

## МИЛЛИЙ КОРПУС ЯРАТИШНИНГ ТЕХНОЛОГИК ЖАРАЁНИ ХУСУСИДА

Гули Ибрагимовна ТОИРОВА

Филология фанлари бўйича фалсафа доктори (PhD), доцент  
Бухоро давлат университети. Бухоро, Ўзбекистон

## О ТЕХНОЛОГИЧЕСКОМ ПРОЦЕССЕ СОЗДАНИЯ НАЦИОНАЛЬНОГО КОРПУСА

Гули Ибрагимовна ТОИРОВА

Доктор PhD

Бухарский государственный университет. Бухара, Узбекистан

## CREATION OF A NATIONAL HOUSING SPECIFICATION OF TECHNOLOGICAL PROCESS

Guli Ibragimovna TAIROVA

Doctor of philosophical science (PhD)

Bukhara State University, Bukhara, Uzbekistan [tugulijon@mail.ru](mailto:tugulijon@mail.ru)

UDC (УЎК, УДК): 81'33

**For citation (иктибос келтириш учун, для цитирования):**

Тоирова Г. И. Миллий корпус яратишнинг технологик жараёни хусусида // Ўзбекистонда хорижий тиллар. — 2020. — № 2 (31). — Б. 57–64.

<https://doi.org/10.36078/1588232876>

Received: February 15, 2019

Accepted: April 12, 2020

Published: April 15, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



**Аннотация.** Мақолада тилнинг интернет, компьютер технологиялари тилига айланиши математик лингвистика, унинг давоми бўлган компьютер лингвистикасининг шаклланганлиги ва ривожланиш даражаси билан боғлиқлиги, айниқса, сунъий интеллект учун табиий тилларни моделлаштириш муаммоси хусусида мушоҳада юритилган. Ўзбек тилининг халқаро мақomini оширишда ўзбек тили миллий корпуси катта аҳамиятга эга. Компьютер лингвистикаси фани доирасида амалга оширилаётган ишлар ўзбек тилидаги мавжуд муаммоларни ҳал этишда муҳим роль ўйнайди. Мақолада миллий корпус яратишнинг заруратлари нимада эканлиги, таълим жараёнида сўзнинг нозик қиррасини ўрганиш, уни нутқда қўллаш кўникмасини шакллантиришда нимага таяниш кераклиги, ўзбек тили миллий корпуснинг фан ва миллат учун нима аҳамияти борлиги, ўзбек тили миллий корпуснинг асосий фойдаланувчилари ким бўла олиши мумкинлиги хусусида асосли фикр-мулоҳазалар билдирилди. Корпуснинг асосий фойдаланувчилари, албатта, турли йуналишларда иш олиб бораётган тилшунос тадқиқотчилардир. Маълум бир даврдаги тилга оид ишончли статистик маълумотлар адабиётшунослар, тарихчилар ва бошқа гуманитар соҳалар вакиллари ўзига жалб этиб келгандир. Тилни ўқитишда ҳам миллий корпуснинг аҳамияти каттадир. Миллий корпус яратишнинг технологик жараёни хусусида фикр-мулоҳазалар баён қилинди. Жумладан, миллий корпус технологик жараёни: танланган матнлар асосида лексема ва сўз шакллариининг такрорланиши луғатини яратиш; олинган такрорланиш луғатининг ҳар қандай бирлиги учун матнни кўриб чиқиш; графикли сўзни бўғинга ажратиш ва бўғинларнинг такрорланиш луғатини тузиш; сўз захираларини саралаш; бир вақтнинг ўзида чекланма-

ган файлларни қайта ишлаш; ташқи белгиларга эга бўлган матнлар корпусларини яратиш кабилардан иборатлиги изоҳлаб берилган.

**Калит сўзлар:** корпус лингвистикаси; компьютер лингвистикаси; корпус; технологик жараён; конверсиялаш; графематик таҳлил; матнни қайта ишлаш; сўз захираларини саралаш

