

“ABDURAUF FITRAT KORPUSI” LEKSIKASINING STATISTIK TAHLILI

Hamroyeva Shahlo Mirdjonovna,
Filologiya fanlari doktori, professor v.b.
shaxlo.xamrayeva@navoiy-uni.uz
ToshDO‘TAU

Alimbekova Mavjuda Xalimjon qizi,
tayanch doktorant
alimbekovamavjuda18@gmail.com
ToshDO‘TAU

Annotatsiya. Ushbu maqolada “Abdurauf Fitrat korpusi” matnlarining leksikasi va ularning qatlamlari statistik tahlili yoritilgan. Korpus matnlari asosida so‘zshakllar korpus yordamida shakllantirilib, ular asosida lemmalar qo‘lda lemmatizatsiya qilingan va qo‘lda leksik qatlamiga ko‘ra teglangan. Teglangan lemmalar bazasi korpusga yuklanib, muallif ijodida qo‘llangan o‘z va o‘zlashgan qatlam so‘zlari aniqlandi. Muallif ijodida qo‘llangan so‘zshakllarning leksik qatlami har bir asarda ko‘rsatilgan. Ular uslub va janrlar kesimida statistik tahlil qilindi.

Abstract. This article covers the lexicon of the texts of the “Abdurauf Fitrat Corpus” and the statistical analysis of their layers. Based on the corpus texts, word forms were formed using the corpus, and lemmas were manually lemmatized and manually tagged according to the lexical layer. The base of tagged lemmas was loaded into the corpus, and the words of the original and borrowed layer used in the author’s work were identified. The lexical layer of word forms used in the author’s work is indicated in each work. They were statistically analyzed in terms of style and genre.

Аннотация. В данной статье рассматривается лексикон текстов “Корпуса Абдурауфа Фитрата” и статистический анализ их слоев. На основе текстов корпуса были сформированы словоформы с использованием корпуса, а леммы были вручную лемматизированы и вручную размечены в соответствии с лексическим слоем. База размеченных лемм была загружена в корпус, и были выявлены слова исходного и заимствованного слоя, используемые в работе автора. В каждой работе указан лексический слой словоформ, используемых в работе автора. Они были статистически проанализированы с точки зрения стиля и жанра.

Kalit so‘zlar: *Abdurauf Fitrat korpusi, lemma, leksik qatlam, o‘z qatlam, o‘zlashgan qatlam.*

Kirish. Fitrat asarlari tilining xususiyatlari, leksikasini tadqiq etish va lug‘atlarini yaratish tilshunoslikning muhim vazifalaridan biridir. Abdurauf Fitrat

milliylikka katta e'tibor bergan va bu asarlarida sezilarli darajada aks etgan. Sof turkiy va shevaga xos so'zlardan foydalanish bilan birga, so'z yasalishiga ham e'tibor berib, tilimizga yot bo'lgan – o'zlashma qo'shimchalardan ham foydalanmaslikka harakat qilgan. Shu sababli Abdurauf Fitrat asarlari tilini tadqiq etish muhim sanaladi. Fitrat asarlari korpusini yaratish o'sha davr tilini tadqiq etish, lug'atlarini yaratish, so'zlarda kuzatiladigan lingvistik hodisalarni qiyosiy-tarixiy usulda o'rganishga yordam beradi. Har bir tilning leksikasi ikkiga: o'z qatlam va o'zlashma qatlamga bo'linadi. O'zbek tili lug'at boyligi turkiy, arabiy, forsiy, yevropaviy qatlamlardan iborat. O'zbek tilining o'z qatlamini o'zbek va boshqa turkiy tillar uchun umumiy bo'lgan, qadimgi davrlardan buyon ishlatilib kelinayotgan so'zlar, o'zbek shevalariga xos so'zlar, o'zbek tilining ichki imkoniyati asosida yasalgan so'zlar tashkil qiladi.

Adabiyotlar sharhi

Dunyo tilshunosligida korpus yordamida muallif asarlari tilining tadqiqiga bag'ishlangan ko'plab ishlarni uchratish mumkin. O.V.Kukushkina, A.A.Polikarpov, E.V.Surovsevalar tomonidan yaratilgan A.P.Chexov badiiy asarlari tilining chastotali grammatik-semantik lug'ati [1] bu sohadagi tadqiqotlarning yorqin namunasidir. Dostoyevskiyning muallif korpusi asosida bir qator lug'atlar, jumladan, muallifning chastotali lug'ati [2] va A.Ya.Shaykevich tomonidan yaratilgan ideomatik so'zlar bazasi [3] yaratilgan.

