

KORPUS LINGVISTIKASIDA MATNLARNI DAVRLASHTIRISHDA JAHON TAJRIBASI

Nuritdinov Abrorbek Sayfiddin o'g'li

University of Business and Science

Til va adabiyot ta'limi kafedrası o'qituvchisi

E-mail: nuritdinovabrorbek@gmail.com

ANNOTATSIYA

Mazkur maqolada korpus lingvistikasida matnlarni davrlashtirish masalasi hamda uning jahon miqyosidagi nazariy va amaliy yondashuvlari tahlil qilinadi. Turli tillarda yaratilgan tarixiy korpuslarda davrlashtirish mezonlari, usullari va avtomatlashtirilgan tahlil texnologiyalari o'rganiladi. Xususan, ingliz tilida yaratilgan diaxron korpuslar misolida davr chegaralarini aniqlash, lingvistik xususiyatlar asosida vaqt belgilash, shuningdek, sun'iy intellekt va mashinali o'rganish vositalari yordamida matnlarni tarixiy jihatdan tasniflash tajribalari ko'rib chiqiladi. Tadqiqot natijalari korpus lingvistikasi doirasida matnlarning tarixiy dinamikasini o'rganish, semantik o'zgarishlarni aniqlash va til taraqqiyotini modellashtirish uchun muhim ahamiyatga ega ekanligi ta'kidlanadi.

Kalit so'zlar: korpus lingvistika, davrlashtirish, diaxron korpus, tarixiy matnlar, lingvistik xronologiya, semantik o'zgarish, mashinali o'rganish, til dinamikasi.

ABSTRACT

This article analyzes the issue of text periodization in corpus linguistics and its theoretical and practical approaches on a global scale. Periodization criteria, methods and automated analysis technologies are studied in historical corpora created in different languages. In particular, on the example of diachronic corpus created in English, the experiments of defining time boundaries, setting time based on linguistic features, as well as historical classification of texts using artificial intelligence and machine learning tools are considered. The results of the study are emphasized in the framework of corpus linguistics for the study of the historical dynamics of texts, the identification of semantic changes and the modeling of language development.

Keywords: corpus linguistics, periodization, diachronic corpus, historical texts, linguistic chronology, semantic change, machine learning, language dynamics.

АННОТАЦИЯ

В данной статье анализируется проблема периодизации текста в корпусной лингвистике и ее теоретические и практические подходы в

глобальном масштабе. Критерии периодизации, методы и технологии автоматизированного анализа изучаются в исторических корпусах, созданных на разных языках. В частности, на примере диахронического корпуса, созданного на английском языке, рассмотрены эксперименты по определению временных границ, установлению времени на основе лингвистических особенностей, а также исторической классификации текстов с помощью искусственного интеллекта и инструментов машинного обучения. Результаты исследования выделены в рамках корпусной лингвистики для изучения исторической динамики текстов, выявления семантических изменений и моделирования языкового развития.

Ключевые слова: корпусная лингвистика, периодизация, диахронический корпус, исторические тексты, лингвистическая хронология, семантические изменения, машинное обучение, динамика языка.

KIRISH

Lingvistik korpus (til korpusi – linguistic corpus yoki text corpus¹) (“corpus” – lotincha “tana” degan ma’noni anglatadi) – elektron shakldagi ma’lum til birliklari majmuasi bo’lib, tadqiqotchilar uchun turli lingvistik yoki ekstralingvistik muammolarni hal etish manbasiga ega tizim sanaladi. Til korpusi – bu, bir tomondan, haqiqiy tilda yaratilgan katta hajmdagi elektron matnlar to’plami, boshqa tomondan esa, u ustida aniq va takrorlanuvchi tahlillar olib borish uchun mo’ljallangan tizimdir. U tadqiqotchilarga obyektiv va statistik metodlar orqali tilni o’rganish, o’quvchilarga esa o’rganishda real misollar asosida bilim olish imkoniyatini beradi. Annotatsiya va qidiruv vositalari orqali esa bu tizim boshqa filologik va ta’limiy platformalardan ustun turadi.

Jahon amaliy tilshunosligida til korpusi, korpus texnologiyasi va korpus lingvistikasi muammolarini o’rganish XX asrning ikkinchi yarmida rivojlanish bosqichiga ko’tarildi. XX asrda dastlabki ingliz tili korpusining paydo bo’lishi boshqa dunyo tillarida ham katta hajmdagi elektron korpuslarning yaratilishida muhim rol o’ynadi. Endilikda tabiiy tilning lisoniy modellari va nutqiy imkoniyatlarini kompyuter tiliga o’tkazish, til bilan bog’liq masalalarni axborot texnologiyalari hamda metodlari yordamida tabiiy tilni o’rganishda korpus obyekt vazifasini bajarmoqda. Tabiiy tillarning rivojlanish barqarorligi, ularning milliy sofligini saqlab qolishga qaratilayotgan bir davrda tillarning elektron korpuslarini yanada takomillashtirish va yangi texnologiyalarni yaratish axborot asrining dolzarb

¹ Кутузов А.Б. Курс «Корпусная лингвистика». – Москва, 2008. – 26 с.

