

## ЎЗБЕК ТИЛИ МИЛЛИЙ КОРПУСИНИ ЯРАТИШ – ЗАМОН ТАЛАБИ

*Г.В.Тоирова*

*Бухора давлат университети*

Сунъий интеллект орқали тилнинг имкониятларидан фойдаланиш борасида замонавий ахборот технологиялари бениҳоя кенг қулайликлар эшигини очди. У инсон онги бажариши мумкин бўлган бир қанча вазифаларнинг бажара олмакда. Сунъий интеллект маҳсули бўлмиш электрон манбалар инсонга зарар етказмаслик ва инсонларнинг оғирини енгиллаштириш мақсадида яратилмокда. Чунончи, ўзбек тилининг интернет ва электрон тилга айланиши, миллий тилнинг электрон ресурсларини (ўзбек тили корпуси, электрон луғатлари, интернет саҳифаларидаги матнлар) ва технологияларини такомиллаштириш долзарб вазифалардандир.

Корпус – бу корпус лингвистикасининг предмети. Ушбу атама илмий адабиётларда турлича таърифланади. Чунончи, инглиз тилида *linguistic corpus* ёки *text corpus* каби атамалар билан қўлланилади[8]. Муайян тилнинг миллий корпусни яратиш, унинг аналитик технологиясини, корпус лингвистикаси соҳасини ривожланиш каби муаммоларни ҳал қилишда А.Н.Хомский, Г.Н.Луч, Ч.Ф.Меер, Ж.Синклер, М.З.Курди кабиларнинг илмий изланишларини эътироф этиш лозим [17,18,21,22].

Рус олими В.П. Захаров «корпус» атамасини қуйидагича изоҳлайди: «корпус –оғзаки ҳамда ёзма матнлар асосида жамланган тил бирликларининг лингвистик маълумотлар мажмуи»[4]. Рус корпус лингвистикасида катта массивли матнлар мажмуаси, корпус тузиш принциплари, лингвистик маълумотлар базаси В.Г.Бритвин, В.П.Захаров, И.А.Мелчук, А.Б.Кутузов, Р.Г.Котов, Л.И.Беляева, Е.В.Недошивина, В.В.Рыков, В.Плунгянларнинг мақсадли тадқиқотларида ўз аксини топган [1,2,3,4,5,6,8,9].

Ўзбек тилшунослигида матнни лингвостатистик таҳлил этиш, унга лексикографик ишлов бериш, автоматик таҳрир қилувчи дастурнинг лингвистик таъминотини, таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари, миллий корпусининг синоним сўзлар базаси, муаллифлик корпусининг лингвистик асослари борасида Х.Исхакова, С.Муҳаммедов, С.Ризаев, С.Муҳаммедова, Б.Менглиев, Д.Ўринбоева,

А.Пўлатов, У.Дысимова, Г.Валиева, Г.Жуманазарова, Н.Абдурахмонова, Ш.Ҳамроева, М.Абжалова, А.Эшмўминовларнинг ишлари диққатга сазовор.

Ўзбек тилшунос олимлари эса «корпус» атамасига куйидагича изоҳ беради: «корпус – маълум мақсадда йиғилган матнлар мажмуини ташкил этувчи тил бирликларининг йиғиндиси» [13], «корпус – исталган табиий (реал) тилдаги электрон шаклда сақланадиган ёзма ёки оғзаки, компьютерлаштирилган кидирув тизимига жойлаштирилган матнлар йиғиндиси»[16] каби таърифларда ўз ифодасини топган. Ўзбек тилшунослигидаги тадқиқотларда корпуснинг моҳияти куйидагича ёритилди: «Корпус – мавжуд маълумотларни матн ҳолида тақдим эта олиш; корпуснинг ҳажмига кўра имкон даражада кўпроқ ахборот беришнинг имконияти; бир марта яратилган корпус маълумотларидан такроран, турли масалаларни ҳал қилишда фойдаланиш имкониятидир» [14].