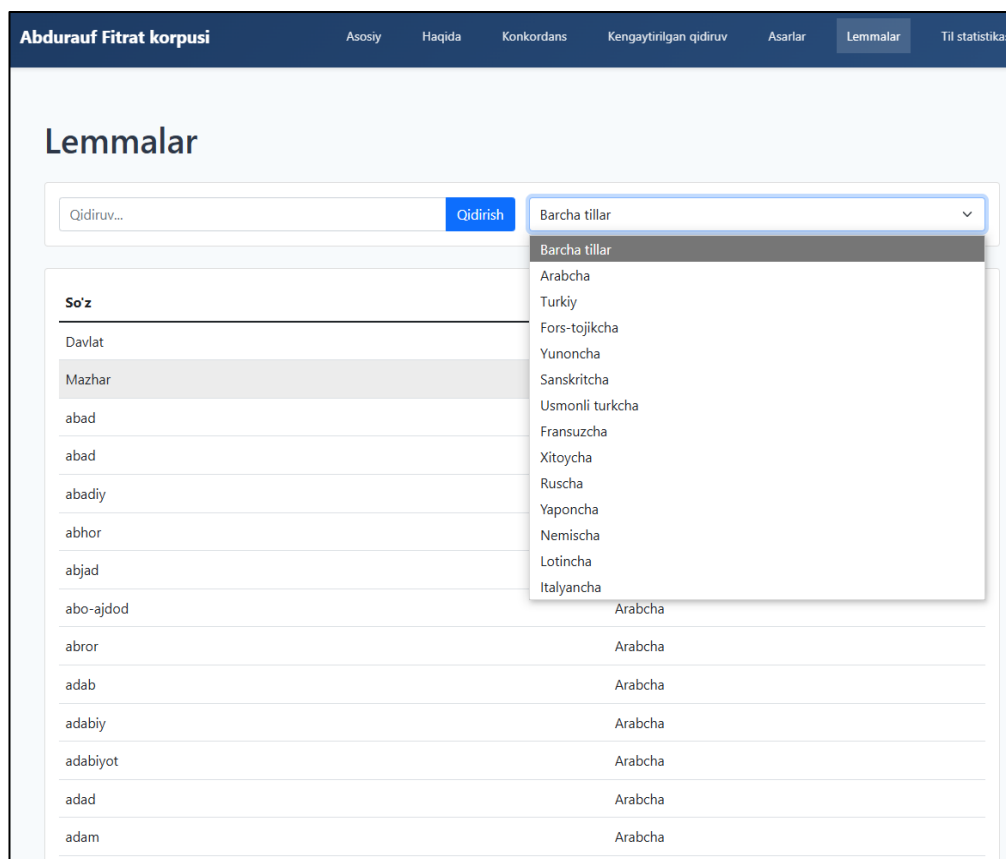
O'zbek tilshunosligida Sh.Hamroyeva o'zbek tili mualliflik korpusining lingvistik asoslarini yaratdi va Abdulla Qahhor mualliflik korpusiga asos soldi [4]. N.G'ulomovanning “Alisher Navoiy muallif korpusi va uning semantik teglar ma'lumotlar bazasi (“Badoye' ul-vasat” devoni asosida)”[5] nomli tadqiqoti hamda M.A.Abjalova hammuallifligida yaratilgan Alisher Navoiy, N.S. G'ulomova, Sh.M. Sa'dullayeva muallif korpusi [6] bu sohada yaratilayotgan ishlar uchun namuna bo'lib xizmat qiladi.

Tadqiqot metodologiyasi

Ushbu tadqiqot ishida statistik tahlilo metodidan foydalanildi. Abdurauf Fitrat asarlarida qo'llangan so'zshakllar korpus yordamida aniqlanib, ular asosida lemmalar qo'lda shakllantirildi. Lemmalarning leksik qatlami o'rganilib, qo'lda teglandi va “Abdurauf Fitrat korpusi”ga yuklandi. Olingan natijalar statistik tahlil qilindi.

Tahlil va natijalar

“Abdurauf Fitrat korpusi” [7] yaratilib, teglangan lemmalar bazasi korpuysga yuklandi. Korpusning “Lemmalar” bo'limida umimiy qidiruv va tillar bo'yicha qidiruv imkoniyatlari yaratildi. (1-rasmga qarang).



