

## **KORPUS LINGVISTIKASI: korpus tuzish va undan foydalanish**

O`quv qo`llanma korpus lingvistikasi konsepsiyasi, shakllanishi, rivojlanishi va bugungi holati bilan tanishtiradi; korpus texnologiyalarini o`rganish, korpuslar bilan ishlash ko`nikmasini hosil qilish, kursning ahamiyati, xususan, korpusning axborot texnologiyalaridagi o`rnini aniqlash imkonini beradi.

Qo`llanma 5A120106 - Kompyuter lingvistikasi magistratura mutaxassisligida tahsil oluvchi talabalarga mo`ljallangan.

Mazkur o`quv qo`llanma Alisher Navoiy nomidagi Toshkent davlat o`zbek tili va adabiyoti universiteti Ilmiy-texnik Kengashining 2020-yil 27-oktabrda 12-sonli yig`ilish qarori bilan nashrga tavsiya etilgan.

Viktor Zaxarov - Sankt-Peterburg davlat universiteti dotsenti, filologiya fanlari nomzodi.

Baxtiyor Mengliyev - Alisher Navoiy nomidagi Toshkent davlat o`zbek tili va adabiyot universiteti professori, filologiya fanlari doktori.

Shahlo Xamroyeva - Alisher Navoiy nomidagi Toshkent davlat o`zbek tili va adabiyot universiteti doktoranti, PhD.

Globe  
EDIT



Victor Zaxarov · Baxtiyor Mengliyev ·  
Shahlo Xamroyeva

## **KORPUS LINGVISTIKASI: korpus tuzish va undan foydalanish**

O`quv qo`llanma



Globe  
EDIT

**Victor Zaxarov**  
**Baxtiyor Mengliyev**  
**Shahlo Xamroyeva**

**KORPUS LINGVISTIKASI: korpus tuzish va undan foydalanish**

FOR AUTHOR USE ONLY

FOR AUTHOR USE ONLY

**Victor Zaxarov  
Baxtiyor Mengliyev  
Shahlo Xamroyeva**

# **KORPUS LINGVISTIKASI: korpus tuzish va undan foydalanish**

**O`quv qo`llanma**

FOR AUTHOR USE ONLY

## **Imprint**

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: [www.ingimage.com](http://www.ingimage.com)

Publisher:

GlobeEdit

is a trademark of

International Book Market Service Ltd., member of OmniScriptum Publishing Group

17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page

**ISBN: 978-620-0-61917-4**

Copyright © Victor Zaxarov, Baxtiyor Mengliyev, Shahlo Xamroyeva

Copyright © 2021 International Book Market Service Ltd., member of OmniScriptum Publishing Group

FOR AUTHOR USE ONLY

**V.ZAXAROV, B.MENGLIYEV, SH.XAMROYEVA**

# **KORPUS LINGVISTIKASI**

**O'quv qo'llanma**

**Toshkent-2021**

**O‘ZBEKISTON RESPUBLIKASI  
OLIIY VA O‘RTA MAXSUS TA‘LIM VAZIRLIGI  
ALISHER NAVOIY NOMIDAGI  
TOSHKENT DAVLAT O‘ZBEK TILI VA ADABIYOTI UNIVERSITETI**

**V.ZAXAROV, B.MENGLIYEV, SH.XAMROYEVA**

# **KORPUS LINGVISTIKASI**

**O‘quv qo‘llanma**

**Toshkent-2021**

Zaxarov, Viktor Pavlovich, Mengliyev, Baxtiyor Rajabovich, Hamroyeva, Shahlo Mirdjonovna.

Korpus lingvistikasi [Matn]: o‘quv qo‘llanma / Zaxarov Viktor Pavlovich, Mengliyev Baxtiyor Rajabovich, Hamroyeva Shahlo Mirdjonovna. – Toshkent: Fan, 2021. – 185 b.

**Taqrizchilar:**

**Saodat Muhammedova** – filologiya fanlari doktori, Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyot universiteti professori

**Ayrat Gatiatullin** – texnika fanlari nomzodi, dotsent (Rossiya Federatsiyasi, Qozon, Tatariston Fanlar akademiyasi)

O‘quv qo‘llanma korpus lingvistikasi konsepsiyasi, shakllanishi, rivojlanishi va bugungi holati bilan tanishtiradi; korpus texnologiyalarini o‘rganish, korpuslar bilan ishlash ko‘nikmasini hosil qilish, kursning ahamiyati, xususan, korpusning axborot texnologiyalaridagi o‘rnini aniqlash imkonini beradi.

Qo‘llanma 70230801 – Kompyuter lingvistikasi magistratura mutaxassisligida tahsil oluvchi talabalarga mo‘ljallangan.

Mazkur o‘quv qo‘llanma Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universiteti Ilmiy-texnik Kengashining 2021-yil 27-oktabrda 12-sonli yig‘ilish qarori bilan nashrga tavsiya etilgan.



## MUNDARIJA

**SO‘ZBOSHI** .....

### **I BOB. KORPUS LINGVISTIKASIGA KIRISH**

Kirish. Korpus lingvistikasi fani obykti, predmeti, maqsad va vazifalari.....

Korpusning amaliy ahamiyati.....

Korpus lingvistikasining asosiy tushunchalari.....

Til korpusi turlari va tipologiyasi.....

### **II BOB. KORPUSDAN FOYDALANISH**

Jahon korpus lingvistikasining bugungi holati: milliy korpuslar va unga qo‘yilgan talablar.....

Rus va ingliz korpus lingvistikasi: tarix va bugun.....

Rus tili og‘zaki nutq korpusi. Rus tili maxsus mantlar korpusi.....

Turkiy korpus lingvistikasi tadriji va bugungi holati.....

Parallel korpuslar .....

Mualliflik korpuslari.....

### **III BOB. KORPUS TUZISH TEXNOLOGIYASI**

Korpus tuzish: bosqichlar va reprezentativlik.....

Tabiiy tilni qayta ishlash. Razmetka texnologiyalari.....

Annotasiya turlari: lingvistik va ekstralingvistik razmetka.....

Korpus qidiruv tizimi sifatida. Korpus menejeri va interfeys.....

Korpus asoslanuvchi leksikogragik, grammatik va diskurs tadqiqotlar.....

### **IV BOB. AMALIY MASHG‘ULOT UCHUN TOPSHIRIQLAR.....**

## I BOB. KORPUS LINGVISTIKASIGA KIRISH.

**Kirish. Korpus lingvistikasi fani obykti, predmeti, maqsad va vazifalari.**

**Reja:**

1. “Korpus lingvistikasi” fanining predmeti, maqsad va vazifalari.
2. Korpus lingvistikasining tilshunoslik fanlari orasida tutgan o‘rni.
3. Korpus lingvistikasi obykti, predmeti.

*Tayanch so‘z va iboralar: kompyuter lingvistikasi, korpus lingvistikasi, til korpusi*

**Korpus lingvistikasi va uning predmeti.** Korpus lingvistikasi jahon kompyuter lingvistikasining juda tez rivojlanib ketayotgan sohasi bo‘lib, bu borada ancha yutuqlarga erishilgan. Dunyo oliy ta‘lim muassasalarida korpus lingvistikasi fan sifatida ham o‘qitiladi. Bu sohaning predmeti korpus yaratish nazariyasi va amaliyoti bo‘lsa, fan sifatida korpusning o‘ziga xosligi, dasturlash asoslari kabi jihatlari o‘qitiladi. Korpus lingvistikasi kompyuter lingvistikasining tarkibiy qismi, til korpusini yaratish, kompyuter texnologiyasi yordamida ulardan foydalanishning umumiy nazariyasi va amaliyoti bilan shug‘ullanadi<sup>1</sup>. Korpus lingvistikasining predmeti–til korpusi. U ingliz tilida *linguistic corpus* yoki *text corpus* kabi atamalar bilan qo‘llanilib, ilmiy adabiyotlarda turlicha ta‘riflangan<sup>2</sup>. “Korpus – ma‘lum maqsadda yig‘ilgan matnlar majmuini tashkil etuvchi til birliklarining yig‘indisi”, “korpus – alohida tilning yozib olingan og‘zaki hamda yozma matnlari asosidagi lingvistik ma‘lumotlar majmuyi”, “korpus – istalgan tabiiy (real) tildagi elektron shaklda saqlanadigan yozma yoki og‘zaki, kompyuterlashtirilgan qidiruv tizimiga joylashtirilgan matnlar yig‘indisi” kabi. Demak, til korpusi<sup>3</sup> ma‘lum tilning belgilangan davrdagi, xilma-xil janr, rang-barang uslub, hududiy hamda ijtimoiy variantdagi matnlarining elektron shakldagi maxsus dasturiy ta‘minot asosidagi yig‘indisidir. Korpus haqida yuqorida keltirilgan ta‘riflar ham u haqida muayyan tilning o‘ziga xos xususiyati va variantlarini aks ettiruvchi bir necha

---

<sup>1</sup> Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 48 с. – С. 27.

<sup>2</sup> Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

<sup>3</sup> Национальный корпус русского языка // Вопросы языкознания. – Москва: Наука, 2006. – С. 149-155.

belgi asosida tanlab olingan elektron shakldagi matn parchalari, lingvistik tadqiqot uchun asos vazifasini o‘taydigan tizim deya xulosa qilishga asos bo‘ladi. Til korpusi bugungi kunda jahonda tez taraqqiy etayotgan soha – korpus lingvistikasi mutaxassislari tomonidan ilmiy tadqiqotlar olib borish, til o‘rgatish maqsadida yaratilmoqda. Jahon tillarining juda ko‘pchiligi mukammallik darajasi, matni (ilmiy) qayta ishlash imkoniyati bilan farq qiluvchi o‘z milliy korpuslariga ega. Zero, “bugungi kunda lingvistik tadqiqot va amaliy topshiriqlar yechimi uchun til korpuslari zamonaviy tilshunoslikning inkor etib bo‘lmas ish quroliga aylandi, chunki korpus turli lingvistik topshiriqlarni yechishga xizmat qiladi”<sup>4</sup>.

Korpus oddiy elektron kutubxona (mas., “Библиотека Максима Мошкова”, “Русская виртуальная библиотека”, “Фундаментальная электронная библиотека”, “ZIYONET kutubxonasi”, “Xurshid Davron kutubxonasi” kabilar)dan farq qiladi. Bunday elektron kutubxonalar xalqning ijtimoiy-siyosiy, ma’naviy, iqtisodiy hayotini aks ettiruvchi badiiy hamda publitsistik asarlarni nisbatan to‘laroq qamrab olishni maqsad qilgan bo‘ladi<sup>5</sup>. Faqat bunda elektron kutubxona matnlari tilshunoslik nuqtayi nazaridan leksik, morfologik, grammatik, semantik belgilar asosida ishlov berilmaganligi sababli lingvistik tadqiqotlarda foydalanish uchun noqulaylik tug‘diradi. Chunki elektron kutubxona ilmiy tadqiqot materialini bazasini tayyorlash uchun tuzilmaydi, balki ma’lum bir millat yoki butun insoniyatga tegishli milliy-ma’naviy merosni to‘plash, markazlashtirishni maqsad qilgan bo‘ladi. Til korpusi esa elektron kutubxonadan farqli ravishda muayyan tilni tadqiq qilish, o‘rganish va o‘rgatish uchun zaruriy, foydali va qiziqarli matnlar to‘planishini nazarda tutadi. Bunday matnlar esa nafaqat badiiy adabiyot namunasi, balki telefondagi oddiy suhbat yozuvi yoki ijara shartnomasi matni ham bo‘lishi mumkin.

Korpusni elektron kutubxonadan ajratib turuvchi birinchi omil undagi matnning xususiyati, qo‘shimcha ma’lumot bilan boyitilganligi hisoblanadi hamda

---

<sup>4</sup> Недошивина Е.В. Программа для работы с корпусами текстов: обзор основных корпусных менеджеров. Работа с системой DDC. // Языковая инженерия: в поиске смыслов/ (электронный ресурс).

<sup>5</sup> [http://wikipedia.org/wiki/Корпусная лингвистика](http://wikipedia.org/wiki/Корпусная_лингвистика). Статья в интернет-энциклопедии Википедия.

bu korpusning alohida qismi – korpus annotatsiyasi (rus., разметка)da o‘z ifodasini topadi.

Korpus – ma’lum maqsadda yig‘ilgan matnlar majmuini tashkil etuvchi til birliklari yig‘indisi, tabiiy tildagi elektron shaklda saqlanadigan yozma yoki og‘zaki, kompyuterlashtirilgan qidiruv tizimiga dasturiy ta’minot asosida joylashtirilgan on-line yoki off-line tizimda ishlaydigan matnlar jamlanmasi. Lingvistik tadqiqot va amaliy topshiriqlar yechimi uchun til korpuslari zamonaviy tilshunoslikning inkor etib bo‘lmas ish quroli. U oddiy elektron kutubxonadan farqlanadi. Elektron kutubxona xalqning ijtimoiy-siyosiy, ma’naviy, iqtisodiy hayotini aks ettiruvchi badiiy va publisistik asarlarni nisbatan to‘liq qamrab olishni maqsad qilgan bo‘ladi. Elektron kutubxona matnlari tilshunoslik nuqtayi nazaridan leksik, morfologik, grammatik, semantik belgilar asosida ishlov berilmaganligi sababli lingvistik tadqiqot uchun noqulaylik tug‘diradi. Chunki elektron kutubxona ilmiy tadqiqot materialini bazasini tayyorlash maqsadida tuzilmaydi, balki milliy-ma’naviy merosni to‘plashni maqsad qilgan bo‘ladi. Til korpusi esa elektron kutubxonadan farqli o‘laroq, tilni o‘rganish uchun zaruriy, foydali va qiziqarli matnlarni to‘plashni nazarda tutadi. Korpusni elektron kutubxonadan ajratib turuvchi birinchi omil undagi matnning xususiyati va qo‘shimcha ma’lumot bilan boyitilganligi hisoblanadi hamda bu belgi korpusning alohida qismi – korpus birliklariga yozilgan izohni tashkil etadi.

Foydalanuvchiga biror so‘z kerak bo‘lsa, buni odatiy matn muharriri ham topib beradi. Lekin so‘zning grammatik ma’nosi va tuzilishini “tushunadigan” matn bilan ishlash juda ko‘p afzallik va qulaylikka ega. So‘zshaklni qidirish kerak bo‘lsa, bunday dasturiy ta’minot, ya’ni korpus tadqiqotchi yoki foydalanuvchiga juda katta yordam berishi mumkin. Agar tadqiqotchi o‘z ishi uchun misolni topish, ularni kartotekaga ko‘chirish (kompyuter texnologiyalari rivojlanishidan oldingi davrda)ga oylab, ba’zan yillab vaqt sarflagan bo‘lsa, bugun dunyo til korpuslari yordamida sanoqli daqiqada yuzlab misolni topishga, ular ustida ishlash imkoniga ega bo‘ldi. Maxsus qidiruv korpusdan ma’lumot olishga mo‘ljallangan bir qancha dasturdan iborat, statistik axborot va qidiruv natijasini foydalanuvchiga qulay

shaklda taqdim etadi. Tilda qanday jarayon kechayotganligini aniq tasavvur qilish uchun korpus qamrovini yanada kengaytirish, nafaqat yozma nutq, balki og‘zaki nutq materialidan foydalanish ham maqsadga muvofiq. Bunday korpus yordamida til taraqqiyoti natijasida unda sodir bo‘lgan va kutilayotgan o‘zgarish haqida aniq xulosa chiqarish mumkin.

### **Mustahkamlash uchun savol va topshiriqlar:**

1. Korpus lingvistikasining maqsad va vazifalari nimalardan iborat?
2. Korpus lingvistikasining tilshunoslik fanlari orasida tutgan o‘rni haqida tushuncha bering.
3. Korpus lingvistikasi obykti, predmeti, metodlarini sanab bering.
4. Korpusning elektron kutubxonadan qanday farqli jihatlari mavjud?
5. Korpus menejerining vazifasi nima?
6. Konkordans haqida ma’lumot bering.
7. Korpus annotatsiyasi nima? Uning qanday turlarini bilasiz?
8. Korpusning hajmiga qanday talablar qo‘yiladi?
9. Korpuslar nimaga asosan klassifikatsiyasiya qilinadi?
10. Korpusning Web-sahifadan qanday farqi bor?
11. On-line va off-line korpus qanday mustarak va farqli jihatlarga ega? Web va prilojeniye shaklidagi korpusning o‘ziga xos xususiyati nimada?

### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. –М.: ФЛИНТА: Наука, 2013. –128 с.
3. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
4. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
5. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.
6. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.

7. [www.ruscorpora.ru](http://www.ruscorpora.ru)

12. <http://www.ruscorpora.ru> . НКРЯ

13. [www.philol.msu.ru/~lex/corpus](http://www.philol.msu.ru/~lex/corpus)

## KORPUSNING AMALIY AHAMIYATI

### Reja:

1. Korpusning lingvistika, lingvodidaktikada ma'lumotlarni qayta ishlash imkoniyati.
2. Korpusning leksikologiya, morfologiya, sintaksis hamda xorijiy tillarni o'qitishdagi ahamiyati.

*Tayanch so'z va iboralar: kompyuter lingvistikasi, korpus lingvistikasi, til korpusi, lingvodidaktika va korpus, lingvistika, matnni avtomatik qayta ishlash, tarjima dasturi, korpusning lingvistik funksiyalari.*

**Korpusning lingvistik ahamiyati.** Korpusda tilni keng, to'liq, obyektiv ko'rsatish imkoniyati bor, shu tufayli u har qanday axborot bankidan afzalligi bilan ajralib turadi. Korpusning qaysi sohadagi amaliy ahamiyati to'g'risida to'xtalishimizdan qat'i nazar, eng avvalo, uning representativlik xususiyati asosiy omil bo'lib qolaveradi. Korpus turli xildagi tadqiqot uchun asos, manba, vosita vazifasini o'tay olganligi uchun ham dunyoda juda ko'p sohalar korpusga tayanib ish ko'radi. Yuqorida aytilganidek, korpus maxsus qo'shimcha ma'lumotlar bilan boyitilgan, lingvistik tadqiqotlar uchun asos bo'la oladigan shunday matnlar massividan iboratki, ular oddiy elektron kutubxonadan farq qiladi. Ta'kidlash joizki, til korpuslari, eng avvalo, tilshunos uchun kerak. Zero, korpussiz bugungi kunning nazariy hamda amaliy filologiyasini tasavvur etish qiyin. Tilshunoslikka oid tadqiqotlar hamisha dalillar asosida ish ko'rishni, muayyan tadqiq mavzusi doirasida faktlar yig'ishni, to'plangan materiallarni sistemaga solishni taqozo etadi. Bunday katta hajmli ishni bajarishda korpus tilshunosning vaqtini, mehnatini tejaydigan bebaho ish qurolidir. Aslida, korpus texnik jarayonni tezlashtiruvchi vosita bo'libgina qolmay, ma'lum til zamonaviy shaklining axborot tizimi, kutilmagan savollarga ham javob bera oladigan, tilshunoslik oldiga avval qo'yilmagan dolzarb muammoni qo'ya oladigan tizimdir<sup>6</sup>.

Korpus asosida ish ko'radigan eng birinchi soha – leksikografiya, u katta hajmli lug'atlarni tuzish uchun asosiy, betakror manba hisoblanadi. Yaratilayotgan

---

<sup>6</sup> Қаранг: [www.ruscorpora.ru/](http://www.ruscorpora.ru/) Плунгян В. Зачем мы делаем корпусы?

barcha zamonaviy, so‘nggi lug‘atlar korpusga asoslangan, ular misollarining haqiqiyliги, ishonarliligi bilan baholanadi. Chunki korpusda til jamiyatda qanday yashasa, shunday aks etadi, natijada, lug‘atdagi misol ishonarli, asosli bo‘ladi<sup>7</sup>. Vaqt o‘tishi bilan korpuslar turli lingvistik yo‘nalishlar uchun ahamiyatli bo‘lishi bilan kuchli informatsion resursga aylandi. Chunki korpus leksikografiya sohasi uchun boy manba hisoblanib, ular asosida kompyuter yordamida lug‘atlar avvalgiga nisbatan tezlik bilan tuziladi, qayta ishlanadi. Shu yo‘l bilan ishning boshlanishidan tugash jarayoni (nashr)gacha tilni aks ettirib turadi, eskirishga ulgurmaydi. Ikkinchidan, korpusning leksikologiya sohasidagi ahamiyati shundan iboratki, so‘zning qo‘llanish davri, chastotasini aniqlashda hech qanday vosita unga tenglasha olmaydi. Korpus asosida ma‘lum so‘zning chastotasini aniqlash uchun berilgan qidiruv natijasida diagramma, grafiklar yordamida so‘zning tartib raqami uning chastotasiga teskari proporsional bo‘ladi, chunki ikkinchi tartib raqamida joylashgan so‘z birinchi raqamli so‘zga nisbatan kamroq, to‘rtinchisi uchinchisiga nisbatan kamroq ishlatilishi aniq<sup>8</sup>. Birorta chastotali lug‘at korpuschalik aniq ma‘lumot berolmaydi, chunki til doim o‘zgarishda, so‘zning chastotasi ham nisbiydir<sup>9</sup>.

Bundan tashqari, korpus tilning lug‘at boyligida bo‘layotgan o‘zgarishlar (neologizm, istorizm, arxaizm hodisalari)ni kuzatishning eng qulay vositasidir. So‘zlarning leksik-semantik birikish imkoniyati tahlilida korpus metodini qo‘llash yangi avlod lug‘at, grammatikalari, xususan, turg‘un birikmalar lug‘atini yaratish imkonini berdi. Korpus yaratilishi, taraqqiy etishi, til korpuslari asri kirib kelishi bilan lug‘atshunoslarda so‘zlarning qo‘llanilish holatlariga doir juda katta matnlar to‘plami bilan ishlash imkoniyati paydo bo‘ldi<sup>10</sup>.

Korpus vositasida so‘zning serqirraligi, bir paytning o‘zida bir necha semantik kategoriyaga mansub bo‘la olishi, bu semantik farqni ajratib olish mumkinligi haqida E.V.Nedoshivina: “Leksikografiya bo‘yicha tadqiqotlar

---

<sup>7</sup> Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

<sup>8</sup> Ўша манба.

<sup>9</sup> Ўша манба.

<sup>10</sup> Қаранг: [www.ruscorpora.ru/](http://www.ruscorpora.ru/)Что такое корпус? Электрон ресурс.



semantika borasidagi izlanishlar bilan chambarchas bog‘liq. Korpusda u yoki bu lingvistik birlikning qurshovini kuzata turib, ushbu birlikni xarakterlovchi ma’lum semantik belgilarni aniqlash mumkin. So‘z bir paytning o‘zida bir necha semantik kategoriyaga mansub bo‘lishi mumkin, shuning uchun so‘zning u yoki bu kategoriyaga tegishlilik darajasiga qarab fikr yuritish lozim. Daraja esa turli kategoriya bo‘yicha taqsimlanish chastotasini sanash yo‘li bilan aniqlanadi”<sup>11</sup>, deb yozadi. Demak, korpus yordamida so‘zlarning qaysi semantik kategoriyalarga mansubligini aniqlash, har bir kategoriyadagi ma’nosi bilan tanishish mumkin.

Til korpusi yordamida o‘tkazilgan ilk lingvistik tadqiqotlar turli birlikning tilda qo‘llanish chastotasini statistik aniqlashni ko‘zlaydi. Ko‘pincha, bunday element so‘z, ba’zida grafema, morfema, so‘z birikmasi ham bo‘ladi. Korpus lingvistikasining keyingi taraqqiyot bosqichida statistik tadqid metodi kompyuter tarjimasi, nutqni sintezlash, tanish, aniqlash, orfografik tekshiruv kabi lingvistik amallarni bajarishda qo‘l kela boshladi. Korpusdan foydalanish leksik birlikni kontekstda o‘rganish imkonini beribgina qolmay, so‘zshakl, lemma, grammatik kategoriyalar chastotasi, ularning birikish xususiyati haqida ma’lumot olish uchun ham muhim. Ibora yoki turg‘un birikmalar semantik jihatdan bo‘linmas birlikni tashkil etgani uchun leksikografiyada matnga avtomatik ishlov berish jarayonida buni inobatga olish masalasiga alohida e’tibor qaratmoq zarur. Korpus materiali asosida statistik metod orqali qaysi so‘zlar doim birgalikda qo‘llanishi, buning natijasida turg‘un birikmaga qanchalik aloqador ekanligini aniqlash mumkin<sup>12</sup>. Rus, Yevropa tilshunosligida til bo‘yicha qilinadigan barcha tadqiqotlarning til korpusiga asoslanishi odatiy holga aylangan, hatto ba’zi tadqiqotlarda majburiy ham sanaladi. Rus tilshunoslari bu sohada AQSH, Evropa, Yaponiyadan orqada qolganligini afsus bilan tan olishadi, vaholanki, rus korpus lingvistikasi ancha taraqqiyot yo‘lini bosib o‘tib, bu borada rivojlangan mamlakatlarga tenglashyapti. Mutaxassislarining fikricha, yaqin orada til o‘rganayotgan o‘quvchi yoki tilning

---

<sup>11</sup> Недошивина Е.В. Программа для работы с корпусами текстов: обзор основных корпусных менеджеров. Работа с системой DDC. // Языковая инженерия: в поиске смыслов/ (электронный ресурс).

<sup>12</sup> Каранг: Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие.–Санкт-Петербург, 2005. – 48 с.

biror jihatini tadqiq etayotgan tadqiqotchi, izlanuvchining lug‘atga bo‘lgan ehtiyoji korpusga ko‘chishi shubhasiz.

Sintaksisni o‘rganishning ahamiyati tilning jonli qurilishini, so‘zlarning o‘zaro birikish imkoniyatini tahlil qila olish bilan xarakterlanishini bilamiz. An’anaviy tilshunoslikdagi misolni olishda badiiy asar tiliga tayanish tajribasidan ko‘ra korpusga tayanish misolning bugungi kun uchun ham ishonarliligini ta’minlaydi<sup>13</sup>.

Uslubiyatni o‘rganishda, jumladan, matnning uslubiy xususiyatini tahlil etishda ham korpusga tayanish mumkin. Bu amaliyotni bajarishda matnlarning statistik holati tahlili(matndagi jumlaning uzunligi, bir so‘zning boshqa so‘z bilan birikish holatining odatiy yoki noodatiyligi)ni aniqlashtiradi. Mazkur usulda korpus yordamida yozma nutq bilan birga og‘zaki nutqni ham o‘rganish mumkin bo‘ladi<sup>14</sup>. Masalan, rasmiy ish qog‘ozlarida hujjat matnida *shartnoma bo‘yicha talablar* shakli ma’qulmi yoki *shartnoma talablari* konstruksiyasi to‘g‘rimi? Korpusga bu tarzda berilgan savolga sanoqli daqiqalarda qaysi shakl ko‘proq qo‘llanilishini aniqlab beradi. Shunday qidiruv natijasida rus tili milliy korpusida *shartnoma bo‘yicha talablar* konstruksiyasi 40 marta, *shartnoma talablari* shakli 3 marta qo‘llanilganligi to‘g‘risida statistik ma’lumot beradi. Bu konstruksiyaning qo‘llanilishi, albatta, rus tiliga xos, ehtimol, u o‘zbek tilida boshqacha natija bersa kerak.

*Korpusning boshqa sohalardagi ahamiyati.* “Korpus tor mutaxassislar uchungina kerakmi yoki yana qaysi soha vakili undan foydalanishi mumkin?” degan savollar kun tartibiga chiqadi. Korpus lingvistikasiga oid qator adabiyotlarda korpusning ahamiyati to‘g‘risida so‘z borar ekan, uning qo‘llanilish doirasining ancha kengligi ta’kidlanadi<sup>15</sup>. Korpus foydalanuvchilari doirasi biz o‘ylaganimizdan ko‘ra ancha keng. Lug‘at, grammatikadan faqat tilshunos foydalanmaganidek, korpus gumanitar fanlar tadqiqotchisi, adabiyotshunos hamda

---

<sup>13</sup> Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

<sup>14</sup> Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

<sup>15</sup> [www.ruscorpora.ru](http://www.ruscorpora.ru). Плунгян В. Зачем мы делаем корпусы?

tarixchi uchun ham birdek zaruriy baza bo‘la oladi. Til korpusiga eng ko‘p ehtiyoj sezuvchi mutaxassis matnni avtomatik qayta ishlash (masalan, tarjima dasturi), turli qidiruv tizimlari bilan ishlaydigan dasturchi. Chunki bu mutaxassis tabiiy til bilan ish ko‘radi hamda ushbu tilda yozilgan barcha matnlar strukturasi (tabiiy, jonli tilda, akademik grammatika, darslik yoki o‘quv qo‘llanmalardagi misollarga tayanib emas!) mukammalroq “tushunishi”, his etishi lozim. Til korpusiga ehtiyoj sezadiganlardan yana biri kundalik turmush tarzida yozuv bilan shug‘ullanuvchi mutaxassislar: gazeta, jurnal muharriri, jurnalist, radio, televideniye xodimi hisoblanadi. Chunki bunday xodim ma’lum bir so‘z, ibora yoki konstruksiyaning qo‘llanilish holati, darajasini, kim, qachon ilk marta shu konstruksiyani qo‘llaganligi, qanday uslub uchun xoslanishini bilishga grammatika bilan shug‘ullanuvchi olimdan ko‘ra ko‘proq ehtiyoj sezadi. Korpusdan tashqari yana qaysi axborot banki shunday savollarga tez, oson javob berishi mumkinligini tasavvur etib bo‘lmaydi. Shuning uchun korpus lingvistikasi tadqiqotchilari muayyan korpuslar jurnalist, muxbir, muharrir, o‘qituvchi hamda dasturchilar uchun maxsus yaratila boshlangan, degan xulosaga ham kelishadi. Biz bugungacha darslik, akademik grammatika, lug‘atga qanday tayangan bo‘lsak, zamonaviy soha vakillari korpusga shu qadar ehtiyoj sezishadi, undan foydalanishadi.

#### **Mustahkamlash uchun savol va topshiriqlar:**

1. Korpus annotatsiyasi nima? Uning qanday turlarini bilasiz?
2. Korpusning hajmiga qanday talablar qo‘yiladi?
3. Korpuslar nimaga asosan klassifikatsiyasiya qilinadi?
4. Korpusning Web-sahifadan qanday farqi bor?
5. On-line va off-line korpus qanday mustarak va farqli jihatlarfga ega? Web va prilojeniye shaklidagi korpusning o‘ziga xos xususiyati nimada?

#### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика.– Иркутск: ИГЛУ, 2013. – 161 с.
2. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. –М.: ФЛИНТА: Наука, 2013. –128 с.

3. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
4. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
5. Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2019. – 140 б.
6. Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти. Монография. – Тошкент, 2018. – 152 б.
7. Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун): Филол. фан. бўйича фалсафа доктори (PhD) диссер. – Фарғона, 2019. – 164 б.
8. [www.ruscorpora.ru](http://www.ruscorpora.ru)
9. <http://www.ruscorpora.ru> . НКРЯ
10. [www.philol.msu.ru/~lex/corpus](http://www.philol.msu.ru/~lex/corpus)
11. <http://www.philol.msu.ru/~lex/corpus>
12. <http://ruscorpora.ru>.
13. <http://ruscorpora.ru/search-murco.html>
14. [http:// corpus.leeds.ac.uk/](http://corpus.leeds.ac.uk/)
15. <http://corpus.leeds.ac.uk/ruscorpora.html>

## KORPUS LINGVISTIKASINING ASOSIY TUSHUNCHALARI

1. Lingvistik korpus va elektron kutubxona.
2. Korpus menedjeri va konkordanslar haqida.
3. Korpus annotatsiyasi, korpus hajmi va klassifikatsiyasi.

**Tayanch so‘z va iboralar:** *lingvistik korpus, elektron kutubxona, korpus menejeri, konkordans, korpus annotatsiyasi, web-sahifa, on-line va off-line korpus.*

*Korpus menedjeri.* Korpus foydalanuvchisi uchun turli imkoniyat va qulaylik yaratadigan, korpusdan zaruriy axborotni olishga yordamlashadigan bir necha maxsus dasturiy vosita – konkordans (qidiruv tizimining nisbatan oddiy ko‘rinishi) yoki korpus menejeri mavjud. Korpus menedjeri maxsus qidiruv tizimi, u korpus ma’lumotlarini olishga mo‘ljallangan bir necha dasturlardan iborat bo‘lib, statistik axborot va qidiruv natijasini foydalanuvchiga qulay shaklda taqdim etadi. Qidiruv natijasi, odatda, konkordans – qidirilayotgan birlik leksik bog‘lanish holatida manbaga ilova qilingan kontekstlar ro‘yxati shaklida namoyon bo‘ladi. Korpus menedjerining qidiruv imkoniyati aniq so‘zshakli, so‘zning *lemma* (korpusda leksema shunday ataladi) shakli, birikmali bo‘lingan va bo‘linmagan sintagma, morfologik belgilarga asoslangan jamlanma shaklini qamrab oladi. Talab qilingan hajmdagi kontekst ko‘rsatkichlari qidiruv natijasi, saralangan materialni alohida faylda saqlash imkonini beradi<sup>16</sup>.

Korpus yordamida bajariladigan amalning samaradorligini oshirish maqsadida korpus menedjeriga quyidagi umumiy talablar qo‘yiladi<sup>17</sup>:

- 1) to‘liq konkordans ro‘yxatini tuzish;
- 2) nafaqat alohida so‘zni, balki murakkab birlik qidiruvini amalga oshirish;
- 3) chiqarilgan ro‘yxat(qidiruv natijasi)ni bir necha mezon asosida saralash;
- 4) foydalanuvchiga qidirilgan so‘z shaklini kengaytirilgan ko‘rinishda taqdim etish;
- 5) korpusning har bir elementi haqida statistik axborot berish;
- 6) lemmalarni aks ettirish;

---

<sup>16</sup> [www.ruscorpora.ru/](http://www.ruscorpora.ru/) Плунгян В. Зачем мы делаем корпуса?

<sup>17</sup> Курузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

- 7) soʻzshakllarning morfologik xususiyati va qoʻshimcha axborotni (bibliografik, tipologik) berish (bu parametr korpusga berilgan qoʻshimcha izohga bogʻliq);
- 8) natijalarni saqlash va nashr etish uchun alohida faylda ishlay olish;
- 9) nafaqat korpus, balki alohida fayllar bilan ham yuqori tezlikda ishlay olish;
- 10) matnning barcha turdagi formatini (txt, rtf, doc, html, xml va h.) oʻzida saqlay olish;
- 11) tajribali va yangi foydalanuvchi uchun birdek engil va intuitiv tushunarli boʻlish.

*Korpus annotatsiyasi.* Turli lingvistik topshiriq/amalni bajarish uchun matnga lingvistik va ekstralingvistik qoʻshimcha maʼlumot bilan ishlov berilgan boʻlishi lozim. Buning uchun matn annotatsiyasi mavjud, matn, uning komponentlariga maxsus izoh (masalan, soʻz turkumi haqida maʼlumot) berilishi lozim. Bu izoh matn tegi (belgi, ishora) yoki *annotatsiya* deb ataladi. Mukammal annotatsiya tadqiqotchiga qidirilayotgan soʻz, soʻzshakl va konstruksiyani topishni osonlashtiradi. Chunki oddiy elektron matnda uning xarakterini ochuvchi alohida maʼlumot (annotatsiya) mavjud emas. Bizga shunchaki soʻz kerak boʻlsa, buni odatiy matn muharriri ham topib bera oladi. Lekin soʻzning grammatik maʼnosi, tuzilishini “tushunadigan” matn bilan ishlash juda koʻp afzallik va qulaylikka ega boʻladi. Faqat leksemani emas, balki soʻzshaklni qidirish kerak boʻlsa, bunday dasturiy taʼminot – korpus tadqiqotchi yoki foydalanuvchiga juda katta yordam berishi mumkin. Annotatsiyaning mukammal, toʻliqroq boʻlishi korpus ahamiyatini belgilovchi asosiy omildir. Masalan, katta hajmli matn (aytaylik, 50-100 sahifa)dan oʻtgan zamoni bildiruvchi grammatik shakllarni topish topshirigʻini qoʻyamiz. Bunday vazifani oddiy matn muharriri, tabiiyki, bajara olmaydi. Berilgan grammatik shaklni topish faqat oldindan ishlov berilgan matndagina bajarilishi mumkin, aks holda, qidiruv faqat qoʻlda bajariladi. Bu esa foydalanuvchidan koʻp mehnat hamda vaqtni talab qiladi<sup>18</sup>. Mukammal ishlov berilgan, katta hajmli, turli janrdagi matnlarni qamrab olgan korpus tadqiqotchi uchun bebaho yordamchi hisoblanadi. Chunki til materiali oʻrganilgan har qanday

---

<sup>18</sup> Курузов А.Б. Корпусная лингвистика. – (Электрон ресурс): Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

tadqiqot materialining boyligi, ishonarliligi, aniq va mukammalligi bilan baholanadi. Agar tadqiqotchi o'z tadqiqoti uchun misolni topish, ularni kartotekaga ko'chirish (kompyuter texnologiyalari rivojlanishidan oldingi davrdagi holat)ga oylab, ba'zan yillab vaqt sarflagan bo'lsa, bugun dunyo til korpuslari yordamida sanoqli daqiqada zarur deb hisoblangan yuzlab misolni topish, ular ustida ishlash imkoniyatiga ega bo'ldi. Til korpuslarida bugungi kunda annotatsiyaning morfologik, sintaktik, semantik, aksentik kabi bir necha ko'rinishi mavjud. Yaqin orada korpus lingvistikasi mutaxassislari tomonidan so'z yasalish annotatsiyasining kiritilishi ham kutilmoqda.

*Korpusning xususiyatlari.* Til korpusi hajm jihatdan imkon qadar katta bo'lishi, 100 millionlab so'z qo'llanish holatini qamrab olishi lozim. Misol tariqasida rus tili milliy korpusi tarkibiga kirgan "Dostoyevskiy asarlari korpusi" uchun yig'ilgan yozuvchining barcha asarlari matni 2 million so'zdan iborat ekanligini keltirish o'rinli. Shu sababli korpusni til modelining kichraytirilgan ko'rinishi deb atash mumkin<sup>19</sup>.

Endi til korpusi yordamida yechimini topadigan muammolardan biri – uning til taraqqiyotini kuzatish, o'rganishdagi ahamiyatiga, hatto mikrotarixiy tilshunoslik muammosi – qisqa davr ichida (aytaylik, oxirgi o'n yillikda) tilda bo'ladigan o'zgarishlarni tadqiq etishdagi qulayligiga to'xtalamiz. Kuzatishimizcha, ilmiy sohada mazkur yo'nalishda hozirgacha deyarli biror tadqiqot amalga oshirilmagan. Bunday qisqa davrdagi o'zgarishni tadqiq etishga rus akademigi V.V.Vinogradov juda qiziqqan. Shu kabi bir-ikki izlanishni hisobga olmasak, tilshunoslikda qisqa davrda bo'ladigan o'zgarishlar maxsus tadqiq manbai bo'lmagan. Ta'kidlash zarurki, bugungi kunga kelib, qisqa davrdagi o'zgarishni maxsus tadqiq etishga hojat ham qolmadi, sababki, bir qadar mukammal yaratilgan korpuslar har qanday masalaga birpasda oydinlik kiritadi.

Ma'lum bo'ladiki, korpus – matn asosida ma'lum vazifani bajaruvchi dastur. Korpus haqida aniqroq tasavvur hosil qilish maqsadida "Bu qanday qidiruv

---

<sup>19</sup> Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 48 с. – С. 8.

dasturi?”, “Biz undan nimani qidiramiz va uning yordamida qanday ishni bajarish mumkin?” kabi savollarga javob izlaymiz.

Bu ijtimoiy yoki lingvistik mazmun kasb etuvchi soʻz va iboraning turli nutqiy vaziyatda qoʻllanish holati yoki uning izohi boʻlishi mumkin. Korpus boʻyicha qidiruv taxminan quyidagi natijalarni beradi<sup>20</sup>:

- 1) belgilangan soʻzning turli kontekstdagi barcha shakldagi koʻrinishi;
- 2) tilning lugʻat tarkibidagi oʻrni va variantlari;
- 3) belgilangan soʻz bilan birikish imkoniyatiga ega soʻzlar roʻyxati;
- 4) u yoki bu yozuvchining ayni soʻzdan foydalanish chastotasi yoki statistikasi;
- 5) soʻzning denotativ, konnotativ maʼnosi;
- 6) leksik qoʻllanishning yashirin modeli (imkoniyati);
- 7) til taraqqiyotining turli davrida qoʻllanish holati.

*Korpus tarkibidagi materialning turi va xususiyati.* Mavjud korpuslar tarkibidagi matnlar nisbatida badiiy adabiyot hissasi 40 foizini tashkil etadi. Bu tarkibga til xususiyati badiiy, publitsistik uslub oraligʻida, jonli tilni oʻrganish uchun ancha qulay sanalgan memuar asarlar ham kiradi. Yevropa tillari korpuslarida badiiy adabiyot materiali 20 foizini tashkil etadi. Zamonaviy yozuvchilar asarlarining til xususiyatini oʻrganishga bagʻishlangan 20 dan ortiq tadqiqot majud boʻlsa-da, ular hali toʻlaligicha mazkur muammoni oʻrganib boʻldi, deyish qiyin. Chunki maʼlum bir yozuvchi asarining til xususiyatidagi oʻzgarishga hali tildagi yangi hodisa deb qarab boʻlmaydi. Ana shunday jarayonlarni kuzatish, tadqiq etishning eng qulay vositasi til korpusidir. Shu boisdan ham bu borada matn mazmuni muhim ahamiyat kasb etadi.

Korpus tarkibiga kiradigan matnlar quyidagi turlarda boʻlishi mumkin:

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

---

<sup>20</sup> Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 48 с. – С. 34.



### **Mustahkamlash uchun savol va topshiriqlar:**

1. Lingvistik korpus va elektron kutubxona tushunchasining farqli jihatlarini tushuntiring.
2. Korpus menejerining vazifasi nima?
3. Konkordans haqida ma'lumot bering.
4. Korpus annotatsiyasi nima? Uning qanday turlarini bilasiz?
5. Korpusning hajmiga qanday talablar qo'yiladi?
6. Korpuslar nimaga asosan klassifikatsiyasiya qilinadi?
7. Korpusning Web-sahifadan qanday farqi bor?
8. On-line va off-line korpus qanday mustarak va farqli jihatlarfga ega? Web va prilojeniye shaklidagi korpusning o'ziga xos xususiyati nimada?

### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
3. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
4. [www.ruscorpora.ru](http://www.ruscorpora.ru)
5. <http://www.ruscorpora.ru>. НКРЯ
6. [www.philol.msu.ru/~lex/corpus](http://www.philol.msu.ru/~lex/corpus)
7. <http://www.philol.msu.ru/~lex/corpus>

## KORPUS TURLARI VA TIPOLOGIYASI

### Reja:

1. Lingvistik korpus va elektron kutubxona.
2. Korpus menedjeri va konkordanslar haqida.
3. Korpus annotatsiyasi, korpus hajmi va klassifikatsiyasi.
4. Korpusning nutqqa ko‘ra turi: og‘zaki, yozma, aralash korpus.
5. Matn tili, parallelligi, matnning ixtisoslashuvi va janriga ko‘ra korpus turlari.
6. Kirish usuli, maqsadi, izoh xususiyati, o‘zgaruvchan va turg‘unligi, strukturasi, matn hajmi, xronologik jihatiga ko‘ra turlari tavsifi.
7. Erkin va yopiq, tadqiqiy va tavsifiy, razmetkalanmagan va razmetkalanmagan korpuslar.

*Tayanch so‘z va iboralar:* on-line va off-line korpus, og‘zaki, yozma, aralash korpus, matn tili, parallelligi, matnning ixtisoslashuvi va janriga ko‘ra korpus turlari, kirish usuli, maqsadi, izoh xususiyati, o‘zgaruvchan va turg‘unligi, strukturasi, matn hajmi, xronologik jihatiga ko‘ra turlari tavsifi.

**Korpus turlari tasnifi.** Korpusning tuzilishi, maqsadi, turg‘un, o‘zgaruvchanligi va boshqa jihatlariga ko‘ra bir qancha turlarga ega bo‘lib, bu mutaxassislar tomonidan turlicha tasniflangan<sup>21</sup> (*Qarang:* 1.1-jadval). Olimlar maqsadi hamda oldiga qo‘yilgan vazifasiga qarab korpus turlarining boshqacha tasnifini ham keltiradi<sup>22</sup>: saqlash shakliga ko‘ra (ovozli, yozma, aralash) (a); matn tiliga ko‘ra (bir va ko‘p tilli) (b); janriy mansubligiga ko‘ra (adabiy, dialektal, og‘zaki, publitsistik, aralash) (v); korpusga kirish imkoniyatiga ko‘ra (erkin, tijorat korpuslari, yopiq) (g); maqsadiga ko‘ra (tadqiqiy, illyustrativ) (d); dinamikligiga ko‘ra (dinamik (monitor), turg‘un) (d); qo‘shimcha axborotga egaligiga ko‘ra annotatsiyalangan (teglangan) va teglanmagan (e). V.V.Rikov korpus tasnifiga boshqacha yondashadi<sup>23</sup>.

### I. Tashkil qilish darajasi va qurilishiga ko‘ra:

<sup>21</sup> Корпусная лингвистика. Статья в интернет-энциклопедии Википедия. <http://wikipedia.org/wiki/>

<sup>22</sup> Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – 48 с.

<sup>23</sup> Қаранг: Корпусная лингвистика. Институт лингвистических исследований РАН, 2008 (электрон ресурс).

1. Elektron arxiv – elektron tashuvchi qurilmadagi matnlar jamlanmasi; faqat shakli standartlanmagan, bir xir shaklga keltirilmagan.
  2. Elektron kutubxona – bir turdagi matnlar keltirilgan; standart shaklga keltirilgan.
  3. Matnlar korpusi – standartlangan, bir xir shaklga keltirilgan; matnlar lingvistik voqelikning ma'lum qismini aks ettirishga mo'ljallangan.
  4. Subkorpus – ma'lum bir korpusning avtonom qismi.
- II. Xronologik belgisiga ko'ra: sinxron, monitor (tilning hozirgi holatini kuzatuvchi), diaxron.
- III. Ko'rsatkichning mavjudligiga ko'ra: oddiy, annotatsiyalangan.
- IV. Tiliga ko'ra: bir, ikki, ko'p tilli.
- V. Qo'llash, foydalanish usuliga ko'ra: tadqiqot korpuslari, illyustrativ, parallel.
- VI. Mavjudlik usuliga ko'ra: dinamik, turg'un.

Barcha tillar uchun korpusni loyihalash, qo'llashning bir xil metodikasi mavjud emas, chunki til, an'ana, texnologik jarayonlar bir-biridan farq qiladi. Shunday bo'lsa-da, bu borada V.V. Rikov quyidagi asosiy talablarni sanab o'tadi:

1. Korpus foydalanuvchisi (yakka shaxs, guruh, lingvistik jamiyat).
2. Korpus ko'zlagan mantiqiy maqsad.
3. Korpus tuzishda ishlanadigan ma'lumotlar bazasi hajmi hamda buning real, zarurlik darajasi.
4. Mantdan foydalanish usuli (parcha, to'liq ko'rinish yoki har ikkisi).

Korpusga matn tanlash amaliyoti har xil maqsadlarda turlicha bo'ladi:

1. Til materialini tadqiq etish.
2. Matnni skanerlash.
3. To'liq shakllantirish, korpusni tuzish.

Yuqoridagi jadvalda ko'rinib turganidek, eng avvalo, yozma, og'zaki manbaga asoslanishiga ko'ra korpus 2 turga bo'linadi<sup>24</sup>. Korpusning dastlabki avlodi to'liq yozma korpus sanalardi. Chunki yozma matnlarni yig'ish ancha qulayroq bo'ladi. Yozma matnlarni kompyuterga kiritishning 3 usuli mavjud:

1. Matnlarni qayta kiritish.

---

<sup>24</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

2. Elektron shaklda mavjud matnlardan foydalanish.
3. Bosma matnni skanerlash (lekin bunda ko‘plab xatoni tuzatish kerak bo‘ladi).

Tajriba shuni ko‘rsatadiki, korpus lingvistikasi kamida 2 tipdagi obyekt (korpus) ustida amaliyot bajaradi. Korpuslarning birinchi tipi universal, ular o‘zida nutqiy faoliyatning barcha xilma-xilliklarini aks ettiradi. Ikkinchi turdagi korpus jamiyatda nutqiy amaliyotning ma‘lum bir lingvistik yoki madaniy fenomenning odatga aylanishi (yoki aylanganligi)ni aks ettirib, ular ma‘lum maqsadni ko‘zlab tuziladi. Bunday korpus sifatida maqollar korpusi, gazeta tilida siyosiy metaforalar korpusini sanab o‘tishimiz mumkin.