masalalardan biri hisoblanadi². Bundan ko‘rinadiki, jahon tilshunosligida korpus lingvistikasi va korpuslarni yaratish jarayoni XX asrning ikkinchi yarmida yangi bosqichga yetgan. Ingliz tili korpusining yaratilishi boshqa til uchun ham elektron korpuslarning yaratilishida dastlabki qadam bo‘lib xizmat qildi. Ushbu korpuslar tabiiy tillarni axborot texnologiyalari yordamida chuqurroq va qulay usulda o‘rganishga imkon yaratdi. Bu orqali boshqa tillar uchun ham shunga o‘xshash resurslarni yaratishga turtki bo‘ldi. Natijada korpus tabiiy tillarni o‘rganish uchun ajralmas vositaga aylandi. Ular axborot texnologiyalaridan foydalangan holda til birliklarini tahlil qilish, nutq xususiyatlarini o‘rganish va turli lingvistik muammolarni hal qilish imkonini beradi.

Adabiyotlar tahlili

Korpus lingvistikasi lokomativi³ ingliz korpus lingvistikasida yaratilgan korpuslar bo‘lganligi tufayli korpusga xos xususiyat, korpus yaratilishi, rivojlanishini ingliz tili korpuslarining yaratilishi, taraqqiy etishini o‘rganishdan boshladik. Bulardan Birinchi avlod korpuslari hisoblangan Braun korpusi – Zamonaviy amerikanacha ingliz tilining Braun universiteti standart korpusi (Brown University Standard Corpus of Present-Day American English⁴) 1961-1964 yillarda tuzilgan. Korpus ingliz tilining amerika variantidagi yozma matnlaridan tashkil topgan, unda dastlab 1 million so‘z qo‘llangan (bunday miqdor dastlabki korpuslarning barchasi uchun hajm mezon bo‘lib qolgan edi)⁵. O‘sha davrda Noam Xomskiy konsepsiyasi⁶ hukmronlik qilganligi sababli Braun korpusi tuzuvchilari Nelson Frensis, Genri Kusera uchun korpus tuzish jarayoni ko‘p qiyinchilik tug‘dirgan. Korpus 2000 so‘zli 500 matndan iborat, qulay, puxta tuzilishga egaligi hamda matn tanlash tamoyilining har tomonlama to‘g‘riligi sababli keyinchalik korpus tadqiqotlari mezonlari yaratilishi uchun asos bo‘ldi. Korpuslardan tadqiqot uchun erkin foydalanish an‘anasini ham shu korpus boshlab berdi. Mazkur korpus asosida 1969-yildayoq “Amerika meros lug‘ati”ga (American Heritage Dictionary) asos solindi⁷. Ingliz korpuslari ustida olib borilgan uzoq muddatli tadqiqotlar korpus lingvistikasidagi muhim bosqich edi. Bu tajribalar asosida korpusshunoslik jadal sur‘atda rivojlana

² Abduraxmonova N. Korpus lingvistikasi. Darslik. – Toshkent, 2023. – 4-bet.

³ Zaxarov V., Mengliyev B., Hamroyeva Sh. Korpus lingvistikasi: korpus tuzish va undan foydalanish. O‘quv qo‘llanma. – Toshkent, 2021.

⁴ <https://www.sketchengine.eu/brown-corpus/>

⁵ Баранов А.Н. Корпусная лингвистика // Баранов А.Н. Введение в прикладную лингвистику. – М., 2003. – С. 112–137.; <https://core.ac.uk/download/pdf/75999967.pdf>

⁶ Chomsky N. Syntactic Structures. – Berlin, 1957.

⁷ Zaxarov V., Mengliyev B., Hamroyeva Sh. Korpus lingvistikasi: korpus tuzish va undan foydalanish. O‘quv qo‘llanma. – Toshkent, 2021.

bordi. D. Juravskiy davrlashtirgan Kompyuter lingvistikasining 1957–1970-yillardagi bosqichi⁸ korpuslarning yaratilishi bilan belgilandi.

Lankaster-Oslo/Bergen (LOB) korpusi (Lancaster-Oslo/Bergen (LOB) Corpus) 1970–1978-yillarda Lankaster, Oslo universiteti hamda Bergen ilmiy markazi loyihasi asosida tuzilgan. Ingliz tilining Britaniya varianti asosida 1 million soʻz qoʻllash holatidan tashkil topgan, tuzilishi jihatidan Braun korpusiga oʻxshaydi. Olimlar past chastotali birliklarni (ular esa tilda juda katta foizni tashkil etadi) tadqiq qilishda 1 million soʻzning yetarli emasligi haqida xulosa qilishdi. Bu xulosa juda oʻrinli edi. Shunday mukammal korpus mavjudligi sababli Braun, LOB korpuslariga asoslangan juda koʻp sifatli, qiziqarli tadqiqotlar amalga oshirilgan⁹. Bu tadqiqotlar natijasida korpuslar uchun keng koʻlamli lingvistik bazalar shakllantirila boshlandi.