**Аннотация.** В статье рассматривается язык в сети интернет, формирование языка компьютерных технологий как продолжение вычислительной лингвистики, ее уровни развития, особенности влияния языка на моделирование нового языка для искусственного интеллекта. Национальный корпус узбекского языка имеет большое значение для повышения международного статуса узбекского языка. Работа, выполняемая в области компьютерной лингвистики, играет важную роль в решении существующих проблем в узбекском языке. Статья основана на том, что значит строить национальный корпус, изучать тонкости слова в процессе обучения, на что опираться, чтобы развить способность использовать его в речи, какое значение имеет узбекский язык для науки и нации и кто может быть основным клиентом/пользователем узбекского языка. Основными пользователями корпуса являются, конечно же, лингвисты, которые работают в разных областях. Надежная статистическая информация о языке определенного периода привлекает также литературоведов, историков и других представителей гуманитарных наук. Национальный язык также важен в преподавании языка. Были высказаны мнения о технологическом процессе создания национального корпуса. В частности, технологический процесс национального корпуса учитывает: создание словарного запаса лексемы и словоформы на основе выбранных текстов; возможность просмотра текста для любой единицы полученного словарного запаса; разбивка графического слова на слоги; подбор слов; одновременная обработка неограниченного количества файлов, создание текстовых файлов с внешними символами.

**Ключевые слова:** корпусная лингвистика; компьютерная лингвистика; корпус; технологический процесс; преобразование; графический анализ; обработка текста; выбор ресурса слова.

**Abstract.** The article deals with the language on the Internet, the formation of the language of computer technology as a continuation of computational linguistics, its levels of development, especially the influence of language on the modeling of a new language for artificial intelligence. The national corpus of the Uzbek language is of great importance in enhancing the international status of the Uzbek language. The work being done in the field of computer linguistics plays an important role in solving the existing problems in the Uzbek language. The article explores the need for building a national corpus, to learn the subtleties of words in the learning process, what to rely on in developing the ability to use it in speech; the importance of the Uzbek language to the science and the nation, There are well-grounded opinions about who can be the main clients and users of the Uzbek language. The main users of the corps are, of course, linguists who work in different fields. Reliable statistical information about a language of a certain period attracts literary scholars, historians, and other humanities. National language is also important in language teaching. Opinions were expressed about the technological process of creating a national corps. In particular, the technological process of the national corps takes into account: the creation of a

vocabulary of the lexeme and word form based on the selected texts; view the text for any unit of the received vocabulary; break a graphic word into syllables and create syllables for repeating syllables; selection of words; The simultaneous processing of unlimited files, the creation of text files with external characters.

**Keywords:** corpus linguistics; computer linguistics; housing; technological process; transformation; graphical analysis; word processing; word resource selection.

**Кириш.** Ҳозирги вақтда инсонлар ўртасидаги мулоқот жараёнлари янада мураккаблашиб, мукамаллашиб бормоқда. Компьютер пайдо бўлиши билан мулоқотнинг Интернет шакли юзага келди. Ўзгараётган шароитларда инсон дунёни тўғри қабул қилиш кўникмасини йўқотмаслиги ва нутқини ўз маданият анъаналари доирасида батафсил ўрганилишини тақозо қилади. Чунончи, улкан матн маълумотларини тезкор қайта ишлаш имконига эга бўлган ҳисоблаш техникаси пайдо бўлди.

Компьютер технологиялари воситасида ахборотни қайта ишлаш, машина таржимаси, электрон луғатшуносликни ривожлантириш, тезауруслар тузиш, тил корпусини яратиш ишларини дунё лингвистикаси цивилизациясига эришган тиллар аллақачон амалга оширди. Инглиз, рус, араб, француз, немис, испан, тожик каби тиллар шулар жумласидандир. Мазкур тилларда интернет тизимида тил корпусини яратиш масаласининг илмий-назарий жиҳатлари ҳам ишлаб чиқилганки, бу ўзбек тилини ҳам интернет “тушунадиган” тилга айлантириш (яъни тил корпусини яратиш) сабъ-ҳаракатларини жадаллаштириш заруратини кун тартибига қўяди.

Бу борада тилшунос олимлар тилни ва у билан боғлиқ бўлган ҳамда илгари маълум бўлмаган соҳаларни ўрганишда жуда муҳим тадқиқот имкониятига эга бўлдилар (2, 7, 8, 11, 12). Чунончи, “WordNet компьютер лексикографиясида эришилган инновацион ютуқлардан биридир. Чунки унга илова этилган сўзликлар машина ўқиши учун мумкин бўлган ҳолатга мослаштирилган. Одатий луғатларда киритилган сўзликларнинг талаффузи, грамматик шаклланиши, ясалиши, этимологияси, изоҳи ҳамда синоним, антоним каби яна бир нечта лингвистик хусусиятлар сингдирилади” (1).