**Korpus: yozma va og‘zaki, bir tilli va ko‘p tilli.** Mutaxassislar nutq turiga ko‘ra korpusning yozma, og‘zaki, aralash shaklini ajratishadi<sup>25</sup>. Yozma korpusda og‘zaki nutq aks ettirilmagan bo‘lsa (Braun korpusi, LOB), og‘zaki korpus faqat og‘zaki matndan iborat bo‘ladi; aralash korpus, odatda, til me‘yorlari aksi bo‘lgan milliy korpuslardir. Zamonaviy katta hajmli korpuslar, odatda, aralash: yozma matn bilan birga og‘zaki matnni ham qamrab oladi. Hatto Britaniya milliy korpusi (BNC) matnlarining 10 foizi og‘zaki matnlar yozuvidan iborat. Til, asosan, og‘zaki shaklda yashaydi, uning yozma shakli ikkilamchi. Shuning uchun og‘zaki korpus juda muhim, hech bo‘lmaganda, korpusning aralash turi til tadqiqi uchun maqsadga muvofiq hisoblanadi.

Adabiylik mezoniga ko‘ra korpus adabiy, dialektal, og‘zaki, terminologik hamda aralash kabi turlar<sup>26</sup>ga ajratiladi. Og‘zaki korpusga Sankt-Peterburgda tuzilgan “Одинъ речевоу денъ” korpusini misol keltirish mumkin. Terminologik korpus – bevosita jonli til asosida terminlar lug‘ati yaratishni maqsad qilgan korpus. Janriga ko‘ra adabiy, folklor, dramaturgiyaga xos hamda publitsistik korpus kabi turlarga ajratiladi<sup>27</sup>. Publitsistik korpus sifatida XX asr oxiri rus tili gazeta matnlari korpusi<sup>28</sup>, siyosiy metaforalar korpuslarini sanash mumkin. Foydalanuvchi uchun eng asosiy mezon korpusdan foydalanish imkonining erkin yoki chegaralanganligi sanaladi. Shunga ko‘ra *erkin, tijorat* (pullik) korpusi hamda

<sup>25</sup> Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – Б.26.

<sup>26</sup> Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – Б.29.

<sup>27</sup> Ўша асар.

<sup>28</sup> <http://www.philol.msu.ru/~lex/corpus>

yopiq korpuslar<sup>29</sup> mavjud. Erkin foydalanish mumkin bo'lgan korpus on-line rejimda xohlagan paytda istalgan matnning to'liq shakli bilan ishlash imkonini beradi. Ba'zan erkin korpus ham korpus ma'lumotlarining ma'lum qismidagina foydalanish bilan chegaralangan bo'ladi. *Tijorat korpusi* (pullik korpus)dan foydalanish on-line rejimda ishlash ruxsatnomasini talab qiladi yoki foydalanuvchi korpusning disk variantini sotib olishi ham mumkin. Korpusni olishdan oldin korpus annotatsiyasi bilan tanishish, hatto sinov variantda korpusning ma'lum qismi bilan ishlab ko'rish mumkin. Yopiq korpuslar o'ziga xos tor (xususiy) maqsadda tuziladi va ular ommaviy foydalanishga mo'ljallanmaydi.

Maqsadiga ko'ra korpusning *tadqiqot*, *tavsifiy* yoki *illyustrativ*<sup>30</sup> kabi ko'rinishlari mavjud. Tadqiqiy korpus til harakati (mavjudligi)ning turli aspektini o'rganish maqsadida tuziladi. Korpusning bu tipi keng miqyosdagi lingvistik topshiriqlarni bajarishga mo'ljallangan. Illyustrativ korpus tadqiqot natijasi sifatida paydo bo'ladi, uning vazifasi faqat yangi faktlarni topish emas, balki u yoki bu tilga oid faktni avvalroq boshqa usulda topilgan misollarni dalillashning lingvistik manbayi sifatida xizmat qiladi. Illyustrativ korpusning tipik namunasi ma'lum bir so'zning aniq matndagi ma'nosini izohlashga mo'ljallangan "Путеводитель по дискурсивным словам русского языка" korpusidir.

Korpusga qo'shimcha izoh mavjudligiga ko'ra qo'shimcha izohli (razmetkalan), izohsiz (razmetkalanmagan) turlari farqlanadi. Korpusning bunday tasniflanishi yana boshqa atamalar (masalan, indekslangan, indekslanmagan, annotatsiyalangan, annotatsiyalanmagan, rus tilida таггированные и нетаггированные<sup>31</sup> kabilar) bilan ham nomlanadi. Teglangan korpusda teg xarakteridan kelib chiqqan holda so'z hamda gapga morfologik, sintaktik, semantik, prosodik va h. izoh (korpus tili bilan *teg*) qo'shilgan bo'ladi.

Korpus tarkibining to'ldirib borilishiga ko'ra *turg'un va dinamik* (o'zgaruvchan) korpus turi mavjud<sup>32</sup>. Dastlabki korpuslar tilning turg'un holati

---

<sup>29</sup> Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. –Иркутск: ИГЛУ, 2011. – Б.22.

<sup>30</sup> Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. –Иркутск: ИГЛУ, 2011. – Б.24.

<sup>31</sup> Инглизча tag (изох) сўзидан олинган.

<sup>32</sup> Қаранг: Курс "Корпусная лингвистика" (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

aksi, tilning ma'lum bir davrini qamrab olgan. Nisbatan ahamiyatli zamonaviy korpus sanalgan Britaniya milliy korpusi (BNC) ham turg'un korpusdir. Bunday korpusning tipik namunasi avtorlik korpusi, ya'ni bir yozuvchi asarlari yig'indisi sanaladi. Ba'zi lingvistik amallarni bajarish uchun (masalan, so'z chastotasini aniqlash) turg'un korpus talabga javob berolmay qoladi. Chunki faqat chegaralangan davr qamrab olingan bo'ladi<sup>33</sup>. Shu ehtiyojdan kelib chiqqan holda keyinchalik to'ldirib boriladigan dinamik (monitoring) korpus ham paydo bo'la boshladi. Dinamik korpusning maqsadi doimiy ravishda kompyuter xotirasini matn bilan to'ldirib, korpusni boyitib borish, asosiy belgisi esa yangilanish, to'ldirilib borilishidir. Bu korpus leksikografiya rivoji uchun xizmat qiladi, chunki til doimiy harakatda, so'zlar yangi-yangi ma'nolar kasb etishi, neologizm, arxaizm, istorizmlarning paydo bo'lishini kuzatib borish imkonini beradi. Dinamik korpus katta davrning yozma manbalarini aks ettirib, ular turli diaxronik tadqiqotlarni bajarishga mo'ljallangan. Bunga Cobuild Project korpusi misol bo'la oladi. Bunday korpuslar "ochiq korpus" deb ataladi va ularning kamchiligi shundaki, har doim ham tilni bir xilda (adekvat holatda) aks ettirmaydi, matn tanlashning aniq bir mezoniga amal qilmaydi.

Yaratilish maqsadiga ko'ra universal, maxsus korpus ko'rinishi ham mavjud. Universal korpus, odatda, turli janrdagi matnlarni qamrab oladi. Ko'pincha milliy korpuslar shunday korpus sirasiga kiradi.

Matn hajmiga ko'ra to'liq matnli yoki lavhali (fragment matnli) korpuslar bor. Braun, Lankaster-Oslo-Bergen korpusiga qo'yilgan dastlabki talab qat'iy ravishda har biri 2000 so'zdan (so'z qo'llash holati) iborat matnlardan tuzilishi lozim edi. Tabiiyki, bu mezonga qat'iy amal qiluvchi korpus yo'q. Shunday ekan, bu korpuslar fragment, ya'ni lavha matnli hisoblanadi. To'liq matnli korpusga ma'lum muallif asarlari korpusi, qisqa matnlar korpusini misol keltirish mumkin. Gazeta sarlavhalari korpusi ham shunga misol bo'ladi.

Korpusga qo'shimcha izoh mavjudligiga ko'ra qo'shimcha izohli (razmetkalanagan), izohsiz (razmetkalanmagan) turlari farqlanadi. Korpusning

---

<sup>33</sup> Каранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. –Иркутск: ИГЛУ, 2011.

bunday tasniflanishi yana boshqa atamalar (masalan, indekslangan, indekslanmagan, annotatsiyalangan, annotatsiyalanmagan, rus tilida таггиrowанный и нетаггиrowанный<sup>34</sup> kabilar) bilan ham nomlanadi. Teglangan korpusda teg xarakteridan kelib chiqqan holda soʻz hamda gapga morfologik, sintaktik, semantik, prosodik va h. izoh (korpus tili bilan *teg*) qoʻshilgan boʻladi.

Korpus tarkibining toʻldirib borilishiga koʻra *turgʻun va dinamik* (oʻzgaruvchan) korpus turi mavjud<sup>35</sup>. Dastlabki korpuslar tilning turgʻun holati aksi, tilning maʼlum bir davrini qamrab olgan. Nisbatan ahamiyatli zamonaviy korpus sanalgan Britaniya milliy korpusi (BNC) ham turgʻun korpusdir. Bunday korpusning tipik namunasi avtorlik korpusi, yaʼni bir yozuvchi asarlari yigʻindisi sanaladi. Baʼzi lingvistik amallarni bajarish uchun (masalan, soʻz chastotasini aniqlash) turgʻun korpus talabga javob berolmay qoladi. Chunki faqat chegaralangan davr qamrab olingan boʻladi<sup>36</sup>. Shu ehtiyojdan kelib chiqqan holda keyinchalik toʻldirib boriladigan dinamik (monitoring) korpus ham paydo boʻla boshladi. Dinamik korpusning maqsadi doimiy ravishda kompyuter xotirasini matn bilan toʻldirib, korpusni boyitib borish, asosiy belgisi esa yangilanish, toʻldirilib borilishidir. Bu korpus leksikografiya rivoji uchun xizmat qiladi, chunki til doimiy harakatda, soʻzlar yangi-yangi maʼnolar kasb etishi, neologizm, arxaizm, istorizmlarning paydo boʻlishini kuzatib borish imkonini beradi. Dinamik korpus katta davrning yozma manbalarini aks ettirib, ular turli diaxronik tadqiqotlarni bajarishga moʻljallangan. Bunga Cobuild Project korpusi misol boʻla oladi. Bunday korpuslar “ochiq korpus” deb ataladi va ularning kamchiligi shundaki, har doim ham tilni bir xilda (adekvat holatda) aks ettirolmaydi, matn tanlashning aniq bir mezoniga amal qilmaydi.

Yaratilish maqsadiga koʻra universal, maxsus korpus koʻrinishi ham mavjud. Universal korpus, odatda, turli janrdagi matnlarni qamrab oladi. Koʻpincha milliy korpuslar shunday korpus sirasiga kiradi.

---

<sup>34</sup> Инглизча tag (изох) сўзидан олинган.

<sup>35</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

<sup>36</sup> Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. –Иркутск: ИГЛУ, 2011.

Matn hajmiga ko‘ra to‘liq matnli yoki lavhali (fragment matnli) korpuslar bor. Braun, Lankaster-Oslo-Bergen korpusiga qo‘yilgan dastlabki talab qat’iy ravishda har biri 2000 so‘zdan (so‘z qo‘llash holati) iborat matnlardan tuzilishi lozim edi. Tabiiyki, bu mezunga qat’iy amal qiluvchi korpus yo‘q. Shunday ekan, bu korpuslar fragment, ya’ni lavha matnli hisoblanadi. To‘liq matnli korpusga ma’lum muallif asarlari korpusi, qisqa matnlar korpusini misol keltirish mumkin. Gazeta sarlavhalari korpusi ham shunga misol bo‘ladi.

## 1.1-жадвал

### Корпус турлари таснифи

Тасниф белгиси	Корпус турлари
Нутқ турига кўра	Оғзаки, ёзма, аралаш
Матн тилига кўра	Русча, инглизча ва ҳ.
Параллеллигига кўра	Бир тилли, икки тилли, кўп тилли
Матннинг ихтисослашувига кўра	Бадий, диалектал, суҳбат, терминологик, аралаш
Жанрига кўра	Адабий, фольклор, драматик, публицистик
Кириш усулига кўра	Эркин, пуллик, ёпик
Мақсадига кўра	Тадқиқий, тавсифий
Изоҳланишига кўра	Қўшимча изоҳли, изоҳсиз
Ўзгарувчанлигига кўра	Ўзгарувчан, турғун
Структурасига кўра	Марказий, архив, маҳаллий
Изоҳ хусусиятига кўра	Морфологик, синтактик, семантик
Матн ҳажмига кўра	Тўлиқ матнли, лавҳа матнли
Хронологик жиҳатига кўра	Синхрон, диахрон
Умумийлигига кўра	Умумий, якка муаллифли

### Mustahkamlash uchun savol va topshiriqlar:

1. On-line va off-line korpus qanday mustarak va farqli jihatlarfga ega? Web va prilojeniye shaklidagi korpusning o‘ziga xos xususiyati nimada?
2. Korpusning nutqqa ko‘ra turi qanday turlari mavjud?



3. Og‘zaki, yozma, aralash korpuslar haqida ma’lumot bering.
4. Matn tiliga ko‘ra korpusning turi qanday turlari mavjud?
5. Parallelligiga korpusning ko‘ra turi qanday turlari mavjud?
6. Matnning ixtisoslashuvi va janriga ko‘ra korpus turlarini sanang.
7. Kirish usuliga ko‘ra korpusning turi qanday turlari mavjud?
8. Maqsadiga ko‘ra korpusning turi qanday turlari mavjud?
9. Izoh xususiyatiga ko‘ra korpusning turi qanday turlari mavjud?
10. O‘zgaruvchan va turg‘unligiga ko‘ra korpusning turi qanday turlari mavjud?
11. Strukturasiga ko‘ra korpusning turi qanday turlari mavjud?
12. Matn hajmiga ko‘ra korpusning turi qanday turlari mavjud?
13. Xronologik jihatiga ko‘ra korpusning turi qanday turlari mavjud?

#### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Грудева В.А. Корпусная лингвистика. Учебное пособие. 2-е изд., стереотип. – М.: Флинта, 2012. – 165 с.
3. Недошивина Э.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб., – 2006. – 26 с.
4. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
5. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
6. [http:// corpus.leeds.ac.uk/](http://corpus.leeds.ac.uk/)
7. <http://corpus.leeds.ac.uk/ruscorpora.html>

**II BOB. KORPUSDAN FOYDALANISH**  
**JAHON KORPUS LINGVISTIKASINING BUGUNGI**  
**HOLATI: MILLIY KORPUSLAR.**

**Reja:**

1. Korpus lingvistikasining shakllanishi, taraqqiyoti hamda bugungi holati. Birinchi va ikkinchi avlod korpuslari.
2. Zamonaviy korpuslarning imkoniyati. Rus, ingliz korpuslari tarixi, rivojlanishi va bugungi holati.
3. Milliy korpusning xususiyatlari.
4. Milliy korpusni yaratishga qo'yilgan talablar.
5. Korpus tuzishning umumiy tamoyillari.
6. Milliy korpusda reprezentativlik masalasi.

***Tayanch so'z va iboralar:** korpus lingvistikasi, zamonaviy korpus, birinchi avlod korpusi, ikkinchi avlod korpusi, milliy korpus, talab, umumiy tamoyil, xususiy tamoyil, reprezentativlik.*

**1. Milliy korpus va uning amaliy ahamiyati. Talab va ehtiyoj.** Korpus – ma'lum maqsadda yig'ilgan matnlar majmuini tashkil etuvchi til birliklari yig'indisi, tabiiy tildagi elektron shaklda saqlanadigan yozma yoki og'zaki, kompyuterlashtirilgan qidiruv tizimiga dasturiy ta'minot asosida joylashtirilgan on-line yoki off-line tizimda ishlaydigan matnlar jamlanmasi. Lingvistik tadqiqot va amaliy topshiriqlar yechimi uchun til korpuslari zamonaviy tilshunoslikning inkor etib bo'lmaz ish quroli. U oddiy elektron kutubxonadan farqlanadi. Elektron kutubxona xalqning ijtimoiy-siyosiy, ma'naviy, iqtisodiy hayotini aks ettiruvchi badiiy va publitsistik asarlarni nisbatan to'liq qamrab olishni maqsad qilgan bo'ladi. Elektron kutubxona matnlari tilshunoslik nuqtayi nazaridan leksik, morfologik, grammatik, semantik belgilar asosida ishlov berilmaganligi sababli lingvistik tadqiqot uchun noqulaylik tug'diradi. Chunki elektron kutubxona ilmiy tadqiqot materiali bazasini tayyorlash maqsadida tuzilmaydi, balki milliy-ma'naviy merosni to'plashni maqsad qilgan bo'ladi. Til korpusi esa elektron kutubxonadan farqli o'laroq, tilni o'rganish uchun zaruriy, foydali va qiziqarli matnlarni

to'plashni nazarda tutadi. Korpusni elektron kutubxonadan ajratib turuvchi birinchi omil undagi matnning xususiyati va qo'shimcha ma'lumot bilan boyitilganligi hisoblanadi hamda bu belgi korpusning alohida qismi – korpus birliklariga yozilgan izohni tashkil etadi.

Foydalanuvchiga biror so'z kerak bo'lsa, buni odatiy matn muharriri ham topib beradi. Lekin so'zning grammatik ma'nosi va tuzilishini "tushunadigan" matn bilan ishlash juda ko'p afzallik va qulaylikka ega. So'zshaklni qidirish kerak bo'lsa, bunday dasturiy ta'minot, ya'ni korpus tadqiqotchi yoki foydalanuvchiga juda katta yordam berishi mumkin. Agar tadqiqotchi o'z ishi uchun misolni topish, ularni kartotekaga ko'chirish (kompyuter texnologiyalari rivojlanishidan oldingi davrda)ga oylab, ba'zan yillab vaqt sarflagan bo'lsa, bugun dunyo til korpuslari yordamida sanoqli daqiqada yuzlab misolni topishga, ular ustida ishlash imkoniga ega bo'ldi. Maxsus qidiruv korpusdan ma'lumot olishga mo'ljallangan bir qancha dasturdan iborat, statistik axborot va qidiruv natijasini foydalanuvchiga qulay shaklda taqdim etadi. Tilda qanday jarayon kechayotganligini aniq tasavvur qilish uchun korpus qamrovini yanada kengaytirish, nafaqat yozma nutq, balki og'zaki nutq materialidan foydalanish ham maqsadga muvofiq. Bunday korpus yordamida til taraqqiyoti natijasida unda sodir bo'lgan va kutilayotgan o'zgarish haqida aniq xulosa chiqarish mumkin.

Korpus bo'yicha qidiruv foydalanuvchiga quyidagi natijani beradi:

- 1) belgilangan so'zning turli kontekstdagi barcha shakldagi ko'rinishi;
- 2) tilning lug'at tarkibidagi o'rni va variantlari;
- 3) belgilangan so'z bilan birikish imkoniyatiga ega so'zlar ro'yxati;
- 4) u yoki bu yozuvchining ayni so'zdan foydalanish chastotasi yoki statistikasi;
- 5) so'zning o'z va ko'chma ma'nosi;
- 6) leksik qo'llanishining yashirin modeli (imkoniyati);
- 7) til taraqqiyotining turli davrida qo'llanish holati.

**2. Milliy korpus yaratish maqsadi.** Korpusning keng ko'lamlı bo'lishi ma'lumotning o'ziga xosligini kafolatlaydi, til hodisalarining barcha qirrasini to'liq namoyish etishni ta'minlaydi. Turli tipdagi ma'lumot til korpusida o'zining

tabiiy kontekstual shaklida joylashadi, bu esa uni har tomonlama va obyektiv o'rganish uchun asos bo'ladi. Bir marta tuzilgan va tayyorlangan axborotlar massivi bir necha tadqiqotchi tomonidan ko'p marta, turli maqsadda ishlatilishi mumkin. Korpus tilni tadqiq etish (tilda so'zning o'zgarishi, istorizm, neologizm, ma'no kengayishi va torayishi; yangi frazeologizmlarning paydo bo'lishini kuzatish), til o'rganish, lug'at tuzishda eng zamonaviy, keng imkoniyatli dasturlashtirilgan tizim. Korpus bugungi kun nazariy va amaliy filologiyasini tasavvur etish qiyin. Tilshunoslikka oid tadqiqotda fakt bilan ish ko'riladigan hollarda material yig'ilishi, sistemaga solinishi lozim. Bunday katta hajmli ishni bajarishda korpus tilshunos uchun vaqt va mehnatni tejaydigan ish quroli. U texnik jarayonni tezlashtiruvchi vosita bo'lib qolmay, ma'lum til zamonaviy shaklining axborot tizimi bo'lib, kutilmagan savolga ham javob bera oladigan, tilshunoslik oldiga avval qo'yilmagan dolzarb muammoni qo'ya oladigan tizim.

**3. Korpus tarkibidagi materialning turi va xususiyati.** Mavjud korpuslar tarkibidagi matnlarning nisbatiga qaraydigan bo'lsak, badiiy adabiyot hissasi 40% ni tashkil etishiga guvoh bo'lamiz. Buning tarkibiga memuar asarlar ham kirib ketadiki, bu janr til xususiyati badiiy va publitsistik uslub oralig'ida bo'lib, jonli tilni o'rganish uchun ancha qulay. Evtora tillari korpuslarida badiiy adabiyot materiali 20% ni tashkil etadi. Masalan, zamonaviy yozuvchilar til xususiyatini o'rganishga bag'ishlangan 20dan ortiq taqiqot mavjud bo'lsa-da, ular hali to'laligicha bu muammoni o'rganib bo'ldi, deyish qiyin. Chunki alohida yozuvchi asarining til xususiyatidagi o'zgarishga hali tildagi yangi hodisa deb qarab bo'lmaydi. Ana shunday jarayonlarni kuzatish va tadqiq etishning eng qulay vositasi til korpusi. Shu sababli matnning mazmuni muhim ahamiyat kasb etadi. Korpus tarkibiga kiradigan matnlar alohida bir muallif yoki bir necha yozuvchi asaridan olingan matn, ma'lum davr (bir necha o'nyillik yoki yuzyillik)ni qamrab olgan matn, belgilangan mavzudagi zamonaviy matn, til va jamiyatning bugungi holatini aks ettiruvchi zamonaviy matnlardan iborat bo'lishi mumkin.

**4. Milliy korpusning tuzilishi va tarkibi.** Mavjud milliy korpuslarni kuzatish asosida korpusning tuzilishi va tarkibini o'rganar ekanmiz, korpus interfeysi va

qidiruv tizimi uning eng asosiy tarkibiy qismi, degan xulosaga kelamiz. Rus tili milliy korpusi (Национальный корпус русского языка) joylashgan [www.ruscorpora.ru](http://www.ruscorpora.ru) saytining birinchi sahifasida korpus va uning tuzuvchilari haqida asosiy ma'lumot, o'ng tomondagi menyuda istalgan sahifaga o'tish imkoniyati mavjud. Menyu to'rt qismdan iborat bo'lib, bosh sahifa, saytning qidiruv resursi, matn haqida unga biriktirilgan qo'shimcha ma'lumot ilovasi, korpus birliklariga izoh yozish prinsiplari, oxirgi blok korpus tuzuvchilari jamoasi, foydalanilgan dastur, matnlarning mualliflik huquqi haqida to'liq ma'lumotlar bazasidan iborat.

O'qituvchi uchun korpus tengsiz xazina, undan har bir o'quv mashg'uloti uchun serqirra va mazmunli o'quv materiallari tayyorlash mumkin. Yuqorida aytib o'tganimizdek, korpusda xilma-xil matnlar jamlangan, undan o'qituvchi istalgan shaklda foydalana oladi. Masalan, o'qituvchi topshiriqni tayyorlash uchun faqat badiiy adabiyot parchasidan (yoki nobadiiy matndan: sport, avtobiografiya yoki shunchaki ma'lum ijodkor ijodi bilan chegaralanmoqchi) foydalanmoqchi. Unda "Mening korpusim" ilovasi orqali o'ziga kerakli matnni ajratib olish imkoni mavjud, ya'ni misollar sirasini chegaralaydi. Korpus turli janr va uslubdagi matnlarni qamrab olganligi sababli har qanday talabni qondirish imkoniga ega. Misol uchun, jurnalistika fakulteti talabalari uchun yozilgan darslikda publitsistik matnlar ancha eskirgan. Har kuni boyitilib boradigan korpusdan yangi misollarni olish, uni talabaga taqdim etish yoki talabaga ham shunday topshiriq berish, albatta, ta'limni bugungi hayotga yaqinlashtiradi. Ko'pincha, qayta-qayta nashr etilgan darslikda misol eskiligicha qolib ketadi va axborot bilan shiddatli yangilanib borayotgan bugungi kun uchun mavzu eskiradi, talaba bilimni davr bilan hamohang o'zlashtirmaydi. Ular ta'lim jarayonida so'zning nozik qirrasini o'rganish, uni nutqida qo'llash ko'nikmasini shakllantirishda nimaga tayanadi? XX asr boshi yoki o'rtasida yozilgan asar bilan ishlash jarayoni unga hozirgi adabiy til uslubini shakllantirish uchun etarli emas. Bugun filolog yoki jurnalist bo'lib etishayotgan mutaxassis tilimizning barcha nozik qirralarini his etishi uchun faqat eski – 20-30 yil oldingi matn namunasi yetarli emas, bugun yozilgan barcha soha,

uslub va janrdagi asarlar bilan ishlashi lozim, shundagina ta'lim ijtimoiy buyurtmani bajargan hisoblanadi. Qaysi fanlar, uning aynan qaysi bo'limida korpusdan qanday foydalanish mumkinligini aniq misollarda ko'rib o'tamiz.

**5. Kutilayotgan natijalar.** Yurtimizda o'zbek tili korpuslari yaratish ishlarini qoniqarli deb bo'lmaydi. Bu sohada olib borilayotgan ilmiy tadqiqot ishlari o'zbek korpus lingvistikasining nazariy asoslarini ishlab chiqish, namuna sifatida bir-ikki korpus lavhasi yaratish jarayoni bosqichida turibdi. O'zbekistonda amaliy tilshunoslikning bir yo'nalishi sanalgan o'zbek korpus lingvistikani rivojlantirish uni ilmiy-nazariy jihatdan tadqiq etish, kompyuter va korpus lingvistikasi markazi va laboratoriyasi tashkil etish hamda ushbu laboratoriyaga mutaxassislarni birlashtirish orqali amalga oshishi mumkin. Korpusning ijtimoiy ahamiyati keng qamrovliligi, undan lingvistik va etno-psixolingvistik tadqiqotlarda, ona tili, adabiyot, xorijiy til ta'limida, matnga avtomatik ishlov berish va tarjima dasturlarini tuzishda foydalanish mumkinligi, korpusning tilshunos, tarjimon, o'qituvchi, dasturchi, jurnalist, muharrir, umuman, kundalik faoliyati jarayonida "ish quroli so'z bo'lgan" har qanday kishi uchun zamonaviy axborot vositasining bir ko'rinishi ekanligi korpus yaratishning dolzarbligini belgilaydi.

## 1.1-jadval

### Korpuslarning qiyosiy tahlili

№		Rus tilining Upsalla korpusi <sup>37</sup>	Rus tili gazetalar kompyuter korpusi
1	Tarkibi	1996-2002 yillar publitsistikasi; XIX-XX asr badiiy adabiyoti	1997 yil gazeta matnlari
2	Hajmi	25 million so'z atrofida	100 ming so'z atrofida
3	So'z yasalishi	+	-
4	sintaktik	element	element
5	semantik	-	element
<b>Qidiruv shakli</b>			
6	So'z shakli orqali	+	+

<sup>37</sup> Ушбу корпус Интернетда "Рус матнларининг Тюбинген корпуси" деган ном билан аталади.

7	Lemma orqali	–	+
8	Bir necha soʻzni baravar	+	–
9	Grammatik belgi asosida	+	+
10	Semantik belgi asosida	–	+
11	Tinish belgi asosida	+	+

### **Mustahkamlash uchun savol va topshiriqlar**

1. Korpus lingvistikasining shakllanishi, taraqqiyoti hamda bugungi holati haqida ma'lumot bering.
2. Birinchi va ikkinchi avlod korpuslarining o'xshash va farqli jihatlarini sanab bering.
3. Zamonaviy korpuslarning imkoniyati nimalarda ko'rinadi?
4. Korpuslarni yaratish davrlari qanday tasnis qilinadi?
5. Korpuslar qaysi xususiyatiga ko'ra davrlarga ajratilgan?

### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Грудева В.А. Корпусная лингвистика. Учебное пособие. 2-е изд., стереотип. – М.: Флинта, 2012. – 165 с.
3. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
4. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
5. Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол. фан. бўйича фалсафа доктори (PhD) диссер. – Қарши, 2019. – 140 б.
6. <http://www.philol.msu.ru/~lex/corpus>
7. <http://ruscorpora.ru>.

8. <http://ruscorpora.ru/search-murco.html>

9. [http:// corpus.leeds.ac.uk/](http://corpus.leeds.ac.uk/)

10. <http://corpus.leeds.ac.uk/ruscorpora.html>



## RUS VA INGLIZ KORPUS LINGVISTIKASI: TARIX VA BUGUN

### Reja:

1. Korpus lingvistikasining shakllanishi, taraqqiyoti hamda bugungi holati.
2. Birinchi va ikkinchi avlod korpuslari.
3. Zamonaviy korpuslarning imkoniyati.
4. Rus korpus lingvistikasining bugungi holati.
5. Ingliz korpus lingvistikasining bugungi holati.

*Tayanch soʻz va iboralar:* korpus lingvistikasi, zamonaviy korpus, birinchi avlod korpusi, ikkinchi avlod korpusi.

*Birinchi avlod korpuslari.* Braun korpusining toʻliq nomi Zamonaviy amerikancha ingliz tilining Braun universiteti standart korpusi (Brown University Standard Corpus of Present-Day American English), 1961-1964 yillarda tuzilgan. Korpus ingliz tilining amerika varianti yozma matnlaridan tashkil topgan, 1 million soʻz qoʻllash holati mavjud (bunday miqdor dastlabki korpuslarning barchasi uchun hajm mezonini boʻlib qolgan edi). Oʻsha davrda Noam Xomskiy konsepsiyasi hukmronlik qilganligi sababli Braun korpusi tuzuvchilari Nelson Frensis, Genri Kusera uchun korpus tuzish jarayoni koʻp qiyinchilik tugʻdirgan. Korpus 2000 soʻzli 500 matndan iborat, qulay, puxta tuzilishga egaligi hamda matn tanlash tamoyilining har tomonlama toʻgʻriligi sababli keyinchalik korpus tadqiqotlari mezonlari yaratilishi uchun asos boʻldi. Korpuslardan tadqiqot uchun erkin foydalanish anʼanasini ham shu korpus boshlab berdi. Mazkur korpus asosida 1969 yildayoq “Amerika meros lugʻati”ga (American Heritage Dictionary) asos solindi<sup>38</sup>.

Lankaster-Oslo/Bergen (LOB) korpusi (Lancaster-Oslo/Bergen (LOB) Corpus) 1970-78 yillarda Lankaster, Oslo universiteti hamda Bergen ilmiy markazi loyihasi<sup>39</sup> 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)

---

<sup>38</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

<sup>39</sup> <http://khnt.hit.uib.no/icame/manuals/lobman/>

tadqiq qilishda 1 million soʻzning etarli emasligi haqida xulosa qilishdi. Bu xulosa juda oʻrinli edi. Shunday mukammal korpus mavjudligi sababli Braun, LOB korpuslariga asoslangan juda koʻp sifatli, qiziqarli tadqiqotlar amalga oshirilgan.

London-Lund korpusi<sup>40</sup> (London-Lund Corpus (LLC) ingliz tilining ogʻzaki nutq korpusi, 1975 yilda tuzilgan. Orfografik transkripsiya, fonetik, prosodik (urgʻu va urgʻusiz, uzun hamda qisqa boʻgʻinlarni talaffuz qilishga oid) tegi mavjud, 500.000 soʻz qoʻllash holatini qamrab olgan. Bu ulkan ish London universiteti kollejida dastlab qogʻoz variantda tayyorlanib, keyinchalik shved shaharlaridan biri Lund tilshunoslari tomonidan elektron koʻrinishga keltirildi. Sanab oʻtilgan korpuslardan tashqari turli maqsadni koʻzlovchi quyidagi korpuslar ham mavjud<sup>41</sup>:

1. Leksikografik tadqiqotlar uchun Amerika meros aloqa, vositachi korpusi (American Heritage Intermediate).
2. Ingliz tili soʻzlashuv tilini oʻrganish uchun Lankaster ingliz tili soʻzlashuv korpusi (Lancaster/IBM Spoken English Corpus, Corpus of Spoken American English, etc).
3. Diaxronik korpus sanalgan Ingliz matnlarining Xelʼsinki korpusi (Helsinki Corpus of English Texts: Diachronic Part). 1,5 million soʻz qoʻllanish holatini qamrab olgan.
4. Lingvodidaktik tadqiqotlar uchun Ingliz tili oʻrganuvchilarining xalqaro korpusi va h.

*Ikkinchi avlod korpuslari.* Kompyuter paydo boʻlishi yozma (noelektron) korpuslarning yoʻqolishiga olib kelmadi. Hozir ham tilshunoslikning bir qancha sohasi uchun karton kartoteka, sopol boʻlaklariga yozilgan material bilan ishlash tadqiqot ishining juda muhim qismi hisoblanadi. Ayrim til, matnlar tadqiqi uchun misol, faktni korpus tipi hamda miqdorida yoyma shaklda oʻrganish muhimligi kuzatiladi. Korpus lingvistikasining lokomativi – ingliz tili. Bugungi kunda hech kim ingliz tilidagi oddiy matnlarni yigʻish haqida oʻylab oʻtirmaydi, chunki bu

---

<sup>40</sup> <http://www.ucl.ac.uk/english-usage>

<sup>41</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

tilning barcha varianti uchun juda katta hajmli, chuqur grammatik tahlilga ega (annotatsiyalangan) korpuslar mavjud. “Annotatsiyalangan korpus” davrigacha korpuslar aniq maqsadli to‘plangan, sistemalashtirilgan matnlar yig‘indisidangina iborat edi. Bunga ilk ingliz tili (Brown corpus, 1960 y.), rus tili (Уппсальский корпус русских текстов, 1980-у.) korpuslari misol bo‘la oladi. Shu tariqa ilk elektron korpuslar o‘z ajdodidan faqat ma’lumotni saqlash shakli hamda sezilarli darajada hajmning kengayishi bilan farq qildi. Zamonaviy korpus u yoki bu til resursini qay darajada ifodalashiga, sifati, miqdoriga qarab matnlarning oddiy elektron kolleksiyasi yoki chuqur annotatsiyalangan korpus kabi turga bo‘linadi<sup>42</sup>. Zamonaviy tilshunoslik ko‘p hollarda matnni elektron shaklda ishlatish ehtiyojini sezadi, bu esa avtomatik ravishda “korpusning birinchi avlodi”ni tashkil etadi. Shuning uchun eski korpuslar “o‘lmaydi”, iste’moldan chiqib ketmaydi, balki to‘ldirilib, chuqurlashib, annotatsiyaga (turli lingvistik tahlil) ega bo‘lib boraveradi.<sup>43</sup> Kompyuter korpuslari yaratishning asosiy davrlari quyidagicha tasniflanadi:<sup>44</sup>

- I. 1960-yillar: Braun korpusi (AQSH) – 1 million so‘z.
- II. 1970-yillar: LOB korpusi (Buyuk Britaniya, Norvegiya) – 1 million so‘z.
- III. 1980-yillar:
  1. Rus tilining mashina fondi (Машинный Фонд русского языка).
  2. Rus tilining Upsall korpusi (Уппсальский корпус русского языка. Швеция) – 1 million so‘z.
- IV. 1990-yillar:
  1. British National Corpus (Britaniya milliy korpusi) – 100 million so‘z.
  2. Milliy korpuslar (venger, italyan, xorvat, chex, yapon) 100 million so‘z.
  3. Ingliz tili banki (The Bank of English, Birmingham (Collins Cobuild) – 600 million so‘z.
- V. 2000-yillar:

---

<sup>42</sup> Қаранг: История корпусной лингвистики. (Электрон ресурс) - <https://www.myfilology.ru/177/istoriya-korpusnoj-lingvistiki/>

<sup>43</sup> Қаранг: Ўша манба.

<sup>44</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

1. American National Corpus (Amerika milliy korpusi) – 100 million so‘z.
2. Zamonaviy Amerikancha ingliz tili korpusi (Corpus of Contemporary American English) – 400 million so‘z.
3. Национальный корпус русского языка (Rus tili milliy korpusi) – 140 million so‘z.
4. Gigaword corpora: ingliz, arab, xitoy tillari – 2 milliard so‘z.
5. Oxford English corpus (Oksford ingliz korpusi) – 2 milliard so‘z.

**Rus korpus lingvistikasi tarixi.** Jahon korpus lingvistikasi taraqqiyoti, o‘ziga xos xususiyatlarini o‘rganishda rus korpus lingvistikasi tajribasini kuzatish foydadan xoli emas. Chunki rus tili korpuslari miqdor, ko‘lam jihatdan jahon korpuslari orasida salmoqli o‘rin egallaydi. “Jahonda korpus lingvistikasi alohida soha sifatida 1990-yillarda shakllangan. Rossiya korpus sohasidagi ishlarni ancha kechikib boshlagan bo‘lsa-da, lingvistikaning mazkur tarmog‘i juda tezlik bilan rivojlanib boryapti”, – deb yozadi rus korpus lingvistikasining peshqadam mutaxassisi V.P.Zaxarov<sup>45</sup>. 1980-yillarda korpus yaratish borasida boshlangan harakatlar 1990-yillarga kelib birmuncha to‘xtab qoldi. Lekin 2000-yillar boshida soha fidoyi tilshunoslarining xatti-harakatlari bilan katta natijalarni qo‘lga kiritdi<sup>46</sup>. 2000-yillarda rus korpus lingvistikasi mutaxassislari bu sohaning jahon miqyosidagi taraqqiyotidan orqada kolayotganligini afsus bilan tan olishgan bo‘lsa-da, Rossiyada tilshunoslikning ushbu tarmog‘i juda qisqa muddatda katta yutuqqa erishdi. Rus korpus lingvistikasi tadqiqotlarining ilk natijasi Rus tili milliy korpusi<sup>47</sup> sanalib, keyinchalik turi, maqsadiga ko‘ra farq qiluvchi bir qancha korpuslar yaratildi. Korpus lingvistikasi shakllanishi, taraqqiyoti, bugungi holati, muammolarini o‘rganish kelajakda qilinishi lozim bo‘lgan vazifalarni belgilab olishning muhim sharti sanaladi. Rus korpuslari shakllanishi, tarixiy taraqqiyot yo‘li haqida Sankt-Peterburg Davlat universiteti matematik lingvistika kafedrasida

---

<sup>45</sup> Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.

<sup>46</sup> Каранг: История корпусной лингвистики. (Электрон ресурс) - <https://www.myfilology.ru/177/istoriya-korpusnoj-lingvistiki/>

<sup>47</sup> Национального корпус русского языка (2004г., <http://ruscorpora.ru>).

tomonidan tashkil etilgan “Korpusnaya lingvistika” konferensiyasi materiallari bilan tanishish bu ishning debochasi bo‘lib xizmat qiladi<sup>48</sup>.

*Ilk rus tili korpuslari.* Rus tilidagi ilk korpusga 1980-yillarda SHvetsiyaning Upsall universitetida asos solindi. Rus tilidagi ilk korpuslargacha 1960-1970-yillarda L.N.Zasorina tomonidan “Rus tilining chastotali lug‘ati” (“Частотный словарь русского языка”) yaratilgan edi. Lug‘at 1 million so‘z qo‘llanish holatiga ega, teng proporsiyaga asoslangan 4 ta janr leksikasini (ijtimoiy-siyosiy matn, badiiy adabiyot, turli sohaning ilmiy hamda ilmiy-ommabop matnlari, dramaturgiya) o‘zida jamlagan sodda matnli fayllarga asoslanib tuzilgan. Bu fayllarni (bugungi kun atamasi bilan aytsak, korpus) tuzish jarayonida zamonaviy korpus lingvistikasining barcha muammolari, korpusga qo‘yilgan talablarga (keng ko‘lam, grammatik tahlil, me‘yorlashtirish hamda lemmatizatsiya) javob beradi. V.P.Zaxarov aslida bizgacha etib kelgan ilk rus tili korpusi shu lug‘at ekanligini ta’kidlaydi<sup>49</sup>.

1985-yilda SSSRda akademik A.P.Yershov tashabbusi bilan “Rus tilining mashina fondi”ni tuzish bo‘yicha ishlar boshlandi. Bu juda keng ko‘lamli loyiha, uni tuzishda 40 dan ortiq tashkilot hamkorlik qildi. Rus tili instituti, Moskva, Sankt-Peterburg, Xarkov, Saratov universitetlari shular jumlasidan. Fondning vazifasiga rus tilini o‘rganish, amaliy ishlanmalarni amalga oshirish uchun kerakli matniy, leksikografik, grammatik ma’lumotlar bazasini yaratish, ularni elektron shaklda saqlash, tashish kabilar kirardi. Ayni vaqtda, lingvistik tadqiqotlar o‘tkazish uchun bir necha dastur ham ishlab chiqildi. 1985-1992 yillarda rus tilining mashina fondi arxitekturasi, konsepsiyasi ishlab chiqildi; XIX–XX rus adabiyotining barcha manbalari, rus tilining asosiy lug‘atlari, qisqa akademik grammatikasi kabilar kompyuterga kiritilib, elektron shaklga keltirildi; she’riyat, badiiy adabiyot, ijtimoiy-siyosiy, texnikaga oid matnlar korpusi yaratildi. 1991-yilga kelib mashina fondni tashkil etish ishlarining iqtisodiy ta’minoti susayib ketdi, keyinroq ish batamom to‘xtab qoldi.

---

<sup>48</sup> Қаранг: <http://corpora.phil.spbu.ru>

<sup>49</sup> Қаранг: Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.

Rus tilining Upsall korpusi (Упсальский корпус русского языка<sup>50</sup>) 1 million soʻz qoʻllanish holatini qamrab olgan turli janrdagi 600 ta matndan iborat. Tuzuvchilarning fikricha, korpus zamonaviy (hozirgi) rus tilining holatini aks ettirib turishi lozim. Ushbu korpusni tuzishning asosiy maqsadi, avvalo, adabiy tilni koʻrsatish sifatida belgilangan, shuning uchun korpusga ogʻzaki nutq namunalari kiritilmagan. Korpusga 1985-1989 yilgi maxsus matnlar (matndan parcha emas, balki toʻliq holda), 1960-1988 yilgi badiiy adabiyot namunalari kiritilgan. Korpus annotatsiyasida koʻrsatilishicha, maxsus matnlarni tanlashda matn mavzui, badiiy adabiyot namunasini tanlashda muallifning mashhurligiga alohida eʼtibor qaratilgan. Matnlar lotin alifbosida kiritilgan boʻlib, bugungi kunda ushbu korpus Tyubingen universitetida 1990-2000 yillarda maxsus SFB 441 ilmiy-tadqiqot sektori doirasida tuzilgan “Rus matnlarining Tyubingen korpusi” (“Тюбингенские корпусы русских текстов<sup>51</sup>”) tarkibida joylashgan. Korpus morfologik annotatsiyalangan, qidiruv tizimi soʻz shakli hamda morfologik belgilar asosida amalga oshirilishi, qidiruv natijasi esa KOI8 yoki Windows-1251 kodi asosida kirillcha hamda lotincha harflar bilan uning transliteratsiyasi asosida boʻlishi mumkin. Qidiruv katta korpuslarni boshqarishga moʻljallangan Shtutgart universiteti Tilni avtomatik qayta ishlash instituti ishlab chiqqan SQR dasturi bilan amalga oshiriladi<sup>52</sup>. Mazkur korpus Internetdagi ochiq, foydalanish erkin boʻlgan, ilk annotatsiyalangan korpus boʻlib, u yakunlangan ish hisoblanadi<sup>53</sup>.

**Ingliz tili elektron lingvistik korpuslari tarixi.** Korpus lingvistikasi lokomativi ingliz tili korpuslari boʻlganligi sababli korpusga xos xususiyat, korpus yaratilishi, rivojlanishini ingliz tili korpuslarining yaratilishi, taraqqiy etishini oʻrganishdan boshladik. *Birinchi avlod korpuslari.* Braun korpusining toʻliq nomi Zamonaviy amerikanacha ingliz tilining Braun universiteti standart korpusi (Brown University Standard Corpus of Present-Day American English), 1961-1964 yillarda tuzilgan. Korpus ingliz tilining amerika varianti yozma matnlaridan tashkil

<sup>50</sup> [www.slaviska.uu.se/korpus.htm](http://www.slaviska.uu.se/korpus.htm)

<sup>51</sup> [www.sfb441.uni-tuebingen.de/b1/rus/korpora.html](http://www.sfb441.uni-tuebingen.de/b1/rus/korpora.html)

<sup>52</sup> Қаранг: Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 — 30. 05.14 < Crimea, Simferopol — Alushta.

<sup>53</sup> Қаранг: История корпусной лингвистики. (Электрон ресурс) - <https://www.myfilology.ru/177/istoriya-korpusnoj-lingvistiki/>

topgan, 1 million soʻz qoʻllash holati mavjud (bunday miqdor dastlabki korpuslarning barchasi uchun hajm mezoni boʻlib qolgan edi). Oʻsha davrda Noam Xomskiy konsepsiyasi hukmronlik qilganligi sababli Braun korpusi tuzuvchilari Nelson Frensis, Genri Kusera uchun korpus tuzish jarayoni koʻp qiyinchilik tugʻdirgan. Korpus 2000 soʻzli 500 matndan iborat, qulay, puxta tuzilishga egaligi hamda matn tanlash tamoyilining har tomonlama toʻgʻriligi sababli keyinchalik korpus tadqiqotlari mezonlari yaratilishi uchun asos boʻldi. Korpuslardan tadqiqot uchun erkin foydalanish anʼanasini ham shu korpus boshlab berdi. Mazkur korpus asosida 1969-yildayoq “Amerika meros lugʻati”ga (American Heritage Dictionary) asos solindi<sup>54</sup>.

Lankaster-Oslo/Bergen (LOB) korpusi (Lancaster-Oslo/Bergen (LOB) Corpus) 1970-78-yillarda Lankaster, Oslo universiteti hamda Bergen ilmiy markazi loyihasi<sup>55</sup> 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 etarli emasligi haqida xulosa qilishdi. Bu xulosa juda oʻrinli edi. Shunday mukammal korpus mavjudligi sababli Braun, LOB korpuslariga asoslangan juda koʻp sifatli, qiziqarli tadqiqotlar amalga oshirilgan.

London-Lund korpusi<sup>56</sup> (London-Lund Corpus (LLC) ingliz tilining ogʻzaki nutq korpusi, 1975-yilda tuzilgan. Orfografik transkripsiya, fonetik, prosodik (urgʻu va urgʻusiz, uzun hamda qisqa boʻgʻinlarni talaffuz qilishga oid) tegi mavjud, 500.000 soʻz qoʻllash holatini qamrab olgan. Bu ulkan ish London universiteti kollejida dastlab qogʻoz variantda tayyorlanib, keyinchalik shved shaharlaridan biri Lund tilshunoslari tomonidan elektron koʻrinishga keltirildi. Sanab oʻtilgan korpuslardan tashqari turli maqsadni koʻzlovchi quyidagi korpuslar ham mavjud<sup>57</sup>:

---

<sup>54</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

<sup>55</sup> <http://khnt.hit.uib.no/icame/manuals/lobman/>

<sup>56</sup> <http://www.ucl.ac.uk/english-usage>

<sup>57</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

5. Leksikografik tadqiqotlar uchun Amerika meros aloqa, vositachi korpusi (American Heritage Intermediate).
6. Ingliz tili so‘zlashuv tilini o‘rganish uchun Lankaster ingliz tili so‘zlashuv korpusi (Lancaster/IBM Spoken English Corpus, Corpus of Spoken American English, etc).
7. Diaxronik korpus sanalgan Ingliz matnlarining Xel’sinki korpusi (Helsinki Corpus of English Texts: Diachronic Part). 1,5 million so‘z qo‘llanish holatini qamrab olgan.
8. Lingvodidaktik tadqiqotlar uchun Ingliz tili o‘rganuvchilarining xalqaro korpusi va h.