Bundan koʻrinadiki, ingliz korpus lingvistikasi korpusga asoslangan ishlarning boy va uzoq yillik anʼanalariga ega. Helsinki korpusi, Braun korpusi va ARCHER kabi manbalar leksik va grammatik oʻzgarishlarni oʻrganish uchun zarur tadqiqot dasturlarini yaratdi. Bundan tashqari, korpus resurslarining davr, janr, uslublardagi diaxronik oʻzgarishlarini tahlil qila olish mumkinligi haqida maʼlumot beradi. Diaxronik korpuslar til taraqqiyoti davomida tildan foydalanishning taqqoslanadigan turlarini ifodalovchi matnli manbalar bilan ishlaydi, shuning uchun kamida ikki davrni oʻz ichiga oladi¹⁰. Masalan, Diachron Corpus of Present-Day Spoken English¹¹ (DCPSE). Hozirda mavjud boʻlgan ingliz korpus lingvistikasidagi korpuslar turli xil matn turlarini qamrab oladi va ularni davr nuqtayi nazaridan farqlaydi, lekin davr oʻtishi bilan matn tarkibidagi leksik va grammatik birliklarning oʻzgarib borishi bilan bogʻliq muammolarni chastotali shaklda doimiy nazorat qilib borish korpus uchun toʻsqinlik qilishi mumkin. Bu kabi holatlar korpus lingvistikasi uchun tadqiqot savollari boʻlib, chastota hisoblari javob berishi mumkin boʻlgan tarzda shakllantiriladi. Diaxronik matnlarning bugungi kun uchun leksik-semantik jihatdan eskirgan tili sinxron tilshunoslik uchun ham ahamiyatlidir. Misol uchun, Szmrecsanyi¹² (2010) 1960- va 1990-yillarda Britaniya va Amerika ingliz tillarining turli matn turlarida genitiv konstruktsiyalardan foydalanish ustida ish olib boradi, amerikalashtirish yoki soʻzlashuv tili sifatida koʻrilishi mumkin boʻlgan oʻzgarishlar mavjudligini oʻrganadi. Uning izlanishlari asosida miqdoriy korpus tadqiqotlariga kiradigan chastota soni koʻpincha token chastotalarini ifodalaydi, ammo muntazam

⁸ Abduraxmonova N. Kompyuter lingvistikasi. Darslik. – Toshkent, 2021. – 9-bet.

⁹ Zaxarov V., Mengliyev B., Hamroyeva Sh. Korpus lingvistikasi: korpus tuzish va undan foydalanish. Oʻquv qoʻllanma. – Toshkent, 2021.

¹⁰ https://www.stgries.info/research/2016_MH-STG_QuantMethInDiachrCorpLing_CUPHb.pdf

¹¹ <https://www.ucl.ac.uk/english/research-projects/2025/may/dcpse>

¹² <https://sites.google.com/site/bszmrecsanyi/>

ravishda turli xil o'lovlar, shu jumladan, tur chastotasi, tarqalish va kollokatsiya o'lovlarini qo'llaniladi.

Tadqiqot metodologiyasi

Korpus lingvistikasining keyingi taraqqiyotida korpus eskirgan matnni va uni o'zgarish chastotasini tahlil qila olish usullari birlashtirildi. Bu orqali leksik birliklarni davriy o'zgarishlari haqidagi quyidagi savollarga javob topish mumkin bo'ldi:

- leksik birliklardagi o'zgarishlar qachon va qanday qilib sodir bo'ladi?
- o'zgarish jarayonini alohida bosqichlarga bo'lish mumkinmi?
- so'zshaklning vazifaviy xususiyatlari bir-biridan mustaqil ravishda o'zgaradimi?
- o'zgarishga olib keladigan omillar qanday, ularning nisbiy ahamiyati nima va vaqt o'tishi bilan ular qanday o'zgaradi?
- o'tmishdagi til o'zgarish holatlari hozirgi kundagi o'zgarish bilan qanday solishtiriladi?

Bu savollardan ko'rinib turibdiki, tadqiqotlar tarixiy o'zgarishlar hozirgi o'zgarishlar bilan nazariy va metodologik jihatdan umumiylikka ega.

Korpus tadqiqotchisi korpus lingvistikasida mavjud bo'lgan usullarni tarixiy korpuslarga qo'llashda ehtiyot bo'lishi kerak. So'zshakllarning o'zgarishi, chastota tahlili, konkordans va avtomatik teglash kabi korpusning lingvistik usullari uchun jiddiy muammolarni keltirib chiqaradi, natijalarning aniqligiga sezilarli darajada ta'sir ko'rsatadi. Bunday vaziyatlarni inobatga olib ingliz korpus lingvistikasida hozirgi ingliz tilining shakllanish davridagi matnlarga alohida e'tibor qaratdi. Hozirgi ingliz tilining shakllanish davridagi matnlarni raqamlashtirishda ko'plab tadqiqotlar olib borildi. Google Books va Early English Books Online kabi tashabbuskor tadqiqotlar raqamlashtirilgan tarixiy matnlar miqdori ortib borayotganligi sababli bunday to'plamlar ichidan yordam beradigan usullardan foydalanish mumkin. EModE va boshqa tarixiy korpuslarni yaratish tadqiqotning muhim yo'nalishlariga aylangan. Ko'plab tarixiy ingliz korpuslari allaqachon ishlab chiqildi. Bularga Helsinki, ARCHER, Lampeter va ZEN kabi korpuslar kiradi. S. Muhammedovanning yozishicha: "Statistik ma'lumot ko'ra, Google, SCIPER, Yandex, Microsoft, ABBYY kabi yirik kompaniyalar o'z mablag'larini axborotni qayta ishlash, mashina tarjimasini, elektron lug'atshunoslik, korpus tadqiqotlariga sarflab, korpus lingvistikasi bo'yicha o'tkazilgan COLLING, DIALOG kabi xalqaro konferensiyalarga homiylik qilmoqda"¹³. Bundan tashqari Ochiq kontent ittifoqi (Open Content Alliance), Google kitoblar qidiruvi (Google Books Search) va Klassik ingliz online kitoblari (Early

¹³ O'zbek tilini dunyoga olib chiqish zarurati//2018-yil 18-iyun/ <http://marifat.uz/marifat/ruknlar/rasmiy/1375.htm>

English Books Online) kabi platformalar orqali tarixiy matnlarning katta qismi raqamlashtirilmoqda.