1-rasm. “Abdurauf Fitrat korpusi”da lemmalar qidiruvi

Fitrat ijodida jami 9157 ta lemma aniqlandi. Ulardan sof turkiy lemmalar soni 5192 ta ekani aniqlandi. O‘zlashgan qatlamdan eng ko‘pi arabcha so‘z 2505 ta, fors-tojikcha so‘z 1039 tani tashkil etdi. Shuningdek, yunoncha, lotincha, italyancha, ruscha, xitoycha, nemischa, sanskritcha so‘zlar ham mavjud:

1) yunoncha: *akademiya, anatomiya, aristokrat, astronomiya, azot, daftar, difteriya, dio‘lo‘g, do‘qtur, drama, energiya, falsafa, fantaziya, faylasuf, faytun, geometriya, gigiena, gram, hipnoz, injil, jo‘g‘rofiya, kapitalizm, kapitol, koinot, komediya, lirik, lirika, mikroskop, lirizm, mo‘no‘lo‘g, neft, papirus, parazit, piragramma, plato‘nizm, poetik, po‘p, sistema, teatr, ritm, telefon, teleg‘rof, tip, tragediya, angel;*

2) fransuzcha: *adres, bilet, bonka, burjuaziya, dinamit, eksplutasiya, idealizm, jandarm, jurnal, kabinet, kamzul, magazin, marsh, miliard, milio‘n, mo‘da, olmon, opportunist, opportunizm, partiya, qo‘mita, raman, ratsionalizm, reforma, reformistik, ro‘l, ro‘mantizm, ro‘mon, ro‘montizm, ro‘montik, sentimentalizm, shovinizm, simvolist, simvolizm, so‘fa;*

3) nemischa: *german, nemis, kapitalist, lampa, mashruta, soldat, tans, yarmarka;*

4) lotincha: *agent, agrar, detsimetr, do‘kto‘r, element, fabrika, fakt, familiya, federatsa, feo‘dal, feo‘dalizm, forma, general, imperator, individ, instinkt, institut,*

iyul, jinerol, kamandir, kanal, kislorod, klassik, klassistizm, kollektiv, kolxoz, komisiya, komissar, kommunist, ko'nsul, ko'piya, mandat, massa, material, materiyol, materialist, metrika, miqro'sqo'f, mobilizm, mo'dernizm, muzey, naturalizm, nota, no'mir, no'yabr, ognistizm, operasiya, pochta, protestant;

5) italyanacha: *banka, gazet, gazit, opera;*

6) ruscha: *etnograficheskaya, gosudarstvennogo, nauk, bolshevik, bronepoyezd, cholvor, choynak, choynak, elektrik, fabrikant, futurist, galos, grajdan, harakteristika, idealistik, ideolog, imperialist, istikan, kalisa, karbon, kartochka, komik, kopiya, marksizm, mashina, mikrob, miqro'b, mistik, muzika, obraz, ochirat, o'ris, o'rus, patnus, pechatniy, plan, poyezd, proletariat, rabo'chi, rumka, samovar, shapka, shkaf, shlyapa, skoriy, skripka, slovo, spiska, stansiya, sud, syezd, ustul, uyezd, valisapid, velosiped, vintovka, xarakteristika, zakon;*

7) usmonli turkcha: *bebak, eksik, gizlamak, indirmoq, kandi, moso, orqadosh, o'rtu, saqin, sayg'i, sevdik, sodacha, terazi, yumruq.*

8) xitoycha: *choy;*

9) sanskritcha: *baxshi, rupiya, buddizm, dakki, fo'ta, varat.*

10) inglizcha: *boykot, folklor, gurji, ingliz, park.*

11) mo'g'ulcha: *anfar, bo'sag'a, inju, juvanvar, lochin, manglay, mergan, o'lja, qurultoy, suyurg'ol, tengri, ulus, yosog'.*

O'zlashgan so'zlardan eng ko'pi arabcha qatlamga mansub, ular 2506 tani tashkil qiladi. Fors-tojikcha so'zlar miqdori 1042 tadan iborat.