*Korpuslarning ikkinchi avlodi.* O‘tgan asrning 90-yillariga kelib saqlash, foydalanish texnologiyasi 100 milliondan ortiq so‘zdan iborat korpuslarni tuzish imkoniyatini berdi. Ingliz tili banki (The Cobuild Project/The Bank of English)ni tuzish 1980-yilda boshlangan. Collins nashriyoti yangi lug‘at yaratish maqsadida korpusni tuzishga kirishdi. 1990 yilda Birmingem universiteti ingliz tili fakulteti, Collins nashriyotining bu boradagi sa’y-harakatlarini “The Bank of English” nomi ostida birlashtirish harakati boshlandi, natijada monitoring korpus sifatida shakllandi. Bu haqda loyiha muallifi Jon Sinkler: “Monitoring korpus doimiy o‘zgarishda, aniq hajmga ega bo‘lmagan, katta til oqimidan iborat korpusdir. Bu oqim lingvistik ma’lumotlarni ajratib oladigan filtdan o‘tadi”, – deb yozadi. 1997-yilda qariyb 300 million so‘z qo‘llash holati bo‘lsa, 2005-yilda esa bu raqam 525 millionga etdi. Har oyda korpusga 2,5 million yangi birikmalar qo‘shiladi. Korpusning 25 foizini og‘zaki nutq, 75 foizini yozma nutq tashkil etadi. Internetda korpusning 56 so‘z million qo‘llanish holatidan iborat fragmenti<sup>58</sup>dan foydalanish mumkin.

The Longman Corpus Network<sup>59</sup> Logman (Longman) kompaniyasi, Lankaster universiteti tomonidan tuzilgan bir necha tijorat korpusi ma’lumotlar bazasidan iborat, 50-100 million so‘z qo‘llash holatini qamrab olgan.

---

<sup>58</sup> <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>

<sup>59</sup> <http://www.pearsonlongman.com/dictionaries/corpus/index.html>



Britaniya milliy korpusi<sup>60</sup> (British National Corpus (BNC)) 100 million soʻz qoʻllash holatidan iborat, ingliz tilini maʼlum janrda emas, balki toʻliq qamrab olgan. Bu korpusning hajmi Cobuild Projectdan farqli oʻlaroq aniq, chegaralangan. Tarkibining 90 foizini yozma manba, 10 foizini ogʻzaki matn tashkil etadi. Korpusni tashkil etishda bir qancha tashkilotlar (Britaniya hukumatidan boshlab) ishtirok etgan va mazkur ish 1995-yilda oʻz nihoyasigiga etgan. Korpus 4124 ta matndan iborat, shundan 863 tasi ogʻzaki suhbat, monologdan yozib olingan. Har bir matn orfografik segmentlangan, matndagi soʻz esa avtomatik tarzda turkumlangan. Soʻzni segmentlash, klassifikatsiyalash Lankaster universitetida ishlab chiqilgan stoxastik teglash dasturi CLAWS asosida amalga oshiriladi. Barcha matnlar eng optimal usullardan biri sanalgan SGML tili hamda TEI dasturi yordamida teglangan. Korpusni tuzishda matn tanlashning yangi usullari qoʻllanilgan, koʻp darajali nazorat tizimidan foydalanilgan<sup>61</sup>.

Xalqaro ingliz tili korpusi<sup>62</sup> (The International Corpus of English (ICE)) bir necha oʻnlab universitetning hamkorlikdagi loyihasi, har biri millionlab soʻz qoʻllash holatini qamrab olgan, 20 ta parallel ichki korpusdan iborat. Ingliz tili davlat tili yoki ikkinchi til hisoblangan davlatlarning (Avstriya, Kanada, Yangi Zelandiya va h.) oʻziga xos xususiyatini oʻrganish mumkin. Bu korpusni tahlil qilish uchun maxsus mukammal dasturiy taʼminot ishlab chiqilgan.

Amerika milliy korpusi<sup>63</sup> (American National Corpus)ning dastlabki koʻrinishi (parchasi) 2003-yilda ishlab chiqilgan. 100 million soʻz qoʻllash holatini qamrab olishi koʻzda tutilgan. Hozircha korpus hajmi 11 millionga etgan boʻlib, u XML formatda va undan foydalanish pullik.

Gigaword corpora<sup>64</sup> – ingliz, arab, xitoy va boshqa tillarning monitoringli korpusi. Evropa Ittifoqi homiyligida, uni Linguistic Data Consortium kompaniyasi tuzgan, korpus ustida uzluksiz ish olib boradi. Hozirda korpus hajmi 1 milliard soʻz holatida. Matnlar, asosan, yangiliklar, publitsistikani qamrab olgan, eng

---

<sup>60</sup> <http://www.natcorp.ox.ac.uk/>

<sup>61</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

<sup>62</sup> <http://www.ucl.ac.uk/english-usage/>

<sup>63</sup> <http://www.americannationalcorpus.org/>

<sup>64</sup> <http://www ldc.upenn.edu>

qimmat korpus sanaladi. Tarkibidagi korpuslar haqida to'liq ma'lumotni <http://www ldc.upenn.edu> saytidan olish mumkin.

### **Mustahkamlash uchun savol va topshiriqlar**

1. Korpus lingvistikasining shakllanishi, taraqqiyoti hamda bugungi holati haqida ma'lumot bering.
2. Birinchi va ikkinchi avlod korpuslarining o'xshash va farqli jihatlarini sanab bering.
3. Zamonaviy korpuslarning imkoniyati nimalarda ko'rinadi?
4. Korpuslarni yaratish davrlari qanday tasnis qilinadi?
5. Korpuslar qaysi xususiyatiga ko'ra davrlarga ajratilgan?
6. Rus tili korpuslari tarixi haqida nimalar bilasiz?
7. Rus tili korpuslarini sanab bering.
8. Ingliz tili korpuslari tarixi haqida gapirib bering.
9. Korpus lingvistikasining bugungi holati haqida gapirib bering.
10. Zamonaviy rus, ingliz, turk tili korpusining jahon korpuslari orasidagi o'rni haqida nima deya olasiz?

### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Грудева В.А. Корпусная лингвистика. Учебное пособие. 2-е изд., стереотип. – М.: Флинта, 2012. – 165 с.
3. Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) // [lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf](http://lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf)
4. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
5. Хамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изохли луғати. – Тошкент: Камалак, 2018. – 96 б.

6. Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол. фан. бўйича фалсафа доктори (PhD) диссер. – Қарши, 2019. – 140 б.
7. <http://ruscorpora.ru>
8. <http://ruscorpora.ru/search-murco.html>
9. [http:// corpus.leeds.ac.uk/](http://corpus.leeds.ac.uk/)
10. <http://corpus.leeds.ac.uk/ruscorpora.html>
11. <http://www.ucl.ac.uk/english-usage/>
12. <http://www.americannationalcorpus.org/>

## **RUS TILI OG‘ZAKI NUTQ KORPUSI.**

### **RUS TILI MAXSUS MANTLAR KORPUSI**

#### **Reja:**

1. Rus tili maxsus mantlar korpusi haqida umumiy ma’lumot.
2. Rus tili maxsus mantlar korpusi tuzilishi va interfeysi.
3. Korpus tarkibi, materiali, razmetkaning o‘ziga xos turi.
4. Rus tili og‘zaki nutq korpusi.
5. Tuzilishi, interfeysi, lingvistik, lingvodidaktik, ijtimoiy ahamiyati.
6. Korpus tarkibi, materiali, razmetkaning o‘ziga xos turi.

***Tayanch so‘z va iboralar:** og‘zaki korpus, yozma korpus, aralash korpus, rus tili og‘zaki nutq korpusi, maxsus og‘zaki korpus, London Lund korpusi, ovoqli korpus, multimodal korpus, rus tili maxsus mantlar korpusi, Sankt-Peterburg XV-XVII asr agiografik matnlari korpusi (SKAT), Agiografik asarlar, rus tilining Regensburg diaxron korpusi, Rudakiy korpusi, “Слово о полку Игореве” parallel korpusi, “Манускрипт” rus elektron meros korpusi.*

Rus tili gazetalar kompyuter korpusi<sup>65</sup> (Компьютерный корпус текстов русских газет конца XX в.) 2000-2002-yillarda Moskva Davlat universiteti filologiya fakultetida Umumiy, kompyuter leksikologiyasi hamda leksikografiyasi laboratoriyasida tuzilgan<sup>66</sup>. Korpus materiali sifatida 1994-1997-yillarda Rossiyaning 13 nomdagi gazetolari to‘liq sonlari asos qilib olingan, matnlardagi so‘z qo‘llanish holati 1 millionni tashkil etadi. Unga kundalik, nokundalik (“MN”, “Новая газета”), “chap” (“Завтра”, “Правда”, “Правда-5”) va “o‘ng” markaziy hamda hududiy, umumiy hamda kasbiy (masalan, “Литературная газета”) gazetalar kirgan. Bu tamoyil gazeta materiali bilan bog‘liq turli xildagi matn (masalan, turli tur, janr) birliklari, ular o‘rtasidagi munosabatni obyektiv, ishonarli kartina asosida kuzatish imkonini beradi. Korpus Moskva Davlat universiteti filologiya fakulteti Umumiy, kompyuter leksikologiyasi hamda leksikografiyasi laboratoriyasining Diktum-1 sistemasi asosida tuzilgan. Bu tizim yordamida korpus birlik va matnlari turli xildagi markerlar bilan avtomatik, yarim

---

<sup>65</sup> [www.philol.msu.ru/~lex/corpus](http://www.philol.msu.ru/~lex/corpus)

<sup>66</sup> <http://www.philol.msu.ru/~lex/corpus>

avtomatik belgi qo‘yiladi: matn – gazeta manbasi, matn hajmi, janri, nashr sanasi va h.; so‘z qo‘llanish holati – grammatik, leksik, morfem va boshqa kategoriya bo‘yicha belgilangan. Korpusning Internetga mo‘ljallangan ko‘rgazmali variantini tayyorlash uchun korpusning avtomatik lemmatizatsiya, morfologik kvalifikatsiya, shuningdek, leksema, so‘zshaklning morfem segmentatsiyalanishi imkoniga ega bo‘lgan 200 000 ta so‘z qo‘llanish holati hajmini qamrab olgan lavha (fragment) ajratib olindi. Korpusning to‘liq varianti ustida ish olib borilyapti. Yuqorida tavsiflangan oxirgi ikki korpusning xususiyatlari qiyosiy tahlilini jadvalda ko‘rish mumkin.

*Zamonaviy rus korpuslari. Rus tili milliy korpusi*<sup>67</sup> (Национальный корпус русского языка) ommabop rus matnlari korpusi 2004-yil 29-aprelda ochilgan. Korpus tuzish 2001-yilda Moskva, Peterburg, Voronej kabi shaharlardagi tilshunoslar guruhi hamkorligida boshlangan, asosiy tuzuvchilar – RFA Rus tili instituti, RFA Tilshunoslik instituti hamda “YAndeks” kompaniyasi. 2007-yilda 140 mln. so‘z qo‘llash holatini qamrab olgan yozma, og‘zaki, badiiy, dialektal matnlar yig‘indisini tashkil etgan<sup>68</sup>. 2013-yil mart oyi holatida hajmi 500 million so‘z qo‘llash holatiga etdi. Ko‘rib turganimizdek, bu korpus doimiy tarzda yangilanib, to‘ldirilib boriladi. Matnning janriy xilma-xilligi uning yana bir xususiyati hisoblanib, rus tilida qo‘llanadigan barcha soha uslub, janrlarini (ilmiy, rasmiy-idoraviy, publitsistik, cherkov, badiiy, kundalik so‘zlashuv, og‘zaki, elektron kommunikatsiya) qamrab olgan<sup>69</sup>. Korpus morfologik, semantik tegga ega; Web sayt orqali erkin foydalanish mumkin. Bu korpusning mashhur, ommabopligi shundaki, bir qator xususiyatlari bilan birga, boshqa korpuslardan farq qiluvchi asosiy belgisi *multimediali korpus*<sup>70</sup> hamda *semantik tegga*<sup>71</sup> ega ekanligi bilan xarakterlanadi.

---

<sup>67</sup> <http://ruscorpora.ru>.

<sup>68</sup> *Қаранг:* Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

<sup>69</sup> *Қаранг:* Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 — 30. 05.14 < Crimea, Simferopol — Alushta.

<sup>70</sup> (<http://ruscorpora.ru/search-murco.html>)

<sup>71</sup> (<http://ruscorpora.ru/corpora-sem.html>)

Rus tilining multimediali korpusi<sup>72</sup> (Мультимедийный русский корпус (МУРКО)), 1930-2000-yillardagi kinofilm lavhalaridan tashkil topgan. Videoryad, audioryad, ovozning matnli rasshifrovkasi, kadrda uchraydigan imo-ishoralar kabi parallel parametrlarni qamrab olgan. Korpus ichidagi qidiruv nafaqat talaffuz qilinayotgan matn, balki imo-ishora (boshni qimirlatish, elka qoqish va h.) hamda nutqiy harakat turi (rozilik, piching, kesatq, kinoya va h.) orqali ham amalga oshirilishi mumkin. Qidiruv natijasi videolavhani tomosha qilish hamda eshitish imkonini beradi.

Xelsinki annotatsiyalangan korpusi (Хельсинский аннотированный корпус<sup>73</sup> (ХАНКО)) 2000-yillarda Xelsinki universitetida “Rus tilining funksional sintaksisi” (rahbar prof. A.Mustayoki) loyihasining bir qismi sifatida ishni boshlagan, u doimiy ravishda to‘ldirib boriladi. Korpus hajmi 100 ming so‘zshaklni tashkil etgan, uning asosini 2001-yil “Sarhisob”i jurnali maqolalari tashkil qilgan. Korpus<sup>74</sup>dan internetda erkin foydalanish mumkin. Morfologik, sintaktik teg hamda shu parametrli qidiruv imkoni mavjud. Korpusning o‘ziga xosligi ma’lumotlar lingvistik tavsifining puxta ishlangan hamda avtomatik teg natijasi to‘liq, birma-bir qo‘lda tekshirish imkoniyati mavjudligidadir. Bu bilan esa grammatik omonimiyani to‘liq ko‘rsatishga erishiladi. Korpusning sintaktik tegi gap bo‘lagi, grammatik aloqani aniqlab berish vazifasini birga olib borishi lozim. Bu korpus boshqalaridan farqli o‘laroq, shu vazifani juda yaxshi bajaradi. Bundan tashqari, korpusda taxminan 2000 birlik atrofida murakkab gap bo‘laklarini hosil qiluvchi turg‘un birikma ham aniqlangan. Semantik kategoriyani belgilaydigan terminlar asosida semantik teglashni amalga oshirish ko‘zda tutilmoqda. Korpus, eng avvalo, ta’limiy maqsadni ko‘zlaydi. 2000-yillarda Lids universiteti (Buyuk Britaniya)ning tarjimonlik tadqiqotlari markazida turli tillarning (ingliz, arab, xitoy, fransuz, nemis, italyan, yapon, ispan, polyak va h.) katta hajmli korpuslari tuzilgan<sup>75</sup>. Ular orasida rus tili korpusi<sup>76</sup> ham mavjud. 116 million so‘zni qamrab

---

<sup>72</sup> <http://ruscorpora.ru/search-murco.html>

<sup>73</sup> <http://h248.it.helsinki.fi/hanco/>

<sup>74</sup> Ўша манба.

<sup>75</sup> <http://corpus.leeds.ac.uk/>

<sup>76</sup> <http://corpus.leeds.ac.uk/ruscorpora.html>

olgan Rus tili milliy korpusining bu versiyasi asosida O.N.Lyashevskaya, S.A.Sharovning “Rus tilining chastotali lug‘ati” foydalanuvchilar qo‘liga etib bordi. Shu bilan birga, bu saytda Rus gazetalari korpusi (2001-2004-yillar, 76 million so‘z), Rus Internet matnlari korpusi (160 million so‘z), biznes-korpus (12 million so‘z) hamda sanab o‘tilganlarni jamlagan yaxlit korpus ham mavjud.

Moshkova kutubxonasi korpusi<sup>77</sup> saytidan 680 million so‘zdan iborat rus matnlari korpusi o‘rin olgan. Korpus Moshkova kutubxonasi matnlari asosida A.Sokirko tomonidan tuzilgan. Leksik birliklarni DDC korpus menejeridan foydalanib so‘z turkumi, morfologik xarakteristikasi asosida qidirish mumkin. Shuningdek, MI mezoni asosida hisoblangan bigrammalar (54 mln.) qidiruvi servisi ham mavjud. Korpus 40 dan ortiq turli til korpuslaridan tuzilgan Ingliz lingvistik xizmati (Lexical Computing Ltd.) tomonidan tijorat operatsiyalari asosida foydalanish huquqini beradi. Ular orasida bir qator rus korpuslari, Internet matnlaridan Wacky texnologiyasi asosida tayyorlangan 20 milliard hajmli korpus e‘tiborlidir. Ingliz tadqiqotchilari va Masarika universiteti xodimlari hamkorligida Sketch Engine korpus menejeri<sup>78</sup> ishlab chiqilgan, mazkur korpus noyob imkoniyatlarga ega. Standart qidiruv tizimi, konkordans bilan birga sintaktik qoliplarni ham aniqlab beradi, chastotali lug‘atni shakllantiradi, leksik birliklarning o‘zaro birikish imkoniyatlarini ko‘rsatib, leksik-semantik maydonda guruhlaydi<sup>79</sup>. Ko‘rinadiki, tarixan qisqa davr ichida rus korpus lingvistikasi ham katta natijalarga erishgan.

Xullas, jahon tilshunosligida ushbu yangi soha tinimsiz rivojlanmoqda.

---

<sup>77</sup> <http://aot.ru/search1.html>

<sup>78</sup> <http://sketchengine.co.uk/>

<sup>79</sup> *Қаранг:* Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.

## Korpuslarning qiyosiy tahlili

№		Rus tilining Upsalla korpusi <sup>80</sup>	Rus tili gazetalar kompyuter korpusi
1	Tarkibi	1996-2002 yillar publitsistikasi; XIX-XX asr badiiy adabiyoti	1997 yil gazeta matnlari
2	Hajmi	25 million soʻz atrofida	100 ming soʻz atrofida
3	Soʻz yasalishi	+	–
4	sintaktik	element	element
5	semantik	–	element
<b>Qidiruv shakli</b>			
6	Soʻz shakli orqali	+	+
7	Lemma orqali	–	+
8	Bir necha soʻzni baravar	+	–
9	Grammatik belgi asosida	+	+
10	Semantik belgi asosida	–	+
11	Tinish belgi asosida	+	+

**Korpus: yozma va ogʻzaki, bir tilli va koʻp tilli.** Mutaxassislar nutq turiga koʻra korpusning yozma, ogʻzaki, aralash shaklini ajratishadi<sup>81</sup>. Yozma korpusda ogʻzaki nutq aks ettirilmagan boʻlsa (Braun korpusi, LOB), ogʻzaki korpus faqat ogʻzaki matndan iborat boʻladi; aralash korpus, odatda, til meʼyorlari aksi boʻlgan milliy korpuslardir. Zamonaviy katta hajmli korpuslar, odatda, aralash: yozma matn bilan birga ogʻzaki matnni ham qamrab oladi. Hatto Britaniya milliy korpusi (BNC) matnlarining 10 foizi ogʻzaki matnlar yozuvidan iborat. Til, asosan, ogʻzaki shaklda yashaydi, uning yozma shakli ikkilamchi. Shuning uchun ogʻzaki korpus

<sup>80</sup> Ушбу корпус Интернетда “Рус матнларининг Тюбинген корпуси” деган ном билан аталади.

<sup>81</sup> Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – Б.26.



juda muhim, hech bo'lmaganda, korpusning aralash turi til tadqiqi uchun maqsadga muvofiq hisoblanadi.

Maxsus og'zaki korpuslar<sup>82</sup>dan London Lund korpusi (London Lund Corpus (LLC, 1975)), Lankaster (Lancaster/IBM Spoken English Corpus (SEC 1992)) korpuslarini alohida ta'kidlab o'tish lozim. Keyingisi 52600 so'z hajmida, audioyozuvi bilan birga CD shaklida foydalanuvchilarga taqdim etilgan. Urg'u, intonatsiya, to'xtam jihatidan to'liq razmetkalangan, ammo ijtimoiy mavqei, respondentning shakllanishi haqida ma'lumotlar bo'lmaganligi korpusning sotsiolingvistikada qo'llanishini cheralaydi.

Og'zaki amerikancha ingliz tili korpusi (Corpus of Spoken American English (1991)), 1 mln. so'z qo'llash holati, 80 soat ovozli yozuv; Map Task Corpus (1991, Glazgo universiteti, Shotlandiya) 147000 so'z qo'llash holati, 16 soat ovozli yozuvdan iborat. Og'zaki korpuslar yozma korpusga nisbatan kam miqdordagi so'zshaklni qamrab oladi. Bu faqat ma'lumotni yig'ishning qiyinligi uchun emas, balki prosodik tadqiqot uchun kam miqdor ham yetarli sanaladi. Masalan, intonatsiyani o'rganish uchun 100.000 so'zdan iborat korpus etarli bo'ladi. Bunday korpuslar monologik, dialogik nutqni o'z ichiga oladi. Material yig'ish uchun radio, televidenie yozuvlari, sotsiologik, sotsiolingvistik so'rov natijalaridan foydalaniladi. Bugungi kunda (XX asrning 70-yillardan farqli o'laroq) ovozni yashirin usulda yozib olish etik me'yorga to'g'ri kelmaydi. Odatda, respondent haqida atroflicha to'liq ma'lumot (yozuv o'rni, respondentning nima bilan shug'ullanayotganligi, vaqti, sana, ishtirokchilar soni, jinsi, yoshi, suhbatning favqulodda (tabiiy)lik darajasi, mavzu, ishtirokchining etnik mansubligi, asosiy (ona) tili, kasbi, ijtimoiy mavqei va h.) yig'iladi. Eng mashaqqatli ish transkripsiyalash jarayoni, bir soatlik ovozli yozuvning orfografik transkripsiyasi, minimal intonatsion razmetkasini tayyorlash uchun 10 soat vaqt kerak bo'ladi. Agar matn TEI (Text Encoding Initiative) asosida to'liq razmetkalansa, bu ishga 25 soat vaqt sarflanishi mumkin. Razmetkasiz og'zaki korpus hech qanday ahamiyatga ega emas, chunki, hech bo'lmaganda, bir vaqtdagi nutqning to'xtam

---

<sup>82</sup> Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

davomiyligi, urg‘u va intonatsiyasi ko‘rsatilishi lozim. Ba‘zan matnga izohlar ham qo‘shiladi. London-Lund korpusi to‘liq razmetkasi og‘zaki korpus razmetkalari uchun standart vazifasini o‘tayapti.

Ovozli korpus<sup>83</sup> namunasi sifatida Rus tili milliy korpusi tarkibidagi og‘zaki nutq korpusi<sup>84</sup>ni sanab o‘tish joiz. Bu korpus multimediali, dialektal, badiiy, aksentologik korpuslar kabi milliy korpus tarkibidan joy olgan. Shu bilan birga, Sankt-Peterburg davlat universiteti filologiya fakulteti filologik tadqiqotlar instituti tomonidan ishlab chiqilgan og‘zaki korpus ham shunday korpus namunasi. “Bir kunlik nutq” deb nomlanuvchi bu korpus – predmeti kundalik so‘zlashuv nutqi, u til egasining kundalik muloqotdagi jonli tilini o‘rganadi. Boshqa ovozli korpuslardan asosiy farqi jonli nutqdan yozib olinganligi, u laborator yoki boshqa maxsus sharoitda yozib olingan nutqdan farq qiladi. Sankt-Peterburgda yashovchi turli yosh, ijtimoiy qatlamga mansub aholidan (30 kishi) iborat informantlar guruhi tanlab olingan, ular bir kun davomida diktofon bilan yurishga rozilik bergan, shunday tabiiy sharoitda ularning kundalik muloqot nutqi yozib olingan. Informant ovoz yozish jarayoni borasida yo‘riqnoma bilan tanishtirilgan, psixologik test, ijtimoiy so‘rovnomadan o‘tkazilgan. Informantlar bilan birga kun davomida muloqotda bo‘lgan shaxs (qarindosh, do‘st, hamkasb, tanish, notanish) nutqi ham kiritilgan, natijada, korpus har xil yoshdagi shaxs, turli soha vakillari nutqini qamrab olgan. Ovozli yozuv 500 soatni tashkil etadi (2013-yil holatida). Bunday ovozli korpus yozib olingan materialning lingvistik dinamikasini o‘rganish imkonini beradi. Jumladan,

- 1) standart statistik metod asosida miqdoriy o‘zgarishlarni o‘rganish;
- 2) leksik, grammatik qator chastotasi, sintaktik struktura, u yoki bu akustik hodisa, prosodik konturning semantika, suhbat mavzuini sutkaning ma’lum qismi, muloqot sharoiti bilan bog‘liq holatni tahlil qilish;
- 3) shaxs kasbining maishiy hayotiga ta’sirini tahlil qilish;

---

<sup>83</sup> Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.

<sup>84</sup> <http://ruscorpora.ru/sbornik2008/07.pdf>

4) til tashuvchisining tabiiy nutq jarayonidagi o‘rtacha artikulyasion tempi haqida ma’lumot olish kabi hodisalar tadqiqi uchun xarakterlidir.

Multimodal korpus<sup>85</sup> muloqot ishtirokchisi videoyozuvini ham o‘z ichiga olgan. Mimika, qo‘l, ko‘z, qosh harakati va h. belgilar asosida razmetkalangan. Korpus o‘zaro emotsional strategiya, ixtilof, muloqot odobi, nutqiy to‘xtam, boshqa hodisalarni o‘rganish vositasi bo‘lishi mumkin. Mijoz bilan ishlaydigan xodim uchun o‘rganish materiali, multiplikator hamda rejissyor uchun emotsional reaksiya ma’lumotlar bazasi sifatida xizmat qiladi. Rus tili emotsional korpusi (RES)<sup>86</sup> shunday korpus namunasi. Irkutsk Davlat tilshunoslik universitetida Ta’limiy multimodal korpus (UMKO) tuzish davom etmoqda. Bu korpus materiali rus, xitoy tili tashuvchisi, ikkinchi tili sifatida ishlatuvchi shaxslarning ma’lum mavzu borasida oldindan tayyorlanmagan ta’limiy dialoglari videoyozuvidir. Korpus ELAN dasturi asosida razmetkalangan, parallel korpus shaklini olgan, dialoglar ma’noviy guruhlariga bo‘lingan holda joylashtirilgan. Masalan, rus tili egasi nutqi bilan rus tilini o‘rganayotgan xitoy tili vakili nutqi ayni bir mavzudagi dialog (“Tanishuv”, “Aeroportdagi registratsiyada”) asosida qiyoslanadi. Ushbu korpus birinchi navbatda lingvodidaktik maqsadni, xususan, tipik xatolarni topish, mashg‘ulotlar davomida shu xatolarni nutqdan chetlashtirish yo‘lini izlash hamda talabaning mustaqil ishiga ko‘maklashishni ko‘zlaydi. Yuqoridagi korpus namunasi multimediali korpus asosida imo-ishora, qo‘shimcha emotsional holatlarning (og‘zaki) nutqqa ta’sirini o‘rganish vositasi ekanligini ko‘rsatadi.

**Maxsus matnlar korpusi.** Maxsus matnlar korpusi<sup>87</sup> – balansga keltirilgan korpus, u hajman u qadar katta bo‘lmagan, alohida tadqiqiy topshiriqqa xizmat qiluvchi, ma’lum maqsadni ko‘zlagan, tuzuvchi rejasiga muvofiq tayyorlangan. Sankt-Peterburg ingliz tilini o‘rganuvchi (SPbEFLLC) maktab o‘quvchilarining ta’limiy matnlari korpusi shunday korpuslar sirasiga kiradi. A.I.Gersen nomidagi pedagogika universiteti amaliy tilshunoslik kafedrasidan yaratilgan bu

---

<sup>85</sup> Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.

<sup>86</sup> <http://www.harpia.ru/rec>

<sup>87</sup> Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.

korpusni tuzishdan asosiy maqsad rus o'quvchisida eng ko'p qiziqish uyg'otgan ingliz matnlarining o'ziga xosligini o'rganish edi. Korpus uchun material vazifasini o'tagan bunday matnlarning asl nusxasi Sankt-Peterburg maktablarida 2007 yil noyabr-dekabr oylarida yig'ilgan. Ushbu matnlarning muallifi 911ta sinfning oldindan maxsus testdan o'tkazilgan ingliz tilini bilish darajasi o'rta (intermediate (26%)), o'rtadan yuqori (upper-intermediate (74%)) sifatida baholangan 78 nafar o'quvchi bo'ldi. Korpus hajmi 50.000 so'z miqdorida. Korpus bazasiga asoslangan tadqiqot shuni ko'rsatdiki, ingliz tili xorijiy til sifatida (English as a Foreign Language) egallangan shaxs nutqida sodda matn, gap tuzilishi hamda qolipi keng qo'llaniladi. Bu EFL (English as a Foreign Language) grammatikasi ingliz adabiy tili grammatikasi bilan aynan bir xil strukturaga ega emasligini ko'rsatadi.

Tuzish jarayoni ancha murakkab hisoblangan korpuslardan biri Sankt-Peterburg XV–XVII asr agiografik matnlari korpusi (SKAT)<sup>88</sup> kabi tarixiy korpusdir. SKAT – Sankt-Peterburg filologiya fakulteti matematik lingvistika kafedrasidan tuzilgan qadimgi rus agiografik adabiy yodgorliklari matni elektron korpusi.

Agiografik asarlar<sup>89</sup> tili ko'p jihatdan XV–XVII asr rus adabiy tili xarakteri, taqdirini belgilab bergan edi. Bu tilni aks ettirishning birlamchi vazifasi o'sha davr turmushiga oid rus matnlari korpusini tuzishdir. Natijada, katta geografik territoriyada turli davrda asos solingan rus agiografik yodgorliklari tili namoyon bo'ladi. 2011 yilda korpus hajmi 500.000 so'zni tashkil etgan edi. Maxsus korpusga bo'lgan ehtiyoj milliy korpusdan qolishmaydi. Xohlagan maxsus korpus aniq bir soha yoki yondosh sohani (kemasozlik, metallurgiya, ekologiya, navigatsiya va h.) o'rganish uchun qo'llanishi tabiiy. Chunki bu korpus mutaxassisga atamani konkret professional qurshovda, u yoki bu muallif shu atama ostida nimani nazarda tutganligini kuzatish imkonini beradi. Maxsus korpuslar sirasida Rus tilining Regensburg diaxron korpusi (qadimgi rus matnlari), Qadimgi

---

<sup>88</sup> <http://www.project.phil.spbu.ru/>

<sup>89</sup> Агиография (грек. *ἀγιος* “муқаддас” ва *ῥάφω* “ёзаман”) – диний ва тарихий-черков муқаддас ёдгорликларини ўрганувчи мустақил фан соҳаси, диний турмуш тарзи лингвистик, диний-тарихий, тарихий, ижтимоий-маданий ва адабий нуктаи назардан ўрганилади.

Rus yozma yodgorliklari korpusi (qo‘lyozma kitoblar, solnoma, qayin po‘stlog‘iga yozilgan qadimgi rus yodgorliklari), “Слово о полку Игореве” parallel korpusi, “Манускрипт” rus elektron meros korpuslarini sanash mumkin.

Rudakiy korpusi<sup>90</sup> misolida korpusning boshqa bir ko‘rinishi haqida fikr yuritamiz. Tojikiston Respublikasi Fanlar akademiyasi Informatson va kommunikatsion texnologiyalar markazi Axborot va kommunikatsion texnologiyalari maktabi hamkorligida Tojik tili milliy korpusi yaratilgan. Korpus tarkibida eng mashhur fors tojik adabiyoti namoyandalari Rudakiy, Firdavsiy, Umar Xayyom, Hofiz, Rumi ijodi korpusi, zamonaviy adabiyotning turli janrlari, tojik radio va televideniesi orqali efirga uzatilgan material matni; turli gazeta, jurnal, informatson byulleten materiallari, ilmiy, ilmiy-ommabop, ishlab chiqarishga oid rasmiy-administrativ, yuridik matnlar, tojik tilining izohli, tarjima lug‘atlari matnlari, ijtimoiy ahamiyatga ega matnlar(yozishma, muloqot)dan tarkib topgan. Ushbu korpus tarkibidagi ikki korpus alohida e‘tiborga molik.

### **Mustahkamlash uchun savol va topshiriqlar**

1. Og‘zaki korpus qanday xususiyatlari bilan ajralib turadi?
2. Yozma korpusnig belgilarini sanang.
3. Aralash korpusning qanday ahamiyati bor?
4. Rus tili og‘zaki nutq korpusi qanday maqsadlarda foydalaniladi?
5. Maxsus og‘zaki korpus qanday afzalliklarga ega?
6. London Lund korpusi haqida ma’lumot bering.
7. Ovozli korpusda birliklar qanday razmetkalanadi?
8. Multimodal korpusning xususiyatlarini sanang.
9. Maxsus matnlar korpusi deganda nimani tushunasiz?
10. Rus tili maxsus mantlar korpusining razmetkasi haqida gapiring.
11. Sankt-Peterburg XV–XVII asr agiografik matnlari korpusi (SKAT) qanday maqsadlarda qo‘llanadi?
12. Agiografik asarlar deb qanday asarlarga aytiladi?
13. Rus tilining Regenbug diaxron korpusi qanday afzalliklarga ega?

---

<sup>90</sup> <http://termcom.tj>.

14. Rudakiy korpusining razmetka darajasi haqida gapiring.
15. “Слово о полку Игореве” parallel korpusining afzalliklarini sanab bering.
16. “Манускрипт” rus elektron meros korpusining xususiyatlarini sanab bering.
17. Rus korpus lingvistikasi tarixi haqida gapirib bering.
18. Elektron va annotatsiyalangan korpuslarning shakllanish bosqichlari qanday tasniflanadi?
19. Rus tili milliy qanday xususiyatlarga ega?
20. Upsalla korpusi haqida tushuncha bering.
21. Rus tili gazetalar korpusi qanday belgilari bilan Milliy korpusdan ajralib turadi?
22. Rus tili Milliy korpusining tarkibida yana qanday korpuslar mavjud?

#### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Грудева В.А. Корпусная лингвистика. Учебное пособие. 2-е изд., стереотип. – М.: Флинта, 2012. – 165 с.
3. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
4. Хамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
5. Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2019. – 140 б.
6. <http://khnt.hit.uib.no/icame/manuals/lobman/>
7. <http://www.ucl.ac.uk/english-usage>
8. <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>
9. <http://www.pearsonlongman.com/dictionaries/corpus/index.html>
10. <http://www.natcorp.ox.ac.uk/>
11. <http://www.americannationalcorpus.org/>

12.<http://www ldc.upenn.edu>

# TURKIY KORPUS LINGVISTIKASI TADRIJI VA BUGUNGI HOLATI

## Reja:

1. Turkiy tillar korpuslari va ularning tavsifi.
2. Turkiy tillar korpus lingvistikasi bo'yicha loyihalar.
3. Turli elektron resurslar: elektron lug'at, tezaurus va korpuslar.
4. Turkiy tillar korpus lingvistikasi: tadqiqot va tadqiqotchilar.
5. Turkiy tillar kompyuter va korpus lingvistikasi bo'yicha tadbirlar.
6. Turkiy tillarning kompyuter instrumentariylari.

Tayanch tushunchalar: *turkiy korpus lingvistikasi, loyihalar, korpus, tezaurus, elektron lug'at, korpus lingvistikasi bo'yicha tadqiqotlar, kompyuter instrumentariylari*

## 1. Turkiy tili korpuslar umumiy tavsifi

Korpus tilshunosligi zamonaviy amaliy tilshunoslikning ustuvor yo'nalishi hisoblanadi. Korpus tilshunosligi zamonaviy kompyuter texnologiyalaridan foydalangan holda lingvistik ma'lumotlar korpusi (matn korpuslari) qurishning umumiy tamoyillarini belgilaydi, yozma va og'zaki matnlarning haqiqiy lingvistik hodisalarini tshplash metodologiyasini, shuningdek, ularni saqlash va tahlil qilish usullarini ishlab chiqadi. Matnli korpuslar bilan ishlash tadqiqotchini sub'ektivlikdan ma'lum darajada mavhumlashtirishga va tilni ob'ektiv o'rganishga imkon beradi.

**Tatar tili yozma matnlar elektron korpusi**<sup>91</sup> – onlayn foydalanish mumkin bo'lgan tatar tilining elektron korpusi. Korpus tatar tili tizimi, holati va kelajagi bilan qiziquvchilarga mo'ljallangan bo'lib, tatar tilini korpus lingvistikasi doirasida o'rganuvchilar uchun muhim. Sayt 2012-yil 15-martda ochilgan. <http://corpus.tatar>. manzilida tatar, rus, turk tilidagi sahifalar mavjud. Korpus hajmi 2014-yilda 10 mlndan ortiq gaplarni tashkil etuvchi 116 mln so'z; turli so'zshakllar miqdori 1,5 mln.ga etadi. Korpus matnlari ko'chirilishning oldini olish

---

91

[https://ru.wikipedia.org/wiki/%D0%9F%D0%B8%D1%81%D1%8C%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9\\_%D0%BA%D0%BE%D1%80%D0%BF%D1%83%D1%81\\_%D1%82%D0%B0%D1%82%D0%B0%D1%80%D1%81%D0%BA%D0%BE%D0%B3%D0%BE\\_%D1%8F%D0%B7%D1%8B%D0%BA%D0%B0?veaction=edit&section=7](https://ru.wikipedia.org/wiki/%D0%9F%D0%B8%D1%81%D1%8C%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D0%B9_%D0%BA%D0%BE%D1%80%D0%BF%D1%83%D1%81_%D1%82%D0%B0%D1%82%D0%B0%D1%80%D1%81%D0%BA%D0%BE%D0%B3%D0%BE_%D1%8F%D0%B7%D1%8B%D0%BA%D0%B0?veaction=edit&section=7)



maqsadida alohida gap shaklida saqlanadi. Korpusga kirish va undan foydalanish bepul. Ushbu korpusni tuzish 2010-yilda ko'ngillilar tomonidan boshlangan: ishning boshlanishiga tatar-ingliz, ingliz-tatar avtomatik tarjima tizimi, avtomatik sintez va ma'lum mavzudagi tatar tilini "tushunish" (raspoznavaniya tatarskoy rechi) ishlab chiqish turtki bo'lgan. Korpus so'z, uning chastotasi va kontekstini ko'rish imkonini beradi. Korpusda statistik-kontekstli qidiruv, murakkab morfologik qidiruv funksiyalari mavjud. 2014-yilda korpusning avtomatik morfologik razmetkasi amalga oshirilgan; xalqaro Apertium loyihasi ishlab chiqqan morfologik teglar asosida turkiy tillar teglari tizimi tayyorlangan. Ushbu loyiha doirasida bir nechta til uchun avtomatik tarjima tizimi ishlab chiqiladi. 2015-2016-yillarda ishlab chiqilgan murakkab morfologik qidiruv korpusda turli kombinatsiyalarda: so'zshakl, lemma, morfologik teglar majmuasi, so'z boshi, o'rtasi, oxiri hamda ikki so'z oralig'ini ko'rsatish orqali qidiruv amalga oshirish qulayliklariga ega.

Tatar tili sintezatori tizimi tatar tilining yozma korpusida topilgan gap va matnlarni eshitish imkonini yaratadi. Mualliflar tomonidan olingan ishlov natijasi korpus saytida statistik ma'lumot sifatida saqlanadi.

Korpusning kamchiliklari sifatida korpusning oflayn versiyasi mavjud emasligi hamda omonimiya farqlangan qidiruvni amalga oshirib bo'lmashligi ko'rsatish mumkin.

Tatar tili yozma matnlari korpusi M.R.Sayxunov (filologiya fanlari nomzodi, TR FA Informatika instituti ilmiy xodimi), T.I.Ibragimov (filologiya fanlari nomzodi, KFU amaliy lingvistika kafedrasida dotsenti), R.R.Xusainov (injener, GDC kompaniyasi) tomonidan tuzilgan.

Mualliflar tadqiqotning ikki yo'nalishini ko'rib chiqdilar:

- 1) tatar tilidan bir-biriga o'xshash tillardan biriga va teskarisiga – shu bilan bog'liq tildan tatar tiliga matnlarni avtomatik tarjima qilish tizimini (MT) ishlab chiqish;
- 2) ma'lum bir mavzu bo'yicha tatar nutqini avtomatik ravishda aniqlash tizimini yaratish.

Adabiyotlar bilan tanishgandan so‘ng, zamonaviy MT tizimlari va nutqni avtomatik aniqlash tizimlari tilning milliy korpusi “gipoteza – tekshiruv” usulidan foydalangan holda ishlaydi.

Tatar tilidagi yozma korpus, asosan, veb-manbalar materiallari asosida yaratilgan. Tatar korpusiga kiritilishidan oldin, matnlar html-teg yorliqlarini tozalash, chet tillaridagi jummalarni olib tashlash, utf-8 ga aylantirish, jumlar chegarasini belgilash kabi dastlabki avtomatik ishlov berishdan o‘tkazilgan.

Tatar tilining ushbu yozma korpusi representativlik talablariga to‘la javob beradi. YOzma korpusga kiritilgan matnlarning aksariyati uchta uslubga tegishli: publitsistika (60%), badiiy adabiyot (35%) va gumanitar sohadagi ilmiy adabiyotlar (5%).

Tatar tilidagi yozma korpus tatar tili olami haqida tizimli tasavvur hosil qiladigan ulkan ma’lumotnoma-lug‘atdir. Tatar tili elektron korpusining asosiy maqsadi – tatar tilining leksikasiga oid ilmiy tadqiqotlarga yordam berish. Bundan tashqari, yozma korpus tilni o‘qitish (lingvodidaktika), shuningdek, har xil hujjatlarni to‘plash uchun ma’lumot sifatida ishlatilishi mumkin. Hozirgi kunda tatar tilining korpusiga kiritish uchun material to‘plash va uni qayta ishlash davom etmoqda. Yozuvchi va olimlar korpusni to‘ldirish uchun o‘z kitoblarining elektron nusxalarini taqdim etishgan. 5-6 million so‘zshakli hajmiga etgach, ushbu materiallar korpusga kiritiladi. Shu bilan birga, ishning funksional imkoniyatlari ham boyitiladi.

Tatar tili matn korpusining dasturiy resurslari quyidagi amallarni bajarishga imkon beradi:

- 1) kerakli so‘zni qidirishni tashkil qilish, ulardan foydalanish chastotasini aniqlash;
- 2) berilgan so‘zning oldida va undan keyin qaysi so‘zga bog‘lanish mumkinligini aniqlash (berilgan so‘zning chap va o‘ng qismlarini aniqlash), so‘z chastotasini aniqlash;
- 3) til o‘rganuvchi uchun zarur bo‘lgan so‘zni topish;
- 4) ma’lum bir tilda berilgan so‘z yoki shakldan foydalanganligini tasdiqlovchi misollarni topish.

Ushbu amallar quyidagi muammolarni hal qilishda imkon beradi:

- 1) tatarcha soʻzlar chastotali lugʻatini tuzish;
- 2) tatar tilidagi matnni ehtimoliy-statistik modellashtirish boʻyicha tadqiqot oʻtkazish;
- 3) tilning leksik/sintaktik birliklarining moslik imkoniyati hamda cheklanishlarini oʻrganish;
- 4) tilning morfologik tizimini oʻrganishda zarur boʻlgan teskari chastotali lugʻatni qurish.

Tatar tilining elektron korpusi yuqorida keltirilgan roʻyxatga nisbatan ancha kengroq va rang-barang. SHunday qilib, hozirgi paytda korpus tilshunosligi doirasida nutqni avtomatik aniqlash va avtomatik tarjima qilish kabi vazifalar echim topmoqda.

**Qrim-tatar tilining lingvistik korpusi**<sup>92</sup> – zamonaviy yozma qrim-tatar matnlari korpusi. Korpus, asosan, XXI asr boshidagi qrim-tatar gazetalarining matnlaridan iborat. Korpusda kirill yozuvidagi matnlar mavjud, shuning uchun soʻrovni kirill yozuvida kiritish talib qilinadi. Foydalanuvchi klaviaturasida kirill alifbosi boʻlmasa, ekrandagi virtual klaviaturadan soʻrov oynasiga kirill harflarini kiritish uchun foydalanish mumkin. Oddiy iboralarga mos keladigan bitta tokenni (soʻzlarni), tokenlar kombinatsiyasini (seriyasini) qidirish imkoni mavjud. mumkin. Qidiruv natijalari har bir satr bittadan mos keladigan KWIC (kalit soʻz kontekstida) koʻrinishida taqdim etiladi. Chap jadvalda ikkita ustun mavjud: tegishli raqamni (korpusdagi hujjat raqamini) bosish bilan tegishli hujjatning izohini olish mumkin, ikkinchi ustundagi ( ) tugmachasini bosgan holda, ajratilgan qidirish chegarasi bilan kengroq kontekst koʻrsatiladi. Pastki satrda boshlash va tugatish, sahifadagi matnlarning umumiy va mos keladigan soni koʻrsatiladi. Hozirgi vaqtda korpusda 521012 token (tinish belgilaridan tashqari) mavjud boʻlib, ular 56752 soʻzshakldan iborat.

### **Tatar tili yozma matnlar elektron korpusi va Qrim-tatar tilining lingvistik korpusi qiyosi (umumiy jihatlar)**

---

<sup>92</sup> <https://korpus.sk/QIRIM/#id9>

№	Tatar tili yozma matnlar elektron korpusi	Qrim-tatar tilining lingvistik korpusi
Umumiy jihatlar		
1.	Tatar tili yozma matnlar elektron korpusi hajmi yilda 10 mlndan ortiq gaplarni tashkil etuvchi 116 mln soʻz; turli soʻzshakllar miqdori 1,5 mln. (2014-y.)	Qrim-tatar tilining lingvistik korpusida 521012 token (tinish belgilardan tashqari) mavjud, ular 56752 soʻzshakldan iborat (2019 y.)
2.	Tatar tili yozma matnlar elektron korpusi soʻz, uning chastotasi va konteksini koʻrish imkonini beradi. Korpusda statistik-kontekstli qidiruv, murakkab morfologik qidiruv funksiyalari mavjud.	Qrim-tatar tilining lingvistik korpusida oddiy iboralarga mos keladigan bitta tokenni (soʻzlarni), tokenlar kombinatsiyasini (seriyasini) qidirish imkoni mavjud. Qidiruv natijalari har bir satr bittadan mos keladigan KWIC (kalit soʻz kontekstida) koʻrinishida taqdim etiladi.
3.	Tatar tili yozma matnlar elektron korpusining <a href="http://corpus.tatar">http://corpus.tatar</a> . Manzilida tatar, rus, turk tilidagi sahifalar mavjud.	<i>Qrim-tatar tilining lingvistik korpusi <a href="https://korpus.sk/QIRIM/#id9">https://korpus.sk/QIRIM/#id9</a> manzilida sahifalar mavjud.</i>

**Tatar tili yozma matnlar elektron korpusi va Qrim-tatar tilining lingvistik korpusi qiyosi  
(farqli jihatlar)**

№	Tatar tili yozma matnlar elektron korpusi	Qrim-tatar tilining lingvistik korpusi
Farqli jihatlar		
1	Tatar tilidagi yozma korpus, asosan, veb-manbalar materiallari asosida yaratilgan. YOzma korpusga kiritilgan matnlarning aksariyati uchta uslubga tegishli: publitsistika (60%), badiiy adabiyot (35%) va gumanitar sohadagi ilmiy adabiyotlar (5%).	Qrim-tatar tilining lingvistik korpusi XXI asr boshidagi qrim-tatar gazetalarining matnlaridan iborat.
2	Tatar tili yozma matnlar elektron korpusi	Qrim-tatar tilining lingvistik korpusida faqat

	matnlari ko‘chirilishning oldini olish maqsadida alohida gap shaklida saqlanadi.	kirill yozuvidagi matnlar mavjud.
3	Tatar tili elektron korpusining asosiy maqsadi: – tatar tilining leksikasiga oid ilmiy tadqiqotlarga yordam berish; – yozma korpus tilni o‘qitish (lingvodidaktika); – har xil hujjatlarni to‘plash	Qrim-tatar tilining lingvistik korpusida foydalanuvchi klaviaturasida kirill alifbosi bo‘lmasa, ekrandagi virtual klaviaturadan so‘rov oynasiga kirill harflarini kiritish uchun foydalanish mumkin.
4	Tatar tili sintezatori tizimi tatar tilining yozma korpusida topilgan gap va matnlarni eshitish imkonini yaratadi.	Qrim-tatar tilining lingvistik korpusi pastki satrida boshlash va tugatish, sahifadagi matnlarning umumiy va mos keladigan soni ko‘rsatiladi.
5	Tatar tili matn korpusi hozirgi paytda quyidagi vazifalarga echim topmoqda: –nutqni avtomatik aniqlash; – avtomatik tarjima qilish.	Mukammallashtirish davom etmayapti. Oxirgi to‘ldirish 2015-yilda amalga oshirilgan.

Jadvaldagi tahlildan ko‘rinib turibdiki, ushbu ikkala til korpusining o‘xshash tomoni sifatida ularning veb-sahifasining mukammal ishlanganligi, korpus hajmining til materialini o‘rganish uchun qulayligi, so‘zshakl va tokenlar asosidagi qidiruvning mavjudligini ko‘rsatish mumkin.