«Корпус – тил бирликларининг хусусиятларини аниқлаш мақсадида кидирув дастурига бўйсундирилган матнлар мажмуи, табиий тилдаги электрон шаклда сақланадиган ёзма ёки оғзаки, компьютерлаштирилган кидирув тизимига дастурий таъминот асосида жойлаштирилган on-line ёки off-line тизимда ишлайдиган матнлар жамланмиси» [11] – деб изоҳланади бошқа бир манбада.

Шунингдек, «таълимий корпус» ҳақида илмий тадқиқот олиб борган Ў.Холиёров ўз ишида куйидаги фикрларни айтиб ўтади: «Таълимий корпус – материаллари тилни ўргатишга йўналтирилган, лингводидактик хусусиятга эга тил корпуси. Ўзбек тилининг таълимий корпуси ўзбек тили имкониятларини ўргатишга мўлжалланган, лингводидактик хусусият касб этган, электрон матнларни қамраб олган, махсус сайт кўринишида амал қиладиган корпусдир»[15].

Юқорида келтирилган таъриф ва изоҳларга қўшилган ҳолда, уни куйидагича таърифлаймиз: Корпус – электрон шаклда мавжуд бўлган матнларни, ҳужжатларни маълумотларни қайта ишлаш, уларни автоматик таҳлил, яъни морфологик синтактик ва семантик таҳлил қилиш, морфологик анализ ва синтез қилиш билан бирга, ярим автоматик режимда маънони бузмасдан ишончли нутқ материаллини мослаштириш даражасини текшириш ушбу соҳадаги тил билан кизиқишни оширишга хизмат қилади. Бугун ўзбек тилидаги матнларни таҳлил қилиш инсон қўл меҳнати билан амалга оширилмоқда, бу эса ўз навбатида матнларнинг таҳлилига кетадиган вақтни кўп сарф қилинишига, иктисодий

харажатларнинг ошишига, инсон томонидан йўл қўйиладиган хатоликларнинг юзага чиқишига олиб келмоқда. Шу жиҳатдан корпусларини тилшунослик бўйича асосий тадқиқот ва амалий топшириқлар ечими учун зарурий материал деб атасак бўлади.

Статистик тадқиқ методи компьютер таржимаси, нутқни синтезлаш ва уни таниш, орфографик текширув каби лингвистик амалларни бажариши корпус лингвистикасининг кейинги тараққиёт босқичида амалга оширилади.

Ҳозирда интернет тизимида фаолиятда бўлган 70 якин тил корпуслари, яъни инглиз, испан, хитой, араб, француз, рус, немис, поляк, поляк-украин, чех, словак, серб, хорват, босния, болгар, болгар-рус, македон, шотланд, нидерланд, нидерланд-француз, швед, дат, норвег, исланд, фарер, ўрта аср француз тили, испан, итальян, португал, румин, литва, латиш, грек, шарқий арман, осетин, албан, ҳинд, хетт, фин, урал тиллари, эстон, вепс, венгер, удмурт, грузин, инглиз-грузин, лазгин, турк, татар, тожик, бошқирд, қрим-татар, қалмиқ, бурят, монгол, араб, иврит, амхар, япон, қадимий япон, баман, эсперанто тилларининг корпуслари яратилган[11]. Юқорида номлари санаб ўтилган корпусларнинг ҳар бири ўзгача устунлик ва камчиликга эга. Масалан, Тожик тили корпуси фақат электрон кутубхона вазифасини бажариб, фото, видео галерея вазифасини бажара олади. У умуман разметкаланмаган, қидирув тизими номукамал ва ноқулайдир. Фақат яхлит асаргагина ишора қила олади, ҳолос.

А.Ғуломов, М.Асқарова, А.Ҳожиев, Ҳ.Неъматов, Р.Сайфуллаева, Б.Менглиев, М.Қурбонова, М.Абузалова, С.Назарова каби ўзбек тилшунослигининг назарий ва амалий жиҳатдан тамал тошини қўйган устозларнинг тадқиқотларни компьютер тизимига мослаштириш ҳамда збек тили миллий корпусини яратишга асос бўла олади.

