

2-TOM, 2-SON
INGLIZCHA ZAMONAVIY DIAKRONIK VA SINXRONIK
KORPUSLARI TAHLILI

Ataboyev Nozimjon Bobojon o'g'li

Filologiya fanlari bo'yicha falsafa doktori (PhD), dotsent

BuxDU Xorijiy tillar fakulteti dekani

Annotatsiya. Maqolada korpus lingvistikasi sohasida tillarning tadrijiy rivojini qayd etish borasida yaratilgan ingliz tilining zamonaviy diakronik korpuslari to'grisida nazariy ma'lumotlar beriladi.

Kalit so'zlar. Korpus lingvistikasi, til korpusi, diakronik korpus, ingliz tili korpuslari

Diakronik korpuslarni yaratish amaliyotida ingliz tilining ancha ildamlab borayotganini inobatga olganda, ushbu tilde yaratilgan shu turdagi korpuslarni sanab o'tish, ularning hajmi va xususiyatlarini inobatga olganda, o'zbek tilining tarixiy/diakronik korpusini yaratishga bo'lgan ehtiyoj aniq planda ko'zga tashlanadi. Bugungi kunda martinweisser.org saytiga ko'ra diakronik korpuslar hajmi soni, korpus yaratilgan davriga mos ravishda me'yorlashtirilgan holda yaratilishi ahamiyatlidir. Saytda sanab o'tilgan inglizcha diakronik korpuslarning ba'zilarini quyida sanab o'tish mumkin¹ (qarang 1-jadval).

1-jadval

Inglizcha tarixiy / diakronik korpuslar

<p><i>ARCHER Corpus (A Representative Corpus of Historical English Registers)</i></p>	<p>Ilk zamonaviy ingliz tilidan misollarni qamrab olgan 1650-1990 yillar oralig'idagi 7 ta tarixiy davrdan olingan yozma va "nutqqa asoslangan" janrlardan shakllangan jami 955 ta matn, ya'ni, 1 789 309 ta so'zni o'z ichiga oladi</p> <p>Uppsala, Xelsinki, Frayburg, Heidelberg, Lankaster, Manchester va Michigan universitetlari bilan korpus qamrovini kengaytirish bo'yicha doimiy hamkorlikdagi tadqiqot ishlari olib borilmoqda.</p>
<p><i>Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old</i></p>	<p>Ingliz matnlarining Xelsinki korpusining qadimiy ingliz bo'limi uzunroq matnlardan namunalar uzunligi 5 000 dan 10 000 so'zgacha bo'lgan qadimiy</p>

¹ *Historical & Diachronic Corpora Available from* http://www.martinweisser.org/corpora_site/historical_corpora.html [Retrieved on 04.02.2024]