Tatar tili elektron korpusi hamda Qrim-tatar tili lingvistik korpusining farqli tomonlari, asosan, korpusning imkoniyatlari bilan bog‘liq. Korpuslar orasidagi birinchi farq korpus materiali va uning reprezentativligida. Tatar tili elektron korpusida matnlar turli uslubga mansub, Qrim-tatar tili lingvistik korpusi materiali faqat gazeta matnlari. Shuningdek, nutqni avtomatik aniqlash, avtomatik tarjima imkoniyatlari bilan tatar tili korpusi ustunlik kasb etadi.

**Xulosa** qilib aytadigan bo‘lsak, Tatar tili yozma matnlar va Qrim-tatar tilining lingvistik korpuslari turkiy tilli korpuslar guruhiga mansub bo‘lib, ular ularning bir qator o‘xshash va farqli jihatlari mavjud. Elektron korpuslar bir qator

muammolarni muvaffaqiyatli hal etib, tilni o‘qitish, shuningdek, har xil hujjat va ma’lumotlarni to‘plash uchun boy xazina sifatida ishlatilishi mumkin.

## **2. Turkiy tilli korpus yaratish bo‘yicha amaliy loyiha natijalari xususida**

**Xakas tili elektron korpusi**<sup>93</sup> umumiy, lug‘at, grammatika, matnlar, badiiy matnlar korpusi, dialekt podkorpusi kabi qismlardan tashkil topgan. Xakas tilining elektron korpusi Rossiya Fanlar akademiyasi Prezidiumining “Korpus tilshunosligi” dasturi doirasida “Rossiya xalqlari tillarida korpus manbalarini yaratish va rivojlantirish” yo‘nalishi bo‘yicha yaratilgan. Hozirda Rossiya Gumanitar jamg‘armasi granti, №15-04-12030-sonli grant, “Rossiyadagi turkiy ozchilik tillarin tarkibini avtomatik morfologik va sintaktik tahlil qilish tizimi”, Umumiy fizika institutining dasturi, “Yevrosiyo merosi va uning zamonaviy ma’nolari”, 4-yo‘nalish Tomsk davlat universiteti (Rossiya Federatsiyasi Hukumatining granti) asosida amalga oshiriladigan filologik tadqiqotlardagi multimedial texnologiyalari (“Rossiyaning oz sonli turkiy tillarini rivojlantirish” loyihalari), shuningdek, “Sibir lingvistik va etnik-madaniy xilma-xilligi, sinxroniya va diaxroniyada: tillar va madaniyatlarning o‘zaro ta’siri” loyihasi doirasida bajarilgan.

A.V.Dibo (filologiya fanlari doktori, Rossiya Fanlar akademiyasining muxbir a‘zosi, Rossiya FA Tilshunoslik institutining Ural-Oltoy tillari bo‘limi boshlig‘i) ushbu yo‘nalishdagi muvofiqlashtiruvchilardan biri bo‘lib, “Minoritar turkiy tillar korpusi” loyihasini boshqaradi (hamkor N.N.Shirobokova, falsafa fanlari doktori, professor, Sibir xalqlari tillari kafedrasini mudiri, filologiya instituti).

Rossiya Federatsiyasi hududida turkiy tillar keng tarqalgan. Ushbu tillarning ba’zilari rasmiy maqomga va adabiy an’anaga ega, bu tillarda juda ko‘p material mavjud. Kichik turkiy til va lahjalarda folklor yozuvlari, tadqiqotchilar tomonidan yozilgan dala yozuvlari (shu jumladan, ushbu loyihada ishtirok etganlar) kabi katta hajmda material to‘plangan. Ushbu tillar, ayniqsa, ularning dialektlarining muhim qismi hozirda yo‘q bo‘lib ketish xavfi ostida. Mavjud materiallar hajmi

---

<sup>93</sup> <https://khakas.altai.ru/>

kompyuterlashtirish va unga kiritishni ta'minlash, ya'ni Rossiya turkiy tillarining ochiq korpusini yaratishning dolzarbligi ko'rsatadi. Korpuslarning ochiqligi ushbu tillarni yanada o'rganishni ta'minlaydi, shuningdek, ushbu tillarni saqlash va rivojlantirishga hissa qo'shadi.

Xakas tili korpusi uchun material, asosan, badiiy janrdagi va epik matnlarga parallel (xakas-rus) badiiy matnlar bo'lib, standart formatga tushiriladi. Korpus muharrirlari O.V.Subrakova (Novosibirsk, 2006) tahriri ostidagi 22 ming so'zdan iborat "Buyuk xakas-rus lug'ati"ning raqamli versiyasi tuzilmoqda.

Bundan tashqari, xakas tili dialektal subkorpusini yaratish rejalashtirilgan. Bugungi kunga qadar loyiha xakas dialektlarida (sagay va bel'tir) dala materiallarini qayta ishlash bilan shug'ullanmoqda. Sagay lahjasi (Kazanovka) RDU (2001,2002-y.), RDU va Tillar instituti (2007-y.) ekspeditsiyalari tomonidan to'plangan matnlarni taxminan qayta ishladi. 12 soatlik ovozli matn (transkripsiya, tarjima, qisman nashr etilgan) yozib olingan; bel'tir lahjasida Tillar instituti (2011) ekspeditsiyasi tomonidan to'plangan matnlar, 6 soatlik ovoz (dekodlash, tarjima) qisman qayta ishlangan.

Iqtisodiyot Oliy Maktabining Tillarni birlashtirish xalqaro laboratoriyasining sahifasida Xakas lahjalari ovozli korpus mavjud. Korpus 2015-2019-yillarda loyiha ishtirokchilari tomonidan to'plangan dala materiallari bilan to'ldirish rejalashtirilgan. Matnlarga qo'shimcha ravishda, turli xil anketa, shifrlash bilan birga kelgan ovozli fayllar (100 so'zli, 200 so'zli, tarixiy fonetik so'rovnomalar, morfologik va sintaktik savolnoma) saytga turkiy tillardagi dialektologik materiallar bilan joylashtirilgan.

Korpusning dialekt qismi Rossiya Ilmiy Jamg'armasining 18-18-00501 "Rossiya turkiy tillarining elektron dialektologik atlasini yaratish" granti asosida amalga oshirilgan.

**Tuva tili elektron korpusi**<sup>94</sup> – "Tuva tilining leksik fondi ma'lumotlar bazasini yaratish" loyiha mahsuli. Loyiha maqsadi – tuva matnlari ma'lumotlar bazasi, grammatik shakllar (affiks, analitik konstruksiya) va boshlang'ich

---

<sup>94</sup> <http://www.tuvancorpus.ru/>

jarayonlar (nominal va fe'l) ma'lumotlar bazasini o'z ichiga olgan tuva tili matn korpusini yaratish. Ma'lumotlar bazasi asosida tuva tili baland chastotali leksemalari elektron lug'ati yaratiladi. Tuva tilidagi lingvistik tadqiqotlar uchun material yig'ish, qayta ishlashni avtomatlashtirish uchun kompyuter dasturlarini yaratish va amalga qo'llanadi. Natijalardan biri sifatida ijrochilar tuva tili elektron korpusi uchun veb-sayt yaratish vazifasini qo'yilgan va amaga oshirilgan. Tuva tili leksik fondining elektron ma'lumotlar bazasi ma'lumotnoma-qidiruv tizimi bo'lib, uning yordamida TTEK matnlaridan kerakli semantik ma'lumotlarni qidirish avtomatlashtirilgan; TTEKning semantik teglarini yaratish uchun asos bo'lib xizmat qiladi. Ishlab chiqilgan elektron ma'lumotlar bazasi tuva tilining barcha leksemalari so'zlarning semantik toifalariga (sinflariga) bo'lishiga asoslanadi. Ma'lumotlar bazasi leksemaning semantik va stilistik xususiyatlari haqida ma'lumotni tez olish, ma'lum bir semantik maydonni tashkil etuvchi leksik birliklarning tarkibini aniqlashga imkon beradi. Tuva tilining leksik fondi ma'lumotlar bazasini rivojlantirishdagi yana bir muhim va ahamiyatli vazifa leksemalarning leksik muvofiqligini aniqlashdir. Avtomatlashtirilgan tizim leksemalarning semantik ruxsat etilgan birikmalarini (ot va ot, sifat va ot, olmosh va ot, atama va fe'l, ot va fe'l) o'z ichiga oladi. Avtomatlashtirilgan tizim semantik jihatdan yaroqsiz kombinatsiyani chiqarib tashlaydi. Semantikani hisobga olgan holda turli xil foydalanuvchi so'rovlarini yaratish, ma'lum leksik birliklarining mosligi qoidalarini aniqlash imkonini beradi. Tuva tili leksik fondining elektron ma'lumotlar bazasi leksik semantika, shuningdek, leksikografiya sohasida ilmiy-tadqiqot ishlarida yangi imkoniyatlar ochmoqda. Bu kerakli ma'lumotlarni qidirishni tezlashtiradi, tadqiqotchining vaqtini tejaydi va foydalanuvchilar safini kengaytiradi; turli lug'atlarni yaratishda ishni soddalashtiradi. Tuva tilining turli lug'atlari: tematik, izohli, sinonim va antonimlar lug'ati, chastotali boshqa turdagi lug'atlarni tuzishda leksik fond ma'lumotlar bazasidan foydalaniladi.

Ma'lumotlar bazasini yaratishda quyidagilar qo'llaniladi: tuva-rus lug'ati (20 000 so'z, 1955), tuvacha-ruscha lug'at (22 000 ga yaqin so'z, 1968, qayta nashri 2014), ikki tilli tuvacha-ruscha lug'at (32 000 so'z, 1980), Tuva tilining



izohli lug‘ati (1-2-jildlari, 2003, 2011), toponimik lug‘at (B.K.Ondar, 2004), Tuva va G‘arbiy Mo‘g‘uliston qush nomlarining olti tilli lug‘ati (1999), Tuva respublikasi o‘simliklari ko‘rsatkichi (2007).

Loyihaning maqsadi – tuva tilining elektron korpusini yaratish, tuva tilshunosligi sohasidagi ilmiy tadqiqot va amaliy ishlanmalarni keyinchalik har tomonlama avtomatlashtirish uchun mo‘ljallangan. Yechilishi kerak bo‘lgan asosiy vazifalar quyidagilardan iborat:

1. Tuva tilining zamonaviy matnlari ma’lumotlar bazasi mahalliy versiyada (Microsoft Access formati) va ma’lumotlar bazasini boshqarish tizimlari;
2. Tuva tili sohasidagi lingvistik tadqiqotlar uchun ma’lumot to‘plashni avtomatlashtirish uchun kompyuter dasturlari (statistik usul);
3. Chastota leksemalari, birlamchi ot va fe‘l turkumi ma’lumotlari va ma’lumotlar bazasini boshqarish tizimlari (birlamchi va chastotali elektron lug‘atlar);
4. Tuva tili matnlari elektron korpusi sayti.

Elektron korpuslarni yaratish tajribasi ba’zi turkiy tillarda mavjud. Germaniya ilmiy-tadqiqot jamiyati (NNIO) va Rossiya fundamental tadqiqotlar fondi (RFFI) tomonidan qo‘llab-quvvatlanadigan «Shorika» loyihasi 1999-2001-yillarda amalga oshirilgan. Xalqaro olimlar guruhining faoliyati natijasida shor tilidagi matnlarning elektron korpusi Shoebox dasturining to‘rtinchi versiyasidan foydalangan holda yaratilgan bo‘lib, u dunyo xalqlarining yozilmagan va erta yozilgan tillarining adabiy shaklini rivojlantirishga ko‘maklashish maqsadida ishlab chiqilgan. Ushbu muammo tuva tilida maxsus o‘rganish obyekti bo‘lmagan. Amalga oshirilgan loyihaning qo‘shimcha imkoniyatlari sifatida masalaga qiziqqan keng doiradagi mutaxassisning erkin (bepul) kirishni ta’minlaydigan audio va vizual fondlarni ko‘rsatish mumkin.

*Loyihada hal etiladigan ilmiy muammolar.* Tilning har qanday qatlamini o‘rganish, shu jumladan, ko‘p jildli lug‘atlarni tuzish, grammatik tadqiqotlar loyiha doirasida hal qilinadi. Tilshunos keng matnlar massivi bilan ishlaydi. Bu nafaqat intellektual kuchni, balki ko‘p vaqtni, ayniqsa, tayyorgarlik bosqichlarida talab etiladigan juda mashaqqatli jarayonni talab qiladi. Shuning uchun material

yig'ishning yangi usullariga o'tish, uni tahlil qilish va lingvistik manbalarning yangi shakllari zarur, bu nafaqat mehnat unumdorligini sezirarli darajada oshirishga, balki tuva tilini o'rganishning yangi usullari va sohalariga yo'l ochib beradi. Ushbu muammoning tahlili tilning elektron korpusini yaratish, uni hal qilishda muhim ahamiyatga ega ekanligini ko'rsatadi.

**Sibir kichik xalqlari tillari korpusini yaratish** bo'yicha loyiha 2011-yildan beri amalga oshirilmogda. 2011-2019-yillarda u RFA Prezidiumining "Korpus tilshunosligi" dasturi doirasida Rossiya Fanlar Akademiyasi Etnologiya va Antropologiya institutining SHimoliy va Sibir bo'limi xodimlari va aspirantlari tomonidan amalga oshirildi. Tomsk davlat universiteti bazasida Moskva davlat universiteti etnologiyasi kafedrasida xodimlari tomonidan 2017-yildan boshlab "Janubiy Sibirning sinxron va diaxron etno-madaniy xilma-xilligi: til va madaniyatning o'zaro ta'siri" loyihasi asosida amalga oshirilgan.

Loyiha maqsadi – folklor materiallarining ommaviy elektron korpusini quyidagi imkoniyatlar asosida yaratish:

1. Matnlarning asl nusxasi (fiksatsiya xususiyatlarini saqlagan holda), shu jumladan, qo'lyozma, audio va video yozuvlar hamda matn yozuvining normallashtirilgan versiyasini bir vaqtda saqlash.
2. Har bir so'zning chap va o'ng qatordoshiga (grafika shaklida qo'shimcha izoh bilan) muhim statistik ma'lumot tanlash, leksemalar/so'zshakldan foydalanish bo'yicha statistik ma'lumotni to'plash, berilgan matn va butun korpusda so'zning ishlatish kontekstini tahlil qilish, matnni korpus vositasida o'rganish. Ushbu qidiruv korpusda keltirilgan barcha matnlardan so'zshakllar ro'yxatlarini – matn, hikoya yoki til, lahja, sheva va tillar bo'yicha taqqoslash.

Yaratilgan noyob manbadan maktab, universitet, mahalliy tillarni o'qitish, shuningdek, ushbu tillar tadqiqotchilari: tilshunos, folklor shunos va etnologlar ilmiy-tadqiqot ishlarida foydalanishi mumkin.

**Korpusning tarkibi va hajmi.** 2011-yilda korpusga folklor materiallari faqat ikki tilda – shor va evenkiys tillarida kiritilishi kerak edi. 2011-yil oxiriga kelib, shor va evenkiys subkorpuslarining hajmi taxminan 120.000 va 6000

tokenlarga etkazilishi rejalashtirilgan. Korpusning boshqa tilda material qo‘shish imkoniyatidan foydalanganib qo‘shimcha ravishda teleut tilida folklor matnlari qo‘shilgan. Hozirgi vaqtda korpus platformasi nenets, teleut, shor, evenk tili korpuslarini ham qamrab olgan. Har bir korpus bo‘yicha joriy statistikasi “Statistika” sahifasida ko‘rish mumkin.

Ushbu loyihalar yakunlangan va tuva<sup>95</sup>, xakas<sup>96</sup>, teleut, shor<sup>97</sup>, yoqut<sup>98</sup> tillari korpusi ishga tushirilgan.

### **“Rossiyaning minoritar turkiy tillari korpuslari uchun avtomatik morfologik va sintaktik tahlil tizimi” loyihasi haqida.**

Loyiha oldiga ozchilik turkiy til va lahjalari matnlariga avtomatik ishlov berish vazifasi qo‘yilgan. Avtomatik morfologik analizator; avtomatik tahlil qilish; avtomatik semantik teglash; belgilangan matnlarni ko‘p faktorli qidirish tizimini yaratish shular jumlasidan. Turkiy tillar uchun mavjud avtomatik morfologik teggerlar agglyutinativ tizim tillarining o‘ziga xos xususiyatlarini hisobga olmaydi, flektiv tillar teggerlari prinsiplardan foydalanadi. Ular har bir so‘z uchun to‘liq paradigmani yaratishga, matnda uchragan so‘z shaklini ushbu paradigma bilan taqqoslashga harakat qiladi. Turkiy tillar uchun bu usul agglyutinativ tilda paradigma so‘zining haddan tashqari murakkabligi va ko‘p qirrali ekanligi sababli maqbul emas, bu ayniqsa, so‘z o‘zgartiruvchi affikslarning ko‘pligi bilan bog‘liq. Mavjud turkiy parserlar qisqartirilgan paradigma hosil qiladi, so‘zshakli sonini uch yuzgacha cheklaydi, bu esa analizator ish sifatiga salbiy ta‘sir qiladi. Bu morfologik analizatorni morfemalarning barcha mavjud birikmalarini hisobga olgan holda maxsus oltoy tipidagi tillar uchun yaratishga majbur qiladi.

Turkiy so‘zshakllarining kompyuter modelini yaratish uchun an’anaviy ravishda agglyutinativ tillarni tavsiflash va “formal grammatikasi”dan foydalaniladi. Ushbu tavsif agglyutinativ tillarni tavsiflash uchun qulay vosita, ularning morfologiyasi quyidagi talablarga javob beradi:

- a) so‘z o‘zgartiruvchi affikslarning qat’iy ketma-ketligi;

---

<sup>95</sup> <http://www.tuvancorpus.ru/>

<sup>96</sup> <http://khakas.altaica.ru/>

<sup>97</sup> <http://corpora.iea.ras.ru/corpora/>

<sup>98</sup> <http://adictsakha.nsu.ru/corpora/corp>

b) ularning grammatik o‘ziga xosligi;

v) ma’lum bir grammatik kategoriya affiksining ma’lum bir so‘z shaklida bir marotaba paydo bo‘lishi.

Morfologik analizatorni qurishda uchta asosiy lingvistik tarkibiy qismlar ishtirok etadi:

1) fonologik qoidalar bilan tavsiflanmagan va alternaturalarni o‘z ichiga olgan lug‘ati;

2) tilni avtomatik tahlil qilishga qaratilgan etarli grammatik tavsifga asoslangan so‘zshakllari kompyuter modeli;

3) so‘zshakldagi affiksga mos kelish qoidalari va ma’lum bir affiksning allomorflarini tanlash uchun fonetik qoidalar to‘plami.

F.Krilov tomonidan ishlab chiqilgan tahlil algoritmi yuqoridagi pozitsiyalarni istalgan turkiy tillardagi materiallar bilan to‘ldirishga imkon beradi.

**Olingan natijalar izohi**<sup>99</sup>. 2015-2016-yillarda “Rossiyadagi minoritar turkiy tillar uchun avtomatik morfologik va sintaktik tahlil tizimi” loyihasi quyidagi asosiy yo‘nalishlar bo‘yicha ish olib bordi:

1. Dasturlar uchun amaliy dasturlar va interfeyslarni ishlab chiqish;
2. Turkiy tillarning kompyuterga asoslangan grammatikasini rivojlantirish;
3. Belgilangan turkiy tillarning elektron lug‘atlarini ishlab chiqish;
4. Rossiyadagi minoritar turkiy tillarning elektron ishlarining matnlari va boshqa materiallarini to‘ldirish.

Ko‘rsatilgan yo‘nalishlarda quyidagi ishlar amalga oshirildi:

1.1. Avtomatik morfologik analizator uchun so‘z o‘zgartiruvchi affiks jadvalining yangi versiyasi kodlangan 4 til modullarida affiks jadvalini morfologik qayd etish bo‘yicha yordam ko‘rsatildi; xakass tili uchun affikslarning muvofiqligi bo‘yicha sinov qilingan va belgilangan qoidalar, saytda amalga oshirilayotgan o‘zgarishlarni aks ettiruvchi texnik hujjatlar saqlandi.

---

<sup>99</sup> <http://www.tuvancorpus.ru/?q=content/o-proekte>

1.2. Mavjud xakas matnlari asosida parallel matnlarni yaratish va ularni MS VBA (MS Visual Basic for Applications) yordamida tarjima qilish mexanizmi ishlab chiqildi. 4 parametrli matnni grammatik analizator bilan ELAN (<https://tla.mpi.nl/tools/tla-tools/elan/>) formatida qayta ishlash natijalarini saqlab qolish uchun S++ tilida dasturiy modul yaratilib, ushbu dasturdan keyingi ishlov berish uchun foydalanildi.

1.3. Yangi soʻzlarning (korpusdan olingan) grammatik belgilariga ega elektron lugʻatga qoʻshish mexanizmining sinov versiyasi ishlab chiqildi;

1.4. Turkiy tillar uchun sintaktik shajara daraxti mexanizmini yaratish boʻyicha tayyorgarlik ishlari olib borildi: asosiy morfologik (nominal va feʼl) koʻrsatkichlar toʻplami aniqlandi, ularga bevosita tarkibiy qismlarning qoidalari asoslanadi.

1.5. Boshqird korpusi uchun avtomatik tahlil qiluvchi sinov versiyasi ishlab chiqildi.

1.6. Sintaktik matnlarni qidirish mexanizmi ishlab chiqildi. Qidiruv bajarilishi kerak boʻlgan parametrlar toʻplami aniqlandi va bir qator grammatik va boshqa parametrlar uchun qidiruv oynasi interfeysi ishlab chiqildi. Qidiruv avtomatik tahlil qiluvchi va qoʻlda tahlil qilingan matnlar yordamida amalga oshiriladi. Matn qidirish mexanizmini sinab koʻrish parametrlar toʻplamini moslashtirishga imkon beradi.

1.7. Belgilangan matnlar uchun koʻp darajali belgi qoʻyilgan omborning arxitekturasi ishlab chiqilgan.

1.8. Sintaktik filtr yordamida morfologik tahlilni avtomatik tahrirlash mexanizmi ishlab chiqildi. Sintaktik filtrning sinov versiyasi (1) grammatika sinfini va (2) umumlashtirilgan morfologik yorliqni belgilab, matnni atom va atom boʻlmagan qismlarga boʻlish asosida tuzilgan. Ushbu ikkala xususiyat ham morfologik belgilarga asoslangan.

2.1. Kompyuterga yoʻnaltirilgan yoqut, oltoy, tuva, chuvash grammatikalarining sinov versiyalari yaratildi. Har bir grammatika uchun soʻz oʻzgartiruvchi affikslari va grammatikasi toʻplamlari ajratilgan, soʻz shaklining

ketma-ket modellari tuzilgan va soʻz oʻzgartiruvchi morfologiya uchun tegishli boʻlgan morfologik qoidalar ishlab chiqilgan.

3.1. Xakas tilining soʻz birikmalarida ish davom ettirildi: soʻzlarni shakllantirish va semantik belgilash ishlari olib borildi. Soʻzlarni shakllantirish va semantik teglar roʻyxati yaratildi; lugʻatda keltirilgan soʻzlarni shakllantirish tahliliga duch kelgan barcha birliklar tahlil qilindi (12 mingdan ortiq lugʻat yozuvlarida “soʻz hosil qilish maydoni” toʻldirilgan); xakas bazasining ogʻzaki va predikativ tokenlari boshqaruv modellari bilan jihozlangan. Ularning 500 tasi uchun ramka tavsiflari ishlab chiqilgan. Xakas tilining elektron maʼlumotlar bazasining 1600 dan ortiq maqolalarida semantik teglar joylashtirilgan. Baza adabiy xakas matnlaridan, shuningdek, xakas tilining dialekt subkorpusidan leksik birliklar bilan toʻldirildi.

3.2. Boshqird tilining belgilangan elektron grammatik lugʻati ishlab chiqilgan.

3.3. Yoqut tilining muhim elektron grammatik lugʻati P.A.Slepsova tahririyati ostida ishlangan. StarLing maʼlumotlar bazasini boshqarish tizimining texnologiyalaridan foydalanib, lugʻatning tahrirlangan elektron versiyasi maʼlumotlar bazasiga kiritildi. Hozirgi vaqtda lugʻat 23608 yozuvdan iborat baza shaklida mavjud. Yakut tilining elektron lugʻatining morfologik izohi yakunlanmoqda.

3.4. Oltoy tilining muhim elektron lugʻati N.A.Baskakov, T.M.Тошакоеvalarning “Oyrot-rus lugʻati” asosida yaratilgan (10000 lugʻat yozuvi). Lugʻat toʻliq rejimda ishlaydigan va ingliz tilidagi tarjimalari bilan jihozlangan Toolbox formatida tuzilgan.

3.5. Belgilangan tuva tilining elektron lugʻati E.R.Tenishevaning “Tuva-rus lugʻati” asosida (22000 lugʻat yozuvi) ishlab chiqildi. Lugʻat faylining shrift belgilarini tekshirish va tahrirlash ishlab chiqildi. StarLing maʼlumotlar bazasini boshqarish tizimining texnologiyalaridan foydalanib, lugʻatdagi matnli fayllar maʼlumotlar bazasiga kiritildi. Tuva tilidagi nutqning nominal qismlarining morfologik koʻrsatkichlari boʻyicha maʼlumotlar bazasi tayyorlangan.

3.6. M.I.Skvorsovning “CHuvash-rus tillari lug‘ati” (M., “Rus tili”, 1982. 40000 lug‘at yozuvi)ga asoslangan chuvash tilining maxsus elektron lug‘ati ishlab chiqilgan.

3.7. Xakas elektron lug‘atini shorga aylantirishda konversiya quyidagi bosqichlarni o‘z ichiga oldi:

1) avtomatik ishlov shor harflarida farq qiladigan harflarni avtomatik ravishda almashtirishdan iborat;

2) etimologik belgilar asosida belgilarning yarim avtomatik almashtirilishi;

3) shor korpusining dastlabki versiyalari bo‘yicha lug‘atni sinovdan o‘tkazish, qisman nashr etilgan.

Hozirgi vaqtda shor korpusi razmetkasi asosida xakas, oltoy, tuva tillarining matnlari annotatsiyalangan<sup>100</sup>.

4.1. 2007 yildagi dala materiallari qayta ishlandi (xakass tilidagi sagay lahjasi, Kazanovka qishlog‘i lahjasi).

4.2. Xakas tili shor lahjasi dala materiallari qayta ishlandi: fonetik kesish (5100 fayl), semantik anketalarni dekodlash (3 soat 48 daqiqa, 3270 Mb).

**Xulosa** o‘rnida aytish mumkinki, turkiy tilli korpus yaratish bo‘yicha amaliy loyiha natijalari salmoqli bo‘lib, Xakas, Tuva, Sibir kichik xalqlari tillari elektron korpusi yaratilgan va hozirgi kunda mukammallashtirilib foydalanilmoqda. Xakas tili elektron korpusi umumiy, lug‘at, grammatika, matnlar, badiiy matnlar korpusi, dialekt podkorpusi kabi qismlardan tashkil topgan bo‘lsa, Tuva tili korpusi matnlar ma’lumotlar bazasi, grammatik shakllar va boshlang‘ich jarayonlar ma’lumotlar bazasini o‘z ichiga olgan. Sibir kichik xalqlari tillari elektron korpusi folklor materiallarining ommaviy elektron korpusini yaratishni o‘z oldiga maqsad qilib qo‘ygan hamda hozirgi vaqtda korpus platformasi nenets, teleut, shor, evenk tili korpuslarini ham qamrab olgani bilan ahamiyatlidir. Ko‘rinadiki, yaratilgan amaliy loyihalar qamrovi o‘xshash va farqli xususiyatlarga ega bo‘lib, bular tilning har qanday qatlamini o‘rganishni, til o‘qitishni va eng asosiysi tillarning “yashab qolishini” ta’minlaydi.

---

<sup>100</sup> <http://dialects.altai.ru/>; <http://khakas.altai.ru/>

### **3. Turkiy tillarda kompyuter ishlovini yo‘lga qo‘yish va korpus lingvistikasi masalalari bo‘yicha tashkil etilgan konferensiya, seminar va portal haqida**

1. “Turkiy korpus tilshunosligi va kompyuter leksikografiyasi” seminari haqida. 2015 yil IPGIT xalqaro dasturi “Turk korpusi tilshunosligi va kompyuter leksikografiyasi” seminari tashkil etilgan. Asosiy masalalar qatorida quyidagilarni muhokama qilingan:

1. Elektron lug‘atlar va kompyuter leksikografiyasi
2. Turkiy korpus tilshunosligi
3. Qrim-tatar leksikografiyasi: muammolar va istiqbollar
4. Korpus tilshunosligi va kompyuter leksikografiyasidan foydalangan holda, qrim-tatar tilshunosligining dolzarb muammolarini hal qilish yo‘llari.

Seminarda turkiy korpus tilshunosligi sohasidagi taniqli mutaxassislar, kompyuter leksikografiyasi va matematik lingvistika mutaxassislari ishtirok etgan.

Dibo Anna Vladimirovna – filologiya fanlari doktori, professor, Rossiya Fanlar akademiyasining muxbir a‘zosi, bosh ilmiy xodim, RAS institutining Ural-Oltoy tillari kafedrasini mudiri.

Zakiev Mirfatix Zakiyevich – filologiya fanlari doktori, professor, Tatariston Respublikasi FA haqiqiy a‘zosi, Leksikologiya va dialektologiya kafedrasini mudiri.

Shirokov Vladimir Anatolyevich – fizika-matematika fanlari doktori, professor, Ukraina NAS akademigi, Ukraina NAS Til va axborot jamg‘armasi direktori, Kiyev.

Krilov Sergey Aleksandrovich – filologiya fanlari doktori, Sharqshunoslik institutining yetakchi ilmiy xodimi (Osiyo va Afrika tillari bo‘limi)

Agranat Tatyana Borisovna – filologiya fanlari doktori, dotsent, guruh rahbari, RAS institutining yetakchi ilmiy xodimi.

Yuqorida nomlari keltirilgan mutaxassislar kompyuter va korpus lingvistikasi mutaxassislari sanaladi.

2. **“Kompyuter va kognitiv tilshunoslikka bag‘ishlangan XV xalqaro konferensiya”**. Anjuman yo‘nalishlari:

1. Turkiy tillar uchun kompyuter lingvistikasi.



2. Kompyuter lingvistik manbalari.
3. Hujjatlarning kompyuter tahlili (tasniflash, qidirish, tonallikni tahlil qilish va boshqalar).
4. Korpus lingvistikasi (korpus tuzish, razmetkalash, qo‘llash usullari va baholash).
5. Kognitiv modellar va ularning kompyuter lingvistikasida qo‘llanilishi.
6. Matn va nutqning avtomatik tarjimasi.
7. Matnni semantik tahlil qilish modellari va usullari.
8. Nazariy va kompyuter leksikografiyasi.
9. Tipologiya va kompyuter tilshunosligi.
10. Formal til modellari va ularning kompyuter lingvistikasida qo‘llanilishi.

**3. Kompyuter lingvistikasi bo‘yicha “TurkLang – Turkiy tillarni kompyuterli qayta ishlovi” doimiy Xalqaro konferensiyasi** 2013 yildan bugungi kungacha har yil turkiy tilli davlatlar (Turkiy, Qozog‘iston, Tatariston, O‘zbekiston, Qirg‘iziston, Qrim, Boshqirdiston)da o‘tkazib kelinmoqda. Konferensiya yo‘nalishlari:

1. Turkiy tillarning formal modellari.
2. Kompyuter tizimi va terminologiyaning milliy lokallashuvi.
3. Korpus lingvistikasi.
4. Matnni qayta ishlovchi ilova.
5. Nutq texnologiyasi.
6. Mashina tarjimasi.
7. Turkiy tillarni o‘rganishda intellect tizimlari va texnologiyalari.
8. Turkiy tillarning lingvistik ontologiyalari.

**4. An’anaviy UniTurk seminari** haqida. Anjuman doirasida **an’anaviy UniTurk seminari** tashkil etiladi. “Uniturk”ning oldingi seminarlarida (Qozon, 2014; Istambul, 2014; Qozon, 2015; Bishkek, 2016) korpuslarda va matnni avtomatik qayta ishlash tizimlarida foydalanish uchun turkiy tillardagi matnlarning yagona morfologik teglarni ishlab chiqish muammolari muhokama qilingan. Belgilangan teglash tizimi, shuningdek, matnli misollarni teglashning universal vositasi bo‘lib xizmat qilishi mumkin. 2014-yil fevral oyida Qozon shahrida

birlashtirilgan morfologik tegning ishchi versiyasi qabul qilindi. Qabul qilingan versiyada turkcha soʻz shaklining morfemik tuzilishi hisobga olingan va turli xil turk tillarining xilma-xilligi toʻliq aks ettirilgan. “UniTurk”ning navbatdagi seminarida muhokama qilish uchun quyidagi savollar taklif etiladi:

1. Grammatik (morfologik) va semantik teglar. Semantik va grammatik teglar oʻrtasidagi farq (raqam kategoriyalari, ovoz va boshqalar), polifunksional affikslarning ifodalanishi.
2. Annotatsiyalashning toʻliqligi va batafsilligi darajasi masalasi.
3. Ayrim toifalarning turli tillarda sintetik va analitik ifoda etish muammosi (masalan, “instrumental kelishik” deb ataladigan narsa, soʻroqni ifodalash usullari, modallik va boshqalar).
4. Feʼl-affikslarning prezentatsiyasi: taqqoslash jadvallarini tayyorlash.

#### **“Turklang” konferensiyasi toʻplamlari:**

1. Компьютерная обработка тюркских языков. Первая международная конференция: Труды. – Астана: ЕНУ им. Л.Н. Гумилева, 2013. – 340 с.
2. Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В2-х томах. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 380 с.
3. Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент: Издательско-полиграфический дом «NAVOIY UNIVERSITETI», 2018. – 390 с.

#### **“TURKLANG” portali haqida**

**“TurkLang” portali** ushbu konferensiyaning sayti sanaladi. Ammo sayt faqat konferensiya yangiliklari emas, balki turkiy tillar kopmyuter lingvistikasi bilan bogʻliq barcha yangiliklarni qamrab olishga moʻljallangan.

**Portal strukturasi:** maqolalar toʻplamlari, rrezentatsiyalar, dasturlar, kitoblardan iborat.

Portalda turkiy tillar uchun quyidagi **elektron manbalar** mavjud: elektron korpuslar, morfologik analizatorlar, tarjima tizimlari, tatar xalq shevalarining atlasining elektron versiyasi.

### **“Turk morfemasi” portali**

Turkologiya va agglyutinativ tillarning tipologiyasi sohasida ilmiy tadqiqotlar o‘tkazish uchun ko‘rib chiqilayotgan tillarning tarkibiy va funksional xususiyatlarini hisobga oladigan dasturiy vositalar kerak.

“Turk morfemasi” portali turk tillarining ushbu xususiyatlarini inobatga olgan holda tadqiqot olib borishga imkon beradigan va hisoblash tilshunosligi. Lingvistik tipologiya sohasidagi tadqiqot faoliyatiga qo‘yiladigan talablarga javob beradigan vositadir.

Ushbu portal turkiy morfemaning tarkibiy-parametrik funksional modeli asosida yaratilgan va turkiy tillarning turli xil lingvistik darajalarida: morfologik, sintaktik, semantik tillarni tavsiflovchi maxsus lingvistik ma’lumotlar bazalarini o‘z ichiga oladi.

Portalning ma’lumotlar bazasidan turkiy tillar uchun ma’lumot tizimi sifatida o‘quv jarayonida ham foydalanish mumkin.

O‘zbek tilshunosligida kompyuter va korpus lingvistikasi yosh soha bo‘lsa-da, bir necha tadqiqotlar amalga oshirilgan, monografiya, qo‘llanma va lug‘atlar nashr qilingan.

1. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) диссер. – Қарши, 2018. – 250 б.
2. Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол. фан. бўйича фалсафа доктори (PhD) диссер. – Қарши, 2019. – 140 б.
3. Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти. Монография. – Тошкент, 2018. – 152 б.
4. Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар

тахрири дастури учун): Филол. фан. бўйича фалсафа доктори (PhD) диссер. – Фарғона, 2019. – 164 б.

5. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.

#### **Tavsiya etilgan adabiyotlar:**

1. Галиева А.М. Возможности использования татарского национального корпуса “туган тел” в научно-исследовательской работе студентов-филологов // <https://dspace.kpfu.ru/>

2. Ибрагимов Т.И., Сайхунов М.Р. Письменный корпус татарского языка: идеи, проблемы, решения // <http://docplayer.ru/>

3. Дыбо А.В., Мудрак О.А. Об историческом принципе при составлении фонетической анкеты для полевого обследования диалекта // III Международная конференция по полевой лингвистике. Тезисы и материалы. М., 2009.

4. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011.

5. Функ Д.А., Мамонтова Н.А., Шаховцов К.Г. Электронный корпус фольклорных текстов на языках малочисленных народов Сибири (на материалах шорского, телеутского и эвенкийского языков): принципы создания и структура // <https://istina.msu.ru/publications/article/5522177/>

6. <http://www.turklang.net/ru/>

7. <https://khakas.altaica.ru/>

8. <https://korpus.sk/QIRIM/#id9>

9. <https://ru.wikipedia.org/wiki>

## PARALLEL KORPUSLAR

### Reja:

1. Parallel korpusning tuzilishi va interfeysi.
2. Parallel korpusning lingvistik, lingvodidaktik, tipologiya va tarjima nazariyasi ta'limidagi ahamiyati.
3. Parallellik xususiyati, tillarning morfologik mutanosiblik darajasi.

*Tayanch so'z va iboralar: parallel korpus, interfeys, lingvodidaktika, tarjima nazariyasi, ta'lim, tipologiya, parallellik xususiyati, morfologik xususiyat.*

Korpus materiali necha tilda berilishiga ko'ra bir va ko'p tilli deb tasniflanadi<sup>101</sup>. Korpus mutaxassisi (asosan, tarjimon)ni doim bir necha tilli korpus yaratish qiziqtirib kelgan. Ilk avlod korpusi davridayoq ingliz, fin, fransuz, nemis, grek, norveg, ispan, shved va boshqa tillar uchun ikki tilli korpus paydo bo'la boshladi. Bunday korpus bitexts deb ham ataladi. Korpusni ikki, uch, to'rt, undan ortiq tilli qilishga hech qanday to'siq yo'q. Sohaning boshqa bir mutaxassislari esa parallellik nuqtai nazaridan korpusni bir, ikki, ko'p tilli<sup>102</sup> kabi turlarga bo'lishadi. Bir tilli korpusda til varianti, shevalar bir-biriga qarama-qarshi qo'yiladi. Masalan, ingliz tili ona tili va xorijiy til sifatida alohida o'rganiladi. Yangi texnologiyalar paydo bo'lguncha ingliz tili xorijiy til sifatida o'rganilmay qolganligi haqida fikrlar mavjud. Ikki, ko'p tilli korpus bir mavzu doirasida turli tilda yozilgan matndan tashkil topadi. Masalan, ma'lum ilmiy muammo borasida bir necha davlatlarda turli tilda o'tkazilgan konferensiya materiallarini qamrab olishi mumkin. Bunday korpuslar, odatda, atama bilan ishlaydi hamda tarjimonlar tomonidan foydalaniladi. Ikki, ko'p tilli korpusning yana bir ko'rinishida original matn hamda tarjima matni mavjud bo'ladi. Korpusning ushbu turi qiyosiy chog'ishtirma tadqiqot olib borishda, tarjima nazariyasi hamda kompyuter tarjimasini o'rganishda muhim, foydali manba vazifasini bajaradi. Ikki tilli korpusning 2 turi mavjud<sup>103</sup>:

---

<sup>101</sup> *Қаранг:* Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

<sup>102</sup> *Қаранг:* Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – Б.20.

<sup>103</sup> *Қаранг:* Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf

- 1) bir-birining tarjimasi bo‘lgan matnli korpus;
- 2) bir mavzuga oid ikki tildagi matnli korpus.

Birinchi tipdagi korpus “parallel korpus” (parallel corpora) deb nomlanib, ma’lum bir tarjimaning turli aspektini o‘rganish uchun qo‘llaniladi. Bunday korpuslar sirasiga Kanada parlamenti yig‘ini (ingliz/fransuz) matnlari korpusini kiritish mumkin. Parallel korpus o‘z navbatida yana 2 turga bo‘linadi. Bu moslangan (aligned), moslanmagan (not aligned) korpusdir. “Moslangan” atamasi korpusda tarjima birliklari orasida bir-birini taqozo etuvchi aniq aloqa mavjudligini bildiradi. Bunday korpusdan u yoki bu gapning qanday tarjima qilinganini darrov topamiz. Bu turdagi korpus tarjimon uchun ahamiyatli, chunki unda noyob resurs – “tarjima xotira”si (translation memory) mavjud.

Moslanmagan korpus “qiyosiy” korpus deb ham ataladi. Matnni uning tarjimasi bilan to‘g‘rilash bu uning tarjimada qaysi birlikka to‘g‘ri kelishini ko‘rsatib turishdan iborat. Moslash avtomatik ravishda yoki qo‘lda bajarilishi mumkin. Birinchi usul oson, lekin xatolari ko‘p. Masalan, tarjima jarayonida sodda gap qo‘shma gap shaklida berilishi mumkin. Bunday paytda qaysi qurilish original ekanligini aniqlash qiyin bo‘ladi. Ko‘p tilli “moslangan” korpus namunasi sifatida Evropa Ittifoqining Acquis Communautaire ma’lumotlar bazasini keltirishimiz mumkin. Parallel korpusning qimmatini uning hajmi, tillarning miqdori bilan belgilanadi. Acquis Communautaire – dunyodagi eng katta parallel korpus. Uning ikki muhim jihati korpusdan foydalanishning bepulligi, malti-eston, sloven-fin kabi kam uchraydigan tillar juftligining mavjudligi bilan baholanadi. Ushbu korpuslardan quyidagi maqsadlarda foydalanish mumkin:

- 1) tipik tarjima usullari, transformatsiyani yuzaga keltirish;
- 2) avtomatik tarjima tizimi statistikasini o‘rganish;
- 3) bir va ko‘p tilli lug‘atlar yaratish;
- 4) ma’lumotni saqlash, uzatish dasturlarini o‘rganish, baholash;
- 5) tarjimaning to‘g‘riligini avtomatik tarzda tekshirish;
- 6) ekvivalent tanlash imkoni kengligi orqali tarjimon mehnatini osonlashtirish.

Ikkinchi xildagi korpus “tarjima korpuslari” (translation corpora) deb atalib, ayni bir fikrning turli tildagi ifodasini o‘rganish uchun muhim. Bundan tashqari, ikki tilli korpus tilga oid bir qancha kuzatuv natijasini (tarjima nazariyasi tadqiqi yoki tarjima mahsulini baholashi) bera olganligi sababli tilshunoslik yo‘nalishini tanlagan talaba uchun malakaviy bitiruv ishini bajarishning juda foydali vositasi hamdir.

Ikkinchi xildagi korpus “tarjima korpuslari” (translation corpora) deb atalib, ayni bir fikrning turli tildagi ifodasini o‘rganish uchun muhim. Bundan tashqari, ikki tilli korpus tilga oid bir qancha kuzatuv natijasini (tarjima nazariyasi tadqiqi yoki tarjima mahsulini baholashi) bera olganligi sababli tilshunoslik yo‘nalishini tanlagan talaba uchun malakaviy bitiruv ishini bajarishning juda foydali vositasi hamdir.

### **Parallel korpusning umumiy xarakteristikasi, tuzilishi va tarkibi**

*Parallel korpusning umumiy xarakteristikasi.* Parallel korpus (Parallel Corpora) – parallel tarjima matnlarining elektron analogi; ko‘plab “original matn va ularning bir/bir necha tarjimasi” bloklaridan iborat. Korpusdagi elektron matnlar original matnning o‘zi yoki uning bir qismi bo‘lishi mumkin<sup>104</sup>. Parallel korpus – original matn va uning tarjimasi yig‘ilgan korpus<sup>105</sup>.

Parallel korpusning tuzilishi, tarkibi va imkoniyati haqidagi ma’lumotlarni D.O.Dobrovolskiy<sup>106</sup>, Yu.Tao, V.Zaxarov<sup>107</sup>, A.A.Kokoreva<sup>108</sup>, E.P.Sosnina<sup>109</sup> larning ishlarida kuzatamiz.

D.Dobrovolskiy parallel korpusga shunday ta’rif beradi: “Parallel korpus – asl va tarjimadagi elektron matnlar yig‘indisidan iborat korpus. Asl va tarjima matnlar shunchaki yonma-yon qo‘yilmaydi, balki bu ikki matndagi gap(sistaktik

<sup>104</sup> Sosnina E.П. Параллельные корпуса в обучении языку и переводу. – (Электрон ресурс): [http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus\\_education\\_translation/](http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_education_translation/)

<sup>105</sup> <https://translationrating.ru/glossary/parallelnyj-korpus/>

<sup>106</sup> Добровольский Д. Какие задачи решают параллельные корпуса? (Электрон ресурс): <https://postnauka.ru/video/>;

Добровольский Д.О. Корпус параллельных текстов в исследовании лексической семантики. (Электрон ресурс): <http://www.dialog-21.ru/media/2503/dobrovolskij.pdf>

<sup>107</sup> Тао Ю., Захаров В.П. Разработка и использование параллельного корпуса русского и китайского языков // Автоматизация обработки текста. (Электрон ресурс): <https://www.researchgate.net/publication/306911853>

<sup>108</sup> Кокорева А.А. Корпус параллельных текстов в обучении иностранному языку. – Вестник ТГУ (Тамбовский государственный университет), выпуск 2 (118), 2013. – С. 57-62.

<sup>109</sup> Sosnina E.П. Параллельные корпуса в обучении языку и переводу (Электрон ресурс): [http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus\\_education\\_translation/](http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_education_translation/)

birlik)lar mazmunan mos kelish darajasida bo‘lakka bo‘linib, mana shu birliklar yonma-yon, bir-biriga ishora/havola qilish holatida turadi. Asliyatdagi matn fragmentiga mos keluvchi tarjima fragment belgilab qo‘yilgan bo‘ladi. Aynan mana shu holat bunday korpuslar yordamida turli lingvistik amallar bajarishga imkon yaratadi”<sup>110</sup>.

Zamonaviy korpus lingvistikasida parallel korpusning ikki ko‘rinishi mavjud:

- 1) ko‘p tilli korpus (Comparable/Multilingual Sorpora);
- 2) tarjima korpus (Translation Sorpora).

Bunday xususiyatga ega korpusning struktur tarkibi uning maqsadidan kelib chiqib turlicha bo‘lishi mumkin:

- 1) tarjimaga havola qiluvchi odatiy matn tarzida;
- 2) qiyoslash uchun qulay bo‘lgan “oynadagi matnlar” shaklida;
- 3) ma’lumotlar bazasi ko‘rinishida.

Tuzilishi va matnlarning joylashuvi, birliklarining moslashtirilishiga ko‘ra parallel korpusning bir necha ko‘rinishi mavjud<sup>111</sup>:

- 1) bir yo‘nalishli (ingliz tilidan rus tiliga tarjima qilingan matn);
- 2) ikki yo‘nalishli (ingliz tilidan rus tiliga tarjima qilingan matn va rus tilidan ingliz tiliga (teskari tarjima));
- 3) ko‘p yo‘nalishli (ingliz tilidan rus, nemis, fransuz tillariga tarjima qilingan matn).

Parallel korpus tarjima tilining struktur tuzilishi, ibora va so‘zlarning ma’lum kontekstdagi imkoniyatlarini kuzatishga xizmat qiladi. Bunday korpuslar quyidagi maqsadlarda foydalaniladi:

- 1) qiyosiy tilshunoslikda (ikki til strukturini tahlil qilish maqsadida);
- 2) tarjima sohasida (original matn ekvivalentini boshqa tillardan qidirish maqsadida);
- 3) avtomatik tarjima sohasida;

---

<sup>110</sup> Добровольский Д. Какие задачи решают параллельные корпуса. (Электрон ресурс): <https://postnauka.ru/video/54851>

<sup>111</sup> <https://translationrating.ru/glossary/parallelnyj-korpus/>



4) lingvodidaktikada;

5) leksikografiyada.

Ishning keyingi qismlarida korpusning qo‘llanishi va maqsadiga ko‘ra turlariga batafsil to‘xtalamiz.

Matn va uning tarjimasini yonma-yon keladigan elektron to‘plamning boshqa ko‘rinishlari ham mavjud. Bitekst dasturi shunday vositalar sirasiga kiradi. Parallel matn (bitekst) – boshqa tildagi tarjimasini bilan beriladigan matnlar to‘plami<sup>112</sup>.