TAHLILLAR VA NATIJALAR

Diaxronik korpuslar tilning vaqt o'tishi bilan rivojlanishini tadqiq qilish uchun ishlatiladi. Ular turli davrlar, mintaqalar va janrlardagi matnlarni o'z ichiga olishi mumkin. Tarixiy matnlar bilan ishlovchi korpuslarning har biri tilshunoslik, adabiyotshunoslik va, qo'yingki, shu davrlar bilan bog'liq tadqiqotlar uchun muhim manba hisoblanadi. Ingliz korpus lingvistikasida tarixiy matnlarni korpuslarga joylash, davrlashtirish bilan bog'liq amaliy tadqiqotlar har jihatdan boyib, mukammallashib bormoqda:

№	Korpus nomi	Yil oralig'i	Xususiyatlari
1.	Helsinki	730–1710	Ingliz tilining qadimiy, o'rta va yangi davrlarini qamrab oladi. Grammatik va leksik o'zgarishlarni o'rganish uchun ishlatiladi.
2.	ARCHER	1600–1999	Britaniya va Amerika inglizlarining ko'p janrli korpusi
3.	Old Bailey Corpus	1720–1913	London sudlarining yozuvlaridan tuzilgan. Og'zaki til va ijtimoiy o'zgarishlarni tadqiq qilish uchun qo'llaniladi.
4.	COHA	1820–2010	Amerikadagi ingliz tilining 200 yillik tarixini aks ettiradi. 475 million so'zdan iborat bo'lib, turli janrlarni o'z ichiga oladi.

Ushbu diaxronik korpuslar o'zing vaqt qamrovi va bu qamrovdagi matnlarning xususiyatlaridan kelib chiqib ichki tuzilmalarni ishlab chiqqan. Bu tuzilmalar subkorpuslar hisoblanib davrlar oralig'ida ishlashda yanada aniqlik beradi.

Helsinki korpusi tarkibidagi uchta asosiy tarixiy davr va ularning bo'linmalari (subkorpuslari) quyidagi jadvalda aniq ko'rsatilgan. Har bir davr tilning grammatik, leksik va uslubiy o'zgarishlarini tahlil qilish uchun maxsus bo'linmalarga ajratilgan:

Asosiy davr	Yillar	Bo'linmalar (subkorpuslar)	Xususiyatlar
Qadimiy ingliz tili (Old English (OE))	750–1150	1. OE1 (750–950) 2. OE2 (950–1050) 3. OE3 (1050–1150) 4. OE4 (qo'shimcha matnlar)	– Anglsakslar davri matnlari (diniy, qonun, she'r). – G'arbiy Saksoniya lahjasi ustunlik qiladi.
O'rta ingliz tili (Middle English (ME))	1150–1500	1. ME1 (1150–1250) 2. ME2 (1250–1350) 3. ME3 (1350–1420) 4. ME4 (1420–1500)	– Normand istilosi ta'siridagi leksik o'zgarishlar. – Dialektal xilma-xillik kuchli.
Hozirgi ingliz tilining shakllanishi	1500–1700	1. EModE1 (1500–1570)	– Standartlashuv boshlanishi. – Shekspir davri matnlari, ilmiy

(Early Modern English (EModE))		2. EModE2 (1570–1640) 3. EModE3 (1640–1700)	asarlarning keng qamrovliligi.
---------------------------------------	--	--	--------------------------------

Mavjud matnlarning soni, sifati davrlar orasida sezilarli darajada farq qilganligi sababli davrlar va kichik davrlar bir xil darajada emas. Matn namunalari bosma nashrlardan kiritilgan, chunki qo‘lyozma matnlardan foydalanish xorijiy kutubxonalarda ancha vaqtni olar edi. Helsinki korpusidagi bo‘linmalar chastotali taqsimot, janrlar balansi va qo‘shimcha materiallar bilan bog‘liq ahamiyatilarga ham ega:

1. Chastotali taqsimot. Har bir subkorpus matnlarni 50–100 yillik davrlar oralig‘ida guruhlaydi. Bu fonetik o‘zgarishlarni (masalan, unlilardagi katta o‘zgarishlar – Great Vowel Shift) aniqroq kuzatish imkonini beradi.

2. Janrlar balansi. Har bir bo‘limda diniy, huquqiy, badiiy va ilmiy matnlar nisbati hisobga olingan (masalan, OE3da *Anglsakslar xronikasi* kabi tarixiy hujjatlar ko‘p).