2-TOM, 2-SON

<i>English</i>	ingliz tili tarkibidan olingan manbalari 106 210 soʻzni oʻz ichiga oladi. Korpus matnlari kompozitsiya, mualliflar va janrlarning bir qator sanalarini ifodalaydi. Matnlar sintaktik va morfologik jihatdan teglangan va har bir soʻz izohlanadi.
<i>Century of Prose Corpus</i>	1680-1780 yillar oraligʻida yashab ijod qilgan jami 120 muallif ijodidan nashr etilgan adabiy va adabiy boʻlmagan ingliz tilidagi yarim million soʻzdan tarkib topgan.
<i>Complete Corpus of Old English</i>	Toronto universitetida tuzilgan barcha qadimiy ingliz tili davrida yaratilgan 3022 ta matnlarni ifodalovchi korpus.
<i>Corpus of English Dialogues, 1560-1760 (CED)</i>	Erta zamonaviy ingliz tilidagi nutq shakllarini ifoda etuvchi 177 ta matndan tashkil topgan korpus 1 million 200 ming soʻzni oʻz ichiga oladi. <i>CED</i> manbalari matn sifatida dialog, sud jarayoni, jumladan, guvohlarning soʻzlari, drama, komediya, didaktik asarlar va badiiy adabiyot namunalaridan iborat.
<i>Corpus of Newsbooks</i>	1653-yil dekabrda 1654-yil maygacha nashr etilgan <i>Tomason</i> nashrlari barcha gazetalaridan olingan taxminan 800 000 soʻzdan tarkib topgan matn.
<i>Corpus of Early English Correspondence (CEEC) & the Parsed Corpus of Early English Correspondence (PCEEC),</i>	Korpus umumiy hisobda jami 2 million 700 ming soʻzdan tashkil topgan boʻlib, u oʻz ichiga 1410-yildan 1681-yilgacha 450 ming soʻz; 1402-1663 yillardan “Erta yozishmalar korpusi” 50 ming atrofida soʻz; 1681-1800 – yillarni qamrab olgan oraliqdan jami 2 million 200 ming soʻz kiritilgan.
<i>Corpus of Early English Medical Writing & Corpus of Middle English Medical Texts (MEMT)</i>	1375 - 1800 – yillardagi tibbiy nashrlar va qoʻlyozmalar korpusi boʻlib, 10-12 ming soʻzli risolalar bilan toʻldirilgan hamda oʻrta asrlar boʻlimi 500 mingga yaqin soʻzlarni oʻz ichiga oladi.



2-TOM, 2-SON

<i>The Málaga Corpus of Early English Scientific Prose</i>	1350-1900 – yillarda ishlab chiqilgan tibbiy materiallarni o'z ichiga olgan tarixiy diaxronik korpus. Korpus deyarli olti yuz yillik davrni o'z ichiga olgan uchta subkorporaga bo'lingan taxminan olti million so'zni o'z ichiga oladi. U tibbiyotga oid qo'lyozmalar, ya'ni nazariy risolalar, jarrohlikka oid yozma ma'lumot va retsept (dori qog'ozi)lar to'plamining taksonomiyasi nuqtai nazaridan me'yorashtirilgan va u diaxronik nuqtai nazardan lingvistik o'zgarishlarni o'rganish uchun ayniqsa foydali vosita sifatida ishlab chiqilgan. Korpus uch xil holatda taqdim etiladi: oddiy matn, me'yolashtirilgan va morfologik teglangan
<i>Newdigate Newsletters</i>	1674-1692 – yillarga tegishli qo'lyozma nashrlardan jami 750 ming so'zdan iborat.
<i>TIME Magazine Corpus</i>	TIME jurnalining 1923-2006 – yillar oralig'idagi nashrlaridan 100 million so'zdan iborat korpus. Vaqt o'tishi bilan so'z va iboralarning qo'llanishi yoki kamayishi, ma'nosi o'zgarganligini ko'rish imkonini beruvchi ma'lumotlar bilan ta'minlaydi.
<i>Zürich Corpus of English Newspapers (ZEN)</i>	1660 – yillardan XX asr boshlarigacha bo'lgan vaqt oralig'idagi London gazeta nashrlaridan tashkil topgan.

Yuqorida sanab o'tilgan lisoniy korpuslar ingliz korpus lingvistikasining yutuqlari sifatida e'tirof etiladi. Shu o'rinda mavjud tajribani qo'llagan holda o'zbek tilining ham diaxronik korpusni yaratishga nazariy asos sifatida tadqiqot ishi amalga oshirilmoqda. Bunda, ikki tilli, ya'ni, o'zbek va ingliz tillarining diaxronik korpusini ishlab chiqish ishimizning maqsadi etib belgilangan. Yozma matnlarning keng ko'lamda topilish mumkin bo'lgan manba bu mediamuloqot vositalari hisoblanar ekan, bizningcha, yaratilajak korpusimizning o'zagini aynan mediamatnlar asosida shakllantirish joiz.

FOYDALANILGAN ADABIYOTLAR RO'YXATI

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