Bitekstlar “moslashtiruvchi asbob” (*alignment tool*) yoki “bitekst uchun moslama” (*bitext tool*) deb nomlanuvchi maxsus kompyuter dasturlari orqali amalga oshiriladi. Bu dasturlar original va tarjima matn mazmunini turli sintaktik birliklar, asosan, sodda gap shaklida moslashtirib beradi. Bitekstlar yig‘indisi *bitekst ma’lumotlar bazasi* yoki *ikki tilli korpus* deb nomlanadi ham turli bog‘lanishlarni kuzatish imkonini beruvchi ma’lumot bazasi (spravochnik) vazifasini bajaradi.

*Parallel korpus: ilk va keyingi ko‘rinish.* Korpus lingvistikasi yo‘nalishi xorijiy tillarni o‘qitish metodikasi va tarjima, kompyuter lingvistikasi yo‘nalishi uchun amaliy ahamiyat kasb etmoqda. Xorijiy til va tarjima nazariyasi ta’limida parallel korpuslarning ahamiyati borasida M.Barlou, Makkeneri, Beyker, Zanneti, Arenberga, Blank, Braun, Cherch, Geyl, Devis, Foster kabi olimlarning fikrlari qimmatli.

Bitekst g‘oyasi Brayan Xerrisga tegishli bo‘lib, u dastlab 1988-yilda bitekst konsepsiyasini ishlab chiqqan; ushbu konsepsiya keyinchalik Monreal univesiteti bir guruh olimlari tomonidan rivojlantirilgan; RALI (*Recherche appliquée en linguistique informatique* yoki *Applied Research in Computational Linguistics* // Прикладные исследования в вычислительной лингвистике”) deb nomlanuvchi uyushmada faoliyat olib borilgan. Guruhda tilshunos va dasturchilar faoliyat yuritgan; matnni qayta ishlash muammolarini o‘rganishgan. Ilk marta bitekst konsepsiyasini Pyer Izabel va Klod Bedarlar amalga oshirishgan<sup>113</sup>.

---

<sup>112</sup> <https://dic.academic.ru/dic.nsf/ruwiki/1085886>

<sup>113</sup> <https://dic.academic.ru/dic.nsf/ruwiki/1085886>

1897-98-yillarda nemis tilshunosi Keding tomonidan birinchi korpus yaratilgan; u qog‘oz shaklda bo‘lib, so‘zlarda tovushlarning taqsimlanish chastotasini tekshirish maqsadida tuzilgan edi. Ammo katta hajmli materialni qo‘lda hisoblab chiqish juda mushkul mashg‘ulot ekanligini hozirgi tadqiqotchilar biladi. Keyinchalik paydo bo‘lgan konkordans yoki undan ham mukammalrog‘i – korpus menejeri bu vazifani ancha engillashtirdi. Korpus tuzish tajribasi shuni ko‘rsatdiki, universal korpus yaratishning iloji yo‘q. Tadqiqotchi maqsadiga ko‘ra korpus imkoniyatlari chegaralanadi va har bir vazifani alohida korpuslar bajaradi.

Korpuslar tarkibidagi material tiliga ko‘ra, dastlab, bir tilli va ko‘p tilli korpus turlariga bo‘linadi. Ikki yoki ko‘p tilli korpuslar esa quyidagi ko‘rinishga ega:

- 1) parallel yoki tarjima korpuslari;
- 2) qiyosiy korpus.

Parallel korpuslar qanday maqsadni ko‘zlashidan qat’i nazar tarkibidagi birliklar moslashtirilgan bo‘lishi lozim. Moslashtirish yuqorida aytilganidek, bitekst dasturi asosida amalga oshiriladi.

Parallel matnlarni to‘g‘rilash – ikkala til (tarjima va asl) matnidagi bir-biriga to‘g‘ri keluvchi gap, abzatlarni identifikatsiyalash jarayoni.

Matn qismlarini bunday moslashtirish turlicha bo‘lishi mumkin. Qismlar muallif maqsadiga qarab har xil bo‘ladi<sup>114</sup>: bob, bo‘lim, abzats, gap va h. Foydalanuvchi uchun eng qulay moslashtirilgan shakl bu asl va tarjima matnlarning gaplar bo‘yicha muvofiqlashtirilishi hisoblanadi. Bu juda qiyin, ko‘p kuch talab etiladigan mehnat. Shuning uchun bo‘lsa kerak, hozirgacha dunyoda yaxshi muvofiqlashtirilgan, namunali sanaladigan parallel korpuslar sanoqli. Chunki asl va tarjimadagi matn lavha(qism)larini muvofiqlashtirishning avtomatik usuli yaxshi natija bermaydi. Eng mukammal ishlangan avtomatik tarjima dasturi ham gaplarning chegarasini aniqlashda aniq echim berolmaydi. Asl matndagi fikr bir gap bilan, ikki-uch gap bilan ifodalanishi mumkin yoki aksincha. Shuning uchun bitta buyruq bilan asl va tarjima matn qismlarini tenglashtirishning iloji yo‘q. To‘g‘ri, shunga mo‘ljallangan texnik vositalar mavjud, ammo ularning hech

---

<sup>114</sup> Добровольский Д. Какие задачи решают параллельные корпуса. (Электрон ресурс): <https://postnauka.ru/video/54851>

biri inson aralashuvisiz muammoni hal etolmaydi. Parallel korpus tarkibidagi matni gaplar bo'yicha mosligiga erishish turli lingvistik amallar bajarishga yo'l ochib beradi. Bilamizki, tarjima jarayonida asl matndagi bitta jumla tarjima matnda ikki jumlagacha bo'linishi, aksincha, original matndagi bir necha gap tarjimada bitta gap holida berilishi, gaplarning izchillik o'rnini almashtirilishi mumkin. Shuning uchun parallel korpus tuzishda eng murakkab jarayon asl va tarjima matndagi moslashtirish jarayoni hisoblanadi<sup>115</sup>.

Parallel korpusni qiyosiy korpusdan farqlash lozim. Qiyosiy korpusda turli tilda bir mavzuda yaratilgan matnlar qiyoslash maqsadida bir korpusga yig'ilgan bo'ladi. Bunday matnlar ham ma'lum tadqiqiy masalalarning yechimi uchun xizmat qiladi. Turli tillarda ma'lum bir diskurs vaziyati, u yoki bu mavzudagi voqea qanday ifodalanishini kuzatish lozim bo'lganda, albatta, qiyosiy korpusga murojaat qilinadi. Agar tilshunos ma'lum bir so'z, birikma, konstruksiya ikkinchi tilda qanday ifodalanishini kuzatmoqchi bo'lsa, unda parallel korpusga murojaat qilgani ma'qul.

Ikki tildagi matn bo'laklari – gaplarni bir-biriga moslashtirish bilan birga matn birliklari teglanishi ham zarur. Bu teglar grammatik, semantik va boshqa xil bo'lishi mumkin. Bu teglar ma'lum bir so'zga yoki butun matnning o'ziga birlashtirilgan bo'lishi mumkin. Agar matn birliklari teglanmagan bo'lsa, unda korpusdan ma'lum bir grammatik shaklni qidirish imkoni bo'lmay qoladi. Demak, matn grammatik teglarga ega bo'lganda bir tildagi so'zshaklning ikkinchi tildagi ekvivalentini qidirish, kuzatish mumkin. Agar so'zshakllar grammatik teglanmagan bo'lsa, ikki til ekvivalentlarini tekshirish faqat sintaktik butunliklar orasidagina amal qiladi.

Semantik teg ham juda muhim omillardan. Masalan, kiyim ma'nosini bildiruvchi so'zlar ishtirok etgan kontekstlarni ajratib berish buyrug'i berilsa, semantik tegga ega matn bu so'rovni qayta ishlay oladi va so'ralgan semali lemma ishtirok etgan kontekst ajratib ko'rsatiladi. To'g'ri, *pidjak*, *ko'ylak*, *jemper* kabi ishlatilish chastotasi yuqori bo'lgan so'zlarning tarjimada berilishini hamma biladi.

---

<sup>115</sup> ўша манба.

Bunday soʻzlarni korpusdan qidirishga hojat yoʻq. Ammo *tulup, armyak, poddelka*<sup>116</sup> kabi past chastotali soʻzlarning turli tarjimonlar tilida qanday ifodalanishini kuzatish foydadan xoli emas, chunki bunday soʻzlar hamma tarjima lugʻatlarida ham uchramaydi<sup>117</sup>.

A.A.Kokoreva “Корпус параллельных текстов в обучение иностранному языку” nomli maqolasida parallel korpusning oʻziga xos xususiyati haqida shunday deydi: “Boshqa til korpuslaridan parallel korpusning tipik farqi uning koʻp tilliligidan. Uning bu xususiyati til oʻrganuvchida leksik koʻnikmalar hosil qilishda qoʻl keladi. Parallel korpus vositasida quyidagi imkoniyatlarni inobatga olib talabga professional leksikani chuqur oʻrgatish mumkin”<sup>118</sup>. A.A.Kokorevaning taʼkidlashicha, professional leksikani oʻrganishda parallel korpusning quyidagi imkoniyatlari qoʻl keladi:

- a) korpusda professional yoʻnaltirilgan yozma manbalar mavjud;
- b) korpusda asl manba va uning boshqa tilga tarjimasi mavjud;
- s) korpus oʻrganilayotgan tildan ona tiliga yoki aksincha, ona tilidan oʻrganilayotgan tilga tarjima qilish koʻnikmasini shakllantiradi;
- d) korpusda ilmiy uslubda yaratiladigan janr namunalari toʻplanadi, shu sababdan ilmiy uslubda tarjima qilish koʻnikmasini parallel korpusdan foydalanish orqali shakllantirish mumkin;
- e) taʼlim jarayoniga illyustrativ korpuslar jalb qilinishi (korpus maqsadiga koʻra tadqiqiy yoki illyustrativ boʻladi) lozim; chunki oʻqitilayotgan til leksikasini professional oʻrganish uchun sanoqsiz illyustrativ misollarning mavjudligi tadqiq etilayotgan masalaga oydinlik kiritadi;
- j) korpusda zamonaviy professional leksika namunalari uchratish oson.

Parallel korpuslar taʼlim jarayoni uchun didaktik xususiyat va metodik funksiyasi bilan qulay taʼlimiy resurs vazifasini oʻtay oladi. Didaktik xususiyat sifatida (a) koʻp tillilik, (b) qidiruv natijasining kontekst shakllidaligi, (v) tilning

---

<sup>116</sup> Мисоллар корпусдаги манбадан айнан олинди.

<sup>117</sup> Добровольский Д. Какие задачи решают параллельные корпуса. (Электрон ресурс): <https://postnauka.ru/video/54851>

<sup>118</sup> Кокорева А.А. Корпус параллельных текстов в обучении иностранному языку. – Вестник ТГУ (Тамбовский государственный университет), выпуск 2 (118), 2013. – С. 57-62.

qo'llanish doirasini chegaralab ko'rsatish imkoniyatlarini ko'rsatish mumkin. Parallel korpusning ushbu didaktik xususiyati talabada maxsus maqsadda qilingan tarjimani o'rgatish, so'z ma'nosini yanada yaxshiroq aniqlash, so'zning aniq nutqiy sharoitda qo'llanishini kuzatish asosida leksik ko'nikmani shakllantirish imkoniyatini beradi. Parallel korpus tarjima matnida sinonimik qatordagi so'zlardan qaysi birini qo'llash o'rinli ekanligini aniq ko'rsata oladi<sup>119</sup>. Tarjimada ma'lum bir so'zning ekvivalentini topish, ularning sinonimlaridan birini tanlash muhim masala. Sinonimlarni qo'llashda turli tarjima matnlari kuzatiladi, so'zdan oldin va keyin qanday instruksiyalar qo'llanilishi kuzatiladi, sinonimik qatordagi so'zning valentlik imkoniyati, valentlikdagi chastota aniqlanadi va aniq xulosaga kelinadi. Ayrim so'zlarning boshqa tilda ekvivalenti mavjud bo'lmaydi, ya'ni o'sha tilda shu tushuncha, voqea-hodisa, predmet bo'lmasligi mumkin. Natijada, bunday birliklar turli xil tarjima qilinadi va lingvospetsifik xususiyatli leksema sifatida qaraladi. Shunday birliklarni aniqlashning eng optimal usuli ham parallel korpusga murojaat qilish. Agar bitta so'z turli tarjimada turlicha keltirilganligi korpus yordamida statistik usul bilan aniqlansa, demak, bema'lol bu leksemaning lingvospetsifik xarakterli ekanligi haqida xulosa qilsak bo'ladi.

---

<sup>119</sup> Добровольский Д. Какие задачи решают параллельные корпуса? (Электрон ресурс): <https://postnauka.ru/video/54851>

D.Dobrovolskiy rus-nemis parallel korpusi tahlili asosida misol keltiradi<sup>120</sup>.

Rus tilida *захолустье* leksemasi lingvospetsifik xarakterda ekanligi taxmin qilinadi. Tarjimada nemis tilidagi *Einode* va *Kaff* soʻzlari yuqoridagi soʻzning ekvivalenti sifatida keltiriladi. *Einode hech kim mavjud emas, boʻsh* degan maʼnoni bildiradi; *Kaff zerikarli va yomon kichik aholi punkti* maʼnosiga ega. *захолустье* esa *kichik aholi punkti* emas, bu yerda *markazdan uzoqlik, madaniyatga egalik* semasi muhimroq. Nemis madaniyatida poytaxtga yaqin yashash juda sivilizatsiyaning muhim omili emas, chunki madaniyat oʻchoqlari boshqa joyda ham boʻlishi mumkin. Rus tilidagi *zaxolustye* leksemasi esa *markazdan uzoqlashish* semasini bildirgan. Demak, shunday birliklar tarjimada qanday berilishini turli matnlarda kuzatishning eng qulay usuli parallel korpusdagi qidiruv asosida ularni kuzatish, bir joyga toʻplash va tahlil qilishdir.

D.O.Dobrovolskiyning fikricha, parallel korpuslar faqat tarjima nazariyasi, ikki tilli leksikografiya, tillarni qiyoslash bilan bogʻliq tilshunoslikning boshqa sohalari uchun foydali ekanligi haqidagi xulosa bilan chegaralanish nooʻrin<sup>121</sup>. Qiyoslanayotgan tillar leksikasidagi soʻzlarning semantik tarkibini yanada chuqurroq tavsiflashda ham muhim ahamiyat kasb etadi.

V.I.Melnik oʻzining “Переводческие корпусы – что они могут сказать нам о переводе и его преподавание?” nomli maqolasida parallel va qiyosiy korpus farqini shunday izohlaydi<sup>122</sup>:

- 1) PK qiyosiy korpusdan farqli ravishda ikki til materialini oʻzida saqlaydi, QPdagi matnlar bir tilga mansub boʻladi;
- 2) PK aniq, mukammal razmetkalangan boʻlsa, QPda razmetka nazarda tutilmaydi;
- 3) PKda tarjima izchilligiga erishiladi, QP tarjima jarayoni bilan bogʻliq emas.

Qiyosiy korpusda tarjima matnlari boʻlmasa, undan tarjima nazariyasi haqida nimani oʻrganish mumkin? Bunday korpus tarjima materialini tahlil qilish

---

<sup>120</sup> ўша манба

<sup>121</sup> Добровольский Д.О. Корпус параллельных текстов в исследовании лексической семантики. <http://www.dialog-21.ru/media/2503/dobrovolskij.pdf>

<sup>122</sup> Мельник В.И. Переводческие корпусы – что они могут сказать нам о переводе и его преподавании? (Электрон ресурс): <https://cyberleninka.ru/article/n/perevodcheskie-korpusy-chto-oni-mogut-skazat-nam-o-perevode-i-ego-prepodavanii>

emas, balki kontrastiv tadqiqot uchun material taqdim eta oladi. Korpus tarkibiga kiruvchi asliyat va tarjimadagi matnlarni qiyoslash korpus birligining pragmatik va funksional xususiyatini tahlil qilish maqsadiga xizmat qiladi. Bunday korpus vositasida tadqiqot olib borish u yoki bu birlikning tarjima matnida asliyatdagi matnga nisbatan qo'llanish chastotasini tekshirish, tahlil qilish maqsadiga xizmat qiladi.

Qaysidir korpus birligining turli chastotaga egaligini tekshirishdan oldin QPning ichki korpuslari orasidagi qiyoslashni qaysi kriteriyalar asosida olib borishni aniqlab olish muhim.

*Birinchidan*, bu korpus hajmi. Qiyosiy korpus tarkibiga kiruvchi ichki korpuslar hajm nuqtai nazaridan relevant holatda bo'lishi kerak. Ichki korpus materiali relevant holatda bo'lganda, statistik metod natijasi noadekvat tus oladi.

*Ikkinchidan*, mavzu ko'lami: o'rganilayotgan korpus materiali mavzu jihatdan mushtaraklik kasb etishi lozim (bir korpusda Tashqi ishlar vazirligi chiqishi, ikkinchi korpusda ilmiy maqolalar to'planmasligi kerak).

*Uchinchidan*, temporallik yoki xronologik kriteriy. Ikki korpus materiali bir xronologik doiraga mansub bo'lishi ahamiyatli (2015-yil Buyuk Britaniya Bosh vaziri va 1910-1915-yillardagi Bosh vazir chiqishlari qiyoslanishi to'g'ri emas).

*To'rtinchidan*, janriy kriteriy. Qiyoslanayotgan korpus birliklari janr jihatdan mos kelishi kerak.

*Beshinchidan*, lingvistik ma'lumot tipi. Bu kriteriy korpus materialining yozma/og'zaki/aralash matn ekanligini nazarda tutadi. Nutq shakllaridan faqat bittasi tanlanishi lozim.

Qiyosiy korpus vositasida original hamda tarjima ichki korpusini qiyoslash asosida tarjima universalialari<sup>123</sup>ni o'rganish (1); tarjima ta'limi (2)ni amalga oshirish mumkin. Bu jarayon tarjima matnining asliyatga mosligini talab qiladi.

Xulosa o'rnida aytish mumkinki, parallel korpus – tarjima va asliyatdagi matnlarning razmetkalangan, o'zaro moslashtirilgan elektron yig'indisi.

---

<sup>123</sup> Ушбу атама остида таржиманинг ўзига хос хусусиятлари назарда тутилади.

Parallel korpusning lingvistik, xorijiy til lingvodidaktikasidagi ahamiyati va tarjimonlik sohasidagi imkoniyatlari.

**Parallel korpusdan universitet lingvistik ta'limida foydalanish.** Korpus lingvistikasi asri kelishi tilshunosning emperik bilish jarayonini ancha yengillashtirdi: tadqiq etilayotgan hodisani kuzatish, misol/dalil yig'ish, taqqoslash, xulosa chiqarish uchun misollar massivining mavjudligi o'rganishning ushbu bosqichini qulaylashtirdi. Til korpusidagi qulay interfeys va qidiruv tizimi tadqiqotchiga tadqiq ob'ekti sifatida xizmat qilmoqda; tadqiq materialini o'rganish katta, ichida oylab, yillab material qidiriladigan kutubxonalarda emas, bitta kompyuter yordamida amalga oshirilmoqda. Grammatik talqinlarning til egalari nutqi misolida dalillanishi korpus lingvistikasi asri uchun yangilik bo'lmay qoldi. Korpusda tipik holatlar, misollarning o'n, yuz, minglab uchrashi tadqiqotchiga xulosa chiqarishga yordam beradi; noodatiy holatlar esa yangi-yangi nazariy qarashlarni ham paydo qiladi. Aniqrog'i, nutqda uchragan yangi qo'llanish, yangi birikma va struktura tilshunoslikda yangi farazlar paydo bo'lishiga imkon beradi<sup>124</sup>.

Korpus lingvistikasi rivojlanishi oliy ta'lim muassasasida RuN-Euro korpusi loyihasi asosida magistrlik dasturlari ishlab chiqishga asosi bo'ldi; shu korpus asosida magistrlik kontrastiv lingvistika bo'yicha tadqiqotlar olib bormoqda. Rus, norveg, ingliz tillarini qiyoslash asnosida magistrlik, doktorlik dissertatsiyalari yoqlanmoqda. Masalan, kvantitativ va kvalitatif yondashuvning farqi borasida qilingan ish shular jumlasidan<sup>125</sup>. Rus, norveg, ingliz tilidagi ravishlarning berilishi parallel korpus misollari asosida kuzatilgan, statistik usul bilan aniqlangan hodisalar asosida umumlashmalar qilingan. Ushbu korpus asosida olib borilgan magistrlik va PhD ilmiy izlanishlari sirasiga “Вид в славянских императивах” (Alvestad, PhD., 2013), “Реалии в переводах с русского на норвежский” (Kharina, PhD., 2013-15), “Дискультурные частицы в сопоставление”, “Интерпритация и перевод местоимений с -то/-нибудь”, “Перевод на

---

<sup>124</sup> Гёрн А. Параллельный корпус в системе университетских лингвистических курсов. – (Электрон ресурс): [http://folk.uio.no/atleg/nizhnij\\_gronn\\_oct2013.pdf](http://folk.uio.no/atleg/nizhnij_gronn_oct2013.pdf)

<sup>125</sup> Krave M. F. Converbs in Contrast: Russian converb constructions and their English and Norwegian counterparts. PhD thesis. – 2011, University of Oslo.



норвежский глаголом начинателного способа действия с пердлогами *по-* и *за*” kabi ishlarni kiritish mumkin. Ushbu tadqiqotlarda korpus tadqiq obyekti manbayi sifatida xizmat qilgan. Ye.Sosnina parallel korpusdan til ta’limida foydalanish haqida shunday fikr bildiradi: “Korpusdan olingan misollar til ta’limida asqotadi. Talabaga amaliy topshiriq va tahlil qilish uchun real nutqiy vaziyatlardan olingan, madaniyatlararo muloqotda foydalanish haqida yorqin tasavvur uyg’otadigan material beriladi. Ko’p darsliklarga xos sanalgan umumiy kamchilik shundaki, ulardagi misollar, odatda, muallif tomonidan o’ylab topiladi, yoziladi, ular haqidagi xulosalar bir yoki ikki kishining fikri asosida shakllantiriladi. Ammo G’arb zamonaviy metodikasida darslikdagi misolning korpus yoki boshqa aniq manbadan olinishi odat tusiga kirgan”<sup>126</sup>.

Korpus xorijiy til ta’limida mavjud materialni tahlil qilish va kamchiliklarni aniqlashda ham asqotadi. Bu ishlar umummetodologik xarakterga ega – u yoki bu struktura/leksika mavjud tipik darslik, lug’at hamda korpus asosida tahlil qilinadi; keyin ikkala manbadagi natija taqqoslanadi. Bunday tadqiqotlarning ko’pchiligi quyidagi xulosaga olib keldi:

- 1) darslik, korpusda ko’rsatilganidek, til egalarning qanday gapirishini “o’rgatadi”;
- 2) ayrim darsliklar muhimroq aspektlarni chetlab o’tadi yoki kamdan-kam qo’llaniladigan oborotlarni namuna sifatida keltiradi (darsliklardagi umumiy kamchilik shunda);
- 3) ta’lim materiali real kontekstni inobatga olmasligi bilan o’quvchini chalg’itadi, til egallash intensivligini pasaytiradi.

Demak, parallel korpus lingvistik tadqiqotlar uchun materiallar massivini taqdim etuvchi zamonaviy o’quv-uslubiy vosita sanaladi.

### **Mustahkamlash uchun savol va topshiriqlar**

1. Parallel korpus deb nimaga aytiladi?
2. Parallel korpusning interfeysi qanday tuzilishga ega bo’ladi?

---

<sup>126</sup> Sosnina E.П. Параллельные корпуса в обучении языку и переводу. – (Электрон ресурс): [http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus\\_education\\_translation/](http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_education_translation/)

3. Lingvodidaktikada parallel korpus qanday ahamiyatga ega?
4. Tarjima nazariyasida parallel korpus qanday ahamiyatga ega?
5. Ta'limda parallel korpus qanday ahamiyatga ega?
6. Tipologik tadqiqotlarda parallel korpus qanday ahamiyatga ega?
7. Parallellik xususiyatiga ko'ra korpus nech tilli bo'lishi mumkin?
8. Tillarning morfologik xususiyati parallel korpus tuzishda qanday ahamiyat kasb etadi?

#### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Грудева В.А. Корпусная лингвистика. Учебное пособие. 2-е изд., стереотип. – М.: Флинта, 2012. – 165 с.
3. Гёрн А. Параллельный корпус в системе университетских лингвистических курсов. – (Электрон ресурс): [http://folk.uio.no/atleg/nizhnij\\_gronn\\_oct2013.pdf](http://folk.uio.no/atleg/nizhnij_gronn_oct2013.pdf)
4. Добровольский Д. Какие задачи решают параллельные корпуса. (Электрон ресурс): <https://postnauka.ru/video/54851>
5. Добровольский Д. Какие задачи решают параллельные корпуса? (Электрон ресурс): [https://postnauka.ru/video/;](https://postnauka.ru/video/)
6. Добровольский Д.О. Корпус параллельных текстов в исследовании лексической семантики. (Электрон ресурс): <http://www.dialog-21.ru/media/2503/dobrovolskij.pdf>
7. Кокорева А.А. Корпус параллельных текстов в обучении иностранному языку. – Вестник ТГУ (Тамбовский государственный университет), выпуск 2 (118), 2013. – С. 57-62.
8. Mengliev B., Karimov R. Theoretical fundamentals of uzbek-english parallel corpus \ JOURNAL OF CRITICAL REVIEWS ISSN- 2394-5125 VOL 7, ISSUE 17, 2020.
9. Соснина Е.П. Параллельные корпуса в обучении языку и переводу. – (Электрон ресурс):

[http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus\\_education\\_translation/](http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_education_translation/)

10. Тао Ю., Захаров В.П. Разработка и использование параллельного корпуса русского и китайского языков // Автоматизация обработки текста. (Электрон ресурс): <https://www.researchgate.net/publication/306911853>

11. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.

12. Хамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.

13. <https://dic.academic.ru/dic.nsf/ruwiki/1085886>

14. <https://translationrating.ru/glossary/parallelnyj-korpus/>

## MUALLIFLIK KORPUSLARI

### Reja:

1. Mualliflik korpusi tuzishning lingvistik asoslari.
2. Mualliflik korpuslarining mushtarak va farqli jihatlarini.
3. A.P.Chexov, A.S.Pushkin, U.Shekspir mualliflik korpusi.
4. Firdavsiy, Rumi, Sa'diy Sheroziy mualliflik korpusi.

*Tayanch so'z va iboralar:* mualliflik korpusi, A.P.Chexov asarlari korpusi, A.S.Pushkin asarlari korpusi, U.Shekspir asarlari korpusi, Firdavsiy asarlari korpusi, Rumi asarlari korpusi, Sa'diy Sheroziy asarlari korpusi.

**Mualliflik korpusi maqsad va vazifalari.** Hozirgi o'zbek tili boshqa tillar kabi o'z tarixiy taraqqiyot yo'lida bir qancha o'zgarishlarga uchradi. Taraqqiyot davridagi har bir bosqichning o'ziga xos xususiyatini tadqiq etish maqsadida tilning taraqqiyot bosqichlari ajratildi hamda o'rganildi, bu tadqiqotlarda davr tilining o'ziga xos xususiyati ma'lum bir adib ijodi misolida yoritildi. Har bir davr ijodkorlari ijodiy merosini to'liq qamrab oluvchi bir qancha lingvistik tadqiqot mavjud bo'lsa-da, ular maxsus tizimlashtirilgan baza shaklini olmagan. Axborot texnologiyalari asrida bunday yaxlit tizimning o'zbek (turkiy) yozma manbalari uchun mavjud bo'lmashligi achinarli holatdir.

Tilshunosligimiz tarixida alohida muallif asari til xususiyatini o'rganishga bag'ishlangan maxsus tadqiqotlar talaygina. Runiy, uyg'ur yozuvi asosida yaratilgan yodgorliklarning til xususiyatiga bag'ishlangan bir qancha asarlar maydonga keldi. Bu tadqiqot ishlarida qadimgi turkiy til xususiyatlarining ayrim morfologik, sintaktik, fonetik xususiyatlari tahlil qilindi. Runiy yozuvdagi yodgorliklar tilini V.V.Radlov, S.E.Malov, A.M.Sherbak, G'.Abdurahmonov, A.Rustamovlar atroflicha tadqiq etishdi<sup>127</sup>. XI-XIV asr yozma yodgorliklari til xususiyati A.K.Borovkov ishlarida<sup>128</sup>, "Devonu lug'otit turk" asari tili A.Fitrat tadqiqotlarida, Yusuf xos Hojibning "Qutadg'u bilig" asari til xususiyati V.V.Radlov, S.E.Malov, V.V.Bartold, E.E.Bertelslar, Ahmad Yugnakiyning

<sup>127</sup> Каранг: Щербак А.М. Грамматический очерк языка тюркских текстов X-XIII вв. из Восточного Туркестана. М.-Л., 1961.; Абдурахмонов Ф., Рустамов А. Қадимги туркий тил. Тошкент: Ўқитувчи, 1982.

<sup>128</sup> Боровков А.К. Лексика среднеазиатского турсира XI-XIII вв. -Москва: ИВЛ, 1963.

“Hibatul-haqoyiq” asari S.E.Malov, E.E.Bertels, Q.Mahmudovlar tomonidan alohida o‘rganildi<sup>129</sup>. O‘g‘uz dialekti belgilari ustun bo‘lgan “Qissayi Yusuf” dostoni M.Brokkelman, o‘g‘uz, qarluq-uyg‘ur dialektining qadimgi yodgorliklaridan hisoblangan “O‘g‘uznoma” afsonasi G.N.Potapin, A.M.Sherbak, XIV asr adabiy tili yodgorligi “Muhabbatnoma” asari til xususiyati ham A.M.Sherbak tomonidan tadqiq etilgan<sup>130</sup>. XI-XIV asrlarda tuzilgan rasmiy hujjatlar uslubida o‘sha davr tiliga xos xususiyatlar o‘z ifodasini topgan. XI-XIV asrlarga oid rasmiy tilni bayon etishda professor S.E.Malov nashr qilgan X-XIII asrlarga tegishli yuridik hujjatlar<sup>131</sup>, O.D.Chexovich nashr etgan XIV asrga oid Buxoro hujjatlari<sup>132</sup> hamda E.R.Tenishev nashr etgan XIII-XIV asrlarga oid xo‘jalik yozuvlari<sup>133</sup> yordam beradi. Eski o‘zbek adabiy tili deb ataluvchi XIV-XIX asr tili xususiyati Atoiy, Sakkokiy, Lutfiy, Navoiy asarlari misolida<sup>134</sup> tadqiq etilgan. Xususan, Alisher Navoiy asarlari tili alohida ahamiyatga ega, eski o‘zbek adabiy tili taraqqiyotida muhim o‘rni borligini A.K.Borovkov tadqiqotlari<sup>135</sup> isbotlaydi. XVII-XIX asr adabiy tili xususiyatlari Abulg‘ozi Bahodirxon, Turdi Farog‘iy, Nishotiy, Munis, Gulxaniy, Maxmur asarlari tili, XX asr o‘zbek adabiy tili leksikasi xususiyatlari Abdulla Qodiriy, G‘afur G‘ulom, Oybek, Abdulla Qahhor asarlari til xususiyatlari<sup>136</sup> misolida tadqiq etilgan. Bu tadqiqotlarda muallif tili leksikasi atroflicha tavsiflangan. Bunday tadqiqotlar tilning rivojlanish jarayoni, o‘zgarishlarni: ma’no torayishi, kengayishi, istorizm, arxaizm, neologizmlar harakatini kuzatish uchun qulay sharoit yaratadi. Mana shunday xususiyat har bir muallif, davr tili uchun tadqiq etilgan bo‘lsa-da, ular alohida-alohida monografiyalarda, tadqiqot ishlarida mavjud. Agar bu tadqiqotlar Milliy

---

<sup>129</sup> Турсунов У. ва б. Ўзбек адабий тили тарихи. –Тошкент: Ўқитувчи, 1995. – 261 б.

<sup>130</sup> Ўша асар. 94-бет.

<sup>131</sup> Малов Е.С. Памятники древнетюркской письменности. М. – Л., 1961.–с.:200-221.

<sup>132</sup> Чехович О.Д. Бухарские документы XIV в. –Ташкент, 1965.

<sup>133</sup> Тенишев Э.Р. Хозайтсвенные записи на древнеуйгурском языке//Сб. “Исследования по грамматике и лексике тюркских языков”.– Ташкент, 1965 год.– с.:37-77.

<sup>134</sup> Юдахин К.К. Материалы к вопросу о звуковом составе чигатайского языка// “Культура и письменность Востока”, кн. 4. – Баку, 1929.; Щербак А.М. Грамматика староузбекского языка. М.–Л., 1962.; Рустамов А. Аlisher Navoiy tilining fonetik-morfologik xususiyatlari. Докторлик дис. Автореферати. –Тошкент, 1966.

<sup>135</sup> Борков А.К. Аlisher Navoiy kаk osnovopolozhnik uzbekskogo literaturnogo yazyka//Сб. “Аlisher Navoiy”. – М. – Л., 1946.

<sup>136</sup> Турсунов У. ва б. Ўзбек адабий тили тарихи. –Тошкент: Ўқитувчи, 1995. – 261 б.

korpusda yaxlit bir tizim asosida o'z aksini topsa, istalgan leksemaning tarixiy taraqqiyot jarayonini turli muallif tili misolida yorqin tasavvur qilish imkoniyati paydo bo'ladi. Tilshunosligimiz oldida turgan endigi vazifa shu tadqiqotlarni birlashtirish asosida korpus yaratishdan iborat. Bu vazifani amalga oshirish ikki bosqichni o'z ichiga oladi:

1. Tilimiz tarixiy taraqqiyoti davrida til xususiyatlari tadqiq qilingan adiblar ijodiy merosi, tadqiqot natijasini korpus sifatida shakllantirish.

2. Shu paytgacha til xususiyati o'rganilmagan yozuvchi, shoirlarning asarlari korpusini tuzish.

Birinchi vazifa o'rganilgan milliy-madaniy merosimizning elektron shaklda yaxlit sistema(korpus shakli)da saqlanishi, undan keyinchalik turli ta'limiy, tadqiqiy maqsadda foydalanish imkonini beradi. Ikkinchi vazifani amalga oshirish til xususiyati o'rganilmagan yozuvchi, shoirlar asariga shoirlarning asarlariga bag'ishlangan turli filologik tadqiqotlar olib borish uchun elektron baza hamda milliy-madaniy meros, tilni elektron shakl (korpus formati)da saqlash, foydalanish, kelgusi avlodga to'liq etkazishni ta'minlaydi. Adabiy til rivojiga salmoqli hissa qo'shgan buyuk yozuvchilar asarlarining leksikasini asrab qolish, nafaqat asrash, balki butun ko'lami bilan tavsiflash, til egalari uchun namuna vazifasini o'taydigan model shakliga keltirish adabiy tilni asrash, rivojlantirish omili bo'lib xizmat qiladi. Bu borada o'z millati milliy, adabiy tili asoschisi hisoblangan butun dunyoga mashhur adiblar Shekspir, Dante, Gyote, Pushkinlarning adabiy merosi asosida qilingan ishlar (tuzilgan korpuslar) namuna sifatida xizmat qiladi<sup>137</sup>. Bu siraga adabiy tilning eng optimal tashuvchisi sifatida tan olingan yozuvchilar Shandor Petefi, Genrik Ibsen, Adam Miskevich, Karel Chapeklarning ishlarini kiritish ham o'rinlidir. Jahon tilshunosligida korpus lingvistikasi ancha yutuqlarga erishgan bo'lsa-da, mualliflik korpusi tuzish tajribasi oxirgi 5 yil ichidagina to'laroq shakllandi, bu borada ancha natijalarga erisha oldi. Bunday korpus tuzish ma'lum davr adabiy tili, xususan, leksik tarkibi, uslubiy xususiyatini tadqiq etish uchun material vazifasini o'tashdek qimmatli vazifani ado etadi. Yuqorida

---

<sup>137</sup> Қаранг: Потемкин С.Б. Авторский корпус и словарь языка Антона Чехова. (Электрон ресурс: // [https:// istina.msu.ru](https://istina.msu.ru).)

ta'kidlanganidek, mualliflik korpusini tuzishning bir qancha afzalliklari bor. Mualliflik korpusi, uning o'ziga xos xususiyati, tuzish tamoyili kabi muammolar ushbu bo'limning tadqiq predmeti hisoblanadi. Mualliflik korpusi tarkibiy qismi, interfeysi, teg masalasi korpus tuzishning asosiy muammosi bo'lganligi sababli Internet tarmog'ida mavjud mualliflik korpuslarining shu jihatiga e'tibor qaratishni maqsadga muvofiq deb hisobladik.

*Korpus uchun material tanlash: muammo va yechim.* Milliy korpusi mavjud tillarning mualliflik korpusini tuzish hech qanday qiyinchilik tug'dirmaydi. Chunki matn (muallif asari) elektron shaklda Internet saytida yoki korpus tarkibida mavjud. Mualliflik huquqi asosida bu asarlar tarmoqda mavjudligi sababli ularni korpusga xom ashyo sifatida olishga monelik yo'q. Milliy korpusi mavjud bo'lmagan tillarda korpus yaratishda esa bu borada biroz muammoga duch kelish tabiiy hol. O'zbek tili milliy korpusi bo'lmasa ham, Ziyonet tarmog'ida mumtoz hamda zamonaviy shoir, yozuvchilarimiz asarlarining anchasi elektron shaklda joylashtirilgan. Shunga asosan, mualliflik korpusi tuzishda ikki manbaga asoslanish mumkin:

1. Har bir muallifning nashr etilgan mukammal asarlari to'plami.
2. Internet tarmog'idagi elektron fayllar.

Birinchi manbani elektron shaklga aylantirish (skanerlash, matnni sun'iy intellekt tushunadigan formatga keltirish) orqali material sifatida ishlatsak, ikkinchi manbadan nisbatan tayyor holda foydalanishimiz mumkin. Ikkala holda ham matnning elektron shakli olingach, uni texnik qayta ishlash – tokenizatsiya, lemmatizatsiya, sintaktik razmetkalash ehtiyoji tug'iladi. Texnik ishlov berishdan oldin matn korpus uchun tayyorlanadi, chunki unda nolingvistik birlik ham uchraydi. Korpus matnining asosiy belgisi unda nolingvistik birlikning (jadval, rasm, grafik chizma) bo'lmasligida. Shundan so'nggina razmetkalash bosqichiga o'tish mumkin. Razmetkalash avtomatik, yarimavtomatik rejimda bajariladi. Muallif qo'llagan neologizm, boshqa alifboda yozilgan so'zshakl lemmatizatsiyasiga alohida e'tibor qaratiladi. Korpus tayyor holga kelgach, tabiiy ravishda muallif asarlari chastotali lug'ati yuzaga keladi. Chunki lemmatizatsiya

jarayonida soʻzshakl hamda leksema (lemma) miqdori aniqlanadi. Masalan, Chexov asarlari korpusi 36 153 lemma yoki leksemani qamrab olgan, lemmalar 1 381 000 qoʻllanish holati (120 000 soʻzshaklda)ni tashkil etgan. Soʻz qoʻllash holatiga qarab, gapning oʻrtacha uzunligi (nechta soʻzdan iboratligi) ham aniqlanadi<sup>138</sup>. Leksema qoʻllanish chastotasi asar yozilish yili, janri, gap uzunligi asosida hisoblanishi ham mumkin. Bu esa foydalanuvchiga hozirgi adabiy til, muallif davri adabiy tili leksikasi chastotasini qiyoslash, xulosa chiqarish imkonini beradi. Bunday qiyosiy tahlil milliy korpus asosida bajarilishi ham mumkin. Demak, bunday korpus tuzishdan yana bir maqsad muallif asarlarining turli xildagi lugʻatlarini yaratishdan iborat. Shu bilan birga, tarix nuqtai nazardan katta davr ichidagi tilning tarixiy-madaniy rivojlanishi, oʻzgarishi ham oʻrganiladi.

**Mualliflik leksikografiyasi va mualliflik korpusi munosabati.** Jahon leksikografiyasida mualliflik leksikografiyasi deb nomlanuvchi alohida bir yoʻnalish mavjud. U muallif tili lugʻati tuzishning nazariy asoslari, mualliflik idiostilini oʻrganish tajribasi, mualliflik elektron lugʻati tuzishning oʻziga xos jihati, milliy korpus, adabiy til korpusosti korpusidan mualliflik leksikografiyasida foydalanish tajribasi, yangi turdagi lugʻatlar tuzish gʻoyalari, mualliflik leksikografiyasi asosida zamonaviy adabiy tildagi oʻzgarishlarni qiyosiy-tahliliy tadqiq etish kabi muammolar bilan shugʻullanadi<sup>139</sup>. Mualliflik leksikografiyasi – jahon, shuningdek, rus leksikografiyasida shiddat bilan rivojlanib borayotgan soha. Uning taraqqiyot natijasi oʻlaroq koʻplab mualliflik elektron lugʻat, korpuslari yaratildi. M.V.Lomonosov nomidagi Moskva Davlat universitetida “Krilov masallari tili lugʻati”, “Griboyedov komediyalari tili lugʻati”, “Gogolning “Revizor” komediyasi tili lugʻati”, “Fonvizin komediyalari tili” kabi lugʻatlar yaratilib, bunday ishlarga namuna sifatida “A.S.Pushkin tili lugʻati”<sup>140</sup> xizmat qildi. Bu tajriba rus leksikografiyasi sohasida ommalashib, boshqa oliy taʼlim muassasalarida ham shunday tadqiqotlar boshlandi. “A.N.Ostrovskiy asarlari tili

---

<sup>138</sup> Потемкин С.Б. Авторский корпус и словарь языка Антона Чехова. (Электрон ресурс: // [https:// istina.msu.ru](https://istina.msu.ru).)

<sup>139</sup> Қаранг: [www.ruslang.ru/seminar\\_aut\\_lexocorg020413](http://www.ruslang.ru/seminar_aut_lexocorg020413) сайтидаги “Теория и практика авторской лексикографии” семинар материаллари.

<sup>140</sup> “Словарь языка Пушкина” в четырех томах, создан в Институте русского языка АН СССР под ред. акад. В.В.Виноградова (издан в 1956–1961 гг.).



lug‘ati”<sup>141</sup>, “M.V.Lomonosov asarlari tili lug‘ati”<sup>142</sup>, “M.Kuzmin she‘rlari konkordansi”<sup>143</sup>, “A.Mitskevichning “Pan Tadeush” poemasi tili ideografik lug‘ati”<sup>144</sup> kabi ishlar shular jumlasidandir. Bu sohadagi tadqiqotlar ichidan O.V.Kukushkina, A.A.Polikarpov, E.V.Surovsevalar tomonidan tuzilgan “A.P.Chexov badiiy asarlari tili lug‘ati”<sup>145</sup> diqqatga sazovor.

*Mualliflik leksikografiyasi asos va hosilasi.* Mualliflik leksikografiyasi mahsuli, odatda, 2 qism: asos hamda mahsulotdan iborat bo‘ladi. Asos – korpus yoki konkordans, mahsulot – lug‘at. Keyingi paytda leksikografiyada (faqat korpus lingvistikasida emas) konkordans tushunchasi keng ommalashdi. *Konkordans* u yoki bu til ifoda vositasining kontekstda barcha qo‘llanish holati manbasiga havola qiluvchi ro‘yxat ma‘nosida korpus lingvistikasida qo‘llaniladi. Shuningdek, konkordans atamasi biror kitob yoki asar kalit so‘zlarining kontekstga havola qiluvchi alifbo tartibidagi ro‘yxati ma‘nosini ham ifodalaydi<sup>146</sup>. Bunday konkordansning ilk namunasi XIII asrda 500 monax ishtirokida tuzilgan. Unda ish hajmi ulkanligi uchun (juda muhim konkordanslar: Veda, Injil, Qur‘on yoki Uilyam Shekspir asarlari) katta jamoalar birlashib ishlashgan. Bundan tashqari, konkordans amaliy tilshunoslikda tarjima jarayonida, leksikografiyada, tilni o‘qitish, o‘rganish hamda matn tahlilida ham qo‘llaniladi<sup>147</sup>. Korpusdan tashqari konkordans so‘zlar ro‘yxatini tuzish, bir so‘zning turlicha ishlatilishini qiyoslash, ibora hamda idiomalarni qidirish, tadqiq etish, kalit so‘zlarni tahlil qilish, turli so‘z, birikmalar tarjimasini izlash yoki qo‘llanish chastotasini aniqlash kabi

---

<sup>141</sup> Хуснутдинов А.А.. Словарь языка А.Н.Островского” как один из вариантов авторского словаря” доклад (Ивановский государственный университет). (Электронный ресурс: [www.ruslang.ru/images/seminar\\_aut\\_lexocorg/husnutdinov\\_doklad.pdf](http://www.ruslang.ru/images/seminar_aut_lexocorg/husnutdinov_doklad.pdf))

<sup>142</sup> Волков С.С., Матвеев Е.М. “О проекте словаря “Риторика М.В.Ломоносова”. Доклад (ИЛИ РАН). (Электрон ресурс: [www.ruslang.ru/seminar\\_aut\\_lexocorg250214](http://www.ruslang.ru/seminar_aut_lexocorg250214))

<sup>143</sup> Гик А.В. “О работе “Конкордансом к стихотворениям М.Кузмина”: том четвертый” (Электрон ресурс: [www.ruslang.ru/seminar\\_aut\\_lexocorg](http://www.ruslang.ru/seminar_aut_lexocorg))

<sup>144</sup> Слободян Е.А. “Идеографический словарь языка поэмы А.Мицкевича “Пан Тадеуш” в авторской и электронной лексикографии”. (Электрон ресурс: [www.ruslang.ru/seminar\\_aut\\_lexicorg](http://www.ruslang.ru/seminar_aut_lexicorg))

<sup>145</sup> Кукушкина О.В., Поликарпов А.А., Суровцева Е.В. “Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова (с электронным приложением) как особый лексикографический продукт”. (Электрон ресурс: [www.ruslang.ru/seminar\\_aut\\_lexicorg](http://www.ruslang.ru/seminar_aut_lexicorg))

<sup>146</sup> Бобунова М.А. Конкорданс XXI в.: новая старая форма // Вопр. лексикографии. –2016. №2 (10) С.41-54.

<sup>147</sup> Якубовский К.И., Якубовская К.А. Обзор современных лингвистических технологий и систем // Вестник МГУП. 2015.– №2 С.315-319.

lingvistik amallarni bajarishda qo‘llaniladi<sup>148</sup>. Konkordans-lug‘at – har bir so‘zga qisqacha kontekst (semantik izoh emas!) keltiriladigan lug‘atning alohida bir tipi hamdir. Matnni o‘rganishda konkordansning qidiruv, evristik, analitik funksiyalaridan foydalanish, indeksatsiya, qiyoslash kabi imkoniyatini ishga solishi mumkin. Qidiruv funksiyasi berilgan so‘z yoki birikmadan foydalangan holda matndagi kerakli lavhani zudlik bilan topishga yordam beradi<sup>149</sup>. A.Stolyarovning fikricha, konkordansning ushbu funksiyasidan sitata tanlab olish yoki ularni tekshirishda, matnning originalligini o‘rganishda ham foydalanish mumkin. Odatda, konkordans xrestomatiya tuzuvchilari tomonidan ko‘proq ishlatiladi: u matnda ibora, idiomalarni qidirish, atamalarni tarjima qilishda muhim. Konkordansning evristik funksiyasi tilshunos yoki oddiy kitobxonga matnning yangi talqini, sharhini ko‘rsatishda, analitik funksiyasi esa tilning turli birliklari: leksema, kalit so‘z, ularning matnda qo‘llanish chastotasi ustida tahlil olib borish, indeksatsiya funksiyasi matnni nashrga tayyorlashda so‘zning ro‘yxati, indeksini tuzish, qiyoslash funksiyasi matn ichidagi barcha konnotatsiyalar, so‘z qo‘llash holatini taqqoslash imkonini beradi. Lug‘at maqolasi oddiy lug‘atdan farqli ravishda konkordans so‘zga misol keltiradi, ya’ni qo‘llanish kontekstini beradi. Bu bilan konkordans turli differensial, shuningdek, izohli lug‘atlar yaratishda asos vazifasini o‘taydi. F.M.Dostoyevskiy, F.I.Tyutchev, A.P.Chexov, A.S.Griboyedov, S.A.Yesenin, S.T.Aksakov, M.V.Lomonosov asarlari konkordansi<sup>150</sup> shunday ishlar jumlasidan hisoblanadi.

---

<sup>148</sup> Столяров А.И. Словарь-конкорданс и его применение в рамках корпусной лингвистики // Гуманитарные научные исследования. – 2017. № 2 [Электронный ресурс]. URL: <http://human.snauka.ru/2017/02/21074> (дата обращения: 29.09.2017).

<sup>149</sup> Труфанов М.В. Применение цифрового словаря с корпусом для работы над лингвистическими проектами по немецкому языку // Социально-экономические явления и процессы. 2011.– №8 С.229-232.