«Ўзбек компьютер лингвистикаси ўзбек тилининг инглиз тилидан тамомила фарқ қиладиган хусусиятлари асосида шакллантирилади. Бу эса ўзбек компьютер лингвистикасини яратишдан олдин ўзбек тилини мукамал даражада системалаштириш, формаллаштириш вазифаларини амалга ошириш зарурияти мавжудлигини кўрсатади. Ўзбек тили каби бой, кенг ва чуқур ривожланган тил масалаларини компьютерда ечиш даражасига олиб чиқиш инглиз тилига қараганда катта ҳажмда иш бажаришни талаб қилади», - дея таъкидлайди Абдумажид Пўлатов. Олимнинг фикрига қўшилган ҳолда, ўзбек компьютер лингвистикасини

яратишда инглиз компьютер лингвистикасидан тўғридан-тўғри фойдаланиб бўлмаса-да, унинг асосий ғояларига таяниш мумкин. Ўзбек тилининг тил корпусини тузишга мўлжалланган лингвистик база ва миллий матнлар банкини тайёрланишда Рус тили миллий корпуси бўйича олиб борилган тадқиқот ишларига мурожаат қилинади.

Бугунги кунда у «виртуал луғатлар, улар билан ишлаш ҳамда яратиш технологиялари» мазмунида оммалашмоқда. Бунда соҳани Интернет тизими доирасида идрок этиш лозим бўлади. Киберлексикография атамаси шу маънода интернет электрон луғатларни – умумий ва махсус турдаги академик, энциклопедик ва лингвистик луғатларни яратишнинг назарий асосларини ўзида ифодалайди. Маълумки, корпус ёзма матнларнинг электрон шаклдаги тўплами бўлиб, махсус ишлаб чиқилган компьютер дастурлари орқали тўпланган луғатлар жамланмаси киритилиши асосида яратилади. Корпус компьютер дастури шаклида тилнинг барча жиҳатларини ёритиб бера олади. Шунинг учун ҳам, корпус лингвистикаси тилнинг хусусиятларини эътиборга олган ҳолда уни қайта баҳолашга олиб келади. Ўзбек тили миллий корпуси ўзбек тилида мавжуд бўлган лексик бирликларнинг, чунончи, ўзлашма сўзлар, синоним, антоним, омоним, сўзларнинг даражаланиши; сўзнинг морфологик таркиби, сўзнинг ясалиши, сўзларнинг маъноси, унинг морфологик хусусиятларини автоматик таҳлил қилиб бера олиши керак. Яъни корпусни тузиш, леммалаш, разметкалаш жараёнида корпус таркибига кирган шундай сўзларни бирма-бир қидирув асосида матнлар ичидан топиш ва уларни махсус изоҳлаш керак бўлади. Бунинг учун эса алгоритм, лингвистик молеллаштириш ишлари амалга оширилиш лозим.

Миллий корпуснинг интерфейси турли дизайн, тузилишга эга 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 олиниши лозим.

Демак, миллий корпусдан фойдаланишда унинг қулай ва энг самарали имкониятларининг намоён бўлишида интерфейсининг мукамал ҳамда тизимли равишда ишлаб чиқилганлиги муҳим саналади. Шу боис интерфейсни замонавий дастурий дизайн талабларига жавоб берадиган, фойдаланувга тушунарли, ишлаш учун қулай шаклда яратилади.