3. Qo‘shimcha materiallar. OE4 kabi bo‘linmalar kam qo‘llaniladigan manbalarni (masalan, arxaik she‘rlar) o‘z ichiga oladi, bu esa tadqiqotchilarga noyob ma‘lumotlar beradi.

Bu tizimli bo‘linish tilshunoslarga aniq solishtirma tahlillar qilish imkoniyatini yaratdi.

ARCHER (*A Representative Corpus of Historical English Registers*) – bu ingliz tilining 1650–1990-yillar oralig‘idagi turli registrlerini (uslubiy turlarini) o‘rganish uchun mo‘ljallangan diaxronik korpus hisoblanadi. Uning asosiy maqsadi – tilning o‘zgarishini nafaqat vaqt, balki janrlar (registrlar) nuqtayi nazaridan tahlil qilish.

ARCHER korpusi 7 asosiy registr (uslubiy tur)ni qamrab oladi, ularning har biri turli davrlarda teng taqsimlangan:

Registr turi	Namuna matnlar	Ahamiyati
Ilmiy yozuvlar	Tibbiyot, fizika, falsafa asarlari	Terminologiya va mantiqiy tuzilmaning rivojlanishini kuzatish
Hujjatlar	Qonunlar, shartnomalar, rasmiy xatlar	Huquqiy va rasmiy tilning o‘zgarishi
Xatlar	Shaxsiy va rasmiy maktublar	Kundalik til va epistolyar uslubning evolyutsiyasi
Gazeta matnlari	Yangiliklar, maqolalar, reklamalar	Ommaviy tilning paydo bo‘lishi va o‘zgarishi
Badiiy adabiyot	Romanlar, hikoyalar, esselar	Adabiy uslubdagi izchil va o‘zgaruvchan elementlar

Diniy matnlar	Va'zlar, Bibliya sharhlari	Diniy diskursning saqlanishi yoki modernizatsiyasi
Nutq transkripsiyalari	Sud yozuvlari, ma'ruzalar (keyingi davrlarda)	Og'zaki tilning yozma tilga ta'siri

Har bir registr 50 yil vaqt oralig'ida (masalan, 1650–1700, 1700–1750) teng miqdorda namunalar bilan ajratilgan. Matnlar AQSh va Britaniya ingliz tili variantlaridan iborat (bu ikki variantdagi ingliz tilining farqlari tahlil qilinadi). Ayrim registrlar (masalan, gazetalar) faqat 1700-yildan keyin keng tarqalganligi sababli, ularning bo'linmalari davrlar bo'yicha moslashtirilgan.

ARCHER korpusida til o'zgarishlari 5 asosiy davrga bo'lingan:

1. 1650–1700-yillar – Erta yangi ingliz (Restoratsiya davri). Bu davr Angliya restoratsiyasi (Charles II taxtga qaytishi, monarxiyaning qayta tiklanishi) davriga to'g'ri keladi. Teatr va adabiyot rivojlanadi, ilmiy inqilob terminologiyani boyitadi. Tildagi lingvistik xususiyatlar: fe'llarning soddalashuvi (“*thou*” (eski 2-shaxs) kamayib, “*you*” keng tarqaladi), passiv konstruksiyalarning ko'payishi (ilmiy matnlarda “*is done*” kabi shakllar keng qo'llanila boshlaydi) va leksik qatlamning (fransuzcha va lotincha so'zlar (masalan, *government, science*) ko'chiriladi) boyishi bilan alohida ajralib turadi.

2. 1700–1750-yillar – Klassitsizm davri. Bu davr klassitsizm ruhi bilan bog'liq bo'lib, til strukturasida aniq, neytral uslubiy jihatlar kuchaygan. Lug'atlar va grammatika qo'llanmalari paydo bo'ladi. Tilda grammatik standartlashtirish (“*whom*” kabi rasmiy olmoshlar qat'iy qo'llaniladi), uslubiy tartiblilik (hukmni ifodalovchi “*shall*” fe'li hujjatlar va qonunlarda ustunlik qiladi) va sintaktik murakkablashuv (uzun, bog'lama gaplar (masalan, Jonatan Swift asarlarida) qo'llana boshlaydi) yuz beradi.

3. 1750–1800-yillar – Ma'rifat davri (Enlightenment). Bu davr ilmiy, ma'rifiy fikrlar va esse janrining kengayishi bilan xarakterlanadi. Amerika va Fransuz inqiloblari yangi siyosiy terminologiyani keltirib chiqaradi. Ensiklopediyalar targ'ib qilinadi. Til strukturasida orqali grammatik xususiyatlar tajribali tahlil qilinadi. “*You*” barcha ijtimoiy tabaqalarda keng qo'llaniladi. London lahjasiga asoslangan standart ingliz tili kuchayadi.

4. 1800–1850-yillar – Industrial inqilob davri. Bu davr sanoat inqilobidagi ijtimoiy va ilmiy rivojlanishning yuzaga kelishi til qurilishlarida ham aks etadi. Shaharlashuv va ishchilar sinfi tilga yangi so'zlar (*strike, union*)ni kiritadi. Texnik leksika kengayib *engine, railway, factory* kabi so'zlar paydo bo'ladi. Gazeta tili rivojlanadi, qisqa, aniq jumlar (*headlines*) keng tarqaladi.