<sup>150</sup> “Частотный словарь-конкорданс публицистики Ф. М. Достоевского”. (Данный словарь создан на основе полного собрания сочинений писателя в 30-ти томах, включая Дневник писателя.); “Частотный словарь-конкорданс словоформ языка Ф.И.Тютчева”. (Создан по изданию: Тютчев Ф.И. Полное собрание стихотворений / Сост., подгот. текста и примеч. А.А. Николаева. – Л.: Сов. писатель, 1987. – 448 с.); “Метапоэтический словарь-конкорданс драматического текста А.П.Чехова”. (В словаре представлено исследование А.П. Чеховым собственного драматического творчества.); “Алфавитно-частотный конкорданс текстов Грибоедова”. (Словарь содержит все слова, встретившиеся во всех опубликованных текстах Грибоедова, взятых из наиболее авторитетных источников.); “Конкорданс к стихам А.С.Пушкина”. (Расширенный словарь языка писателя, в котором на каждое слово и словоформу Пушкина приводится полный текст всех стихотворных произведений.); “Частотно-семантический конкорданс служебных слов в поэзии Сергея Есенина”. (Словарь представляет собой первое полное лексикографическое описание служебных слов в лирике великого русского поэта С.А.Есенина.); “Конкорданс произведений С.Т.Аксакова”. (Может быть применим при реконструкции художественного мира С.Т.Аксакова,

Aytish mumkinki, konkordans an'anaviy leksikografiyaga nisbatan katta hajmli ma'lumotlar bazasini o'zida saqlash imkoniyatiga ega. Konkordans birligi matndagi paradigmatic munosabatni kuzatish, ijodkor merosining turli parchalarida so'z qo'llanish holatini miqdoriy tahlil qilish, unga xos sintaktik qurilmalar haqida xulosalar chiqarish kabi vazifalarni bajara oladi. Lug'at-konkordans – lingvistik, adabiyotshunoslikka oid tadqiqotlar bajarishda zamonaviy, samarali vosita, uni korpus lingvistikasining muhim ahamiyatli qismiga aylantiradi<sup>151</sup>.

**Mualliflik korpusining o'ziga xos xususiyati: tuzilishi, tarkibi va hosilasi (A.P.Chexov badiiy asarlari korpusi misolida).**

*A.P.Chexov asarlari korpusi razmetkasi.* Korpus tarkibiga A.P.Chexovning barcha tugallangan badiiy asarlari kiritilib, matnlarga yozuvchining mukammal asarlar to'plami<sup>152</sup> asos qilib olingan. Korpus tarkibida 600 ta matn (asar): 17 ta p'esa, 583 ta nasriy asar mavjud, 1 271 664 so'z qo'llash holati – kontekst (176 308 so'z p'esa, 1 095 356 so'z nasriy asardan), 101 282 turli so'zshakllar (23 093 tasi dramatik; 96 141 tasi nasriy asardan), 36 419 leksema (11 802 tasi dramatik; 34 763 tasi nasriy asar) mavjud. Bitta nom bilan atalgan dramatik, nasriy asar matni<sup>153</sup> alohida sanalgan. A.P.Chexov asarlari korpusi *lingvistik*: so'zshakl (1); so'zning boshlang'ich shakli (2); so'z turkumi (3); maxsus qo'llanishlar (4) hamda *ekstralingvistik*: asar nomi (1); sarlavha, bag'ishlov (2); asar yozilgan yil (3); davrlashtirish (4); matn tipi: nasriy yoki dramatik (5); imzo (6); asar janri (7); asar janrining muallif nuqtayi nazaridan turi (8); nutqiy janr (9); yozilish joyi (10); hikoyachi tipi (11); asar mavzulari ko'lami (12); asarning asosiy mavzusi (13); semantik guruh (Geo; Shaxs; Ism; Rang; Tovush) (14); mukammal asarlar to'plamida jild, sahifa raqami (15) kabi belgilar asosida razmetkalangan. *So'zshakl* sirasiga raqam, ajratib yozilgan xorijiy so'zlar (o'zlashmagan, boshqa tilda) ham

---

тематического своеобразия, мотивно-образной системы прозаических произведений писателя.); “Конкорданс к текстам Ломоносова”. (Является частью проекта по созданию электронного научного издания “Ломоносов”).

<sup>151</sup> Столяров А.И. Словарь-конкорданс и его применение в рамках корпусной лингвистики // Гуманитарные научные исследования. –2017. № 2 [Электронный ресурс]. URL: <http://human.snauka.ru/2017/02/21074>

<sup>152</sup> Чехов А.П. Полное собрание и писем в 30 т. – М.: Наука, 1974–1983.

<sup>153</sup> Чехов А.П. “Иванов” (комедия; пьеса). Полное собрание и писем в 30 т. Сочинения. Том 12. М., 1986.; “Иванов” пьеса. (1887)

kiradi. *Soʻzning boshlangʻich shakli* “soʻz” oynasi orqali topiladi. Undagi soʻzlarning berilishi lugʻat<sup>154</sup> dan biroz farq qiladi: 1) korpus soʻzligiga lugʻatdan farqli ravishda “Ivanov” (1987) dramasi ham kiritilgan; 2) funksional omonimiya (ravish/predikativ) indeks bilan ajratilmagan, lugʻatda esa bu indeks bilan belgilangan; qaysi turkumga kirishini bilish uchun soʻz turkumi haqidagi maʼlumotga qarash kerak boʻladi; 3) variantlilik haqidagi maʼlumot quyidagi usulda: asosiy variant, ikkilamchi variant yonma-yon keltirilgan. Masalan, *b*, *b*bi**, *b*. Asosiy variant keltirilsa, ikkilamchi variant keltirilmaydi. Qidirilayotgan birlik “soʻz” oynasidan topilmasa, “soʻzshakl” oynasiga murojaat etish maqsadga muvofiq. “Asar nomi” oynasi maʼlum bir asar bilan bogʻliq barcha maʼlumotni: asar yozilgan yil va joy, asar haqidagi fikrlar, asarga A.P.Chexov mukammal asarlar toʻplamida keltirilgan izohlar kiritiladi. “Sarlavha (qoʻshimcha sarlavha) va bagʻishlov” oynasidan asardagi ikkinchi sarlavha (agar mavjud boʻlsa), asarning kimga bagʻishlanganligi haqida maʼlumot olish mumkin.

*Tizimning asosiy vazifasi va imkoniyatlari.* Korpusning shartli ravishda “ИСТОК”<sup>155</sup> deb ataluvchi axborot-tadqiqiy tizimi mavjud. Bu tizim lugʻat, konkordansni koʻrib chiqish, alohida faylga nusxa koʻchirish (a); korpus razmetkalangan u yoki bu maʼlumot tipi boʻyicha foydalanuvchi tomonidan tanlangan pozitsiya yuzasidan yangi lugʻat, konkordans roʻyxati yaratish (b); barcha asarning toʻliq matni bilan ishlash rejimi, u haqida turli maʼlumot olish (v) kabi amallarni bajara oladi. Tizim uch rejimda ishlaydi: (1) “Konkordans”, (2) “Matn”, (3) “Turli maʼlumot beruvchi material”. Eng asosiy rejim “Konkordans”. SHu rejim orqali korpus boʻyicha qidiruv amalga oshiriladi – lugʻat, matn bilan ishlanadi, asarning asosiy matniga oʻtiladi, u orqali korpus maʼlumotlariga – razmetkaga yoʻl topiladi. Konkordansning tarkibiy qismi – lugʻat, matn turli oynalarda joylashgan, ular oʻzaro bogʻlangan. Konkordans roʻyxatidagi har bitta qator qidiruv natijasidagi birlik, u alohida matndan olingan illyustrativ misol. Konkordansda namoyon boʻlgan misol (kontekst) hajmi foydalanuvchi xohishiga

---

<sup>154</sup> Кукушкина О.В. и др. Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова с электронным приложением. – М., 2012.

<sup>155</sup> “ИСТОК” – “Исследование Словаря, Текстовых Особенностей, Конкордансов” сўзларининг бош харфидан олинган кискартма.

ko‘ra uzayishi, qisqarishi mumkin. Har bir konkordans korpus materialini u yoki bu shaklda so‘ralgan axborot asosida yuzaga chiqaradi. Ro‘yxatda misol keltirilgan matn (asar)ning korpusda joylashuvi ham aks ettiriladi. Konkordans lug‘ati 3 usulda: alifbo tartibi, chastotali, teskari tartibda ifodalanadi. Bu esa turli tipdagi lug‘at: alifbo-chastotali, chastotali, ters lug‘at tuzish imkonini beradi. Shu bilan birga kontekstlar foydalanuvchi talabiga ko‘ra turli tomondan saralanishi, lug‘at, kontekst, matnning to‘liq shakli buferga joylashtirilishi, alohida fayl xotirasiga saqlanishi mumkin. Ma‘lum bir matnga o‘tish ham shu rejim bilan amalga oshiriladi, buning uchun kerakli kontekst belgilanib, “*matn*” ilovasiga o‘tish yetarli. Agar aynan bir asar matniga o‘tish ehtiyoji tug‘ilsa, interfeys oynasi<sup>156</sup> dan “*Asar nomi*” oynasidan kerakli asar nomini tanlab, “*matn*” tugmasi bosilishi lozim. Ba‘zi konkordans birliklariga (“*so‘z*”, “*asar nomi*”, “*alohida qo‘llanishlar*”) avtomatik tarzda ma‘lumotlar ichidan qo‘shimcha ma‘lumot – lug‘at maqolasi keltiriladi. Konkordansda katta ro‘yxat paydo bo‘ladi; “*Axborot tipi*” menyusidan kerakli axborotni tanlasak, yangi konkordans kelib chiqadi, buning uchun “*filtr*” funksiyasini ishga tushiramiz. Bu foydalanuvchiga korpusdan faqat o‘z tadqiqoti uchungina kerakli ma‘lumotni saralab olish uchun yaxshi imkoniyat; katta ro‘yxatni o‘qib, saralab o‘tirishga ortiqcha vaqt, kuch sarflashga ehtiyoj qolmaydi. Tizimning ushbu funksiyasi bir-biriga o‘xshash elementlar tahlili imkonini beradi. Foydalanuvchi tadqiq etilayotgan birlikning o‘ng, chap tomonida joylashgan barcha elementlardan iborat konkordansni qo‘lga kiritishi mumkin. Bunday sintaktik qolip asosidagi aloqani “*Matnlarni saralash*” funksiyasidan foydalanib saralash mumkin. “*Ma‘lumot beruvchi materiallar*” (“Справочные материалы”) rejimi “izoh, personallik”, “semantik izoh” kabi ikki xil ma‘lumotlar bazasi bilan ish ko‘radi. Birinchisida A.P.Chexov asarlariga izohlar, yozuvchi bilan bog‘liq ayrim shaxslar haqidagi ma‘lumot jamlangan bo‘lsa, ikkinchisi asarning ajratilgan leksik birliklariga semantik izohni qamrab oladi. Bu semantik izohlar lug‘at<sup>157</sup> dagi semantik izoh bilan aynan bir xil.

---

<sup>156</sup> <http://www.philol.msu.ru/~lex/>

<sup>157</sup> Кукушкина О.В. и др. Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова с электронным приложением. – М., 2012.

*A.P.Chexov badiiy asarlari tili grammatik-semantik chastotali lug'ati.* Lug'at va elektron korpus yig'indisi yangi tipdagi leksikografik mahsulot. Bu mahsulotning ahamiyati lug'at bilan ishlashning yangicha uslubi, qulayligi, foydalanuvchining yozuvchi ijodi til xususiyatlari bo'yicha keyinchalik turli mustaqil tadqiqot olib borishi mumkinligi bilan belgilanadi<sup>158</sup>. Ushbu elektron nashr elektron formatdagi (diskda) lug'at hamda ilovadan iborat. Bu elektron nashr tarkibiga A.P.Chexov badiiy asarlari tili grammatik-semantik chastotali lug'ati<sup>159</sup>, A.P.Chexov badiiy asarlari matnlari korpusi<sup>160</sup>, korpusning dasturiy ta'minoti<sup>161</sup> hamda korpus, "ИСТОК" tizimi bilan ishlash uchun metodik ko'rsatma<sup>162</sup>lar kirgan. Korpus va lug'at Moskva Davlat universitetida<sup>163</sup> ishlab chiqilgan. Lug'at va Korpus tarkibiga A.P.Chexovning barcha tugallangan badiiy asarlari kiritilgan.

Lug'at 4 bo'limdan iborat. Birinchi bo'limda A.P.Chexov badiiy asarlaridagi leksik tarkib tavsifi berilgan, bu bo'lim 34000 dan ortiq birlikni (shuningdek, son, ro'yxatni raqamlovchi birlik, xorijiy tildagi ibora tarkibiy qismi ham) qamrab olgan. Leksik tarkibni tahlil qilishda leksik guruhlashning grammatik-semantik prinsipiga amal qilingan, tahlilning bu usuli so'zning muayyan turkumga mansubligi, semantik xususiyatiga asoslanadi. Lug'at boshida atoqli otlar berilgan, u shaxs nomi, hayvon laqabi, kema nomi, toponim, etnonim kabi guruhlariga bo'lingan. (1) Turdosh ot, (2) fe'l, kesimlik belgisiga ega bo'lgan so'z (predikativ), (3) sifat, ravish, ot predikativ, (4) son so'z turkumi. Asar matnida uchraydigan harf, nota nomi, so'z qismi kabi birliklar alohida guruhni tashkil qilgan. Lug'at chastotali xarakterga ega, chunki har bir birliklarning qo'llanish darajasi, miqdori umumiy hamda dramatik, nasriy asarlar kesimida ko'rsatiladi.

---

<sup>158</sup> Кукушкина О.В., Поликарпов А.А., Суровцева Е.В. Электронный корпус текстов художественных произведений А. П. Чехова: принципы организации и возможности лексикографического использования// Слово и словарь. Vocabulum et vocabularium. Сборник научных трудов по лексикографии. Вып. 12. Харьков – Клагенфурт – 2011. Под ред. В.В.Дубичинский.

<sup>159</sup> "Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова". Авторы: Кукушкина О.В., Суровцева Е.В., Лапонина Л.В., Рюдигер Д.Ю.;

<sup>160</sup> Электронный корпус художественных текстов А.П.Чехова, на базе которого создавался словарь. Создатели: Кукушкина О.В., Суровцева Е.В., Лапонина Л.В.;

<sup>161</sup> Программное средство для работы с корпусами текстов «Компьютерная информационно-исследовательская система для работы с электронными корпусами текстов "ИСТОК"». Разработчики: Федотов В.В., Кукушкина О.В., Поликарпов А.А.;

<sup>162</sup> Методическое руководство по работе с системой «ИСТОК» и корпусом текстов А.П.Чехова. Авторы: Кукушкина О.В., Суровцева Е.В., Поликарпов А.А., Варламов А.А., Федотов В.В.

<sup>163</sup> Лаборатория общей и компьютерной лексикологии и лексикографии (зав. лаб. А.А.Поликарпов; сотрудники кафедры русского языка филологического факультета МГУ (зав. каф. М.Л.Ремнёва).

Ikkinchi bo‘lim alohida semantik guruhlar tavsifiga bag‘ishlanadi. Bunday so‘zlar guruhini tuzishda lingvistik, adabiyotshunoslik hamda nutqiy vaziyat (kognitiv) nuqtayi nazaridan qiziqish uyg‘otadigan birliklarga (chunonchi, “ovozni bildiruvchi birlik”, “shaxs nomi” va h.) e‘tibor berilgan. Uchinchi bo‘lim A.Chexov asarlari leksik tarkibining miqdoriy tahlil natijasi ifodasidan iborat. Eng ko‘p (100 va undan ortiq asarda) uchraydigan birliklar ifodalangan, ular yozuvchi “leksikasi yadrosi” sifatida baholanadi. Shuningdek, ushbu bo‘limda, odatda, chastotali lug‘atda beriluvchi ko‘llanish miqdori katta bo‘lgan so‘zlar jadvali ham keltirilgan. To‘rtinchi bo‘limda lug‘atning elektron ilovasi tavsiflangan.

*Lug‘at maqolasining tuzilishi va tarkibi.* To‘liq lug‘at maqolalari birinchi bo‘limda berilgan. U quyidagi ma’lumotlarni o‘z ichiga olgan: (1) bosh so‘z; (2) uning grammatik xarakteristikasi; (3) miqdoriy ko‘rsatkichi; (4) varianti haqida ma’lumot (agar variantlari ham bir maqola tarkibida bo‘lsa); (5) semantik izoh. Birlik ma’nosini izohlash qiyinchilik tug‘dirgan paytda asar nomi ko‘rsatilgan holda qo‘llanish konteksti keltiriladi. Familiyalar semantik izohi sifatida shaxs haqida qisqacha ma’lumot hamda u tilga olingan asar nomiga havola beriladi.

**A.S.Pushkin, F.M.Dostoyevskiy, A.S.Griboyedov mualliflik korpuslarining o‘xshash va farqli tomonlari.** A.A.Polikarpov “Pushkin tili elektron ensiklopediyasi”<sup>164</sup>ning yaratilishi haqida to‘xtalar ekan, shungacha (2006-yil) nafaqat Pushkin asarlari korpusi, balki biror muallif ijodining shunga o‘xshash nashri yo‘qligini, ushbu nashr bu yo‘ldagi ilk tajriba ekanligi hamda mualliflik leksikografiyasi sekin rivojlanayotganligini afsus bilan ta’kidlaydi. Uning fikricha, mualliflik korpusi – jahon korpus lingvistikasida ham e‘tibordan chetda qolib ketayotgan soha. Shu kungacha yaratilgan mualliflik korpuslari (Petrozavodsk Davlat universiteti olimlari tomonidan ishlab chiqilgan F.M.Dostoyevskiy publitsistikasi korpusi) tahlilidan ma’lum bo‘ladiki, ular muallif til xususiyatining bir-ikki qirrasini namoyon etuvchi hamda elektron korpus sifatida ham tor imkoniyatli (asosan, ma’lumot beruvchi) sanaladi. Bu muammo jahon kompyuter leksikografiyasida ham o‘z yechimini topa olgani yo‘q. Jahon

---

<sup>164</sup> Қаранг: Поликарпов А.А.Корпусная информационно-исследовательская система// Вып. 1. Поэзия и драматургия А.С.Пушкина. Путеводитель по Пушкину”. – М., 2006.

mualliflik korpuslarida qidiruvning asosiy (hatto yagona) birligi – tavsiflanmagan (razmetkalanmagan), omonimlik hodisasi farqlanmagan so‘zshakl. A.S.Pushkin asarlari korpusi<sup>165</sup> mualliflik korpusi tuzishda muallif asarlari tili kompleks xususiyatlarini belgilashda yagona tajribaligicha qolib kelayotgan edi, “Pushkin tili elektron ensiklopediyasi”<sup>166</sup> esa bu borada jiddiy qadam bo‘ldi. A.Pushkin asarlari tili hozirgi rus adabiy tili, grammatik me‘yorlari shakllanishida asos vazifasini o‘taganligi sababli batafsil, chuqurroq o‘rganilishiga sabab bo‘ldi. Pushkin asarlari til xususiyatini aks ettiruvchi mavjud elektron nashrlar faqat so‘zshakl asosidagi qidiruv tizimi<sup>167</sup> yoki Pushkin asarlarining ba’zi izoh, illyustratsiyasigina mavjud elektron to‘plam<sup>168</sup>dan iborat. Lingvistik ma’lumotlar bazasini tuzish ehtiyoji shu paytgacha A.Pushkin asarlari tili yaxlit, har tomonlama tadqiq etilmaganligidan kelib chiqdi. Nafaqat A.Pushkin leksikasi, frazeologiyasi semantik xarakterlanmagan, balki A.Pushkin tili grammatik-morfologik, morfem, so‘z yasaliş, sintaktik xususiyatlari ham tizimlashtirilmagan. So‘z, ibora, umuman, so‘zlikning A.Pushkinning turli janrlari matnida ishlatilishi o‘ziga xosligi ham yaxlit tizim sifatida o‘rganilmagan. Bu muammolarni yaxlit holda A.Pushkin asarlari korpusini tuzish (har tomonlama lingvistik tahlil – muallif asarlari matnida uchraydigan so‘z, morfema, birikmani grammatik, semantik tasniflash) orqali yechish mumkin. Shunday xususiyatlarni o‘zida birlashtirgan A.Pushkin asarlari korpusi barcha mutaxassislar foydalanishi, mustaqil tadqiqot olib borishi hamda zaruriy ilmiy xulosalar chiqarishda ko‘maklashuvchi mukammal ma’lumotlar bazasi bo‘lishi zarur. Bunday xususiyatli korpus birligining (so‘z, hatto o‘zakning ma’nosi ham izohlanishi) keng lingvistik axborotga ega bo‘lishi lingvistik, til tarixi, til egasining tilni bilish, anglash jarayoni shakllanishi, evolyutsiyasi sohasidagi tadqiqotlar doirasini kengaytiradi. Ushbu korpus ma’lumotlar bazasi sifatida, shuningdek, Pushkin tili xususiyatlari tadqiqining yangi bosqichi,

---

<sup>165</sup> “КИИСа. Поликарпов А.А.Корпусная информационно-исследовательская система// Вып. 1. Поэзия и драматургия А.С.Пушкина. Путеводитель по Пушкину”. – М., 2006.

<sup>166</sup> “Электронная энциклопедия языка А.С.Пушкина”

<sup>167</sup> “Александр Сергеевич Пушкин. Полное энциклопедическое собрание сочинений”, изд. ИДДК.

<sup>168</sup> Мас., Мультимедиа-энциклопедия “А.С.Пушкин. В зеркале двух столетий”, изд. 1С.



A.Pushkin matnlari so‘z birikmasi, sintaktik birliklari va strukturasi tadqiqi uchun asos bo‘ladi.

*F.M.Dostoyevskiy matnlari korpusi.* F.Dostoyevskiy asarlari tili lug‘ati tuzish manbai sifatida yaratilgan, korpus tuzish parametrlariga lug‘at maqolasi tuzilishi, tarkibi asos bo‘lgan. Lug‘at F.Dostoyevskiy leksikasini to‘liq qamrab olishni maqsad qilganligi sababli korpusga muallifning barcha badiiy, publitsistik asarlari kiritilgan. Unda matn saqlanish birligi alohida asar (romandan tortib xatgacha) shaklida belgilangan. Bugungi kunda barcha elektron tashuvchi vositalarda uning badiiy asarlari, publitsistikasi, epistolyar merosi mavjud; asar, sahifaga havola qiluvchi minimal teg bilan ASCII formatida saqlanadi. Korpus asosida bir necha lug‘at yaratilgan, A.Ya.Shaykevich tomonidan tuzilgan chastotali lug‘at shular jumlasidan. Muallif idiomalari lug‘atiga asos bo‘lgan idiomatik ma’lumotlar bazasi ham tuzilgan. F.Dostoyevskiy korpusi CD-rom shakli (“Достоевский: тексты, исследования, материалы”) ga ham ega.

*Lug‘at tuzish tamoyillari.* F.M.Dostoyevskiy lug‘at-konkordansida muallifning 30 jildlik mukammal asarlari to‘plamidagi barcha matnlar (badiiy asar, publitsistika, yozuvchi “kundaligi”)dagi istalgan so‘z, so‘zshaklning ishlatilish konteksti, chastotali xususiyati, ularning asosiy matniga o‘tish imkoniyati mavjud. Matn sifatida barcha tugallangan asarlari olingan, tugallanmagan yoki muallif qalamiga mansubligi aniq bo‘lmagan asarlar (mas., “Ряд статей о русской литературе” – “Вопрос о университетах”), 30 jildlik mukammal asarlar to‘plamidagi muharrir izoh, havolalari, yozuvchi yondaftari bitiklari, qoramalar kiritilmagan. A.P.Chexov asarlari korpusida esa mukammal asarlar to‘plamidagi muharrir izohlariga ham korpus birligi sifatida qaralgan. Ushbu lug‘at-konkordans F.M.Dostoyevskiy asarlari korpusi, chastotali lug‘ati uchun asosiy manba vazifasini o‘tagan. Korpus uchun matnni tayyorlashning barcha bosqichi bajarilgan; uning asosiy birligi so‘zshakl va lemma hisoblanib, yozuvchining o‘z so‘zi hamda sitata sifatida asar matniga kiritgan so‘zlar chastotasi aniqlangan. So‘z birikmasi sifatidagi atoqli otlar bitta birlik sifatida lemmalashtirilgan (“Русский вестник” jurnali nom sifatida beriladi; qidiruv ham shu tartibga amal qiladi), bu

esa muallifning u yoki bu nashr, asarga murojaat qilish chastotasini aniqlashga yordam beradi. Lug‘at-konkordansning yana bir afzalligi so‘z konteksti, uning manzilini aniqlay olishidadir. Foydalanuvchi qidiruvga berilgan so‘zshaklni o‘ng, chap tomonidan o‘nlab so‘zlar kontekstida ko‘rishi, tezda asosiy matni (korpusga asos bo‘lgan 30 jildlik asar matni) tanlab olishi mumkin. Bu esa matni elektron nashr matni bilan qiyoslash imkonini beradi. Korpusning o‘ziga xos xususiyati sifatida quyidagilarni ta’kidlash mumkin:

1. Asosiy reestrda so‘z bosh harflarda berilgan (“Белинский” so‘zi “БЕЛИНСКИЙ” shaklida), ammo so‘zni qidiruvga berishda harf shakli ahamiyatsiz.

2. Texnik sabablarga ko‘ra matnning asl manbasida uchraydigan kursiv so‘zlar neytrallashtirilgan.

3. Leksik omonimlar turli lug‘at maqolasida emas, bitta maqolada berilgan (foydalanuvchi kontekstga qarab omonimni bemalol aniqlay oladi).

4. Murakkab bog‘lovchi, iboralar tarkibga ajratilgan.

5. Qisqartma, abbreviatura izohlanmagan, raqamlar so‘zga aylantirilmagan.

6. Ishlatilish chastotasi baland bo‘lgan, qisqartilib yozilgan jurnal, gazeta nomi izohlangan, lemmalashtirilgan (Mas. Sovr. – “Современник”, Вест. Европы – “Вестник Европы”).

7. Defisli so‘z normallashtirilgan: bez-dar-nogo – “bezdarный”.

8. So‘zlar orfografiyasi o‘zgartirilmagan (muallif davri orfografik shakli saqlab qolingan: “parke”, “Toloza” lemmasi (zamonaviy “parket”, “Tuluza”).

9. Xorijiy so‘zlar (o‘zlashmagan, boshqa imlodagi) lemmalashtirilmagan (m., fr.n'est), so‘z turkumi razmetkasi qo‘yilmagan.

### **avol va topshiriqlar**

1. Mualliflik korpusi tuzishning lingvistik asoslari nimalardan iborat?
2. Mualliflik korpuslarining mushtarak va farqli jihatlarini sanab bering.
3. A.P.Chexov, A.S.Pushkin, U.Shekspir mualliflik korpuslarining o'ziga xos xususiyati nimada ko'rinadi?
4. Firdavsiy, Rumi, Sa'diy Sheroziy mualliflik korpuslari razmetka darajasi bilan G'arb mualliflik korpuslaridan qanday farq qiladi?
5. Qaysi mualliflik korpuslari o'z parametrlari bilan ustunlikka ega?
6. Mualliflik korpusining maqsad va vazifalarini sanab bering.
7. Korpus uchun material tanlashda qanday muammolar mavjud? Mutaxassislar ularga qanday yechim taklif qilmoqda?
8. Mualliflik leksikografiyasi va mualliflik korpusining qanday bog'liq tomonlari bor?
9. Mualliflik leksikografiyasi korpusning hosilasimi yoki asosimi?
10. F.M.Dostoyevskiy, F.I.Tyutchev, A.P.Chexov, A.S.Griboyedov, S.A.Yesenin, S.T.Aksakov, M.V.Lomonosov asarlari konkordansi qanday xususiyatlarga ega?
11. Mualliflik korpusining o'ziga xos xususiyati: tuzilishi, tarkibi va hosilasi haqida gapiring.
12. A.P.Chexov asarlari korpusi razmetkasi qanday afzalliklarga ega?
13. A.P.Chexov badiiy asarlari tili grammatik-semantik chastotali lug'ati qanday tuzilishga ega?
14. A.S.Pushkin, F.M.Dostoyevskiy, A.S.Griboyedov mualliflik korpuslarining o'xshash va farqli tomonlari nimalarda ko'rinadi?

#### **Tavsiya etilgan adabiyotlar:**

1. Бобунова М.А. Конкорданс XXI в.: новая старая форма // Вопр. лексикографии. – 2016. №2 (10) С.41-54.
2. Волков С.С., Матвеев Е.М. “О проекте словаря “Риторика М.В.Ломоносова” // [www.ruslang.ru/seminar\\_aut\\_lexocorg250214](http://www.ruslang.ru/seminar_aut_lexocorg250214))
3. Гёрн А. Параллельный корпус в системе университетских лингвистических курсов. [http://folk.uio.no/atleg/nizhnij\\_gronn\\_oct2013.pdf](http://folk.uio.no/atleg/nizhnij_gronn_oct2013.pdf)

4. Гик А.В. “О работе “Конкордансом к стихотворениям М.Кузмина””: том четвертый” (Электрон ресурс: [www.ruslang.ru/seminar\\_aut\\_lexicorg](http://www.ruslang.ru/seminar_aut_lexicorg))
5. Грудева В.А. Корпусная лингвистика. Учебное пособие. 2-е изд., стереотип. – М.: Флинта, 2012. – 165 с.
6. Добровольский Д. Какие задачи решают параллельные корпуса. (Электрон ресурс): <https://postnauka.ru/video/54851>
7. Добровольский Д.О. Корпус параллельных текстов в исследовании лексической семантики: <http://www.dialog-21.ru/media/2503/dobrovolskij.pdf>
8. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
9. Кокорева А.А. Корпус параллельных текстов в обучении иностранному языку. – Вестник ТГУ (Тамбовский государственный университет), выпуск 2 (118), 2013. – С. 57-62.
10. Кукушкина О.В., Поликарпов А.А., Суровцева Е.В. “Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова (с электронным приложением) как особый лексикографический продукт” // [www.ruslang.ru/seminar\\_aut\\_lexicorg](http://www.ruslang.ru/seminar_aut_lexicorg)
11. Потемкин С.Б. Авторский корпус и словарь языка Антона Чехова. (Электрон ресурс: // <https://istina.msu.ru>.)
12. Слободян Е.А. “Идеографический словарь языка поэмы А.Мицкевича “Пан Тадеуш” в авторской и электронной лексикографии”. (Электрон ресурс: [www.ruslang.ru/seminar\\_aut\\_lexicorg](http://www.ruslang.ru/seminar_aut_lexicorg))
13. Соснина Е.П. Параллельные корпуса в обучении языку и переводу// [http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus\\_education\\_translation/](http://ling.ulstu.ru/linguistics/resources/literature/articles/corpus_education_translation/)
14. Столяров А.И. Словарь-конкорданс и его применение в рамках корпусной лингвистики // Гуманитарные научные исследования. – 2017. № 2 [Электронный ресурс]. URL: <http://human.snauka.ru/2017/02/21074> (дата обращения: 29.09.2017).

15. Тао Ю., Захаров В.П. Разработка и использование параллельного корпуса русского и китайского языков // Автоматизация обработки текста. (Электрон ресурс): <https://www.researchgate.net/publication/306911853>
16. Труфанов М.В. Применение цифрового словаря с корпусом для работы над лингвистическими проектами по немецкому языку // Социально-экономические явления и процессы. 2011.– №8 С.229-232.
17. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
18. Хуснутдинов А.А. Словарь языка А.Н.Островского” как один из вариантов авторского словаря” //
19. Хамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
20. Якубовский К.И., Якубовская К.А. Обзор современных лингвистических технологий и систем // Вестник МГУП. 2015. – №2 С.315-319.
21. [www.ruslang.ru/seminar\\_aut\\_lexocorg020413](http://www.ruslang.ru/seminar_aut_lexocorg020413) сайтидаги “Теория и практика авторской лексикографии” семинар материаллари.

### III BOB. KORPUS TUZISH TEXNOLOGIYASI

#### Korpus tuzish: bosqichlar va reprezentativlik.

##### Reja:

1. Korpusni loyihalash va tuzish texnologik jarayoni.
2. Matnni avtomatik o‘qilish shaklida qayta ishlash.
3. Tahlil matnga dastlabki ishlov berish. Konversiyalash va grafematik tahlil.
4. Til korpusida reprezentativlik.
5. Korpus birliklarining janriy va miqdoriy ko‘rsatkichi.
6. Badiiy, dialektal, suhbat matnlari. Folklor, dramatik, publitsistik, she‘riy matnlar praporsiyasi.

**Tayanch so‘z va iboralar:** *reprezentativlik, janriy xususiyat, miqdoriy ko‘rsatkich, badiiy, dialektal, suhbat matnlari, matnlar praporsiyasi, korpusni loyihalash, texnologik jarayon, matnni avtomatik o‘qilish shaklida qayta ishlash, matnga dastlabki ishlov berish, konversiyalash, grafematik tahlil.*

**Korpusni loyihalash va tuzish bosqichining texnologik jarayoni.** Korpus loyihasi – uni tuzish bosqichi, keyinchalik takomillashtirish yo‘llarini qamrab olishi mukammal korpus yaratishning asosiy bosqichi. Korpus tushunchasi, ta’kidlaganimizdek, tilshunos uchun an’anaviy kartotekalarning yangi shakli; ular XX asrga kelib kompyuterlashtirildi, undan ommaviy foydalanish imkoniyati paydo bo‘ldi. Kartotekalarning korpusga aylanishida, albatta, internet tarmog‘i salmoqli ahamiyat kasb etdi. Natijada, turli lingvistik tadqiqotlar olib borish imkonini beruvchi katta hajmli matnlarning umumiste‘mol varianti paydo bo‘ldi. Bu borada lug‘at, grammatikalar uchun asos vazifasini o‘taydigan til materialining ko‘lami hamda balansi masalasi kun tartibiga chiqib, xususan, milliy korpuslar yaratish jarayonida ko‘ndalang turdi. Korpusning reprezentativlik masalasi matnlar yetarliligi, xilma-xilligi bilan hal etildi. V.P.Zaxarov va S.Yu.Bogdanovanning fikricha, korpusning janriy-mavzuviy tuzilishi ko‘rib chiqilayotganda korpus matni sifatida qanday birlikning olinishi muammosiga alohida e’tibor qaratish lozim bo‘ladi<sup>169</sup>. Masalan, gazetalardagi kichik reklama matni alohida matn sifatida

---

<sup>169</sup> Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – Б.36.

qaraladimi yoki ularni bir matnga birlashtirish lozimmi? Gazeta maqolasi matn sanaladimi yoki gazetaning bitta sonini yaxlit matn sifatida baholash kerakmi? Har bir she'r bitta matnmi yoki she'riy to'plamni yaxlit holda kiritish kerakmi? Bir-biriga javob tarzida yozilgan, mohiyatan bir mavzu muhokama qilingan nashr etilgan maktublar bitta matnmi yoki alohida korpus birligi sifatida yondashish lozimmi? Bu savollarga tuzuvchi korpusning turi, keyinchalik bajaradigan vazifasidan kelib chiqib javob beradi. Milliy korpus yoki maxsus korpus ekanligiga qarab korpus birligi belgilanadi. V.P.Zaxarov, S.Yu.Bogdanova korpusni loyihalashtirish jarayonining muhim jihati sifatida xronologiya masalasini ham keltiradilar. Masalan, tilning zamonaviy korpusi deganda nima tushunilishi lozim? Turli janrlarda korpusning xronologik chegarasi turlicha bo'lishi tabiiy. Korpus keng omma foydalanishi hamda xilma-xil topshiriqlar bajarilishi uchun (jumladan, boshqa grafika asosida rus tilida yozilgan matnlarni o'rganish uchun) tuziladi. Korpusda matnning boshlang'ich shaklidan qanday qism olinib, nimalar chiqarib tashlanishi – yana bir e'tiborga molik masala. Matn tarkibida mavjud bo'lgan rasmlar til materialiga tegishli bo'lmaganligi uchun korpus tarkibiga kirgan matndan chiqarib tashlash, jadvallarni korpusga moslab qayta ishlash ham muhim. Ular matnning mazmunini ifodalashda ahamiyatli, lekin korpus tarkibida qoldirilsa, razmetkalashda qiyinchilik tug'diradi. Sitata, ko'chirma gap, o'zlashma birlik (atama), o'lchov birlik ham alohida e'tibor talab qiladi. Sanab o'tilgan masalalar loyihalashtirish bosqichida ma'lum prinsip asosida hal etilsa, ayrimi korpus tuzish jarayoni, korpusdan foydalanishda hal etiladi. Shu bilan birga, korpusni ishga tushirishdan oldin foydalanuvchi bilan qaytar aloqani ham nazarda tutish lozim. Mutaxassislar korpus tuzishning texnologik jarayonida quyidagi bosqichlarni ajratishadi<sup>170</sup>:

1. Belgilangan manbaga muvofiq holda matnning korpusga kirishini ta'minlash.

2. Matnni avtomatik o'qilish shaklida qayta ishlash. Korpusga kiritiladigan elektron shakldagi matn turli usul bilan olingan bo'lishi mumkin: qo'lda terilgan,

---

<sup>170</sup> ҚараиГ: <http://www.uni-koeln.de/phil-fak/englisch/bald/corpora>.  
[http://www.essex.ac.uk/linguistics/clmt/w3c/corpus\\_ling/content/history.html](http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/history.html).

skanerlangan, mualliflik nusxasi, hadya, ayirboshlash, internet, nashriyotlar tomonidan korpus tuzuvchisiga beriladigan original-maketlar.

3. Tahlil, matnga dastlabki ishlov berish. Ushbu bosqichda turli manbalardan qabul qilingan matnlar filologik tekshiruv, tahrirdan o'tadi.

4. Konversiyalash, grafematik tahlil. Ba'zi matnlar qayta kodlashtirish jarayoni amalga oshadigan ilk mashina ishlovidan qayta-qayta o'tadi, nomatniy qismlar (rasm, jadval) o'chiriladi yoki o'zgartiriladi. Matndagi bo'g'in ko'chirish, chegaralar (MS-DOS matnlarida) bekor qilinadi, tire, boshqa belgilar bir xilligiga erishiladi. Grafematik tahlil korpusga kiruvchi matnni qismga (so'z, bog'lovchi) ajratish, nomatniy elementni o'chirish kabi amallarni bajarishdan iborat.

5. Nostandart (noleksik) elementni belgilash, rasmiylashtirish, maxsus matniy elementni (qisqartma asosida yozilgan nom (ism, familiya), boshqa alifboda yozilgan o'zlashma leksema, rasmga berilgan nom, izoh, zarvaraq, adabiyotlar ro'yxati va b.) bir xil mezon asosida qayta ko'rib chiqish. Albatta, bu amallar avtomatik ravishda matn muharriri tomonidan bajariladi.

Korpusni loyihalashtirishning keyingi bosqichi manbani saralash hisoblanadi. Korpusning ahamiyati uning bir tildagi keng ko'lamli matnlarni bir joyga yig'ib, tartib berilganida emas, shu sababli uni tuzishda bir necha mezon asosida ish ko'riladi. Korpus materialini saralashda korpusning asosiy birligi nimadan iboratligi, uning hajmi qanday bo'lishi (unda qancha so'z bo'lgani ma'qul), yozma matn qaysi manbaga asoslanishi, qancha miqdorda bo'lishi, unga kiruvchi matn tilning qaysi sohasiga tegishli bo'lishi kabi masalalarga yechim topiladi. Ushbu savolning ilk javobi 1965-80-yillar oralig'ida R.G.Piatrovskiy, uning shogirdlari tomonidan berilgan edi. Ular chastotali lug'at, lingvostatistik tadqiqot o'tkazish uchun matn tanlash tamoyillarini tuzib chiqishgan. Bu muammo L.N.Zasorina tahriri ostidagi chastotali lug'at<sup>171</sup> so'z boshisida ham ko'tarilgan. O'shanda ilk marotaba matn tanlashning statistik usuli, hajmi, miqdori kabi omillar sanab o'tilgan. Korpusning asosiy birliklari so'zshakl, o'zak (negiz, lemma) va gap. Tuziladigan korpus hajmi korpusning maqsadidan kelib chiqib belgilanadi.

---

<sup>171</sup> Частотный словарь русского языка. (Под ред. Л.Н.Засорина.) – М.: Русский язык, 1977.– 936 с.



Agar u harf, harfiy birikma, tovush, diftonglarni tadqiq etishni maqsad qilgan bo'lsa, u qadar katta bo'lishi shart emas. Matnning leksik birliklari, morfologik hodisa, sintaktik, uslubiy xosligini tadqiq etish maqsadida tuzilsa, katta hajm talab etiladi. S.A.Sharovning fikricha<sup>172</sup>, saralash jarayonida qaysi janrga oid matni (nasr, drama, she'riyat, ilmiy matn, gazeta, jurnal material va h.) tanlash, matnning qaysi davrni (zamonaviy, 10 yillik, 50 yillik va mumtoz matn) qamrab olishi, matn faqat adabiy tilda bo'lishi yoxud boshqa manbalar ham kirishi kabi masalalar ham muhim ahamiyat kasb etadi. Korpus tuzuvchisi bu jarayonda, albatta, tilshunos, lingvostatistika mutaxassisi yoki anketa metodiga murojaat etadi. Korpus tuzish jarayonida muallif o'z tajribasiga tayangan holda korpusning umumiy hajmi, matnning nashr vaqti, matn soni, elementar tanlov hajmi, tanlanadigan janr xili, turini asosiy omil hisoblaydi. So'rovnoma usuli "Amerika meros korpusi" ("The American Heritage Intermediate Corpus") tuzuvchilari tomonidan qo'llangan. 5 million so'zshakl hamda ingliz tilida 22 turdagi bolalar, o'smirlar janriga oid matnlar kiritilgan. AQSHning 221 ta maktabiga qanday matni tanlash maqsadga muvofiqligini aniqlovchi so'rovnoma yuborilgan. So'rovnoma natijasi o'rganilgach, 19 000 nomdagi kitoblar ro'yxati tuzilgan. Bu asosda har biri 500 so'zshakldan iborat 1045 matn tanlab olingan. Xulosa sifatida korpusni loyihalashtirish jarayonida material (matn) tanlash, saralash, uni texnik jihatdan korpusga moslashtirish eng asosiy bosqich ekanligini qayd etish joiz.

**Korpusning eng asosiy xususiyati reprezentativligi.** *Reprezentativlik* atamasi ostida zaruriy hajm, miqdor, janr, uslub, mualliflarga tegishli matnlar proporsiyasining to'g'ri tanlanishi tushuniladi. Reprezentativlik xususiyatini korpus tuzish jarayonida ham, undan foydalanish jarayonida ham saqlab turish korpusning ahamiyatini ta'minlovchi asosiy omil hisoblanadi. Korpus ma'lum tilning belgilangan davrda mavjud bo'lgan barcha turdagi matn namunalarini o'zida jamlagan bo'lishi, shu bilan birga, hajm jihatdan to'g'ri proporsiyani tashkil etishi lozim. Demak, korpus tilning xususiyatlarini o'rganishda qanchalik ahamiyatli bo'lmasin, uning faqat badiiy asarlar bilan chegaralanib qolishi to'g'ri

---

<sup>172</sup>Каранг: Шаров С.А. Представительный корпус русского языка в контексте мирового опыта. (Электрон ресурс. <https://lamb.viniti.ru>)

emas. U turli mavzudagi (ijtimoiy-siyosiy mavzudan toki sport mavzusigacha) gazeta, jurnal materiali, fanning barcha sohasiga tegishli ilmiy, ilmiy-ommabop asarlar, reklama, shaxsiy yozishmalar matnidan tarkib topgan bo‘ladi. Tilda qanday jarayon kechayotganligini aniq tasavvur qilish uchun korpus qamrovini yanada kengaytirish, yozma nutq bilan birga og‘zaki nutq materiallaridan ham foydalanish maqsadga muvofiq. Bunday korpus yordamida til taraqqiyoti natijasida sodir bo‘lgan, kutilayotgan o‘zgarish haqida aniq xulosa chiqarish mumkin bo‘ladi<sup>173</sup>.

Korpus ma’lumotining ishonarliligini ta’minlovchi omil – uning representativligi. Bu xususiyatga korpus hamjmining kattaligi, matnlarning turli janrga mansubligini ta’minlash bilan erishib bo‘ladi. Nutqiy faoliyat, haqiqatan ham, rang-barang, turli shakllarda (og‘zaki, yozma, bosma va h.) namoyon bo‘ladi hamda nutqiy voqelanishlarning cheksizligini ta’minlaydi. XX asrning 60-yillarida yaratilgan korpuslar birinchi turga mansub, ular universallikka intilgan. Chunonchi, Braun korpusi shunday korpuslar sirasiga kiradi. Representativ korpus uchun tanlanadigan matnlar, ta’kidlaganimizdek, 15 turdagi janr(registr)ni qamrab olishi, ulardan 6 tadan 80 tagacha elementar belgilarga ko‘ra tanlanishi lozim. Matnlar, odatda, quyidagi janrlarda bo‘ladi:

- 1) matbuot: reportaj;
- 2) matbuot: bosh maqola;
- 3) matbuot: obzor;
- 4) diniy matnlar;
- 5) ko‘nikma, mashg‘ulot, xobbi;
- 6) ilmiy-ommabop adabiyot;
- 7) belletristika, biografiya, esse;
- 8) turli (boshqaruv hujjatlari, korxonalar, tashkilot, ishlab chiqarish hisoboti),
- 9) ilmiy asar;
- 10) badiiy adabiyot;
- 11) mistika, detektiv;
- 12) ilmiy proza;

---

<sup>173</sup> [www.ruscorgora.ru/](http://www.ruscorgora.ru/) Плунгян В. Зачем мы делаем корпусы?

- 13) sarguzasht, memuar;
- 14) ishqiy romanlar;
- 15) hajviy asarlar.

Ikkinchi tipdagi korpuslarning reprezentativlik mezoniga tuzuvchilarni qiziqtiruvchi hodisani maksimal obyektiv ko'rsatish omili xizmat qiladi. Masalan, ingliz maqollari korpusi ingliz tili egasi nutqida maqollarning maksimal qo'llanishini ma'lum davr hamda geografik region bo'yicha ko'rsatadigan reprezentativlik holati ingliz siyosiy metaforasi korpusi uchun bu xususiyatni aks ettirmaydi. Korpusning tur, xillari qanchalik ko'p bo'lmasin, ularni ikki tomondan tasnif qilish kerak.

1. Korpusni butun tilga (odatda, tilning ma'lum davriga), ma'lum bir tilning ichki (janr, uslub, ma'lum bir ijtimoiy yoki yoshga oid guruh, yozuvchi yoki olim tili) korpusiga qarshi qo'yish.
2. Lingvistik teg turiga asosan guruhlash. Razmetkaning bir qancha ko'rinishi bo'lishiga qaramay, juda ko'p korpuslar morfologik, sintaktik razmetkaga ega.

### **Savol va topshiriqlar**

1. Til korpusining reprezentativligi deganda nima tushuniladi?
2. Korpus tarkibiga kiruvchi mantlarning janriy xususiyati nimalardan iborat bo'ladi?
3. Korpus birliklarining miqdoriy ko'rsatkichi va ularning tengligi korpusning reprezentativligida nechog'lik ahamiyatga ega?
4. Badiiy, dialektal, suhbat matnlari proporsiyasiga qanday erishiladi?
5. Folklor, dramatik, publitsistik, she'riy matnlar proporsiyasi qanday tuziladi?
6. Korpusni loyihalash va tuzishning texnologik jarayoni qanday bosqichlardan iborat?
7. Matnni avtomatik o'qilish shaklida qayta ishlash qanday natija beradi va bu jarayon uchun qanaqa dasturiy ta'minot kerak?
8. Tahlil matnga dastlabki ishlov berish jarayoni nimalarni talab qiladi?
9. Konversiyalash va grafematik tahlil bosqichini tavsiflang.

### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Недошивина Э.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб., – 2006. – 26 с.
3. Структурно-функциональная компьютерная модель татарских морфем. – Казань: ФЭн, 2003. – 220 с.
4. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. –М.: ФЛИНТА: Наука, 2013. –128 с.
5. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
6. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.
7. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.
8. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
9. Шаров С.А. Представительный корпус русского языка в контексте мирового опыта // <https://lamb.viniti.ru>
10. [www.ruscorpora.ru](http://www.ruscorpora.ru)// Плунгян В. Зачем мы делаем корпусы?
11. <http://www.ruscorpora.ru>. НКРЯ
12. [www.philol.msu.ru/~lex/corpus](http://www.philol.msu.ru/~lex/corpus)
13. <http://www.philol.msu.ru/~lex/corpus>
14. <http://ruscorpora.ru>

## TABIIY TILNI QAYTA ISHLASH. RAZMETKA TEXNOLOGIYALARI.

### Reja:

1. Tokenizatsiya, lemmatizatsiya, stemming, parsing.
2. Teglashning korpus tuzishdagi ahamiyati.
3. Teg turlari va xususiyatlari.
4. Matnning avtomatik razmetkasi: tagging va parsing.

