

ЎЗБЕК ТИЛИ МИЛЛИЙ КОРПУСИНИ ЯРАТИШ МЕТОДОЛОГИЯСИ ВА УНИНГ МУАММОЛАРИ

*Тоирова Гули Ибрагимовна,
ўзбек тилшунослиги ва
журналистика кафедраси
профессори, филология фанлари
доктори*

Аннотация. Мақолада ўзбек тили миллий корпусини яратишда интерфейс турлари, корпуснинг қидирув ойнасининг аҳамияти муҳокама қилинади. Миллий корпусининг интерфейси турли дизайн ва тузилмалардан иборат бўлиб, унинг тўлиқлиги учун муаллиф жавобгар бўлади, интерфейс жозибали ва таъсирли бўлиши керак. Интерфейсни яратиш миллий ёки замонавий хусусиятларга асосланган бўлиб, интерфейсда миллий коллоритга эътиборни қаратиш керак. Тил корпуси - бу жуда катта ютуқларга эришган, ҳисоблаш лингвистикаси дунёсининг жуда тез ривожланаётган тармоғи. Интерфейс технология ва фойдаланувчи ўртасидаги алоқа тизими ҳисобланади. Интерфейс визуал, имо-ишоравий, овозли каби турлари таҳлил қилинган.

Калит сўзлар: интерфейс, муаллиф корпуси, математик моделлаштириш, морфологик ва семантик изоҳ, маълумот, лингвистик база, сунъий интеллект, компьютер лингвистикаси, корпус тилшунослиги, тил корпуси, махсус дастурий таъминот, электрон кутубхона, лексик, морфологик, грамматик, семантик белгилар, лингвистик белгилар билан боғлиқ муаммолар.


Корпус лингвистикаси – дунё фанининг, жумладан, тилшуносликнинг энг ривожланган соҳаси, корпус эса тилшуносларнинг зарурий иш қуроли; оғзаки, ёзма ёдгорликлар, миллий-маданий меросни акс эттирувчи ахборот

манбаи. У қидирув дастурига бўйсундирилган матнлар йиғиндиси, мукаммал разметкага эга корпус лингвистик тадқиқотлар самарадорлигини таъминлашда барқарор лингвистик база вазифасини бажаради[1,2,4,5]. Сунъий интеллект маҳсули сифатида: электрон луғатлар, таржима порталлари, терминологик маълумотлар банки, виртуал (электрон) кутубхона, электрон матнли корпуслар, электрон ҳукумат, электрон нашр, электрон дарслик ва қўлланмаларнинг барчасини санаш мумкин. Сунъий интеллект маҳсули бўлган лингвистик электрон манбалар муайян тил корпусини яратиш учун хом ашё саналади.

Ўзбек тили миллий корпусини яратишда визуал интерфейс турини таклиф қиламиз. Интерфейснинг бу тури илк корпус яратиш жараёнида қийинчилик туғдирмайди[20]. Чунки, мониторда намоиш этиладиган визуал тасвирлар ёрдамида маълумотларни узатадиган стандарт компьютер интерфейси ҳисобланади. Келажакда кўзи ожиз инсонларни ҳисобга олган ҳолда овозли турини таклиф этиш мумкин.

Ушбу соҳадаги энг янги йўналишлардан бири бу сенсорли интерфейс саналади. Унинг ишлаш принципи маълум объектлар орқали амалга ошириладиган фойдаланувчи ва машинанинг жисмоний ўзаро таъсирига асосланади[34]. Айтишимиз мумкинки, бу монитрдан фойдаланган ҳолда фойдаланувчига илгариги маълумотни графикавий равишда олган маълумотларни тақдим этишга уринишдир.

O'ZBEK TILI MILLIY KORPUSI



Alisher Navoiy hayoti va ijodi
• O'zbek mumtoz adabiyoti

FOTO ALBOM
+ Buyuk siymolar

TAQDIMOT
+ Taqdimot

ROMANLARI
HIKOYALARI
PYESA
PORTRET
FELVETON

Korpus nima...