Тилни ўрганишда тилшунослар учун нутқ фаолиятини анча тўлиқ акс эттирадиган қулай ва ихчам нутқ материалига эҳтиёж сезади. Бу эҳтиёж эса электрон шаклда тақдим этилган матнларни, яъни лингвистик корпусларни яратишни тақозо қилади. Лингвистик корпуслар ўзида тил тизимининг тилшунослик изоҳларини ва бошқа тилшунослик фанларида қўлланилиши мумкин бўлган матнларни корпусида ўзига хос акс этадиган нутқ материали сифатида тадқиқ этади. Анъанавий тилшуносликдан фарқли равишда корпус тилшунослиги тил ўрганиш билан шуғулланмайди, балки нутқнинг корпусида махсус танланган матнлари барча кўринишларини (бадий, илмий, муомала тилини) ўрганади. Бу ўринда мантикий фикрлаш ва анъанавий тилшунослик дедукциясига нисбатан эҳтимоллик, статистик ва индуктив усуллар тез қўлланилади. Бунда узоқ йиллар давомида йиғилган тажрибалар асосида олинган натижаларга, баҳоларга ва амалларга эмас, балки эмпирик жиҳатдан тўпланган материалларга кўпроқ таянилади. Корпусли тилшунослик учун асосий восита

бир ёки бир неча тиллар доирасидаги нутқ фаолияти ҳамда корпусдаги кейинги тадқиқотларни ихчамлаштириш учун зарур бўлган корпус материалларини кодлаштириш ҳисобланади. “Илк корпуснинг яратилиши 1812 йилга бориб тақалади, бунда немис олими Кадлинг ўзининг немис сўзларидаги ундош ҳарфлар дистрибуциясини таҳлил қилган. Ваҳоланки, у даврда ҳали компьютер терминининг ўзи ҳам бўлмаган. Кейинчалик, замонавий инглиз тили корпуси намуналаридан бири сифатида Бровн корпуси 1960–1961 йилларда яратилди ва у илк бор босма ҳолда чоп этилди. Орадан бир йил ўтибгина, ушбу корпус электронлаштирилди. Кўриниб турибдики, илк корпуслар компьютер технологияларисиз ҳам мавжуд бўлган ва изланишлар олиб борилган. Кейинчалик фан ривожиди кидирув методи(конкорданс)ни электрон форматдаги матнларда ўтказишни тақозо этди ва компьютер лингвистикаси билан соҳалараро муносабатга эҳтиёж туғилди” (3). Куйидаги хусусиятлар корпус учун муҳим омил саналади: тўпланган материалларни, албатта, машина сақловчиларига (компьютерларга) жойлаштириш; электрон кидирувга (морфологик, синтактик кидирувлар) имкон берадиган ўзига хос белги; якуний ўлчов бирлигига эга ва репрезентативлик (тилдаги кўплаб жанрларнинг асл ҳолида тўлиқ акс этиш).

Ўзбек тили миллий корпуси — ўзбек тилининг бадиий асарлари ҳамда вариантлари, кенг кўламли лингвистик ва метаматнли маълумотлари билан таъминланган, турли хил услубларни акс эттирувчи электрон матнлар мажмуасини ифодалай олиши керак.

Корпус билан ишлашда юқори самарадорликка эришиш учун, тўлиқ матнли кидирув тизими зарур. Мазкур тизим улкан матнлар массивида контекст кидирув воситаси учун махсус оптималлаштирилган. Бундай тизимлар ўртасида энг машҳурлари — Яндекс интернет кидирув — серверлари ва Google, шунингдек, индексация ва интернет-манбалар кидируви каби тизимлар ҳам. Табиий равишда ўзбек тили корпуси учун Яндекс-сервер тизими жуда мос. Яндекс-сервер тизими жуда юқори самарадорлик ва кенг кўламлилик, контекст кидирувни тўлиқ амалга ошириш, қудратли тил сўровномасига эга бўлиш, матнларнинг ностандарт турлари учун эгилувчан индексатор созлагичини ўрнатиш, ўзбек тилини тўлиқ кўтариш каби имкониятларига эга. Яндекс-сервер қисқа фурсат ичида улкан матнлар массивида мураккаб сўровларни амалга ошириш имкониятига эга бўлиб, бунда кидирув тезлиги корпус ҳажмига умуман боғлиқ эмас.