**ROMANLARI**  
**HIKYOVALARI**  
**PYESA**  
**PORTRET**  
**FELYETON**

**● Alisher Navoiy hayoti va ijodi**  
**● O'zbek mumtoz adabiyoti**

**FOTO ALBOM**  
+ **Buyuk siymolar**

**TAQDIMOT**  
+ **Taqdimot**

**Korpus nima...**

Korpus — til birliklarining xususiyatlarini aniqlash maqsadida qidiruv dasturiga bo'yundirilgan matnlar majmuidir, tabiiy tilning elektron shaklida saqlanadigan yozma yoki ovozli, kompyuterlashtirilgan qidiruv tizimiga dasturiy ta'minot asosida joylashtirilgan on-line yoki off-line tizimda ishlaydigan matnlar jamlamasidir[1]. U elektron shaklda mavjud bo'lgan matnlarni, hujjatlarini ma'lumotlarni qayta ishlash, ularni avtomatik ta'biy, ya'ni morfologik sintaktik va semantik ta'biy qilish, morfologik analiz va sintez qilish bilan birga, yarim avtomatik rejimda ma'noli buzmasdan ishonchli nuzul materialini moslashtirish darajasini tekshirish ushbu sohadagi til bilan qiziqishni oshirishga xizmat qiladi. Korpus — tilni tadqiqot qilish, til o'rganish, lug'at tuzishda eng zamonaviy, keng imkoniyatli dasturlashtirilgan tizimdir. Korpus - milliylik so'z kontekstlariga havola qiluvchi, maxsus qidiruv tizimi asosida ishlaydigan elektron matnlar yig'indisi. U Internet tizimida elektron kutubxona, lug'at va lingvistik grammatikadir. Korpus - tabiiy (real) tilning elektron shaklida bo'lib, uni to'ldirish, tuzatish, ta'bir qilish imkoniyatiga ega[2]. Korpus bu so'zlar, so'z birliklari, grammatik shakllari ma'nosi ma'lum bir izlash tizimi orqali topilgan 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 ifoda etadi. Masalan, o'zbek tili Milliy korpusi Internetda amalga o'ralsa, 140 mln so'z izlashmalaridan iborat shu tilning barcha turdagi matnlari o'zida muassamlashtirila oladi. Kelajakda o'zbek tili Milliy korpusi o'zida 300 mln so'z izlashmalarini jamlashi lozim bo'ladi. Boshqa milliy tildagi korpuslari kabi o'zbek tili Milliy korpusi xam ikkita muxim xususiyatga ega. Birinchidan u ancha salohiyat va barcha yo'nalishlar uchun (turi badiiy janrlar: publitsistik o'quv ilmiy, ish yuritish, nutk - so'zlashuv, shevalar va boshqalar uchun) bir xilda muvofiqlashtiradi. Bu matnlar o'z davrida tilning ta'sir kuchiga qarab proporsional ravishda korpusda jamlangan bo'ladi. Ikkinchidan korpus matnlariga xos bulgan alohida ijtihatini qo'shimcha ma'lumot sifatida o'zida aks ettiradi (masalan ma'lum bir ixtatdan belgilanishlar va annotatsiyalar). So'zlar va boshqa birlikmalar belgilanishi - korpusning bosh xarakteristikasidir (tasnifi); u korpusi 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 olinish kerak bo'ladi: metamatni (u matni muallif qarashi, janri va boshqa xususiyatlari bo'yicha to'liq ifodalaydi), morfologik va semantik belgilanishlar (morfologik va semantik belgilanishlar matni emas balki alohida bir so'zni tasnif beradi). ADABYOTLARI: 1. Grudeva V.A. Korpusnaya lingvistika. Uchebnoe posobie. 2-e izd., stereotip. - M.: Flinta, 2012. - 165 s. 2. Zakharov V.P. Korpusnaya lingvistika: Uchebno-metod. posobie. - SPb., 2005. - 48 s. 3. Zakharov V.P., Bogdanova S.Yu. Korpusnaya lingvistika. - Irkutsk: IGLU, 2011. - 161 s. 4. Torova 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 hqqa matni asosida amalga oshirildi. Korpus

Хуллас, ўзбек тили миллий корпусининг умумий кўриниши бир неча ойналарга ҳамда ўнг ва чап устунчаларга ажратилган. Унда қуйидаги ойналар мавжуд бўлади: «Лексик қидирув», «Морфологик қидирув», «Синтактик қидирув». Ундан сўз ва гаплар бир неча сония вақт ичида автоматик таҳлил қилиб берилади.

## «ЎТКАН КУНЛАР» РОМАНИДА ҚЎЛЛАНГАН ПРАГМАТИК ХУСУСИЯТЛАР

*Д.Б.Жумаева*

*Навоий давлат педагогика институтини*

Тил кишилар ўртасида фақат алоқа воситасигина бўлиб қолмай, куммулятив ва эстетик вазифани ҳам бажаради. Ижтимоий ҳаётда юз бераётган ҳар қандай ҳодиса, қайси даврда бўлишидан қатъи назар, бадий адабиётда акс этади ва келажак авлодларга етказилади. Бу жараён