Korpus — til birliklarining xususiyatlarini aniqlash maqsadida qidiruv dasturiga bo'ysundirilgan matnlar majmui, 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[1]. U elektron shaklda mavjud bo'lgan matnlarni, hujjatlarni ma'lumotlarni qayta ishlash, ularni avtomatik tahlil, ya'ni morfologik sintaktik va semantik tahlil qilish, morfologik analiz va sintez qilish bilan birga, yarim avtomatik rejimda ma'noni buzmasdan ishonchli nutq materialini moslashtirish darajasini tekshirish ushbu sohadaagi til bilan qiziqishni oshirishga xizmat qiladi. Korpus — tilni tadqiq etish, til o'rganish, lug'at tuzishda eng zamonaviy, keng imkoniyatli dasturlashtirilgan tizimdir. Korpus — millionlab so'z kontekstlariga havola qiluvchi, maxsus qidiruv tizimi asosida ishlaydigan elektron matnlar yig'indisi. U internet tizimidagi elektron kutubxona, lug'at va lingvistik grammatikadir. Korpus - tabiiy (real) tilning elektron shaklda bo'lib, uni to'ldirish, tuzatish, tahrir qilish imkoniyatiga ega[2]. Korpus bu so'zlar, so'z birkimlari, grammatik shakllari ma'nosini ma'lum bir izlash tizimi orqali topishning elektron shakldagi matnlar to'plamidir. Korpuslarning har xil turlari mavjud. Masalan, bir muallif korpusi[3], bir kitob korpusi (jumladan dastlabki korpuslar "Bibliya" uchun qilingandir). Ma'lum bir til Milliy korpusi shu tilning xavotining barcha qirralarini, janrlarini, usullarini, xududiy va ijtimoiy variantlarini o'zida ifodalaydi. Masalan, o'zbek tili Milliy korpusini Internetda amalga qo'yilsa, 140 mln so'z ishlatmalaridan iborat shu tildagi barcha turdagi matnlarni o'zida mujassamlashtira oladi. Kelajakda o'zbek tili Milliy korpusi o'zida 300 mln so'z ishlatmalarini jamlashi lozim bo'ladi. Boshqa milliy tillar korpuslari kabi o'zbek tili Milliy korpusi xam ikkita muxim xususiyatga egadir. Birinchidan u ancha salobati va barcha yo'nalishlar uchun (turli badiiy janrlar: publitsistik, o'quv, ilmiy, ish yuritish, nutk - so'zlashuv, shevalar va boshqalar uchun) bir xilda muvofiqlashandir. Bu matnlar o'z davrida tildagi ta'sir kuchiga qarab proporsional ravishda korpusda jamlangan bo'ladi. Ikkinchidan korpus matnlariga xos bulgan alohida jihatlarni qo'shimcha ma'lumot sifatida o'zida aks ettiradi (masalan ma'lum bir jihatdan belgilanishlar va annotatsiyalar). So'zlar va boshka birkimlar belgilanishi - korpusning bosh xarakteristikasidir (tasnifidiri); U korpusni hozirda Internetni to'ldirib yuborgan boshqa matnlar kolleksiyalari va kutubxona matnlaridan farqlaydi. Matnlar belgilanishi qanchali boy va turli tuman bo'lsa korpusning ilmiy ahamiyati shuncha yuqori bo'ladi. O'zbek tili Milliy korpusini belgilanishlarning uch turi e'tiborga olinishi kerak bo'ladi: metamatni (u matnning muallif qarashi, janri va boshka xususiyatlariga bo'yicha to'liq ifodalaydi), morfologik va semantik belgilanishlar (morfologik va semantik belgilanishlar matnning emas balki alohida bir so'zni tasniflab beradi). ADABIYOTLAR: 1. Grudeva V.A. Korpusnaya lingvistika. Uchebnoe posobie, 2-e izd., stereotip. - M.: Filinta, 2012. - 165 s. 2. Zaxarov V.P. Korpusnaya lingvistika: Uchebno-metod. posobie. - SPb., 2005. - 48 s. 3. Zaxarov V.P., Bogdanova S.Yu. Korpusnaya lingvistika. - Irkutsk: IGLU, 2011. - 161 c. 4. Toirova G.I. Importance of Interface in Creating Corpus. International Journal of Recent

SO'Z QIDIRISH

Hikoya nomi:
Kartina
Qidiruv

LEKSIK-GRAMMATIK QIDIRUV

So'z
 Grammatik holati
 Omonim
 Qo'llanilishi
 Paronim
 Varianti
 Sinonimi
 Antonimi
 Uslubi
 Qo'llanilish davri
Qidiruv