*Tayanch so‘z va iboralar:* tokenizatsiya, lemmatizatsiya, stemming, parsing, teglash, razmetka, annotatsiya, razmetka turlari.

**Lemmatizatsiya** (lemmatization) – so‘zshakllarni uning lug‘atdagi oddiy shakli – lemmaga keltirish jarayoni, ya’ni so‘zshakl lemma deb atalmish ta’minotdagi so‘zning asosiy yoki lug‘at shakliga qaytariladi. Bu jarayon ayrim rus va xorij manbalarida normallashtirish deyiladi. O‘zbek kompyuter lingvistikasining lingvistik tahlil yo‘nalishida esa otlar uchun bosh kelishik, birlik son (kitoblarim – kitob), fe’llar uchun aniq nisbat, bo‘lishli, fe’lning II shaxs buyruq-istak mayli (kelmagan – kel) oddiy shakllar hisoblanadi. Flektiv tillarda lemmatizatsiya qiyin jarayon sanaladi, chunki flektiv tillarda asos va affiks chegarasi yo‘qoladi. Shu bois muayyan so‘zshaklning lemmasi (asosi)ni topish birmuncha qiyinlashadi. Masalan, *водитъ – вожу, нуть – нью – ньѣщ – ней; man – men, tooth – teeth* kabi.

Stemming (stemming) – kiritilgan so‘zning asosi (o‘zak)ni topish jarayoni. Bunda topilgan so‘z asosi morfologiyada qabul qilingan so‘z o‘zagiga mos kelishi talab qilinmaydi. Zotan, dastur ta’minotiga “tayyor qolipli so‘zlar” asos sifatida alohida kategoriya qilib kiritilish ehtimoli yuqori bo‘ladi. Masalan, *huquq, tadqiq* kabi leksemalarga egalik affikslari qo‘shilganda leksema so‘ngidagi **q** undoshi o‘zgarishga uchramagani bois bunday istisnoli leksemalar “tayyor qolipli so‘zlar”, ya’ni “Avtomatik hosila” kategoriyasiga *huquqi, tadqiqi* tarzida kiritiladi va ular stemmingda asos, deb qabul qilinadi. Kompyuter lingvistikasida so‘zning o‘zagini topish vazifasi uzoq yillar davom etayotgan muammo hisoblanadi. Stemming, asosan, foydalanuvchining so‘rovi bo‘yicha izlanishni kengaytirish maqsadida qidiruv tizimlari uchun qo‘llaniladi, matnni me’yorlashtirish jarayonining qismi

hisoblanadi. Soʻz asosini topishning muayyan usuli stemming algoritmi, uning amalga oshirilishi, yaʼni dasturning oʻzi **stemmer** deyiladi.

Maʼlumki, matnda bir leksemaning turli soʻzshakllari qoʻllanilishi mumkin. Masalan, ingliz tilidagi *organize, organizes, organizing*, oʻzbek tilidagi *kelyapti, kelgan, kelmoqdalar* va h.k. Shuningdek, ifoda va mazmun planiga koʻra yaqin soʻzlar oilalari ham mavjud. Masalan, *democracy (demokratiya) – democratic (demokratik) – democratization (demokratlashtirish)*. Lemmatizatsiya va stemming yaqin ifodaga ega soʻzlarga morfemalarning birika olish imkoniyatini ularning asosini topish orqali aniqlaydi. Stemming va lemmatizatsiyaning maqsadi soʻzshakllarini umumiy asos shaklga keltirishdir. Masalan, ingliz tilidagi *cars, car's, cars'* soʻzshakllarini *car* asos (lemma, stem)ga keltirish kabi. Jarayon bir xil koʻrinsa-da, stemming lemmatizatsiyadan farq qiladi. Stemming, odatda, taxminiy jarayon deyiladi. Zero, stemmingda soʻzshakllar oxiridan boshlab lingvistik taʼminotga kiritilgan asos shaklga qadar kesib kelinadi. Koʻp hollarda bu holat oʻzini oqlagan. Lemmatizatsiya soʻzlar leksikonidan foydalanib, ularning morfologik tahlilini amalga oshiruvchi aniq jarayon boʻlib, bunda faqat fleksiyaga uchragan affikslar oʻchiriladi va lemma deb atalmish taʼminotdagi soʻzning asosiy yoki lugʻat shakliga qaytariladi.

Yuqorida aytilgan farqni faktik misolda koʻramiz: ingliz tilidagi *saw* leksemasi stemmingda *s* harfiga qadar kesib, tekshiriladi, lemmatizatsiya esa uni *see* (feʼl yoki otga mansubligiga qarab) feʼlining oʻtgan zamon shakliga qaytaradi. Koʻrinadiki, stemming bir asosdan yuzaga kelgan soʻzshakllardagi belgilar bilan ishlaydi, lemmatizatsiya esa bir lemmaning flektiv (affiks qoʻshilishi natijasida oʻzgarishga uchragan) shaklini eʼtiborga oladi. Tokenizatsiya ingl. *tokenizing* soʻzidan olingan boʻlib, informatiklar tomonidan qoʻllanishga kiritilgan termin hisoblanadi va tilshunoslikda leksik tahlil birikmasi bilan tushuntiriladi. Tokenizatsiya – elektron matndan birliklarni, yaʼni kiritilayotgan belgilar, ramzlar va soʻzlarni ajratib chiqish jarayoni. Mazkur texnologiya grafematik tahrir bosqichida muhim oʻrin tutadi. Lingvistik tahlil dasturlash tillarining manba kodi tarjimalari va kompilyator (oʻzgarlar tadqiqot natijalarining mustaqil yoki

qo‘shimcha ishlovlarsiz berilishi)larida, shuningdek, tabiiy tillarning turli parser – dasturlarning LTLarini yaratishda qo‘llaniladi. Parser (ingl. parser; parse – tahlil qilish) yoxud sintaktik analizator – dastur qismi yoki faqat sintaktik tahlilni amalga oshirish uchun yaratilgan maxsus dastur. Kirish ma’lumotlarini (matn) muayyan formatga keltirish orqali tahlil qiladi. Parserda quyidagilar asosiy o‘rinda turadi: tasniflangan leksemalar ketma-ketligi, ierarxik tuzilmalar, ma’lumotlar jadvali, tezauruslar va h.k. Sistema yoxud texnologiyaning qanday nomlanishidan qat’i nazar, foydalanuvchi uchun matni qayta ishlovchi mahsulotning yaratilgani va uning sifat darajasi muhimdir. Chunki matn ustida ishlash, ya’ni rasmiy hujjatlarni yuritish barcha turdagi faoliyatlarning asosi hisoblanadi.

**Teglashning korpus tuzishdagi ahamiyati.** Teg, uning tur, xususiyati korpus lingvistikasi mutaxassislari tomonidan keng tavsiflangan. Korpusni oddiy elektron to‘plam yoki virtual kutubxonadan farqlovchi, matn ustida turli lingvistik amallarni bajarishga imkon beruvchi eng asosiy omil – uning razmetkasi. Avvalgi bobda korpus yordamida bajarilishi mumkin bo‘lgan ishlar haqida aytib o‘tildi. Bu vazifalarni bajarish uchun korpusning o‘zi mavjud bo‘lishi yetarli bo‘lmay, korpus materialiga turli qo‘shimcha lingvistik izoh ham bo‘lishi kerak. V.P.Zaxarov, B.Kutuzov, S.A.Sharovlar bu boradagi ishlari<sup>174</sup>da teg, uning korpusdagi ahamiyatini atroflicha yoritib berishgan. Shuningdek, Z.P.Zaxarov rahbarligi ostida L.V.Severyuxina ishlari<sup>175</sup>da, V.V.Rikovning shu mavzudagi ma’ruzalari<sup>176</sup>da ham teglash masalasi atroflicha yoritilgan. Ishlarning har birida razmetkaning alohida jihati ta’riflangan. Barchasini umumlashtirgan holda lingvistik annotatsiya yoki korpus razmetkasi (ing. linguistic markup) atamasi korpusga matnning qismi bo‘lmagan, lekin shu matnga biriktirilgan qo‘shimcha ma’lumot (metaaxborot) deb xulosa qilish o‘rinli. Bu qo‘shimcha ma’lumotning eng oddiy ko‘rinishi so‘z turkumi haqidagi axborot. U quyidagi shaklda berilishi mumkin: Osmon (*ot*) ostidagi (*ot*) hamma (*olmosh*) narsa (*ot*) omonatdir (*sifat*).

---

<sup>174</sup> Каранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

<sup>175</sup> Каранг: Л.В. Северюхина. “Моделирование логико-понятийной области корпусной лингвистики”. (Дип.раб; рук. доц. В.П. Захаров).

<sup>176</sup> В.В.Рыков маърузалари билан <http://rykov-cl.narod.ru/c.html> сайтида танишиш мумкин.

(Laosszi). Teglash korpusni avtomatik tahlil qilishda qo‘l keladi. Matndagi so‘zlarni bir marta turkumga ajratsak, istalgan tadqiqotni bajarishda (masalan, korpusdagi barcha sifatlarni ajratishda) foydalanish uchun qo‘l keladi. Teglash masalasiga to‘xtalishdan oldin teg tizimlari tarixiga nazar tashlash foydadan xoli bo‘lmaydi.

O‘tgan asrning 80-yillarida SGML (Standard Generalized Markup Language) nomi ostidagi elektron matnlar razmetkasi standarti qabul qilingan edi. Bu standart dastlab tipografiya sanoati uchun ishlab chiqilgan bo‘lsa-da, tez orada boshqa sohalarda ham qo‘llanila boshlandi. SGMLning mohiyati turli matn muharrirlarida terilgan hujjatni tahrirlash, tahlil qilish, o‘zgartirishdan iborat.

*Teg lingvistik vosita sifatida.* SGML teg konsepsiyasini olib kirdi. “Teg” (ingl. tags) – matndagi ishchi izoh, u matn haqidagi ma’lumotni qamrab oladi. Korpus yordamida statistik hisob jarayonida tilimizda mavjud so‘zlarning faqat chastotasini aniqlash emas, balki yana bir qancha ma’lumotlarni olishimiz mumkin. Masalan, har bir so‘z bilan yonma-yon uning turkumi belgilangan bo‘lsa, tilda turli nutqiy vaziyatda so‘z turkumlarining qo‘llanilish darajasini aniqlash ham mumkin. Lingvistik teglash har bir so‘zning ma’lum kodga ega bo‘lishi bilan xarakterlanadi. Ushbu kod *teg*, so‘zni kodlash esa *tegg* (ingl. tagging)<sup>177</sup> deyiladi. Bugungi kunda matnga lingvistik, boshqa ma’lumotlarni qo‘shishning umume’tirof qilingan standarti mavjud emas. Lekin Text Encoding Initiative (TEI) maxsus xalqaro loyihasi razmetkaning standart vositasini ishlab chiqishga mo‘ljallangan. Buning uchun hujjat razmetkasining butun xalqaro qabul qilingan tili – SGML va XML mavjud. XML kengaytmasi<sup>178</sup> XML (ing. Extensible Markup Language) bazasi asosida qurilgan hamda lug‘at teg va atributi, qoidalar bazasini qamrab olgan aniq grammatika tili. SGML<sup>179</sup> esa hujjat uchun teglash tilini aniqlaydigan metatil (метаязык).

An’anaviy teglar juft holda qo‘llanuvchi (ochiluvchi, yopiluvchi) uchburchak qavsdan iborat bo‘ladi. Masalan, <f> ochiluvchi, </f> yopiluvchi teg,

---

<sup>177</sup> <http://rykov-cl.narod.ru/c.html>

<sup>178</sup> <https://ru.m.wikipedia.org/wiki/XML>.

<sup>179</sup> <https://ru.m.wikipedia.org/wiki/SGML>.



ochiluvchi teg berayotgan ma'lumotning yopilganligini bildiradi. `<ds> Bu, <em>birinchi navbatda </em> sizga bog'liq </ds>`, dedi u. Bu, *birinchi navbatda sizga bog'liq*, -dedi u matni “ko'chirma gap” (ing., *direct speech*) ma'nosini beruvchi `<ds>`, uning ichida kelgan *birinchi navbatda* kiritmasi ajratuvchi `<em>` belgisi (teg) orqali ajratilgan. Teglar yakka holda (juftlashmay) ishlatilishi ham mumkin. Masalan, og'zaki korpusda ma'lum o'rinda to'xtamni bildirish uchun `<pause>` tegi ishlatiladi. Bu yakka teg hisoblanadi. Odatda, tegning o'zi foydalanuvchiga ko'rinmaydi, razmetkalangan matnni ko'rsatayotgan dastur tegni qabul qilingan, ayni shu matn uchun qo'llashga kelishilgan shaklda ko'rsatishi mumkin.

*Matn turli tomondan (alternative views)*. V.P.Zaxarov SGML tipi razmetkasining o'ziga xos imkoniyatidan biri matnning turli ko'rinishini namoyish eta olish ekanligini ta'kidlaydi. Bir marta razmetkalangan matn qo'yiladigan topshiriqqa binoan turli shaklda berilishi mumkin. Korpusdan faqat ko'chirma gapli qo'shma gapning muallif gapini ajratib olmoqchi bo'lsak, matnni ko'rib chiqayotgan dasturimiz `<ds>` belgisi ichiga kirgan qismni yashiradi, natija quyidagicha ko'rinadi: – *dedi u*. Yoki biz ko'rayotgan matnda muhim qism (tadqiq manbasi) ko'k rangda, muallif gapi qora rangda berilishini buyuramiz. Natija quyidagicha ko'rinadi: *Bu, birinchi navbatda sizga bog'liq*, – *dedi u*. Bundan ham murakkabroq “alternative views”ni ko'rishimiz mumkin. Masalan, dramatik asarda turli qahramon nutqi alohida teglar bilan belgilangan, kerakli holda bir qahramon nutqi yoki ma'lum ikki qahramon dialogi ruyxatini bitta buyruq bilan ajratib olishimiz mumkin.

*Matnning avtomatik razmetkasi*. Ko'rinyaptiki, katta hajmli korpusni qo'lda razmetkalash uzoq muddatli, qimmatli mehnatni talab etadi. Shuning uchun o'tgan asrning 70-yillaridayoq bu vazifani sun'iy intellektga yuklatish bo'yicha bir necha loyihalar ishlab chiqildi<sup>180</sup>. O'shanda TAGGIT dasturi Braun korpusidagi so'zlarning 77 foizini so'z turkumiga to'g'ri ajratgan edi, qolganini esa 10 yil davomida qo'lda bajarishga to'g'ri kelgandi. Lekin 80-yillarga kelib, CLAWS

---

<sup>180</sup> Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс)

(Constituent Likelihood Automatic Word-tagging System) dasturi Braun korpusi soʻzlarini 95 foiz toʻgʻri tahlil qildi. Bugungi kunga kelib Yevropa tillari uchun soʻz turkumi avtomatik razmetkasi (word-class tagging), gap boʻlaklari avtomatik razmetkasi (parsing) ishlab chiqilgan. Bu ishning natijasi esa avtomatik tarjima, internet-qidiruv tizimining ishlashida namoyon boʻladi. Shu oʻrinda “Matnga avtomatik ishlov berish” ustida ish olib borayotgan olimlar guruhining (sayt <http://www.aot.ru>) bu borada qilayotgan ishlarini taʼkidlab oʻtish joiz. Ular, asosan, nazariy lingvistikani zamonaviy axborot texnologiyalarida qoʻllash ustida ish olib borishmoqda. Bu borada rus, ingliz va nemis tillaridagi matnni tahlil qilishning grafem (soʻz chegarasini aniqlash), morfologik (soʻz turkumini aniqlash), sintaktik (gap boʻlaklarini aniqlash) hamda semantik (soʻzlararo semantik munosabat) modullarini yaratishdi. Oʻzbek tili soʻz turkumlarini oʻqiy oladigan dastur – tagging ishlab chiqish kompyuter lingvistikasining oldida turgan dolzarb vazifalardan biridir.

AskNet. Intell Servis. Semantik savol-javob tarzidagi qidiruv sistemasi AskNet va ishlov berish instrumentlari rus va ingliz tillaridagi matnlarni lingvistik tahlil qilishni toʻliq amalga oshiradi. Lingvistik tahlil qilish moduli oʻz zimmasiga morfologiya (lugʻatli va lugʻatsiz), sintaksis, semantikani oladi. Intell Servisning dasturiy mahsulotlarida toʻla versiyalar bilan korporativ(uyushma)li, saytli va shaxsiy (maxsus) qidiruv tizimlari taklif qilinadi. Internet boʻyicha savol-javobli qidiruv metaqidiruvli tizimlar bazasida ([www.asknet.ru](http://www.asknet.ru)) amalga oshiriladi. Hozirda analitik qidiruv tizimi AQUA ishlab chiqilmoqda. Bu tizim matnni avtomatik tekshirishda semantik nuqtayi nazardan savollarga javob topadi. Dasturlar savdo-sotiqda keng yoyilmoqda. Lingvistik tahlil darajalari: grafematik, morfologik, sintaktik va semantik.

Ontos. Ushbu dastur matnli hujjatlarni tahlil qilish, annotatsiyaning tuzilishini tekshirish, kiritilgan matnlarga (OntosMiner, LightOntos for Workgroups, Ontos, Ontos SOA, TAIS Ontos) ishlov berishga moʻljallangan. Dastur algoritmlari grafematik, morfologik va semantik tahlilga qaratiladi. Sistemalar morfologik lugʻatlar va semantik kartrijlardan asosiy tillar (ingliz,

nemis, fransuz, rus) uchun foydalanadi. Matnda yo‘l qo‘yilgan xatolarning to‘g‘ri javoblari (variantlari) ko‘rsatiladi, grafalarning bog‘lanishi beriladi. Dasturlar Java.netda amalga oshiriladi va ular Microsoft SQL Serverda ishlatiladi. Lingvistik tahlil darajalari: grafematik, morfologik, semantik tematik sinonimlar.

Google Mini, Google Desktop, Google. Dasturli apparat kompleksi Google Mini va Google Desktop dasturlari tarmoq ichi, ya‘ni internetda kompyuter vositasida xabarlarni qidirishga mo‘ljallangan. Asosiy tillarda (rus, ingliz va h.k.) kalit so‘zlar bo‘yicha morfologik jihatdan tekshiruvni amalga oshiradi. Lingvistik tahlil imkoniyatlari: grafematik va morfologik.

Yandex.Server. Yandeks. Dastur (NTTR) serveri uchun mo‘ljallangan, bir yoki bir necha saytlar bo‘yicha qidiruv va raqamlashni amalga oshiradi. Qidiruv rus, ingliz va ukrain tillarining morfologiyasi bo‘yicha ishlaydi. Qidiruv natijasi hujjatlar ro‘yxatining relevantlik yoki sana asosida tartiblanishi hisoblanadi. Relevantlikda hujjatdagi leksemalarning miqdori, ularning chastotali tavsifi va leksemalararo joylashishi inobatga olinadi. Matnlarning sintaktik va semantik tahlili hisobga olinmagan. Dastur versiyalari Windows va Unix uchun mo‘ljallangan. Lingvistik tahlil darajalari: grafematik va morfologik.

Galaktika-ZOOM. Galaktika. Dastur hujjatning ma‘noga ega so‘z va so‘z birikmalarini aniqlaydi, foydalanuvchi tomonidan kiritilgan kalit so‘zlar hamda ularning sinonimlari yuzasidan tekshiradi, qidiruv ishini olib boradi. Shuningdek, hujjatda uchraydigan so‘zlarning chastotasi bo‘yicha statistikani formallashtiradi. Dastur rus tilidagi matnlarning tahlilini ta‘minlaydi. Algoritmalar morfologik va statistik tahlildan foydalanishga asoslanadi. Lingvistik tahlil darajalari: grafematik va morfologik.

Quintura Searchcrystal. Qidiruv natijalarini grafik tarzda taqdim qiluvchi metaqidiruvli tizim. Natijalar statistik mezonlar asosida klasterlanadi. Dastur morfologik tahlil qilishdan foydalanadi. Masalan, “yuvuvchi” so‘zi qidiruvga berilganda Quintura klasterga “ajoyib enaga” so‘zini kiritadi, uning o‘zagi bilan bog‘liq “yuvayotib” so‘zi va qidirilayotgan “yuvuvchi” so‘zini beradi.

### **Savol va topshiriqlar**

1. Tokenizatsiya qanday jarayon?
2. Lemmatizatsiyaning razmetkadagi ahamiyati nimada?
3. Stemming nega kerak?
4. Parsing nima va uning qanday ko‘rinishlari mavjud?
5. Teglashning korpus tuzishdagi ahamiyatini izohlang.
6. Teglarining qanday turlari mavjud?
7. Teg tizimiga nimalar kiradi?
8. Matnning avtomatik razmetkasi deganda nimani tushunasiz?
9. Matnning avtomatik razmetkasi qanday amalgam oshiriladi?
10. CLAWS (Constituent Likelihood Automatic Word-tagging System) dasturi qanday dastur?
11. Tagging va parsing uchun jahon korpus lingvistikasida qanday yutuqlarga erishilgan?
12. Tagging va parsingning qanday shakllarini bilasiz?

#### **Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.
2. Недошивина Э.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб., – 2006. – 26 с.
3. Структурно-функциональная компьютерная модель татарских морфем. – Казань: ФЭН, 2003. – 220 с.
4. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. –128 с.
5. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
6. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.
7. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.
8. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.

9. [www.ruscorpora.ru/](http://www.ruscorpora.ru/) Плунгян В. Зачем мы делаем корпуса?
10. [http://www.ruscorpora.ru.](http://www.ruscorpora.ru/) НКРЯ
11. [www.philol.msu.ru/~lex/corpus](http://www.philol.msu.ru/~lex/corpus)
12. <http://www.philol.msu.ru/~lex/corpus>
13. [http://ruscorpora.ru.](http://ruscorpora.ru)

**ANNOTASIYA TURLARI:  
LINGVISTIK VA EKSTRALINGVISTIK RAZMETKA.**

**Reja:**

1. Teg lingvistik vosita sifatida.
2. SGML va XML kengaytmasi.
3. Lingvistik va ekstralingvistik razmetka.
4. Lingvistik razmetka turlari.

***Tayanch so‘z va iboralar:** teg, lingvistik vosita, SGML, XML kengaytmasi, lingvistik razmetka, ekstralingvistik razmetka.*

Korpus lingvistikasida razmetkaning lingvistik hamda ekstralingvistik turlari ajratiladi<sup>181</sup>. Ekstralingvistik razmetkaning quyidagi turlari farqlanadi:

1. Matn formatining o‘ziga xosligini aks ettiruvchi (bob, xatboshi, qism va h.) teg.

2. Matn, uning muallifiga tegishli ma’lumotni ifodalovchi teg. Chunki muallif haqidagi ma’lumot nafaqat nom, balki yosh, jins, u yashagan yil kabi ma’lumotlarni ham bildirishi mumkin. Matn haqidagi ma’lumot esa matn (asar) nomidan tashqari tili, yozilgan hamda nashr etilgan yilini qamrab oladi. Bunday ma’lumotlarning mavjudligi bazada detallashtirilgan qidiruvni amalga oshirish imkonini beradi. Ekstralingvistik teg yoki metaaxborot (ruscha: метаданные) – tashqi-intellektual ma’lumotlarni qamrab oluvchi, bibliografik, tipologik, tematik, sotsiologik tavsifni; formal-struktur teg (matn, bo‘lim, bob, qism, abzats, gap)ni, shuningdek, texnik-texnologik teg (kodirovka, manbaning elektron versiyasini qayta ishlash sanasi)ni birlashtiruvchi teg turi hisoblanadi. Metaaxborot majmui foydalanuvchiga uning imkoniyatlarini aniqlab beradi. Bu ma’lumotlarni tanlashda korpus tuzuvchisi tadqiqot maqsadi, tilshunoslar talabi hamda matnga u yoki bu qo‘shimcha belgini qo‘shish imkoniyati bilan tanish bo‘lmog‘i lozim. Tashqi intellektual teg, birinchidan, tilning o‘zaro aloqasi, mavjudligini aniqlash; ikkinchidan, tilning o‘ziga xos xususiyatini o‘rganish uchun kerak. Tilga ta’sir

---

<sup>181</sup> <http://rykov-cl.narod.ru/c.html>.

qiladigan ikkita: tashqi, ya'ni nolisoniy hamda ichki omil mavjud. Dj.Sinkler tashqi omilning uch guruhini ajratadi<sup>182</sup>:

1. Muallif tomonidan matn yaratishga aloqadorlik omili.
2. Matnning tashqi belgilariga tegishli omil.
3. Matnning yaratilishiga sabab bo'luvchi, auditoriyaga ta'sir qiluvchi omil.

Ichki omilning ikki guruhi – matnning mavzu doirasi, uslubiy xususiyati (uslub, janr) ajratiladi. Masalan, rus tili milliy korpusida quyidagi metaaxborotlar majmui mavjud: Birinchi blok:

1. Matn muallifi: ismi, jinsi, tug'ilgan vaqti (taxminiy yoshi).
2. Matn nomi.
3. Yozilish vaqti, joyi (aniq shahar yoki davlat bo'lishi mumkin).
4. Matn hajmi: badiiy asarlar uchun me'yor sifatida hikoya hajmi kamida 5000 so'z; qissa hajmi 5000dan 15.000 so'z gacha; roman 15.000 so'zdan ortiq bo'lishi odat tusiga kirgan.

Ikkinchi blok: metaizoh korpusdagi matnning 3 asosiy ko'rinishi – badiiy matn, nobadiiy matn, dramatik asarlarni farqlashga mo'ljallangan. Chunonchi, rus tili milliy korpusida badiiy asar uchun quyidagi ma'lumotlar ko'rsatiladi:

1) matn janri: avtobiografik proza, detektiv, bolalar adabiyoti, tarixiy, kriminal adabiyot, sarguzasht, fantastika, yumor, satira;

2) matn tipi: avtobiografik proza, latifa, detektiv, ocherk, adabiy maktub, qissa, masal, pyesa, hikoya, roman, ertak, triller, epopeya, esse va h.;

matn xronotopi: tasvirlanayotgan voqeaning taxminiy vaqti, joyi.

### **Teglash – korpusdan foydalanish imkoniyati darajasi va mezoni.**

*Lingvistik teglash muammolari.* Internetdagi oddiy elektron fayldan korpusning farqi razmetkaning mavjudligidadir. Bu muallif asari yoki uning parchalariga lingvistik, ekstralingvistik sharh qo'shish bilan amalga oshiriladi. Agar muallifning asari to'liq razmetkalangan bo'lsa, keyinchalik bu korpusdan faqat elektron asarlar yig'indisi sifatida emas, balki turli lingvistik tadqiqotlar olib borishda foydalanish

---

<sup>182</sup> *Қаранг:* Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – Б.36.

mumkin bo‘ladi. Shuning uchun korpus razmetkasining mukammalligi undan foydalanish darajasiga to‘g‘ri proporsional. Korpus tarkibidagi matn (muallif asari)ga quyidagi qo‘shimcha ma’lumot(teg)lar yozish mumkin. Asar yozilish vaqti, janriy xususiyati to‘liq asarga nisbatan yoziladi, bu ekstralingvistik razmetkadir. Asar tarkibidagi har bir gapga uning sintaktik strukturasi haqida ma’lumot beriladi, matndagi leksemalarga esa alohida morfologik belgilar qo‘yiladi. Korpus tuzish tajribasini o‘rganar ekanmiz, rus, slavyan tili korpuslari, roman-german korpuslari uchun muallif sintaktik razmetkaga juda katta e’tibor qaratganining guvohi bo‘lamiz<sup>183</sup>. Matnni avtomatik razmetkalashdagi birinchi muammo omonimlarning noto‘g‘ri izohlanishi bo‘lib, shu sababli omonim so‘zlar qo‘lda razmetkalanadi. Omonim bir turkum doirasida bo‘lganda, matnda so‘zshakl holida turganda ham so‘z ma’nosi oydinlashmaydi. Natijada, foydalanuvchi ayni bir ma’noni ifodalovchi bir necha so‘zshaklni qidirganda omonim xalaqit beradi. Matnda bir leksema so‘zshakl holida boshqa leksemaga teng kelib qolishi mumkin. Masalan, *boshda, birga, birdan, pastdan* kabi kelishik shaklidagi ot shu shakldagi ravish bilan bir shakl ega bo‘lib qoladi. Bunday so‘zlar ham korpus muallifi tomonidan qo‘lda razmetkalanadi. Agglyutinativ tillarga qaraganda flektiv tildagi so‘z razmetkasi qiyinroq va ahamiyatliroq. Turkiy tillarda har bir grammatik ma’no alohida grammatik shakl bilan berilganligi sababli teglash jarayoni osonroq kechadi. Teglash natijasi o‘laroq, korpus bo‘yicha qidiruvni amalga oshirish uchun “matn muallifi”, “matn nomi”, “matn yaratilish vaqti” (ba’zi elektron kutubxonalarda matn yaratilish, elektron formatga keltirish sanasi ma’lum), “matn janri” kabi qidiruv oynalarini yaratish kerak. Oldingi bo‘limda lingvistik razmetkaning muhim bosqichi: parsing, taggingning xususiyati, vazifasi, ularni amalga oshiruvchi dastur, tillarning ahamiyati ham yoritilgan.

Lingvistik razmetkaning ham o‘z navbatida bir qancha ko‘rinishi mavjud:

1. Morfologik teglash keyingi – sintaktik va semantik teglashga asos bo‘luvchi ahamiyatli teg hisoblanib, ingliz tilida part-of-speech tagging deb

---

<sup>183</sup> Қаранг: Сичинава Д.В. К задаче создания корпусов русского языка. (Электрон ресурс: // [https:// istina.msu.ru](https://istina.msu.ru).)



ataluvchi, soʻzlarni turkumga ajratuvchi razmetkadir. Razmetkaning ushbu turi teglar yordamida amalga oshirilib, tegning matnda mavjudlik darajasi, koʻlami korpusning xususiyatidan kelib chiqib, har xil boʻladi. Teg qancha koʻp boʻlsa, korpusning lingvistik amallarni bajarish imkoniyati shuncha keng boʻladi. Lekin keyingi avlod korpuslari hajmining kattaligi sababli tegni soddalashtirish yoʻli qulay deb topildi. Kodirovkaning soddalashtirilgan tizimi ortiqcha xatoliklarning oldini oladi, morfologik koʻp maʼnolilikni keltirib chiqarmaydi, bir necha million soʻzni qamrab oluvchi katta massivli matnlarning razmetkalanishini tezlashtiradi.

2. Sintaktik teg – sintaktik tahlil, parsing (ingl. parsing) natijasi, u komponentlarining grammatik strukturasi asoslanadi. Gapdagi boʻlaklar orasidagi sintagmalar grafik, shajara tarzida, matnda esa ular oʻtli, feʼlli hamda murakkab birikmalarni, sodda, qoʻshma gapni koʻrsatuvchi ochiluvchi, yopiluvchi qavslar yordamida koʻrsatiladi. Sintaktik razmetkaga ega korpuslar treebanks nomi bilan ommalashgan. Xuddi morfologik razmetkada boʻlgani kabi keyingi paytda tahlilni tezlashtirish maqsadida sintaktik teglash ham soddalashtirildi, natijada bu usul skeleton parsing nomini oldi.

3. Semantik tegning ham, boshqa razmetkalarda boʻlganidek, yagona standart shakli boʻlmasa ham, harf, raqam yoki faqat raqamdan iborat kodlardan foydalaniladi. Birinchi harf yoki raqam umumiy semantik maʼnoni, keyingi belgi esa soʻz maʼnosini yanada maxsuslashtiruvchi kichik semantik guruhni ifodalaydi. Semantik teg nafaqat soʻz, balki koʻplab birikmalarni ham semantik guruhlariga birlashtiradi, bunday paytda turli birikuvdagi bir maʼnoni bildiruvchi birikmalar bitta belgi bilan kodlanadi. Idiomatik birlik (ibora) tarkibidagi soʻzlar miqdorini bildiruvchi axborot ham razmetkadan joy oladi. Semantik teg korpusdagi soʻz maʼnosining ixtisoslashuvi, omonimlik, sinonimlik, maʼnoviy guruhga ajratish kabi muammolarni hal qiladi. V.P.Zaxarov, S.YU.Bogdanovalar rus tili milliy korpusini tuzishda semantik razmetkalashning oʻz variantini taklif qiladi<sup>184</sup>. Bu

---

<sup>184</sup> Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – Б.45.

korpusda har bir soʻzga uch xil – soʻz razryadi, leksik-semantik tavsif, derivatsion izoh beriladi.

4. Anaforik teg. Matnga ishlov berishda katta qiyinchilik tugʻdiradigan turkum bu olmosh, chunki matndagi qaysi soʻzga ishora qilishiga qarab turli maʼnoni bildiradi. Ishora soʻzning matndagi maʼnosini ajratib olish uchun razmetkaning alohida koʻrinishiga ehtiyoj tugʻiladi. Anaforik teg shu xildagi maʼlumotni qoʻshish uchun kerak boʻladi. Maʼnosi olmosh bilan ifodalanayotgan soʻz alohida kodlanib, keyingi oʻrinda shu soʻzga ishora qilayotgan olmosh yoniga shu kod biriktiriladi. Natijada olmoshning matndagi maʼnosi aniqlanadi hamda kerakli tadqiqotlarda korpusga asoslanish imkoniyati paydo boʻladi.

5. Prosodik teg. Ovoz transkripsiya qilingan korpusda urgʻu, ohangni ifodalovchi izoh mavjud boʻladi. Razmetkaning diskurs deb ham ataluvchi ushbu turi sharh, izoh, eslatma, takrorlardagi toʻxtamlarni bildirish uchun ishlatiladi.

Korpusni teglash (annotatsiyalash) dasturlashtirilgan yoʻllar bilan amalga oshiriladi. Bunda, avvalo, vaqtni tejash, mehnatni kamaytirish nazarda tutilsa, ikkinchidan, matnga avtomatik ishlov berish muammosiga yechim topiladi. Hozircha anaforik, prosodik teg qiyinligicha qolib ketyapti hamda teglash faqat qoʻlda bajarilyapti, keyinchalik bu ham dasturlashtirilishi zarur, albatta. Morfologik, sintaktik teglash esa tegger, parsing yordamida amalga oshirilsa-da, bu dasturlarning ham aksariyati avtomatik razmetkadan keyingi tuzatishni talab qiladi. Chunonchi, morfologik omonimiya (koʻproq flektiv tillarga xos), sintaktik koʻpmaʼnolilik holatida dastur xulosaning bir necha koʻrinishini taklif qiladi, tadqiqotchi esa keraklisini tanlaydi. Yangi avlod korpuslari hajmining favqulodda kattalashgani mutaxassislar oldiga razmetkaning toʻliq avtomatlashtirilgan turiga oʻtish, yangi, mukammal tegger, parsinglar yaratish vazifasini qoʻyadi. Avtomatik morfologik tahlil (tegger) yordamida har bir leksik birlikka (soʻz turkumi, lemma, grammema guruhi) alohida grammatik xarakteristika (shaxs-son, kelishik, boshqa grammatik kategoriya) beriladi. Masalan, Braun korpusida soʻzning chastotasini aniqlash oson. Faqat bu soʻzshaklning (korpus tilida *token*) chastotasi boʻladi. Leksemaning chastotasini aniqlash uchun esa har bir soʻzga uning lemmasi

biriktirilgan bo‘lishi kerak. Korpusni avtomatik razmetkalashning oddiy usuli so‘zning leksik kategoriyasi ko‘rsatilgan hajman katta elektron lug‘atni razmetkalangan korpus bilan birlashtirishdir. Shunda elektron lug‘atdagi izoh (grammatik kategoriya tavsifi) razmetkalanmagan korpusdagi so‘zga teg sifatida o‘zlashtiriladi. Masalan, korpus, elektron lug‘atda *axborot*, *siyosat* so‘zlari mavjud bo‘lsa, lug‘atdagi “ot” tegi avtomatik tarzda korpusga ko‘chadi. Lekin bu usul bilan ham korpusni to‘liq razmetkalashning imkoni yo‘q. Chunki ba‘zi so‘z, birikmalar bir vaqtning o‘zida bir necha kategoriyaga mansub bo‘lishi mumkin. Bu holat morfologik ko‘pma’nolilik (ambiguity) muammosi bilan bog‘liq. *Olma*, *matematik*, *etik*, *surma*, *suzma*, *bo‘lmoq*, *qo‘llar*, *bog‘lar* kabi so‘zshakllar birdan ortiq grammatik kategoriyaga tegishli, shu sababli bunday so‘zlarni razmetkalash faqat elektron lug‘at yordamida amalga oshmaydi. Tabiiyki, kontekstda so‘zshakl faqat bitta kategoriyaga tegishli bo‘lib qoladi, shuning uchun razmetkaning yana ham mukammalroq ko‘rinishi: morfologik teglash uchun sintaktik teg, sintaktik teg uchun semantik teglash qilish korpusni to‘liq, to‘g‘ri teglashga olib keladi. V.P.Zaxarov fikricha, lingvistik razmetkaning barcha (morfologik, sintaktik, semantik, anaforik, prosodik) turlari quyidagi tamoyillar asosida amalga oshiriladi:

- 1) teg sxemasini tavsiflash (asoslash);
- 2) umumiy lingvistik tushunchalar tizimini aniqlash;
- 3) foydalanuvchi uchun ma’lum bo‘lgan tahlil sxemasini shakllantirish;
- 4) teg sxemasining nazariy an’anaviyligiga erishish;
- 5) xalqaro andozalarga amal qilish.

## Экстралингвистик разметка натижаси

диалектный  
позитический  
устный  
акцентологический  
мультимедийный  
мультипарк  
исторический

ользование корпуса

**Основные параметры текста** ?

Название

Автор текста

Пол:  любой  мужской  женский

Год рождения: от  до   Точное вхождение

Год создания: от  до   Точное вхождение

**Версии**

Дата добавления документа: с  самого начала ▼ по  текущий момент ▼

Дата изменения документа: с  самого начала ▼ по  текущий момент ▼

**Жанр и тип текста** ?

1. Художественные тексты

Жанр текста [выбрать](#)

### Savol va topshiriqlar

1. Teg lingvistik vosita sifatida sifatida qanday vazifa bajaradi?
2. SGML va XML kengaytmasining korpus matnlarini saqlashdagi ahamiyati nimadan iborat?
3. Lingvistik razmetka nima? Uning qanday ko‘rinishlari mavjud?
4. Ekstralingvistik razmetka nima? Uning natijasi interfeysda qanday o‘z aksini topadi?

### Tavsiya etilgan adabiyotlar:

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика.– Иркутск: ИГЛУ, 2013. – 161 с.
2. Недошивина Э.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб., – 2006. – 26 с.
3. Сичинава Д.В. К задаче создания корпусов русского языка. (Электрон ресурс: // [https:// istina.msu.ru](https://istina.msu.ru).)
4. Структурно-функциональная компьютерная модель татарских морфем. – Казань: ФЭН, 2003. – 220 с.
5. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. –М.: ФЛИНТА: Наука, 2013. –128 с.

6. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
7. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.
8. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.
9. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
10. [www.ruscorpora.ru/](http://www.ruscorpora.ru/) Плунгян В. Зачем мы делаем корпусы?
11. [http://www.ruscorpora.ru.](http://www.ruscorpora.ru/) НКРЯ
12. [www.philol.msu.ru/~lex/corpus](http://www.philol.msu.ru/~lex/corpus)
13. <http://www.philol.msu.ru/~lex/corpus>
14. [http://ruscorpora.ru.](http://ruscorpora.ru)

## KORPUS QIDIRUV TIZIMI SIFATIDA. KORPUS MENEJERI VA INTERFEYS.

### Reja:

1. Korpus menejerining qidiruv, saralash, filtrlash vazifasi.
2. BONITO, XAIRA, SARA, SQR, DDS korpus menejerlari.
3. Konkordans va interfeys, uning turlari haqida.

*Tayanch soʻz va iboralar:* korpus menejeri, qidiruv, saralash, filtrlash, BONITO, XAIRA, SARA, SQR, DDS korpus menejerlari, konkordans va interfeys.

**Korpus menejeri qidiruv tizimi sifatida.** Lingvistik korpusning ajralmas, eng asosiy qismi uning qidiruv tizimi – korpus menejeri, u matn, lisoniy birliklarni boshqaruvchi sistema. Korpus menejeri – korpus maʼlumotlari ustida ishlash uchun moʻljallangan maxsus qidiruv tizimi, u statistik maʼlumot, qidiruv natijasini foydalanuvchiga qulay koʻrinishda koʻrsatib beruvchi dasturiy taʼminot hisoblanadi. E.V.Nedoshivina korpus menejeriga qoʻyiladigan talablarni sanar ekan, ularning eng asosiysi sifatida matnning kalit soʻzlari roʻyxati, toʻliq konkordans roʻyxatni yarata olish; faqat soʻzni emas, balki soʻz birikma holdagi soʻrovga ham javob bera olish; shablon asosida (murakkab soʻrov) qidiruvni amalga oshirish; olingan natija (chiqarilgan roʻyxat)ni bir necha mezon asosida saralay olish; soʻzshaklga berilgan soʻrovni cheklanmagan miqdordagi kontekstda aks ettirish; korpusning alohida elementlari boʻyicha statistik maʼlumot bera olish; korpus razmetkasidan kelib chiqqan holda lemma, soʻzshaklning morfologik xususiyati hamda metaaxborot (bibliografik, tipologik)ni toʻliq ifodalay olish; natijalarni saqlash, chop etish; fayl, korpusning cheklanmagan hajmi bilan ishlay olish; qidiruvni tez amalga oshirish, natijalarni chiqarish; turli matn formatlarini (txt, doc, rtf, html, xml va b.) oʻqiy olish hamda shu format bilan ishlash; malakali hamda yangi foydalanuvchi uchun birdek qulay boʻlish kabilarga toʻxtaladi<sup>185</sup>.

SARA, XAIRA (BNC), Manatee/Bonito, SQR, DDC kabi korpus menejerlari nisbatan universal sanalib, korpus tuzish tajribasida shu menejerlardan

---

<sup>185</sup> Каранг: Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. работа с системой DDC. // Языковая инженерия: в поиске смыслов. (электрон ресурс: <https://docplayer.ru>).

ko'proq foydalaniladi. Korpus ma'lumotini qayta ishlash uchun ma'lumotlar bazasini boshqarish yoki qidiruvning yangi tizimini ishlab chiqish mumkin. Informatsion so'rov<sup>186</sup> – ma'lum axborotga bo'lgan ehtiyojni so'z orqali ifodalash usuli. So'rov o'zining predmetlik belgisi, shakliy mundarijasi asosida tahlil etiladi, korpusda ishlovchi dasturning so'rov tili terminlari lug'ati asosida izohlanadi. Qidiruv amaliyoti korpusdagi har bir element bilan navbatma-navbat qiyoslash, ularning moslik darajasini aniqlash bilan yakunlanadi. Umuman olganda, so'rov tili modeli, asosan, qidiruv elementi (informatsion ehtiyojni ifodalovchi atama), matn qidiruv elementini normallashtiruvchi vosita, operator yo'naltiruvchisi (konyunksiya, dizyunksiya, inkor), chiziqli grammatika vositasi (oralik masofa operatori, pozitsiya operatori) kabi tarkibiy qismlarni o'z ichiga oladi. E.V.Nedoshivina qidiruvning qo'shimcha shartlari sirasiga korpusning alohida maydonlaridagi qidiruvi (masalan, teglar ichida), qidiruv maydonining chegaralanishi (aniq bir muallif asari, hujjatni yaratish sanasi va h. bo'yicha), topilgan natijalarni saralashga qo'yiladigan talab, qidiruv natijasini ifodalash shakliga qo'yilgan talab, topilgan natijalarning ko'rinishi, topilgan hujjat miqdori kabi omillarni kiritish mumkin. Quyida nisbatan samarali, yaxshi ishlaydigan korpus menejeri Bonito/Manateening qidiruv tilining o'ziga xosliklariga to'xtalamiz. Lingvistik korpus bilan ishlash jarayonida ko'p ishlatiladigan qidiruv tili elementi hamda qidiruv bo'yicha misollarni izohlashni maqsad qildik. Bonito – menejer nomi, Manatee – korpus tagtizimini boshqaruvchi dasturiy ta'minot. Bonito korpus menejeri korpus matni bilan ishlashni to'liq qamrab olgan dasturlar yig'indisi bo'lib, bu tizim ikki qismdan: server, grafik foydalanish interfeysi (GUI – graphical user interface)dan iborat. Ushbu interfeys Chexiyaning Masarika universiteti informatika fakul'tetida P.Rixli va NLPlab (Natural Language Processing Laboratory) guruhi tomonidan tuzilgan. Tizim ishi bilan tanishtirish uchun ingliz matnlarining SUSANNE (Surface and Underlying Structural

---

<sup>186</sup> Каранг: Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. работа с системой DDC. // Языковая инженерия: в поиске смыслов. (электрон ресурс: <https://docplayer.ru>)

Analysis of Natural English) korpusi<sup>187</sup> dan foydalaniladi. Ushbu korpus Buyuk Britaniyaning Sasseks universitetida yaratilgan, ingliz tilining amerika variantini ifodalovchi Braun korpusining 130.000 soʻzini oʻzida mujassamlashtirgan, SUSANNE sxemasiga muvofiq razmetkalangan. Bonito tizimining asosiy xususiyati soʻrov tilining alohida atributlar qidiruviga (soʻzshakl, lemma, teg) egaligi, odatiy ifodalarning qoʻllanishi, mantiqiy operatori, strukturani ishlab chiqish vositasining (gap chegarasi va b.) mavjudligi, murakkab soʻrovni tezlik bilan ishlab chiqa olishi bilan xarakterlanadi. Konkordans roʻyxati oʻzida foydalanuvchi soʻrovi tarixi, soʻzshaklning morfologik xususiyatini koʻrib chiqish, lemmani aks ettirish kabi imkoniyati bilan ajralib turadi. Konkordans ustida ishlash jarayonida roʻyxatni faylda saqlab qolish, roʻyxatni bosma shaklda chop etish, kontekstdagi kalit soʻz asosida saralash, kontekstni chegaralanmagan darajada kengaytirish, filtrlash (konkordansning keraksiz qismini oʻchirish), “takror”ni oʻchirish kabi amallarni bajarish mumkin. Chastotali taqsimot nuqtai nazaridan korpus yoki kontekstdagi soʻz yoxud boshqa atributlar chastotasi, chegaralanmagan darajada guruhlashni ham amalga oshira oladi. Bundan tashqari, kodlashtirish tanlovi ham mavjud, foydalanuvchi xususiy korpusini yaratishi mumkin. Shuningdek, teglarni ixtiyoriy tanlay olish, boshqa tilni qoʻshish imkoniyati mavjudligi kabi xususiyatga ham ega. Korpus menejerining asosiy vazifasi foydalanuvchi soʻroviga javob berishdan iborat ekan, soʻrov, uni amalga oshiruvchi buyruq ham ahamiyatli. Yuqorida tilga olingan korpus menejerida foydalanuvchi qidiruv tizimi tiliga mos ravishda qidiruv oynasiga oʻzining shaxsiy savolini kiritishi yoki shablon soʻrovdan foydalanishi mumkin. Uning tarkibida bir nechta oyna, ularning vazifasi quyidagilardan iborat:

1) tasdiqlovchi filtr (P-filter) – soʻrov bilan mos keluvchi satr konkordans roʻyxatida beriladi;

2) inkor qiluvchi filtr (N-filter) – soʻrov bilan mos keluvchi satr konkordans roʻyxatidan oʻchiriladi;

---

<sup>187</sup> <http://www.grsampson.net>.



3) soʻz birikma (Collocations) – soʻrovni qanoatlantiruvchi birikma (konkret soʻzshakl) konkordansda alohida rangda belgilanadi.

Konkordansda shunga mos keluvchi pozitsiyani topa olishiga imkon yaratish uchun tasdiqlovchi, inkor qiluvchi filtr hamda soʻz birikmasi uchun maxsus interval belgilash kerak. Foydalanuvchi interval chegarasini belgilaydi (“From:” va “To:” kiritish oynalari). Mazmun tasdiqlovchi boʻlsa, qidiruv chiqish nuqtasidan oʻng tomonda qarab amalga oshiriladi, agar inkor qiluvchi boʻlsa, buning aksi amalga oshiriladi. Boshlangʻich nuqta sifatida soʻzshaklning boshlanishi, soʻzshaklning tugallanishi olinishi mumkin. Muhimi shuki, kiritilgan barcha savollar soʻrovlar tarixida (Query History) saqlanadi, soʻrov aynan bir xil boʻlganda, soʻrovlar tarixida saqlanmaydi. Soʻrov tarixini ochish uchun soʻrov oynasidagi “pastga” tugmasini bosish etarli. Kerak boʻlganda, avvalgi soʻrovga qaytish mumkin. Xulosa oʻrnida shuni aytish muhimki, korpus menejeri, yaʼni uning qidiruv tizimi korpusning muhim tarkibiy qismidir.