Ushbu teglangan 9157 ta lemma “Abdurauf Fitrat korpusi”ga yuklandi va bu teglangan lemma orqali korpusi matnlaridagi 60185 ta so'zshaklning qatlami aniqlandi, statistikasi chiqarildi (1-jadvalga qarang). 1-jadvalda faqat lemmalarning leksik qatlami bo'yicha statistikasi ko'rsatilgan. So'zshakllar bo'yicha leksik qatlam statistikasi har bir asarda alohida ko'rsatilgan.

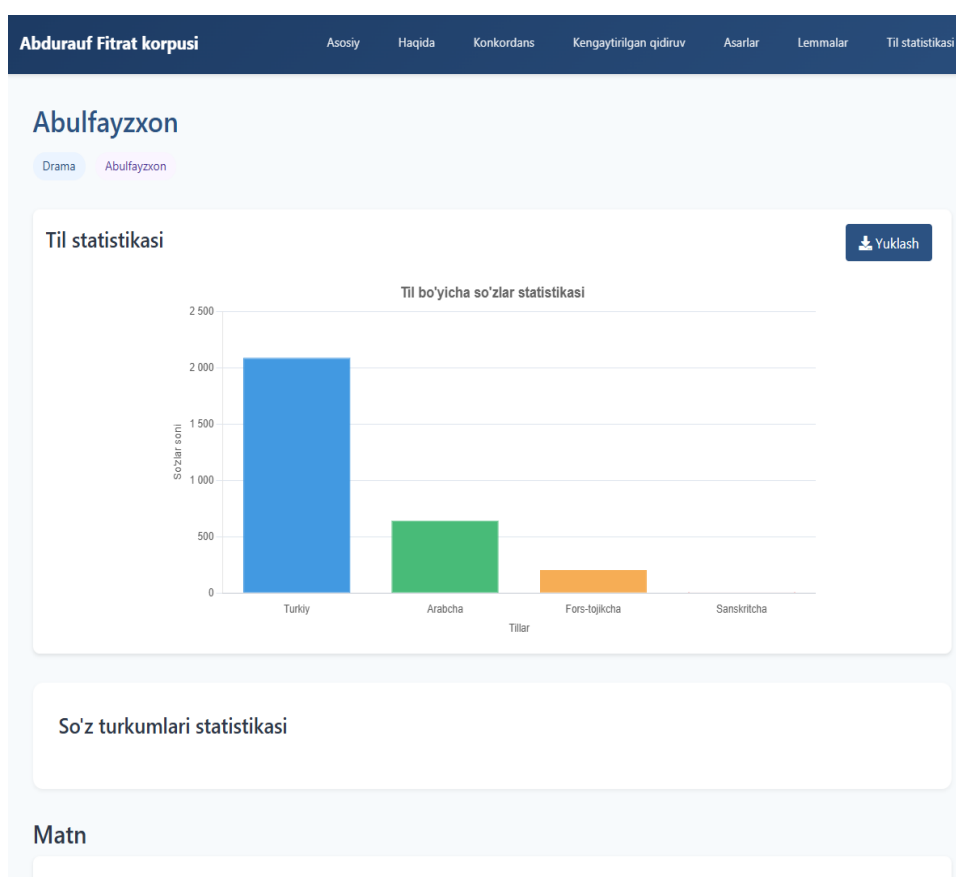
1-jadval. “Abdurauf Fitrat korpusi” lemmalarida leksik qatlam statistikasi

Til	Miqdor	Foiz
Turkiy	5192	56,7%
Arabcha	2506	27,4%
Fors-tojikcha	1042	11,4%
Noma'lum qatlam	118	1,3%
Lotincha	66	0,7%
Ruscha	66	0,6%
Yunoncha	53	0,4%
Fransuzcha	39	0,2%
Mo'g'ulcha	13	0,2%
Usmonli turkcha	14	0,1%
Nemischa	9	0,1%
Sanskritcha	6	0,1%

Inglizcha	5	0%
Italyancha	4	0%
Xitoycha	1	0%
Yaponcha	1	0%

Ushbu jadval ma'lumotlaridan ko'rinib turibdiki, turkiy lemmalarning soni 5192 tani tashkil etgan. O'zlashgan qatlama oid so'zlarning asosiy qismini arabcha va fors-tojikcha so'zlar tashkil qilgan. Ruscha, lotincha, yunoncha, fransuzcha lemmalar terminlar, ilm-fan taraqqiyoti natijasida kirib kelgan so'zlardir. Nemischa, italyancha, sanskritcha, xitoycha, yaponcha so'zlar juda kam miqdorni tashkil etadi. Leksik qatlami aniqlanmagan so'zlar 118 tani tashkil etdi. Ular “noma'lum qatlam” nomi ostida birlashtirildi.