SINTAKTIK QIDIRUV

Gapning tuzilishiga ko'ra turlari
 Gapning maqsadiga ko'ra turlari
Qidiruv

MUALLIFLAR

Mualliflar

DASTUR HAQIDA

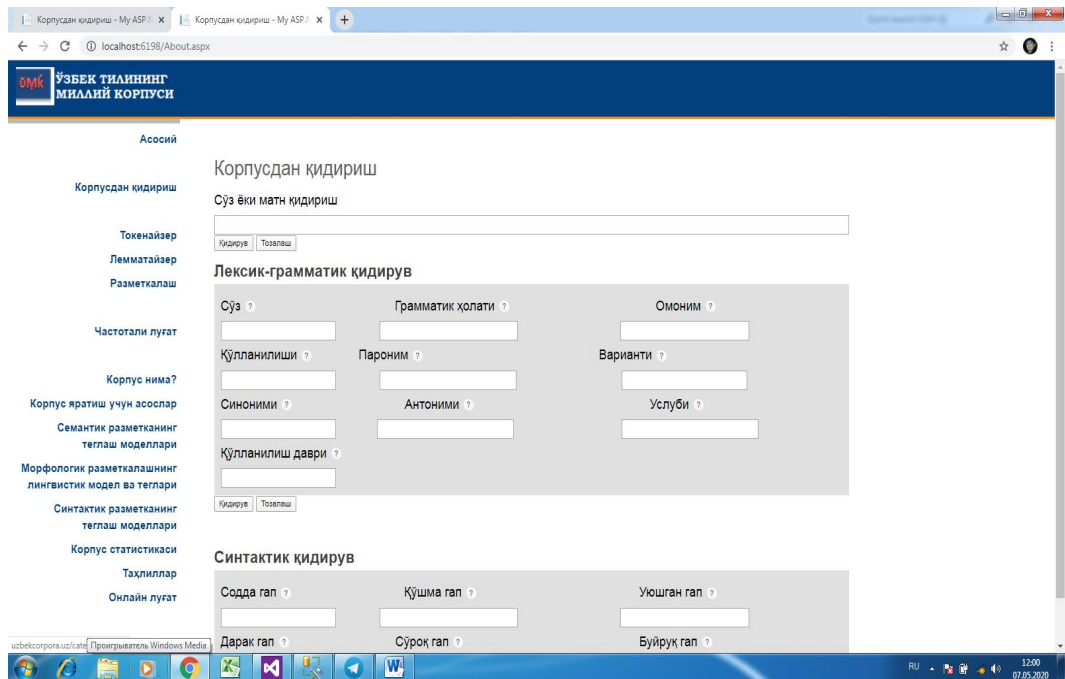
Ushbu korpus qidiruv natijasi shu haqiqiy matni asosida amalga oshiriladi. Korpus



1-расм.

Korpuslarining interfeysi turli dizayn, tuzilişga ega бўлиб, унинг мукамаллиги корпус яратувчи муаллифнинг зиммасига юкланади. Чунки интерфейс корпус ҳақида илк таассурот қолдирувчи, ўзига жалб этувчи умумий кўринишдир. Интерфейс яратишда миллий коллоритни акс эттирувчи безаклар ҳамда мумтоз ёки замонавийликни акс эттирувчи белгилар эътиборга олинishi лозим.

Бу корпуснинг қидирув ойнаси, яъни интерфеysi саналади. “Корпусдан қидириш” – бу жойда сўз ёки гапни қидириш имкони мавжуд. Бунда натижа қуйидаги кўринишга эга бўлади. Корпуснинг “Лексик-грамматик қидирув” тугмасидан сўзнинг қуйидаги хусусиятларини билиб олиш мумкин: Сўзнинг маъноси, грамматик ҳолати, синонимлиги, омонимлиги, паронимлиги, антонимлиги, сўзнинг варианты, сўзнинг қўлланилиш даври, сўзнинг қўлланилиш услуби кабилар[25, 28].



2-расм

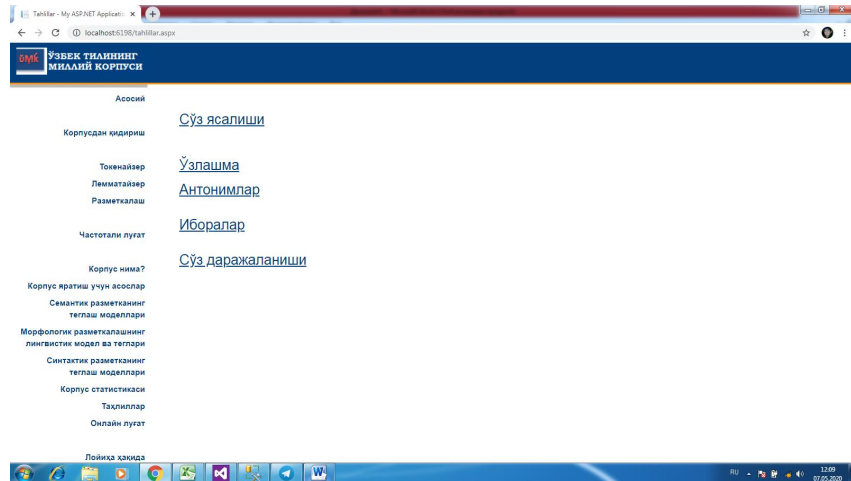
Жумладан, “ингичка” сўзи ёзилганда, қуйидаги маълумотларга эга бўламиз:

- **сўзнинг маъноси:** 1.Кўндаланг кесими меъёрдан кичик; 2 Чийилдоқ, ўткир (товуш ҳақида);
- **грамматик ҳолати:** [сифат];
- **синоними:** майин;
- **омоними:** йўқ;
- **пароними:** йўқ;
- **антоними:** йўғон;
- **варианти:**йўқ;
- **қўлланилиш даври:** замондош;
- **қўлланилиш услуби:** бетараф;

“Синтактик қидирув” тугмасидан биз гапнинг мақсадига кўра: дарак, сўроқ, буйруқ турини ва гапнинг тузилишига кўра: содда ва қўшма, уюшган гапларни ажратиб оламиз.



3-расм



4-расм

“Сўз ясалиши” тугмасида 200га яқин сўзнинг ясалиши таҳлил этилган.

илож	но	лан	лик					
илтифот	сиз	лик	бе	лик				
инсоф	ли	сиз	лик	бе	лик	но	лик	она
инсон	лик	парвар	лик	сифат				
интизом	ли	лик	сиз	лик	бе	лик		
ипак	ли	лик	фуруш	чи	лик			
иқлим	лаш	шунос	лик					
иқтидор	ли	сиз						
иқтисод	ий	от	чи					
исбот	ла	сиз	бе-					
ил	гак	гич						
илгари	ги	ча	ла					
жабр	ла	ли						
жадал	ла	лик						
жадид	изм	чилик						
жафо	каш	лик	кор	ли	сиз			
жаҳолат	ли	параст						
жаҳон	бахш	гир	лик	намо	шумул			
жилов	бардор	дор	ла	хона				
.....	---	---	---				

5-расм

Маълумотлар базасини яртиш жараёнида “Еxсel жадвалларидан фойдаланилади, бунда биринчи устунда файл номи (аниқ кўрсатилган йўл), бошқа устунларда эса метаматнли атрибутлар ва технологик маълумот келтирилади. Мазкур амал Еxсel дастурининг ўрнатилган воситаларидан самарали фойдаланиш имконини беради ва қидирув тизимида анча қулайликлар туғдиради. Масалан, қидирув, фильтрация, таҳлил ва

маълумотларни қайта ишлаш (амаллар рўйхати, автотўлдириш, статистика). Бунда жадваллар матн форматида сақланиши лозим ва бу форматни Excel тушуниши шарт. Мазкур амал орқали жадвал кўринишда сақланган файл, нафақат Excel, балки бошқа жадвалли дастурлар қабул қилиши ва ишлаш вазиятининг самарадорлигини ошириш имкониятини беради.

Хулоса ўрнида шуни айтиш лозимки, ўзбек тили миллий корпусининг онлайн варианты тузилишда “Ўзбек тили изоҳли луғат”да мавжуд сўзлар таҳлилга тортилади. Интерфейс - замонавий дастурий дизайн талабларига жавоб берадиган, фойдаланувчига тушунарли, ишлаш учун қулай шаклда бўлиши керак. Интерфейс технология ва фойдаланувчи ўртасидаги алоқа тизими ҳисобланади. Интерфейс визуал, имо-ишоравий, овозли каби турларда бўлади. Миллий корпус тизим ва амалий дастурлаш интерфейсидан ташкил топиши керак бўлади.

Адабиётлар рўйхати

1. Charlez, Meyer, (2004). English corpus linguistics: An introduction. *Cambridge University Press*, UK, 168 p.
2. Eshmuminov, A., (2019). Synonymous database of the Uzbek language national corpus. *Dissertation of PhD in Philology*, Tashkent.
3. Fries, Ch.C., (1969). The structure of English. *An introduction to the construction of English sentences*, London.
4. Hamroeva, Sh., (2018). Linguistic bases of creation of the author's corpus of the Uzbek language: *Author's Abstract of the Dissertation of PhD in Philology*, Tashkent.
5. Zakharov, V.P., (2011). Corpus linguistics: a textbook for students of humanitarian universities, Irkutsk, 161 p.
6. Karimov, R., (2021). Linguistics and programming issues of creating a parallel corpus of Uzbek and English, *Author's Abstract of dissertation of PhD*, Bukhoro, 151 p.
7. Leech, G., (1991). The State of Art in Corpus Linguistics, *English Corpus Linguistics*, London.