5. 1850–1990-yillar – Zamonaviy davr. ARCHER korpusining bu davrida ingliz tilining Britaniya va Amerika variantlarida ko'p tahlillar amalga oshiriladi.

Radio, televideniye va internet tizimlari tilga tezroq o'zgarishlar kiritadi. Og'zaki nutq va ilmiy janrlar hali ham markaziy rol o'ynaydi, leksik yangilanishlar va kollokatsiya tendensiyalari ortadi. "I'm", "won't" kabi qisqartirishlar nutqda ustunlik qiladi.

ARCHER korpusi registrarlari ingliz tilining rivojlanishini o'rganish uchun noyob manba bo'lib, uning tuzilishi va davrlashtirish metodologiyasi til o'zgarishlarini ijtimoiy kontekstda tushunishga yordam beradi.

Bailey tarixiy korpusi (Old Bailey Corpus) – bu 1720–1913-yillar oralig'idagi Londondagi Old Bailey sudlarining yozma transkriptlaridan tuzilgan diaxronik korpus. U og'zaki nutq, huquqiy til va ijtimoiy o'zgarishlarni o'rganish uchun noyob manba hisoblanadi. Korpusda matnlarni davrlashtirish (periodizatsiya) yillar, ijtimoiy va lingvistik o'zgarish tamoyillari asosida amalga oshirilgan.

Old Bailey korpusi har 20–30 yillik bo'limlarga ajratilgan, bu esa til o'zgarishlarini aniq kuzatish imkonini beradi. Asosiy davrlar:

Davr	Yillar	Muhim lingvistik va ijtimoiy o'zgarishlar
Dastlabki davr	1720–1750	– sud jarayonlarida arxaik huquqiy atamalar ko'plab uchraydi; – nutqda kuchli sinfiy xilma-xillik kuzatiladi.
O'rta davr	1750–1800	– standart ingliz tilining shakllanishi bosqichi hisoblanadi; – jinoyat atamalarida aniqroq ifodalar paydo bo'ladi.
Industrial davr	1800–1850	– fabrika ishchilari bilan bog'liq yangi jinoyat turlari paydo bo'ladi; – matnlarda sintaktik soddalashtirilgan qurilmalar ko'payadi.
Viktoriya davri	1850–1913	– huquqiy tilning ixtisoslashuvi kuzatiladi; – jinoiy psixologiya terminlari paydo bo'ladi.

Dastlabki davrda (1720–1750) 1720-yillardanoq stenografiya yordamida sud nutqlari, deyarli, so'zma-so'z saqlana boshladi, bu nutqning real shaklini qayd etishda inqilobiy yangilik bo'ldi. Arxaik huquqiy atamalar keng qo'llanilgan. Sud kotiblar so'z ohangini aniqlashtirish uchun lahjalar va dialektlarni ham transkripsiyada yozishgan.

O'rta davrda (1750–1800) London lahjasiga asoslangan standart til shakllana boshlaydi. Jinoyat atamalarida aniqroq farqlar paydo bo'ladi. Nutqda qisqa, aniq gap tuzilmalari ko'payadi. Ma'rifat davri ta'sirida huquqiy tizimda sistemalik ortadi. Matbaa, nashriyot taraqqiyoti til standartlashuviga yordam beradi.

Industrial davrda (1800–1850) sanoat inqilobi yangi ijtimoiy muammolar va jinoyat turlarini keltirib chiqaradi. Leksika yangi so'zlar bilan boyiydi. Sintaktik soddalashish yuz beradi, ya'ni tezkor yozish talabi tufayli gaplar qisqaradi. Shaxs va

vaqtlarga oid murakkab jumlar oddiyroq tuzilmalar orqali ifodalanadi, bu esa sud ma'ruzalarini o'qishni osonlashtiradi.

Viktoriya davri (1850–1913) – qirol nomi bilan ishlab chiqilgan bu davr (1837–1901) davr o'ziga xos ijtimoiy, iqtisodiy va ilmiy o'zgarishlar bilan xarakterlangan, nom zahirida esa bu o'ziga xos davr haqida tushuncha mujassamlangan. Viktoriya davrida sud, ma'muriy-huquqiy tizimning so'z boyligi kengaydi, yangi terminlar paydo bo'ldi. Hujjatlar yanada aniq va tushunarli bo'lishi uchun shakl ishlab chiqiladi. Sud tizimida psixologik usullar qo'llanila boshlaydi va bu leksikaga o'z ta'sirini ko'rsatadi.

COHA (*Corpus of Historical American English*) – bu 1810–2009-yillar oralig'idagi Amerika ingliz tilining rivojlanishini o'rganish uchun yaratilgan diaxronik korpus hisoblanadi. U 400 million ortiq so'zdan iborat bo'lib, turli janrlardagi matnlarni (badiiy adabiyot, gazetalar, ilmiy asarlar va boshqalar) o'z ichiga oladi. COHA korpusida matnlarni davrlashtirish har 20 yillik bo'limlarga ajratilgan, bu esa til o'zgarishlarini aniq kuzatish imkonini beradi. Unda jami 10 ta davr mavjud:

Davr	Yillar	Muhim lingvistik va ijtimoiy o'zgarishlar
1-davr	1810–1829	– Amerika mustaqillik davri tilining so'nggi izlari qo'llangan; – arxaik so'zlar (“ <i>thou</i> ”, “ <i>hath</i> ”) hali ishlatiladi.
2-davr	1830–1849	– g'arbiy kengayish davri: yangi geografik atamalar paydo bo'lishi; – lotincha o'zlashmalar ko'payadi.
3-davr	1850–1869	– Fuqarolar urushi davri: siyosiy va irqiy atamalar faollashadi.
4-davr	1870–1889	– industrial inqilob: texnik atamalar leksikani boyitadi.
5-davr	1890–1909	– immigratsiya to'liqini: italyan, yahudiy va boshqa tillardan o'zlashmalar ko'payadi.
6-davr	1910–1929	– Birinchi jahon urushi: hududlarda dialektal leksika aralashuvi yuz beradi.
7-davr	1930–1949	– Buyuk depressiya va Ikkinchi jahon urushi: harbiy terminologiya rivojlanadi.
8-davr	1950–1969	– Sovuq urush davri: yadroviy qurollar, kosmik poygalar bilan bog'liq so'zlar tarqaladi.
9-davr	1970–1989	– kompyuter inqilobi: dasturiy ta'minot, internet paydo bo'ladi.
10-davr	1990–2009	– globallashuv: terrorizm, ijtimoiy mediya kabi yangi tushunchalar paydo bo'ladi.

COHA korpusi o'zida mavjud manbalarni davrlashtirishda har bir davr matnlar miqdorini teng qamrab olgan (har bir 20 yillik davrda taxminan 20–25 million so'z mavjud), bu esa til o'zgarishlarini solishtirishda statistik aniqlikni ta'minlaydi. Har bir davrda 4 ta asosiy janr (badiiy adabiyot; gazetalar va jurnallar; ilmiy asarlar; maqolalar va nutqlar) muvozanatli tarzda tartiblangan. Bular orqali leksik qatlamning

boyishi yoki eskirishi (*“telegraph”* (1830–1970) → *“email”* (1990–2009)), leksemalarning semantik taraqqiyoti, dialekt va kasb-hunarga xos terminlarning kengayishi, grammatik strukturaning soddalishishi kabi lingvistik xususiyatlarni kuzatish imkonini beradi. Korpus katta hajmliligi bilan (400 milliondan ortiq soʻz) batafsil statistik tahlillarga imkon beradi. 200 yil atrofidagi keng davr oraligʻi orqali tilning uzluksiz evolyutsiyasini koʻrish mumkin. Toʻgʻridan toʻgʻri foydalanish imkoniyatiga ega boʻlganligi uchun (online interfeys¹⁴) soʻzlar chastotasini tekshirish mumkin. Korpus faqat yozma matnlardan iborat, ogʻzaki matnlar mavjud emas.

XULOSA

Shunday qilib, Helsinki, ARCHER, OBC (Old Bailey Corpus) va COHA kabi diaxronik korpuslar til tarixining turli bosqichlarida yozilgan matnlarni oʻz ichiga oladi va tilshunoslarga grammatik, leksik va semantik oʻzgarishlarni vaqt boʻyicha nazorat qilish imkonini beradi. Masalan, Helsinki korpusi (730–1710) VIII asrdan XVII asrgacha boʻlgan matnlardan tashkil topib, “long-diachrony” yondashuvi orqali (tilshunoslikda til oʻzgarishlarini keng vaqt oraligʻida, yillar, asrlar yoki hatto ming yillik tarixiy kontekstda tahlil qilish usuli) til evolyutsiyasini keng vaqt oraligʻida tahlil qilishda qoʻllaniladi. Bu korpus grammatik va leksik faol birliklarni tahlil qilishda, xususan, sintez darajasidagi oʻzgarishlarni aniqlashda noyob asos boʻlib xizmat qiladi.

Boshqa yondashuv — “short and fat” strategiyasi (diaxronik korpus dizaynida qoʻllaniladigan yondashuv boʻlib, bu yondashuvda “short” – har bir davr uchun tanlangan kichik vaqt oraligʻi (masalan, 50 yil); “fat” – bu vaqt oraligʻida kattaroq, turli janrlardan toʻplangan boy matnlar toʻplami) bilan ishlab chiqilgan ARCHER korpusi (1600–1999) boʻlib, unda Britaniya va Amerika inglizlari har 50 yillik davrda turli janrlardan — badiiy, ilmiy, ogʻzaki — 3,3 million soʻz hajmda matnlarni taqdim etadi. Bu koʻp janrli tuzilma ARCHERni grammatika va uslub oʻzlashuvini zamon kesimida tahlil qilishda muhim vositaga aylantiradi.

Old Bailey korpusi (1720–1913) esa London sudlari yozuvlari asosida tuzilgan korpus boʻlib, ayniqsa sotsiollingvistik va ogʻzaki nutq jihatidan izlanishlarda katta ahamiyatga ega. Sud protseslari matnlari orqali XIX asr registrlaridagi leksik va pragmatik xususiyatlar aniq koʻrinadi.