Fitrat asarlarining qaysi janrlarida o'z va o'zlashgan qatlam so'zlarining ulushi qancha ekani korpus yordamida aniqlandi. Har bir janrning so'zshakllari bo'yicha leksik qatlami aniqlanib, bar graph ko'rinishida statistikasi chiqarildi. (2-rasmga qarang).



2-rasm. “Abulfayzxon” asarida qo'llangan so'zshakllar leksik qatlami

Ushbu rasmda ko'rsatilgani kabi muallifning har bir asaridagi so'zshakllar va ularning leksik qatlami diagramma shaklida berilgan. Korpusga yuklangan muallifning barcha asrlaridagi til statistikasi asosida janrlar va uslublar kesimida leksik qatlamning ifodalanishi statistikasi ko'rsatilgan. (2-jadvalga qarang). Bunda o'z qatlamga oid so'zlarni turkiy so'zlar deb nomladik va o'zlashma qatlamga oid so'zlardan eng ko'p ko'rsatkichga ega bo'lgan arabcha, forsha so'zlar olingan. Boshqa kam sonli so'zlardan iborat qatlamlar olinmadi.

2-jadval. “Abdurauf Fitrat korpusi”da so'zshakllarning leksik qatlami janr va uslublar kesimida

Badiiy uslub matnlarida leksik qatlamning aks etishi			
janrlar	Turkiy	Arabcha	Fors-tojikcha
she'r	364	60	29
g'azal	8		1
mansura	206	34	6
qissa	3694	2780	732
hikoya	1728	581	217
drama	12 696	3402	1172
jami	18696	6857	2157
Ilmiy uslub matnlarida leksik qatlamning aks etishi			
ocherk	1278	591	157
ilmiy maqola	7317	4715	1144
darslik	1726	358	66
majmua	296	153	18
risola	5249	3699	780
qo'llanma	6571	2630	669
jami	22437	12146	2834
Publitsistik uslub matnlarida leksik qatlamning aks etishi			
publitsistik maqola	1683	985	126
maktub	564	393	37
risola	14086	10928	2495
munozara	2168	2410	574
jami	18501	14716	3232

Badiiy uslub janrlarida muallif turkiy so'zlardan ko'p foydalangan. Arabcha va fors-tojikcha so'zlar boshqa uslubdagi ushbu qatlam so'zlariga nisbatan ancha kam qo'llanilgan. Badiiy uslub matnlarida turkiy so'zlar 18696 ta, arabcha so'zlar 6857 ta, fors-tojikcha so'zlar 2157 tani tashkil qiladi. Badiiy asar xalq tilining boyligini, go'zalligini ifodalovchi uslubdir. Unda xalq shevalari, xalq tiliga xos bo'lgan ifodalar ko'pligi bilan boshqa uslubdan ajralib turadi. Fitratning badiiy uslubda yozilgan asarlarida turkiy so'zlarning ko'pligidan ham bilish mumkin. Badiiy uslub janrlaridan qissa janrida muallifning “Hind sayyohi bayonoti” qissasi nomli asari mavjud. Ushbu asarda badiiy uslubning boshqa janrlariga nisbatan



arabcha soʻzlar koʻproq qoʻllanilgan. Buning sababi qissaning qahramoni tilidan hadislar, oyatlar keltirilgan oʻrinlar koʻp uchrashidadir.

Ilmiy uslub janrlarida qoʻllanilgan oʻz qatlamga mansub soʻzlarning soni 22437 tani tashkil qiladi. Arabcha soʻzlar 12146 ta, fors-tojikcha soʻzlar esa 2843 ta. Ilmiy uslub matnlarida arabcha, fors-tojikcha oʻzlashmalarning koʻp uchrashining asosiy sababi arabcha terminlarning mavjudligi, Fitrat tomonidan oʻrganilgan shoir-yozuvchilarning ijodida qoʻllanilgan soʻzlarning koʻpligidir.