8. Melchuk, I.A., (1985). Word order in the automatic synthesis of the Russian word (preliminary messages), *Scientific and technical information*, 12:12-36.
9. Mengliev, B., (2018). Is the Uzbek language corpus being created? *Ma'rifat newspaper*. April 3, 2018, retrieved from: http://marifat.uz/marifat/v_pomosh_uchitelu-marifat/savol/1142.htm.
10. Mengliev, B., Bobojonov, S., Hamroeva Sh., (2018). Uzbek National Corpus. April 26, 2018, retrieved from: <http://marifat.uz/marifat/ruknlar/fan/1241.htm>.
11. Pulatov, A. Q., (2011). Computer Linguistics. Tashkent, Akademnashr, 520 p.
12. Toirova, G., (2019). The Role of Setting in Linguistic Modeling. *International Multilingual Journal of Science and Technology*, 4(9):722-723, available at: <http://imjst.org/index.php/vol-4-issue-9-september-2019/>.
13. Toirova, G., (2020). About the technological process of creating a national corpus. *Foreign languages in Uzbekistan*, 2(31):57-64, available at: <https://journal.fledu.uz/uz/2-31-2020>.
14. Toirova, G., (2020). The importance of the interface in the creation of the corpus. *Internauka*, 7, DOI: <https://doi.org/10.25313/2520-2057-2020-7-5944>.
15. Тоирова Г. Корпус лингвистикасининг атамалар луғати. Изоҳли луғат. Германия: «GlebeEdit» номли халқаро нашриёт. 2020, –68 б.
16. Тоирова Г. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари. Халқаро монография. Германия: «GlebeEdit» номли халқаро нашриёт, 2020. – 169 б.
17. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // *International Journal of Recent Technology and Engineering (IJRTE)* ISSN: 2277-3878, Volume-8 Issue-2S10, September 2019. –P.352-355. (scopus)
18. Toirova G., Jurayeva O., Abulova Z., Norova M., Norova F. Application of Innovative Technologies in Teaching Process. // *International Journal of*

Psychosocial Rehabilitation, Vol. 24, Special Issue 1, 2020. ISSN: 1475-7192.– P.386-390. (scopus)

19. Toirova G., Hamroyeva N. The importance of linguistic models in the development of language bases // Sciences of Europe. vol 2, No 59 (2020) ISSN 3162-2364. –P. 57-64

20. Тоирова Г. Важность интерфейса в создание корпуса. International Scientific Journal «Internauka», // *Международный научный журнал «Интернаука»*. – 2020. – №7. Онлайн журнал. <https://doi.org/10.25313/2520-2057-2020-7-5944>(Impact Factor –8,758)

21. Менглиев Б., Тоирова Г., Хамроева Ш. Ўзбек тилининг миллий ва муаллифлик корпуси. Guvohnoma № DGU 05735 Ўзбекистон Республикасининг интеллектуал мулк агентлиги. – Тошкент, 2018

22. Toirova G. Importance of interface in creating corpus. // Хоразм Маъмур академияси ахборотномаси. –Урганч, 2020, – №2/2. –Б. 49-52.

23. Тоирова Г. Миллий корпусни яратишда интерфейснинг аҳамияти. //Қарақалпақ давлат университети хабаршысы. – Нукус, 2019, ISSN 2010-9075–№ 4(45). –С.195-198.

24. Тоирова Г. Миллий корпус яратишнинг технологик жараёни хусусида //Ўзбекистонда хорижий тиллар. Электрон илмий-методик журнал. – Тошкент. 2020, –№ 2 (31), –Б.57– 64.

25. Тоирова Г. Ўзбек тили миллий корпусни яратишда интерфейснинг аҳамияти. // Сўз санъати халқаро журнали, – Тошкент, 2020, № 3, – Б.100-105.

26. Toirova G. The importance of linguistic module forms in the national corpus// Zamonaviy fan, ta’lim va tarbiyaning dolzarb muammolari (Mintaqada zamonaviy fan, ta’lim va tarbiyaning dolzarb muammolari) (Elektron ilmiy jurnal), – Урганч. 2020, –№ 5 , –Б.155– 166.

27. Toirova G. The importance of linguistic models in the development of language bases. // Бухоро Давлат университети илмий ахбороти. – Бухоро, 2020. –№6. – Б.98-106.

28. Тоирова Г. Лингвистик базани тузишда модуллаштиришнинг аҳамияти // Наманган давлат университети илмий ахборотномаси. –Наманган, 2021. – №3. -Б.377-386.