Amerika ingliz tilining evolyutsiyasini 1820–2010-yillarga qamrab olgan COHA (Corpus of Historical American English), yaʼni 475 million soʻzli tuzilma matnlardan iborat boʻlib, leksik, semantik va sintaktik birliklarning vaqt bilan bogʻliqligini aniqlashda juda qulay hisoblanadi. Masalan, “to-V” konstruktsiyasidan

¹⁴ <https://www.english-corpora.org/coha/>

“V-ing” tuzilmalariga o‘tish tendensiyalarini aniqlashda COHA tahlillari muhim natijalar beradi.

Korpuslarning bu turlari turli yondashuvlarini sintezlab, integrativ model hosil qilish mumkin: Helsinki korpusining boy tarixiy qamrovi, ARCHERning annotatsiyalangan boyligi, Old Baileyning ijtimoiy konteksti va COHANing yillik kontnualligi birlashtirilsa, diaxronik tilshunoslikda – ayniqsa morfologik, sintaktik, leksik va sotsiollingvistik o‘zgarishlarni chuqur tahlil qilishda – mukammal tahliliy bazani tashkil etadi.

FOYDALANILGAN ADABIYOTLAR

1. Mengliev, D., Nabiyeva, D., Abdurakhmonov, A., Makhmudov, K., Nuritdinov, A., & Otemisov, A. (2025, June). Educational Text Analysis in Uzbek: Developing an NER Algorithm for Academic and Pedagogical Content. In *2025 IEEE 26th International Conference of Young Professionals in Electron Devices and Materials (EDM)* (pp. 2100-2103). IEEE.
2. Nuritdinov, A. (2025). MATNLARNI LINGVOSTATISTIK TAHLIL QILISHDA KORPUS USULLARIDAN FOYDALANISH. *Молодые ученые*, 3(19), 93-97.
3. Nuritdinov, A. (2025). KONKORDANS–LINGVISTIK TAHLIL VOSITASI SIFATIDA. *Теоретические аспекты становления педагогических наук*, 4(13), 173-178.
4. Nuritdinov, A. (2025). Korpus lingvistikasida lingvostatistik tahlil metodi. *MAKTABGACHA VA MAKTAB TA'LIMI JURNALI*, 3(5).
5. Nuritdinov, A. (2024). JADID DAVRI ADABIY MUHITIGA DOIR ASARLARDAN KORPUSDA FOYDALANISH. *COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS*, 1(1).
6. Nuritdinov, A. (2022). O ‘ZBEK TILI KORPUSI UCHUN ABDURAUF FITRATNING LINGVISTIK ASARLARINI MANBA SIFATIDA OLINISHI. *COMPUTER LINGUISTICS: PROBLEMS, SOLUTIONS, PROSPECTS*, 1(1).
7. Nuritdinov, A. S. O. G. L. (2022). O‘zbek tili milliy korpusi uchun jadid tilshunoslarining lingvistik asarlarini manba sifatida olinishi. *Science and Education*, 3(4), 2048-2057.
8. Xabibillayev, N. (2024). IZOFALI BIRIKMALARNING TARKIBGA AJRATILISHI BORASIDAGI QARASHLAR VA ULARNING AYRIMLARI TAHLILI (2-QISM). *Прикладные науки в современном мире: проблемы и решения*, 3(9), 18-20.

9. Xabibillayev, N. (2024). IZOFALI BIRIKMALARNING TARKIBGA AJRATILISHI BORASIDAGI QARASHLAR VA ULARNING AYRIMLARI TAHLILI. *Общественные науки в современном мире: теоретические и практические исследования*, 3(11), 79-80.
10. Xabibillayev, N. (2024). HOZIRGI O 'ZBEK TILSHUNOSLIGIDA RAVISH MASALASI (1-qism). *Молодые ученые*, 2(22), 17-22.
11. Xabibillayev, N. (2024). AZIM HOJIYEVNING MORFEMA BORASIDAGI QARASHLARI. *Педагогика и психология в современном мире: теоретические и практические исследования*, 3(8), 21-26.
12. Habibullayev, N. N. O. G. L., & QOZOQ, V. O. Z. T. M. (2023). SO'ZLAR CHOG'ISHTIRMASI.
13. Xabibillayev, N. (2024). Mumtoz Adabiyotni Targ 'Ib Etishda Adabiy Moslashtirishning Ahamiyati (—Lison Ut-Tayrl Ning Moslashtirilgan Matni Misolida). *Ta'lim innovatsiyasi va integratsiyasi*, 16, 65-71.
14. Xabibillayev, N. (2024). KEYINGI DAVRDA O 'ZBEK TILSHUNOSLIGIDAGI SO 'Z TARKIBI BO 'YICHA QILINGAN ILMIY TADQIQOTLAR TAHLILI. *Наука и инновация*, 2(22), 73-77.
15. Nosirjon o'gli, N. K. (2024). TRADITIONAL INTERPRETATION OF THE CONCEPT OF MORPHEME IN UZBEK LINGUISTICS. *American Journal of Philological Sciences*, 4(10), 53-56.
16. Ogli, N. K. N. (2024). ANALYSIS OF SOME WORDS IN UZBEK WITH COMPLEX MORPHOLOGICAL COMPOSITION. In *Next Scientists Conferences* (pp. 61-63).
17. Habibullayev, N. N. O. G. L. (2023). Qozoq Va O'zbek Tilidagi Modal So'Zlar Chog'Ishtirmasi. *Oriental renaissance: Innovative, educational, natural and social sciences*, 3(5), 27-33.