**Interfeys – korpus mundarijasi.** Korpus interfeysi – foydalanuvchini uning butun mazmuni bilan tanishtiruvchi, mundarija vazifasini bajaruvchi juda muhim tarkibiy qismi. Mualliflik korpusi interfeysi muallif ijodiy merosi janriy tasnifini bir qarashdayoq tushunish uchun yaxshi imkoniyat, korpusning boshqa bazalardan farqi ham aslida shunday imkoniyatning mavjudligidadir. Fors-tojik mumtoz allomalari, zamonaviy adiblari korpusi majmuasini tashkil qilgan bosh korpusning interfeysi oʻziga xos koʻrinishga ega, bosh korpus tarkibiga alohida mualliflar korpusi<sup>188</sup> birlashtirilgan (*Qarang:* 1-rasm). Interfeys tasviriga eʼtibor qaratsak, mualliflik korpusi davr boʻyicha tasniflanganligi, mumtoz, zamonaviy shoirlar korpusiga ajratilganligiga guvoh boʻlamiz. Mualliflik umumiy korpusidan mumtoz adabiyot namoyandalari Rudakiy, Firdavsiy, Saʼdiy Sheroziy, Mavlono Rumi, Hofiz Sheroziy hamda muosir adabiyot vakillari Loyiqali Sherali, Moʻmin Qanoat, Habibullo Fayzullolar korpusi joy olgan. Muallif korpusiga kirar ekanmiz, yangi interfeys bilan tanishamiz. Interfeysda esa muallif ijodiy merosi tasnifi yoki muallif ijodining qanday qirrasini haqida maʼlumot topa olishimiz haqida ilk

---

<sup>188</sup> www.sit.tj.

ma'lumotga ega bo'lamiz. Mumtoz adabiyot vakillari korpusi<sup>189</sup> interfeysi tarkibi qiyosiy tahlilda ko'rinadi (*Qarang: 2.1-jadval*). Hofiz korpusi tahlilida shoir biografiyasi haqida ma'lumot berishi; g'azallari to'plami, devoniga ishlangan rasm, miniatyura hamda Hofiz g'azallari lug'atidan tashkil topganligi ko'rinadi. Interfeysda tilni tanlash imkoniyati mavjud, foydalanuvchi o'z xohishiga ko'ra 3 ta til: tojik, rus, ingliz tilida korpusdan foydalanishi mumkin. "Biografiya" qismida Hofiz haqida to'liq ma'lumot, "G'azallar" bo'limida Hofizning tanlangan 31ta g'azali keltirilgan, korpusning bu qismi 31 sahifadan iborat. Ushbu korpusning eng ahamiyatli bo'limi – "Lug'at" qismi; chastotali lug'at xarakteriga ega ushbu sahifa o'ziga xos tuzilish, dizaynga ega (*Qarang: 2-rasm*).

Zamonaviy ijodkorlar korpusi<sup>190</sup>dan Loyiqali Sherali korpusini o'rganar ekanmiz, korpus interfeysi biografiya, she'riyat, fotogalereya, she'rlarning muallif ovozi bilan aytilgan mp3 formatdagi yozuvi ("Ustoz ovozi"), lug'at, video kabi qismlardan iboratligini ko'ramiz (*Qarang: 3-rasm*). Ushbu korpus tarkibidagi lug'at alifbo tartibli, chastotali, ters lug'at sifatida ishlatilishi mumkin. Korpusda jami 21.536 ta so'z mavjud. Zamonaviy ijodkorlar korpusi interfeysi tuzilishi qiyosiy tahlili korpus haqida umumiy xulosa chiqarishga yordam beradi (*Qarang: 2.2-jadval*).

Interfeys korpusning yuzi, mundarijasi ekan, u juda mukammal ishlanishi, bir qarashda muallif shaxsi, ijodiy merosi, tasnifi foydalanuvchi ko'z o'ngida to'la namoyon bo'lishi nihoyatda muhim. Shuningdek, korpus bo'yicha qidiruvni amalga oshirish oynasi ham shu erda joylashadi. Korpusning ilk yutug'i interfeysining mukammal, o'ziga xos usulda tuzilganligidadir. Yuqorida tahlil qilingan korpus interfeyslari rus, ingliz, boshqa tillar milliy korpusi bilan taqqoslanganda oddiy, lekin qulay va tushunarli ekanligi bilan ajralib turadi.

**1.2-жадвал**

---

<sup>189</sup> [www.termcom.tj](http://www.termcom.tj).

<sup>190</sup> [www.cit.tj](http://www.cit.tj).

Муаллиф	Хофиз корпуси (Пайкараи Хофиз)	Саъдий Шерозий корпуси (Пайкараи Саъди)	Рудакий корпуси (Пайкараи Рудаки)	Фирдавсий корпуси (“Шоҳнома”и Фирдавси)	Румий корпуси (Пайкараи Мавлоно)
Интерфейс таркиби	1.Биография 2.Ғазаллар 3.Расмлар 4.Луғат	1.Таржимаи ҳол 2.Қасидалар 3.Ғазаллар 4.Рубойлар 5.Китоблар 6.Расмлар 7.Биз билан алоқа	1.Қасида, ғазал, парчалар 2.Рубой 3.Парчалар 4.Рудакий асарлари бўйича Online тест	1.“Шоҳнома”нинг 9 жилди 2.Қидирув 3.Луғат	1.Биография 2.Маснавий 3.Рубой 4.Танланган рубойлар 5.Луғат 6.Суратлар 7. Линклар

### 1.3-жадвал

Муаллиф	Лойиқали Шерали корпуси	Мўмин Қаноат корпуси	Ҳабибулло Файзулло корпуси
Интерфейс таркиби	1. Биография 2.Шеърят 3.Фотогалерея 4.Шеърларнинг муалиф овози билан айтилган мр3 форматдаги ёзуви (“Устоз овози”) 5.Луғат 6.Видео	1. Ижодиёт (“Достонлар”, “Шероз куйлари”, “Шеърлар”) 2. Галерея (“Расм”, “Видео”, “Овоз”) 3. Қидирув	1.Асосий саҳифа. 2.Хотира. 3.Биография. 4.Ижодиёт. 5.Расм. 6.Видео. 7.Қўшиқлар.

### Savol va topshiriqlar

1. Korpus menejerining qidiruv, saralash, filtrlash vazifasi nimalardan iborat?
2. BONITO, XAIRA, SARA, SQR, DDS korpus menejerlarining xususiyatlarini sanab bering.
3. Konkordans nima? Uning vazifasi va turlarini tushuntiring.
4. Iinterfeys va uning turlari haqida gapiring.

### Tavsiya etilgan adabiyotlar:

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2013. – 161 с.

2. Недошивина Э.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб., – 2006. – 26 с.
3. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. работа с системой DDC // Языковая инженерия: в поиске смыслов. (электрон ресурс: <https://docplayer.ru>)
4. Структурно-функциональная компьютерная модель татарских морфем. – Казань: ФЭН, 2003. – 220 с.
5. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. –М.: ФЛИНТА: Наука, 2013. –128 с.
6. Ҳамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изоҳли луғати. – Тошкент: Камалак, 2018. – 96 б.
7. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.
8. Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.
9. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
10. [www.ruscorpora.ru](http://www.ruscorpora.ru)// Плунгян В. Зачем мы делаем корпусы?
11. <http://www.ruscorpora.ru>. НКРЯ
12. [www.philol.msu.ru/~lex/corpus](http://www.philol.msu.ru/~lex/corpus)
13. <http://www.philol.msu.ru/~lex/corpus>
14. <http://ruscorpora.ru>.

## KORPUS ASOSLANUVCHI LEKSIKOGRAFIK, GRAMMATIK VA DISKURS TADQIQOTLAR.

### Reja:

1. Korpus lingvistikasi va leksikografiya munosabati.
2. Korpusga asoslangan leksikografik mahsulotlar.
3. Korpusdan leksikografik mahsulot olish.

*Tayanch soʻz va iboralar:* korpus lingvistikasi, ideografik lugʻat, leksikografik mahsulot, asos, hosila.

**Ideografik teglash – ideografik lugʻat yoki tezaurus asosi.** Oldingi bobda korpusning lingvistik ahamiyati sanalgan, uning hosilasi sifatida turli, xususan, chastotali lugʻatlar yaratilishi haqida mulohaza yuritgan edik. Ushbu oʻrinda ideografik razmetkaning ahamiyatini ochib berish uchun lugʻatlardan biri – ideografik lugʻat xususiyatiga toʻxtalish joiz. Buyuk lugʻatshunoslardan biri V.I.Dal shunday yozadi<sup>191</sup>: “Алфавитный способ расположения слов крайне туп и сух. Самые близкие и сродные речения разносятся далеко врозь и томятся тут и там в одиночестве; всякая живая связь речи разорвана и утрачена..., читать такой словарь нет сил, на десятом слове ум притупеет и голова вскружится, потому что ум наш требует во всем какойнибудь разумной связи, постепенности и последовательности”. Yaʼni, V.I.Dalning fikricha, soʻzning lugʻatda joylashish tamoyili lugʻatning oʻquvchi uchun foydalanish qulayligi bilan belgilanadi.

Til ongimizda sistema sifatida yashashini bilamiz. N.V.Krushevskiy soʻzlarni lugʻatda joylashtirish masalasiga toʻxtalar ekan: “Hech qachon tilning asosiy xarakterini unutmaslik lozim. Soʻz–narsaning belgisi. Soʻz, narsa haqidagi tasavvur assotsiatsiya haqidagi qonun asosida juftlik hosil qilgan. Shunday ekan, soʻzlar bizning ongimizga ham oʻzi anglatadigan jismlar guruhi sifatida kirishi kerak”<sup>192</sup>, – degan edi. Leksikani ideografik nuqtayi nazardan oʻrganishning asosiy maqsadi soʻzning turli maʼnolarini bus-butun, yaxlitligicha oʻrganish, ularning oʻz oʻrnida qoʻllash va qoʻllanishini taʼminlashdir. Soʻzlarni semantik maydonga

<sup>191</sup> Идеографический словарь русского языка.(Сост. О.С.Баранов.) – М., 1995. – 4166 статей.

<sup>192</sup> Каранг: Крушевский Н.В. Избранные работы по языкознанию. – М.: Наследие, 1998.–296 с.

yig'ib o'rganish ularning turli xususiyati, leksik ma'nosining emotsional ta'siri, boshqa jihatini kuzatib, leksik ma'nosini o'zaro qiyoslab o'rganishda muhim omil sifatida xizmat qiladi. Semantik maydonning aniq chegarasi mavjud emas, bu tabiiy hol, chunki bir so'z turli pozitsiyada turlicha talqin qilinishi mumkin. Masalan, **olma** *mevalar nomi* semantik maydoni va **olma** – *daraxtlar nomi* semantik maydoni yoki *o'simliklar dunyosi* mavzuviy qurshovi tarkibida ham ishtirok etishi mumkin.

Har bir tilning leksikasi o'sha tilda gapiruvchi uchun faol, nafaol holatda bo'ladi. Faol qatlamga kundalik so'zlashuv so'zlari kiradi, nutq jarayonida u so'zlarni miyamizdan qidirib o'tirishga hojat yo'q, bu so'zni aytishga ehtiyoj tug'ilgan zahotiy oq talaffuz etamiz. Nafaol so'z esa biz biladigan, lekin kundalik nutqimizda kam ishlatadigan so'zlarimizdir. Faol so'zda ma'no, talaffuz orasidagi munosabat ikki tomonlama:  $M \leftrightarrow T$ . Nafaol so'z orasidagi aloqa bir tomonlama, chunki uni o'qiymiz yoki biror kishi nutqidan eshitamiz, uning nutqimizda faollashuvi tez sodir bo'lmaydi, uni eslash kerak. Demak,  $T \rightarrow M$ . Bundan kelib chiqadiki, eshitganimiz, o'qiganimizni tushunishga yordam beruvchi lug'at (alifbo lug'ati) foydali, so'zni butun mohiyati, sistemasi bilan ko'z oldimizga keltiruvchi lug'at (ideografik lug'at) esa ma'lum bir til imkoniyatidan bema'lol, oson foydalanish imkoniyatini beradi. Shuning uchun bunday lug'at ham muhim axborot banki sirasiga kiradi. Yu.N.Karaulov ideografik lug'atga shunday ta'rif beradi: "Barcha tashkil qiluvchi birliklari orasida kuchli semantik munosabat yaqqol ko'rinuvchi lug'atni tezaurus deb ataymiz"<sup>193</sup>. Oddiy izohli lug'atdan biror so'zni qidirmoqchi bo'lsak, uning moddiy tomoni (harflar birikmasi)ga e'tibor beramiz. Bunday lug'atga kalit bitta: alifbo. Yu.N.Karaulov alifbo lug'ati, tezaurus haqida shunday misol keltiradi: "D.N.Ushakov lug'atida"<sup>194</sup> **dissoniruyushiy** so'zi *dissonirovat*, *dissotsiatsiya* so'zlari orasida keladi. Bu so'z musiqa sohasiga oid, bu sohaga tegishli barcha so'zlar haqida bilimga ega bo'lish

---

<sup>193</sup> Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Прометей, 1990.

<sup>194</sup> "Толковый словарь русского языка". В 4 т.. (Под ред. Д.Н.Ушакова.) – М.: Государственный институт СЭ, 1935-1940.

uchun butun lug‘atni o‘qib chiqish lozim. X.Kasares tezaurusi<sup>195</sup> da esa bu so‘z “Musiqqa” maqolasi ichida quyidagi sifatlar qatorida keladi: “... мелодичный, гомофонический, консонирующий, созвучный, аккордийный, гармонический, диатонический, полухроматический, энгармонический, полифалический, характерный, фугированный...”. Bu maqola 508 so‘zni qamrab olgan. Qisqacha aytganda, butun lug‘atni emas, undagi atigi 1-2 betni varaqlasak, kerakli so‘z topiladi. Demak, tezaurusda sistema, semantika markazda turadi. Kalit so‘zlar esa alifbo tartibida bo‘ladi. Ideografik lug‘atning yana bir ahamiyatli tomoni shundaki, u turli tillar leksikasini qiyosiy o‘rganishda muhim ahamiyatga ega, tarjima lug‘ati tuzishda qo‘l keladi. Chunki ma‘lum tildagi narsa bir so‘z bilan atalsa, boshqa tilda bir necha so‘z bilan izohlanuvchi birlikka to‘g‘ri kelishi mumkin. Yevropada ideografik lug‘atga qiziqish uyg‘otgan “Ingliz so‘z, iboralari tezaurusi” (“Тезаурус английских слов и выражений”) muallifi P.M.Rodjet o‘z lug‘ati<sup>196</sup> so‘z boshida aytadi: “Bog‘li matn tuzayotgan yozuvchi yoki xohlagan kishi nima demoqchiligini yaxshi bila turib, ayni paytda o‘rinli ibora yoki so‘zni topolmasa, boy, kerakli so‘z tez topiladigan, to‘g‘ri sistemalashtirilgan, guruhlarga bo‘lingan bir kitobdan yordam oladi. Bir so‘z bilan aytganda, tezaurusga murojaat etadi... Biz aytmoqchi bo‘lgan fikrimizni qanchalik aniq tasavvur qilmaylik, ba‘zan shunday bir vaziyatga tushib qolamizki, bizga lug‘at boyligimizdagi so‘z etishmay qoladi. O‘z fikrimizni aniq ifodalash uchun bitta so‘z o‘rniga bir necha ijobiy, salbiy, kuchli, kuchsiz, ba‘zan esa keraksiz so‘zlarni ishlatib yuboramiz. Ideografik lug‘atning eng muhim xususiyati shundaki, u foydalanuvchiga bir fikrning barcha qirralarini ochib beruvchi so‘z, iboralarning boy to‘dasini tortiq qiladi”. Demak, ideografik lug‘at maqolalari odatdagidek alifbo tartibida emas, balki ma‘no (bosh so‘z, so‘z birikmaning leksik ma‘nosi) asosida tartib berilgan lug‘atdir. Alifbo tartibli lug‘atdan aynan bir so‘z ma‘nosini bilish mumkin bo‘lsa, ideografik lug‘at biror tushuncha haqida, bir ma‘noni qanday so‘zlar bilan ifodalash mumkinligi borasida ma‘lumot beradi<sup>197</sup>. Leksikografiya ideografik lug‘atlar tuzish borasida

---

<sup>195</sup> Casares J. Diccionario ideológico de la lengua Española. Barcelona.–1951.

<sup>196</sup> Караңг: Roget P.M. Thesaurus of English words and phrases. Lnd. – 1852.

<sup>197</sup> <https://ru.m.wikipedia.org>.

katta yutuqlarga erishdi, uning ideografik tezaurus<sup>198</sup>, analogik lug‘at<sup>199</sup>, o‘quv-tematik lug‘at kabi turlarini yaratdi. Leksikografiya tarixida ingliz, fransuz, nemis, ispan tillari materiali asosida tuzilgan lug‘atlar anchagina<sup>200</sup>. Bu siraga rus leksikograflari K.Plots, K.Babov, A.Vorgulev, L.Rozkovsova, Z.Ganusova, V.V.Morkovkin, Yu.N.Karaulov, nemis leksikografi G.O.Fogt, ingliz leksikograflari Dj.Roudel, P.M.Rodjetlar ishlarini kiritish mumkin.

Ideografik lug‘atning ahamiyati korpusda ideografik razmetkaga ehtiyoj borligini ko‘rsatadi. Muallif asari mavzularini tasnif qilish natijasida birliklar orasidagi mantiqiy aloqadorlik asosida birlashgan so‘zlar to‘plamini aks ettiruvchi ideografik lug‘at paydo bo‘ladi. A.P.Chexov nasriy, dramatik asarlarining mavzuviy razmetkasi xususiyatlari E.V.Surovseva tomonidan alohida tadqiq etilgan<sup>201</sup>. Bunday mavzuviy razmetkani amalga oshirish uchun uning 583 ta nasriy, 17 ta dramatik asari matni tanlangan. Tahlilga olingan matnlar uchun muallifning mukammal asarlari to‘plami (Чехов А.П. Полное собрание сочинений и писем. В 30 томах. М.: Наука, 1974-1983.) asos qilib olingan. Ideografik teglash uchun avtomatik algoritmi – maxsus dastur bo‘lmaganligi sababli teglash qo‘lda amalga oshiriladi. Chunki har bir muallif asari mavzu doirasi turlicha bo‘ladi. Avval muallifning barcha asarlari matni tahlil qilinib, keyin umumiy xulosaga kelinadi. Mavzuviy teg matn mavzusi asosida leksik-semantik guruhlar shakllantirishni maqsad qilib belgilaydi. Matn tanlanib, undan ma’lum bir mavzu atrofida birlashgan so‘zlar belgilanadi. So‘zlarni belgilashda matnni qayta aytib berish uchun butun matn mazmunini to‘liq ifodalashga xizmat qiluvchi so‘z, iboralar ajratiladi. Aynan shu matndan boshqa bir leksik-semantik guruhga (LSG) mansub so‘zlar alohida rang bilan belgilanadi. Masalan, E.V.Surovseva A.P.Chexov asarlari ideografik razmetkasi uchun yozuvchining “Semiz va oriq”

---

<sup>198</sup> *Қаранг*: Roget P.M. Thesaurus of English words and phrases. Lnd. – 1852.

<sup>199</sup> March F.A. March’s Thesaurus Dictionary. N.Y.– 1958.

<sup>200</sup> Масалан, Lorange R. Le mot et l’idée. Paris. – 1965.

<sup>201</sup> Суровцева Е.В. Словарь тем писателя как разновидность идеографических словарей. На материале художественной прозы и драматургии А.П.Чехова. // Acta linguistica, Vol. 4 (2010), 2; Лапони́на Л.В., Суровцева Е.В. Тематическая разметка в корпусе художественной прозы А.П.Чехова // Чехов и мировая культура: взгляд из XXI века. Тезисы докладов международной научной конференции (Москва, 29 января – 3 февраля 2010 года) / Сост. и ред. Р.Б.Ахметшин, М.О.Горячева, В.Б.Катаев (отв. ред.). М.: Издательство Московского университета, 2010.



hikoyasini obyekt sifatida tanlaydi. Birinchi bosqichda hikoyadan *uchrashuv*, ikkinchi bosqichda *bolalik*, *xotira* mazmuniy guruhiga kiruvchi soʻzlarni alohida ranglar bilan belgilaydi, *doʻstlik* mazmuniy guruhiga kiruvchi soʻzlarni kursivda ajratadi. Shunday tahlildan keyin koʻrinadiki, LSG bir xil emas. Bu guruhlar *holat*, *mavhum tushuncha*, *qahramon tipi* yoki *voqea joyi* kabi guruhlar boʻlishi mumkin. Demak, bitta hikoyada bir necha LSGga kiruvchi soʻzlar guruhi aniqlandi. Faqat bitta LSGga kiruvchi soʻzlardan iborat asar (hikoya) kam uchraydi. Masalan, “Bahorni kutish”da faqatgina *uchrashuv*, “Odamning qisqacha anotomiyasi”da *ilm*, “Parcha”da *provinsiya* kabi LSGlar mavjud. Baʼzi hikoyalarida esa oʻnlab LSG aniqlangan. Asarda hikoyachi, qahramonda xayol surish, fikr yuritish, tahlil qilishga moyillik boʻlsa, oʻsha asarni mavzuviy razmetkalash oson kechadi. Barcha asarlar tahlil qilingandan soʻng mavzular ikkiga: makro, mikromavzuga ajratilib, baʼzi mavzular bir vaqtning oʻzida ikki makromavzuga tegishli boʻlib qolishi mumkin. Deylik, *hayit* LSGi mazmuniga koʻra ham *din*, ham *vaqt* makromavzusiga kiritilishi mumkin. E.V.Surovsevaning fikricha, korpusda bu tipdagi razmetkaning mavjudligi rus tili (umuman olganda, barcha tilni)ni xorijiy til sifatida oʻqitishning samaradorligini oshirishda, korpus materialidan taʼlimiy maqsadda foydalanishda ahamiyatli. Chunki mavzulashtirilgan matn xorijiy til taʼlimi jarayonida oʻquv materialini tayyorlashda axborot manbai vazifasini bajaradi. Hatto bu usulda razmetkalangan korpus yordamida jahon mualliflik kartinasini kuzatish qulayligi ham paydo boʻladi. Ideografik razmetkali korpus yordamida bir muallif asarlarini yaxlit holda gʻoyaviy-badiiy tahlil qilishda, u yoki bu mavzuni qaysi usul va vosita bilan ifodalaganligini oʻrganishda ham asqotadi. Masalan, A.Qahhorning inson, maʼnaviyat mavzui koʻtarilgan barcha asarlarini bir prizmadan oʻtkazish orqali u yoki bu til ifoda vositalarini qoʻllash mahoratini ham oʻrganish mumkin. Shuningdek, stilistik, lingvo-poetik, kognitiv, pragmatik, lingvomadaniy, konseptologik, diskursiv nuqtai nazardan oʻrganishda qoʻl keladi.

### **Savol va topshiriqlar**

1. Korpus lingvistikasi va leksikografiya oʻzaro qanday munosabatga ega?
2. Korpusga asoslangan leksikografik mahsulotlar nimalarni tashkil qiladi?

3. Korpusdan leksikografik mahsulot olish jarayonini tavsiflang.
4. Korpusdan qanday lug‘atlar tuzishda foydalanish mumkin?
5. Lug‘at tuzishda korpusning qaysi imkoniyatlari qo‘l keladi?

**Tavsiya etilgan adabiyotlar:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика.– Иркутск: ИГЛУ, 2013. – 161 с.
2. Идеографический словарь русского языка.(Сост. О.С.Баранов.) – М., 1995. – 4166 статей.
3. Информационные технологии в лингвистике: учеб. Пособие / Л.Ю.Щипицина. – М.: ФЛИНТА: Наука, 2013. –128 с.
4. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Прометей, 1990.
5. Крушевский Н.В. Избранные работы по языкознанию. – М.: Наследие, 1998.–296 с.
6. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. работа с системой DDC. // Языковая инженерия: в поиске смыслов. (электрон ресурс: <https://docplayer.ru>)
7. Недошивина Э.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб., – 2006. – 26 с.
8. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.
9. Толковый словарь русского языка. В 4 т.. (Под ред. Д.Н.Ушакова.) – М.: Государственный институт СЭ, 1935-1940.
10. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Монография. – Германия: GlobeEdit, 2020. – 250 б.
- 11.Хамроева Ш.М. Корпус лингвистикаси атамаларининг қисқача изохли луғати. – Тошкент: Камалак, 2018. – 96 б.
- 12.Casares J. Dissionario ideologico de la lengua Espanola. Barselona.–1951.
13. Loriche R. Le mot et l’idee. Paris. – 1965.
14. March F.A. March’s Thesaurus Dictionary. N.Y.– 1958.

15. Roget P.M. Thesaurus of English words and phrases. Lnd. – 1852.

16. <http://ruscorpora.ru>.

17. <https://ru.m.wikipedia.org>.

## **IV BOB. AMALIY MASHG‘ULOT UCHUN TOPSHIRIQLAR**

### **1-amaliy mashg‘ulot: Lingvistik korpus imkoniyatlarini o‘rganish**

**Mashg‘ulot maqsadi:** Talabalarda Rus tili milliy korpusi (НКРЯ) (ruscorpora.ru), Хельсинки (<http://h248.it.helsinki.fi/hanco/>), Aranea (unesco.uniba.sk) korpuslari misolida lingvistik korpuslar imkoniyatlari haqida umumiy tasavvur hosil qilish; Milliy korpusi qidiruv tizimi ustida ishlash.

**Tayanch tushunchalar:** milliy korpus, korpus menejeri, interfeys.

**Topshiriq mazmuni:** НКРЯ sayti (ruscorpora.ru) haqida ma’lumot berish. Rus tili milliy korpusi imkoniyati va afzalliklarini o‘rgatish. Rus tili milliy korpusi qidiruv tizimi hamda interfeysi haqida bilim hosil qilish.

**Topshiriqni bajarishda foydalaniladigan vositalar:** НКРЯ sayti (ruscorpora.ru)

#### **Topshiriq:**

1. Rus tili milliy korpusi (НКРЯ) (ruscorpora.ru), Хельсинки (<http://h248.it.helsinki.fi/hanco/>), Aranea (unesco.uniba.sk) saytlariga kiring.
2. Korpus xaritasi yordamida korpus tarkibi bilan tanishing va shu korpuslar asosida qiyosiy jadval tayyorlang.
3. НКРЯ, Хельсинки, Aranea korpuslari interfeysini qiyoslang, o‘xshash va farqli tomonlarini izohlang.
4. НКРЯ, Хельсинки, Aranea korpuslari qidiruv oynalarini tavsiflang.

### **2-amaliy mashg‘ulot: Rus tili milliy korpusi (НКРЯ) misolida lingvistik korpuslarga xos xususiyatlarni o‘rganish**

**Mashg‘ulot maqsadi:** Talabalarda Rus tili milliy korpusi (НКРЯ) misolida lingvistik korpuslarga xos xususiyatlar haqida umumiy tasavvur hosil qilish; Rus tili milliy korpusi (НКРЯ) imkoniyati va afzalliklarini tahlil qilish. Rus tili milliy korpusi (НКРЯ) qidiruv tizimi hamda interfeysini tahlil qilish. Lingvistik ma’lumot hamda metama’lumotlar asosidagi qidiruvni, tilshunoslikning turli bo‘limlariga oid topshiriqlarni bajarishni o‘rgatish.

**Tayanch tushunchalar:** Rus tili milliy korpusi, interfeys, qidiruv tizimi.

**Topshiriq mazmuni:** НКРЯ sayti (ruscorpora.ru) haqida ma'lumot berish. Rus tili milliy korpusi imkoniyati va afzalliklarini o'rgatish. Rus tili milliy korpusi interfeysi haqida bilim hosil qilinadi. Rus tili milliy korpusi (НКРЯ) misolida lingvistik korpuslarga xos xususiyatlar haqida umumiy tasavvur hosil qilinadi; Rus tili milliy korpusi (НКРЯ) qidiruv tizimi hamda interfeysini tahlil qilish. Lingvistik ma'lumot hamda metama'lumotlar asosidagi qidiruvni, tilshunoslikning turli bo'limlariga oid topshiriqlarni bajarishni o'rgatish.

**Topshiriqni bajarishda foydalaniladigan vositalar:** НКРЯ sayti (ruscorpora.ru)

### **Topshiriq:**

1. Бодрый so'zining mujskoy roddagi hayvon nomlari bilan birikmalarini toping. Alifbo bo'yicha saralang, tartiblashtiring.
2. Asosiy korpusdan inson xulq-atvori va ijobiy xatti-harakatini ifodalovchi –сть qismi bilan tugaydigan birliklarni toping (исть qismli so'zlar mustasno).
3. Rus tili milliy korpusi (НКРЯ)dan читать, слушать, приходиться fe'llari bilan qo'llanuvchi mujskoy, jenskiy roddagi ravishlarni toping. 10 ta kontekstni alohida faylga ko'chiring.
4. Rus tili milliy korpusi (НКРЯ)dan «выбор» (yoki «выборы») so'zi qanday ma'noda kelishini aniqlang. Ayni bitta ma'nosiga 10 ta misol keltiring.
5. «медвежья услуга» birikmasi qaysi frazemalar bilan qo'llanilishini izohlang. (Yechim; javob namunasi; kontekstlar)
6. Rus tilida *пра-* pistavkasi borligini rad etuvchi qoidaning o'rinsiz ekanligini isbotlang.
7. Til rivojlanishining qaysi davrida «набитый дурак» frazemasini «круглый дурак» frazemasiga nisbatan ko'p/kamroq qo'llanganligini tavsiflang.

### **Uyga vazifa. Topshiriqni bajaring.**

1. Ichki korpus tuzing: 1990-2000-yillar publitsistikasi. His-hayajon fe'llarini toping.
2. <https://ruscorpora.ru/new/search-poetic.html> manzilidan foydalanib ichki korpus tuzing: Rossiyaning sovet davridan keyingi hayotni tasvirlovchi matnlar.

3. Quyidagilarni toping: 1) romanslar; 2) erkak muallif tomonidan yozilgan; 3) 4 misradan iborat band.

### **3-amaliy mashg‘ulot: Rus tili milliy korpusi (НКРЯ) misolida lingvistik korpuslarga xos xususiyatlarni o‘rganish**

**Mashg‘ulot maqsadi:** Talabalarda Rus tili milliy korpusi (НКРЯ) misolida lingvistik korpuslarga xos xususiyatlar haqida umumiy tasavvur hosil qilish; Rus tili milliy korpusi (НКРЯ) imkoniyati va afzalliklarini tahlil qilish. Rus tili milliy korpusi (НКРЯ) qidiruv tizimi hamda interfeysini tahlil qilish. Lingvistik ma’lumot hamda metama’lumotlar asosidagi qidiruvni, tilshunoslikning turli bo‘limlariga oid topshiriqlarni bajarishni o‘rgatish.

**Tayanch tushunchalar:** Rus tili milliy korpusi, interfeys, qidiruv tizimi.

**Topshiriq mazmuni:** НКРЯ sayti (ruscorpora.ru) haqida ma’lumot berish. Rus tili milliy korpusi imkoniyati va afzalliklarini o‘rgatish. Rus tili milliy korpusi interfeysi haqida bilim hosil qilinadi. Rus tili milliy korpusi (НКРЯ) misolida lingvistik korpuslarga xos xususiyatlar haqida umumiy tasavvur hosil qilinadi; Rus tili milliy korpusi (НКРЯ) qidiruv tizimi hamda interfeysini tahlil qilish. Lingvistik ma’lumot hamda metama’lumotlar asosidagi qidiruvni, tilshunoslikning turli bo‘limlariga oid topshiriqlarni bajarishni o‘rgatish.

**Topshiriqni bajarishda foydalaniladigan vositalar:** НКРЯ sayti (ruscorpora.ru)

#### **Topshiriq:**

1. Insonning belgi-xususiyatini bildiruvchi salbiy ottenkali sifatlarni toping. Kontekstlar namunasini keltiring. (Yechim; javob namunasi; kontekstlar; Bo‘lim: leksikografiya, semantika)
2. Adj + Adj + N sxemali «кот» so‘zi bilan hosil bo‘luvchi yuqori chastotali kollokatlarni toping. (Yechim; javob namunasi; kontekstlar; Bo‘lim: frazeologiya (kollokatsiya), sintaksis, morfologiya)
3. «с моей точки зрения» oborotining uslubiy xoslanishini (badiiy/ilmiy) aniqlang. (Yechim; javob namunasi; kontekstlar; Bo‘lim: frazeologiya)

4. Kichraytirish-erkalash shaklidagi плюшевый leksemasining mebel jihozlari bilan birikuviga (predlogsiz qoʻllanish) misollar toping (Masalan, плюшевый диванчик).

5. Rus tili milliy korpusi (НКРЯ)dan «пере-» приставkasi bilan yasalgan feʼllarni toping. 20 ta kontekstni alohida faylga koʻchiring.

6. Rus tili milliy korpusi (НКРЯ)dan «однако» soʻzshakli kirish soʻz boʻlgan holatlarni toping. Turli janrdagi misollar namunasini shakllantiring.

#### **Uyga vazifa. Topshiriqni bajaring.**

1. Rus tili milliy korpusi (НКРЯ)dan «язык» soʻzi “ovqat” maʼnosini bildirgan kontekstlarga misollar toping. Badiiy uslubga mansub turli janrlardan 10 ta misollar kontekstini namuna sifatida keltiring.

2. Rus tili milliy korpusi (НКРЯ)dan «язык» soʻzi “tana aʼzosi” maʼnosini bildirgan kontekstlarga misollar toping. Badiiy uslubga mansub turli janrlardan 10 ta misollar kontekstini namuna sifatida keltiring.

3. <https://ruscorpora.ru/new/reqsem.html> manzilidan koʻrsatilgan oyna boʻyicha turli kollokatsiyadagi qidiruvni amalga oshiring hamda natijani fayl koʻrinishida taqdim eting.

#### **4-amaliy mashgʻulot: Ingliz-rus parallel korpusi ustida ishlash**

**Mashgʻulot maqsadi:** Talabalarda Rus tili milliy korpusi (НКРЯ) ([ruscorpora.ru](https://ruscorpora.ru)) tarkibidagi rus-ingliz, ingliz-rus parallel korpusi (<https://ruscorpora.ru/new/search-para-en.html>) misolida parallel korpus haqida umumiy tasavvur hosil qilish; rus-ingliz, ingliz-rus parallel korpusi ustida ishlash.

**Tayanch tushunchalar:** parallel korpus, rus-ingliz parallel korpusi, ingliz-rus parallel korpusi, korpus menejeri, interfeys.

**Topshiriq mazmuni:** Rus tili milliy korpusi (НКРЯ) ([ruscorpora.ru](https://ruscorpora.ru)) tarkibidagi rus-ingliz, ingliz-rus parallel korpusi haqida maʼlumot berish; parallel korpuslar qidiruv funksiyasi bilan tanishish.

**Topshiriqni bajarishda foydalaniladigan vositalar:** НКРЯ sayti ([ruscorpora.ru](https://ruscorpora.ru)), <https://ruscorpora.ru/new/search-para-en.html>

### **Topshiriq:**

1. Rus tili milliy korpusi (НКРЯ) tarkibidan rus-ingliz parallel korpusini toping.
2. Korpusda “вода”, “ребёнок”, “water”, “child” so‘zlari qidiruvini amalga oshiring. Natijalarni o‘rganish va saralang.
3. 1p,persn,m gammemali, ya’ni ism, mujskoy rod, I shaxs grammatik belgilari asosida qidiruv bering, natijalarni tahlil qiling, saralang va fayl ko‘rinishida saqlang.
4. Taksonomiya bo‘limidan “растения”, mereologiya qismidan “части растений” kabi semantik belgi asosida qidiruv tashkil qiling, natijalarni tahlil qiling, saralang va fayl ko‘rinishida saqlang.

### **9-amaliy mashg‘ulot: Turkiy tillardagi korpuslar bilan ishlash**

**Mashg‘ulot maqsadi:** Talabalarda turkiy tillarda yaratilgan korpuslar haqida umumiy tasavvur hosil qilish; tatar, turk, qozoq, tuva, yoqut tili milliy korpusi haqida umumiy tasavvur hosil qilish. Turkiy tillar korpuslari interfeysi, imkoniyatlari, qidiruv tizimining afzalliklari, o‘xshash va farqli tomonlari haqida tushuncha berish.

**Tayanch tushunchalar:** turkiy tilli korpuslar, TurkLang (turklang.tatar) platformasi, tatar tili milliy korpusi, qozoq tili korpusi.

**Topshiriq mazmuni:** Talabalar turkiy tillarda yaratilgan korpuslar bilan tanishadi; tatar, turk, qozoq, tuva, yoqut tili milliy korpuslari ustida amallar bajarish bilan ularning interfeysi, imkoniyatlari, qidiruv tizimining afzalliklari, o‘xshash va farqli tomonlarini biliib olishadi.

**Topshiriqni bajarishda foydalaniladigan vositalar:** TurkLang platformasi (<http://www.turklang.net/ru/>), turkiy tillar korpuslari, [https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic\\_uz](https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic_uz)

<http://altay2.gasu.ru/>

**Корпус башкирского языка**

<http://web->



(поэтический)	<a href="http://corpora.net/bashcorpus/search/">corpora.net/bashcorpus/search/</a>
Корпус казахского языка	<a href="http://kazcorpus.kz/">http://kazcorpus.kz/</a>
Корпус крымскотатарского языка	<a href="http://korpus.juls.savba.sk/QIRIM/">http://korpus.juls.savba.sk/QIRIM/</a>
Корпус татарского языка ‘Туган тел’	<a href="http://www.corpus.antat.ru">http://www.corpus.antat.ru</a>
Письменный корпус татарского языка	<a href="http://www.corpus.tatar/">http://www.corpus.tatar/</a>
Корпус тувинского языка	<a href="http://www.tuvacorporus.ru">http://www.tuvacorporus.ru</a>
Корпус турецкого языка	<a href="http://www.tnc.org.tr/">http://www.tnc.org.tr/</a>
Корпус хакасского языка	<a href="http://khakas.altai.ru/">http://khakas.altai.ru/</a>
Корпус шорского языка	<a href="http://corpora.iea.ras.ru/corpora/">http://corpora.iea.ras.ru/corpora/</a>
Корпус якутского языка	<a href="http://adictsakha.nsu.ru/corpora/corp">http://adictsakha.nsu.ru/corpora/corp</a>

### **Topshiriq:**

1. Quyida manzili ko‘rsatilgan turkiy tilli korpuslar bilan tanishing:

<http://korpus.juls.savba.sk/QIRIM/>

<http://www.corpus.antat.ru>

<https://www.corpus.tatar/>

<https://search.corpus.tatar/index.php?of=search/search.php>

[https://bonito.corpus.tatar/run.cgi/first\\_form](https://bonito.corpus.tatar/run.cgi/first_form)

<http://kazcorpus.kz/>

2. <http://korpus.juls.savba.sk/QIRIM/> korpusida so‘zshakllar qidiruvini amalga oshiring. Misollarni alohida faylda saqlang.

3. [http://bonito.corpus.tatar/run.cgi/first\\_form](http://bonito.corpus.tatar/run.cgi/first_form) manzilida so‘zshakllar qidiruvini amalga oshiring va natijani fayl shaklida saqlang.

4. <http://search.corpus.tatar/index.php?of=search/search.php> manzilidagi tatar tili yozma korpusidan “йокы” so‘zining o‘ng va chap tomondagi birikuvchilarini toping. Qaysi so‘zlar bilan qo‘llanish chastotasi balandligini aniqlang.

### **10-amaliy mashg‘ulot: O‘zbek tilidagi korpus bilan ishlash**

**Mashg‘ulot maqsadi:** Talabalarda SketchEngine platformasidagi o‘zbek tili korpusi haqida bilim hosil qilish, o‘zbek tilida so‘rovni amalga oshirish ko‘nikmasini shakllantirish. Ushbu korpusdagi so‘zshakllar qidiruv imkoniyati, so‘zshakllar miqdori haqida ma’lumot berish.

**Tayanch tushunchalar:** o‘zbek tili korpusi, SketchEngine platformasi.

**Topshiriq mazmuni:** Talabalar SketchEngine platformasi bilan tanishadi, o‘zbek tilida qidiruvchi amalga oshiradi, natijani alohida fayl shaklida saqlaydi, “SketchEngine”da so‘zshakllar qidiruv imkoniyati, so‘zshakllar miqdori haqida ma’lumot oladi.

#### **Topshiriqni bajarishda foydalaniladigan vositalar:**

1. SketchEngine platformasi
2. [https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic\\_uz](https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic_uz)

#### **Topshiriq:**

[https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic\\_uz](https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic_uz)

Qarang: Uzbek.PNG

### **11-amaliy mashg‘ulot: O‘zbek tilidagi matnlar asosida korpus tuzish**

**Mashg‘ulot maqsadi:** Talabada mustaqil ravishda korpus tuzish ko‘nikmasini shakllantirish: korpus tuzish bosqichlari haqida ma’lumot berish; korpus tuzish algoritmini ishlab chiqish; korpus materiali – matnlarni tayyorlash, matnlarni qidiruv tizimiga bo‘sondirish.

**Tayanch tushunchalar:** lingvistik korpus, korpus menejeri, interfeys.

**Topshiriq mazmuni:** Talaba korpus tuzish bosqichlari haqida ma’ruza mashg‘ulotlarida olgan bilimlarni mustahkamlaydi; korpus tuzish algoritmi bo‘yicha mustaqil reja tuzadi; korpus uchun matn yig‘adi, reprezentativlik talablari

asosida matnlarni janriga ko‘ra saralaydi; korpusga mos qidiruv tizimi tanlaydi hamda matnlarni korpusini shakllantiradi.

**Topshiriqni bajarishda foydalaniladigan vositalar:**

1. English corpora (Araneum Anglicum)
2. [https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic\\_uz](https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic_uz)

**Topshiriq:**

1. Korpus tuzish bosqichlari haqida ma‘ruza mashg‘ulotlarida olingan bilimlarni takrorlang.
2. O‘zingizga ma‘qul mavzuda 10000-15000 so‘z atrofidagi matnlarni yig‘ing.
3. Matnlarni korpus talabiga moslang: rasm, jadval, noleksik birliklarni o‘chiring.
4. Korpus qidiruv tizimini ishga tushirish uchun o‘zingiz biladigan korpus menejrlari orasidan mosini tanlang.
5. Korpusni ishga tushiring hamda test rejimida sinab ko‘ring.
6. So‘zshakl va lemma qidiruvini amalga oshiring, natijani fayl ko‘rinishida saqlab oling.

**12-amaliy mashg‘ulot: Internet matnlari asosida o‘zbek tili korpusini tuzish**

**Mashg‘ulot maqsadi:** Talabada internet materiali asosida mustaqil ravishda korpus tuzish ko‘nikmasini shakllantirish: korpus tuzish bosqichlari haqida ma‘lumot berish; korpus tuzish algoritmini ishlab chiqish; korpus materiali – matnlarni tayyorlash, matnlarni qidiruv tizimiga bo‘sondirish.

**Tayanch tushunchalar:** internet, lingvistik korpus, korpus menejeri, interfeys.

**Topshiriq mazmuni:** Talaba korpus tuzish bosqichlari haqida ma‘ruza mashg‘ulotlarida olgan bilimlarni mustahkamlaydi; korpus tuzish algoritmi bo‘yicha mustaqil reja tuzadi; korpus uchun internetdan matn yig‘adi, representativlik talablari asosida matnlarni janriga ko‘ra saralaydi; korpusga mos qidiruv tizimi tanlaydi hamda matnlarni korpusini shakllantiradi.

**Topshiriqni bajarishda foydalaniladigan vositalar:**

1. English corpora (Araneum Anglicum)

2. [https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic\\_uz](https://app.sketchengine.eu/#dashboard?corpname=preloaded%2Fturkic_uz)

**Topshiriq:**

1. Korpus tuzish bosqichlari haqida ma'ruza mashg'ulotlarida olingan bilimlarni takrorlang.
2. Internetdan tanlangan mavzuda 10000-15000 so'z atrofida mant yig'ing.
3. Matnlarni korpus talabiga moslang: rasm, jadval, noleksik birliklarni o'chiring.
4. Korpus qidiruv tizimini ishga tushirish uchun o'zingiz biladigan korpus menejerlari orasidan mosini tanlang.
5. Korpusni ishga tushiring hamda test rejimida sinab ko'ring.
6. So'zshakl va lemma qidiruvini amalga oshiring, natijani fayl ko'rinishida saqlab oling.



**Захаров Виктор Павлович** – Санкт-Петербург давлат университети филология факультети математик лингвистика кафедраси доценти, филология фанлари номзоди. Россия Фанлар академияси Лингвистик тадқиқотлар институти катта илмий ходими.

**Номзодлик диссертацияси:** «Ахборот қидирувининг тилга оид воситалари» -1997.

**Қизиқишлар соҳаси:** компьютер лингвистикаси, корпус лингвистикаси, ахборот қидируви тиллари, компьютер лексикографияси.

**Илмий нашрлари:** «Информационные системы (документальный поиск): Учебное пособие» (2002), «Информационные технологии: Основы работы с программными средствами и информационными ресурсами Интернет: Практикум» (2003), «Корпусная лингвистика: Учебно-метод. пособие (2005), Информационно-поисковые системы: Учебно-метод. пособие» (2005), «Корпусная лингвистика: Учебно-метод. пособие» (2018).

**Захаров Виктор Павлович** – Кандидат филологических наук, доцент кафедры математической лингвистики филологического факультета Санкт-Петербургского государственного университета, старший научный сотрудник Института лингвистических исследований РАН.

**Кандидатская диссертация:** «Языковые средства информационного поиска» - 1997.

**Научные интересы:** компьютерная лингвистика, корпусная лингвистика, информационный поиск, компьютерная лексикография

**Научные публикации:** «Информационные системы (документальный поиск): Учебное пособие» (2002), «Информационные технологии: Основы работы с программными средствами и информационными ресурсами Интернет: Практикум» (2003), «Корпусная лингвистика: Учебно-метод. пособие (2005), Информационно-поисковые системы: Учебно-метод. пособие» (2005), «Корпусная лингвистика: Учебно-метод. пособие» (2018).



**Менглиев Бахтиёр Ражабович** – Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёт университети Ўзбек тилшунослиги кафедраси профессори, филология фанлари доктори.

**Докторлик диссертацияси:** «Лисоний тизим яхлитлиги ва унда сатҳлараро муносабатлар» - 2003.

**Қизиқишлар соҳаси:** амалий филология, компьютер лингвистикаси, корпус лингвистикаси, ўқув луғатчилиги,

**Илмий нашрлари:** «Тил яхлит система сифатида», «Она тили», «Замонавий она тили», «Ҳозирги ўзбек адабий тили», «Ўзбек тилидан универсал қўлланма», «Замонавий ўзбек тили. Морфология», «Замонавий ўзбек тили. Синтаксис», «Modern uzбек language» каби монография ва ўқув қўлланмалар.

**Менглиев Бахтиёр Ражабович** – Профессор Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои, доктор филологических наук.

**Докторская диссертация:** «Целостность языковой системы и межуровневые отношения» – 2003.

**Научные интересы:** прикладная лингвистика, компьютерная лингвистика, корпусная лингвистика, учебная лексикография.

**Научные публикации:** «Язык как целая система», «Родной язык», «Современный родной язык», «Современный узбекский язык», «Универсальное пособие по узбекскому языку», «Современный узбекский язык. Морфология», «Современный узбекский язык. Синтаксис», «Modern uzбек language».



**Хамроева Шахло Мирджоновна** – Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёт университети доценти, филология фанлари доктори

**Докторлик диссертацияси:** “Ўзбек тили морфологик анализаторининг лингвистик таъминоти” – 2021.

**Қизиқишлар соҳаси:** амалий филология, компьютер ва корпус лингвистикаси, электрон таълим ресурслари.

**Илмий нашрлари:** «Морфологиядан машқлар тўплами», «Ҳозирги ўзбек адабий тили фанидан таълим технологияси. Академик лицей ва коллеж ўқитувчилари учун методик қўлланма», «Она тили таълимида замонавий педагогик технологиялар», «Корпус лингвистикаси атамаларининг қисқача изоҳли луғати» каби луғат, ўқув-услубий қўлланмалар.

**Хамроева Шахло Мирджоновна** – Доцент Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои, доктор филологических наук.

**Докторская диссертация:** “Лингвистическое обеспечение морфологического анализатора узбекского языка” – 2021.

**Научные интересы:** прикладная, компьютерная и корпусная лингвистика, электронные образовательные ресурсы.

**Научные публикации:** «Сборник упражнений по морфологии», «Педтехнология по курсу «Современный узбекский язык» Уч. пособ. для учителей академического лицея и коллежа», «Современные педагогические технологии в преподавание родного языка», «Краткий толковый словарь терминов корпусной лингвистики».