Publitsistik uslub janrlarida 18501 ta turkiy soʻz, 14716 ta arab tilidan oʻzlashgan soʻz, 3232 ta fors-tojikcha soʻzlar qoʻllanilgan. Ushbu uslub matnlarida arabcha soʻzlar boshqa uslub matnlaridagiga nisbatan koʻproq qoʻllanilgan. Publitsistik uslubda yozilgan “Oila”, “Rahbari najot” nomli asarlarida arabcha kalimalar, oyatlar, hadislar koʻp qoʻllanilgan. Millat ravnaqi uchun, oila mustahkamligi uchun yagona najot yoʻli islom dini ekani yoritilgan ushbu asarlarda arabcha soʻzlarning ulushu katta boʻlishi tabiiy. Munozara janrida yozilgan “Hindistonlik bir farangi ila buxorolik mudarrisning jadid maktablari xususida qilgʻan munozarasi” nomli asarida boshqa janrlardan farqli tarzda arabcha soʻzlar turkiy soʻzlardan koʻp qoʻllanilgan. Turkiy soʻzlar 2168 ta, arabcha soʻzlar 2410 ta, fors-tojikcha soʻzlar 574 tani tashkil qilgan. Ushbu asarda hindistonlik sayyoh va buxorolik mudarris suhbatida Qurʼon oyatlari, hadislar, arabcha jumlar koʻp qoʻllanilgan. Bu arabcha soʻzlarning ulushi turkiy soʻzlardan koʻp boʻlishiga sabab boʻlgan.

Xulosa. Abdurauf Fitrat ijodida jami 9157 ta leksik birlikdan foydalangani maʼlum boʻldi. Ulardan oʻz qatlamga mansub lemmalarning soni 5192 tani tashkil etadi. Oʻzlashgan qatlamga oid soʻzlarning asosiy qismini arabcha va fors-tojikcha soʻzlar tashkil qiladi. Arabcha soʻzlar 27.4 ta, fors-tojikcha soʻzlar 11.4 foiz, boshqa talam soʻzlari 5 foizni oʻz ichiga oladi. Ruscha, lotincha, yunoncha, fransuzcha lemmalar terminlar, ilm-fan taraqqiyoti natijasida kirib kelgan soʻzlardir. Nemischa, italyanacha, sanskritcha, xitoycha, yaponcha soʻzlar juda kam miqdorni tashkil etadi.

Badiiy uslub janrlarida muallif turkiy soʻzlardan koʻp foydalangan. Arabcha va fors-tojikcha soʻzlar boshqa uslubdagi ushbu qatlam soʻzlariga nisbatan ancha kam qoʻllanilgan. Badiiy uslub matnlarida turkiy soʻzlar 18696 ta, arabcha soʻzlar 6857 ta, fors-tojikcha soʻzlar 2157 tani tashkil qiladi. Ilmiy uslub janrlarida qoʻllanilgan oʻz qatlamga mansub soʻzlarning soni 22437 tani tashkil qiladi. Arabcha soʻzlar 12146 ta, fors-tojikcha soʻzlar esa 2843 ta. Ilmiy uslub matnlarida arabcha, fors-tojikcha oʻzlashmalarning koʻp uchrashining asosiy sababi arabcha terminlarning mavjudligi, Fitrat tomonidan oʻrganilgan shoir-yozuvchilarning ijodida qoʻllanilgan soʻzlarning koʻpligidir. Publitsistik uslub janrlarida 18501 ta turkiy soʻz, 14716 ta arab tilidan oʻzlashgan soʻz, 3232 ta fors-tojikcha soʻzlar qoʻllanilgan.

Foydalanilgan adabiyotlar:

1. Кукушкина О.В., Суровцева Е.В., Лапоница Л.В. Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова с электронным приложением. – М.: МАКС Пресс, 2012. – 571 с.
2. Заботкиной В.И. “Методы когнитивного анализа семантика слова компьютерно-корпусной подход”, Москва: Языка словянской культуры, п. 348, 2015.
3. Словарь языка Достоевского. Идеоглоссарий.// Российская академия наук институте русского языка им В.В.Виноградова, 2008.
4. Hamroyeva Sh.M. “O‘zbek tili mualliflik korpusini tuzishning lingvistik asoslari”, Buxoro, – 259 bet, 2018.
5. G‘ulomova N.S. “Alisher Navoiy mualliflik korpusi va uning semantik teglari bazasini yaratish (“Badoye’ ul-vasat” devoni asosida)”, Toshkent, – 190 bet, 2022.
6. <https://navoiykorpusi.uz/>
7. <https://fitratkorpusi.uz/>