan olingan misollar fe'l shakllari va ularning o'zaro grammatik munosabati hamda sintaktik qurilmalar xususida keng ma'lumot manbaidir. Media matnlar diaxronik korpusida frazeologik birliklar tadqiqi uchun zarur KWIC tizimi orqali birikishi mumkin bo'lgan so'zlar va kelishi mumkin bo'lgan orfografik belgilar tahlili imkoniyatlari mavjud. Morfologik birliklar tadqiqida Media matnlar diaxronik korpusi amaliy jihatdan morfologik birliklarning shaklan va birikma holatida ingliz tilida mavjud yoki mavjud emasligini ko'rsatadi.

Zamonaviy ingliz tili korpuslaridan bo'lmish ikkinchi avlod vakili Media matnlar diaxronik korpusi tarkiban linvomadaniy birliklarning tildagi ifodasini aniqlashga doir miqdoriy misollar manbaidir. Korpus tahlillarida muayyan tushuncha (konsept) tanlab olinadi va unga oid konkordans misollari bir tilli korpuslararo qiyosiy o'rganib chiqiladi. Madaniy birlikni ifodalashga kerak bo'lgan barcha ma'lumotlar yig'ilib, ikki madaniyatda «oila» tushunchasini anglashdagi o'zaro o'xshash va farqli jihatlar aniqlandi.

#### ADABIYOTLAR

1. Gries S.Th. Quantitative designs and statistical techniques/ English Corpus Linguistics edited by Douglas Biber and Randi Reppen. – United Kingdom: Cambridge, 2015. – P. 52.
2. Bobojon o'g'li, N. A. Corpus-Based Research On The Language Features Of Corpus Linguistics: In The Example Of ECOCL. Language, 3(2139), 950.
3. <https://www.english-corpora.org/googlebooks/>
4. <https://mediatextcorpus.com/>