Корпус яратишнинг технологик жараёни тўғрисида В. В. Рыков қуйидаги асосий талабларни санаб ўтади: “Корпус фойдаланувчиси (якка шахс, гуруҳ, лингвистик жамият); корпус кўзлаган мантикий мақсад; корпус тузишда ишланадиган маълумотлар базаси ҳажми ҳамда бунинг реал, зарурлик даражаси; мантидан фойдаланиш усули (парча, тўлиқ кўриниш ёки ҳар иккиси)” (9). Муаллифлик корпусини яратиш бўйича Ш. Ҳамроева корпус яратишнинг технологик жараёни тўғрисида қуйидаги таклифни беради: “Тил материалларини тадқиқ этиш; матнни сканерлаш; тўлиқ шакллантириш, корпусни тузиш” (12). Корпус яратишнинг технологик жараёни хусусида, чунончи, “Рус тили миллий корпуси” матнлари устида иш олиб борган олим С. О. Савчук қуйидаги жараёнга ажратади: “Матнларни қайта киритиш; электрон шаклда мавжуд матнлардан фойдаланиш; босма матнни сканерлаш (лекин бунда кўплаб хатони тузитиш керак бўлади)” (10).

Ўқитувчи учун корпус тенгсиз хазина, ундан ҳар бир ўқув машғулоти учун серқирра ва мазмунли ўқув материали тайёрлаш мумкин. Юқорида айтиб ўтганимиздек, корпусда хилма-хил матнлар жамланган, ундан ўқитувчи исталган шаклда фойдалана олади. Мунтазам тўлдирилиб бориладиган корпусдан янги мисолларни олиш, уни талабага тақдим этиш ёки талабага ҳам шундай топшириқ бериш, албатта, таълимни бугунги ҳаётга яқинлаштиради. Тадқиқотчига жуда катта ҳажмдаги турли-туман материал билан ишлаш имконини беради. Корпуснинг асосий фойдаланувчилари, албатта, турли йуналишларда иш олиб бораётган тилшунос тадқиқотчилардир. Маълум бир даврдаги тилга оид ишончли статистик маълумотлар адабиётшунослар, тарихчилар ва бошқа гуманитар соҳалар вакиллари ўзига жалб этиб келгандир. Тилни ўқитишда ҳам миллий корпуснинг аҳамияти каттадир.

Бизнингча, “Ўзбек тилининг миллий корпус”ни яратишнинг технологик жараёни қуйидаги босқичлар ёки қадамлар шаклида намойиш этилиши мумкин:

- манбалар рўйхатини аниқлаш;
- матнларни рақамлаштириш (компьютер шаклига ўтказиш).

Айтиш керакки, матнларни компьютерга киритиш олдингидек қийин ва кўп вақт талаб қилар эди, бугунги кунга келиб, бу муаммо, ҳеч бўлмаганда замонавий имло замонавий матнларига нисбатан жуда осон ҳал қилинди. Ушбу енгиллик оптик кириш (сканерлаш) ва матнни аниқлаш ва замонавий ҳаётни глобал компьютерлаштириш, шу жумладан матнни қайта ишлаш билан боғлиқ соҳаларда эришилган ютуқларга асосланади. Корпусларни яратиш учун электрон шаклдаги матнларни турли хил усуллар билан киритиш мумкин. Чунончи, қўлда киритиш, сканерлаш, муаллифлик нусхалари, совғалар ва биржалар, Интернет, муҳаррирларга тақдим этилган асл макетлар ва бошқалар.

Мутахассислар К. Ф. Мейер ва И. А. Мельчук корпус тузишнинг технологик жараёнида қуйидаги босқичларни ажратишади:

- белгиланган манбага мувофиқ ҳолда матннинг корпусга киришини таъминлаш;
- матнни автоматик ўқилиш шаклида қайта ишлаш. Корпусга киритиладиган электрон шаклдаги матн турли усул билан олинган бўлиши мумкин: қўлда киритилган, сканерланган, муаллифлик нусхаси, ҳада, айирбошлаш, Интернет, нашриётлар томонидан корпус тузувчисига бериладиган оригинал-макетлар;
- таҳлил, матнга дастлабки ишлов бериш. Ушбу босқичда турли манбалардан қабул қилинган матнлар филологик текширув, таҳрирдан ўтади;
- конверсиялаш, графематик таҳлил. Баъзи матнлар қайта кодлаштириш жараёни амалга ошадиган илк машина ишловидан қайта қайта ўтади, номатний қисмлар (расм, жадвал) ўчирилади ёки ўзгартирилади. Матндаги бўғин кўчириш, чегаралар (MS-DOS матнларида) бекор қилинади, тире, бошқа белгилар бир хиллигига эришилади. Графематик таҳлил корпусга кирувчи матнни қисмга (сўз, боғловчи) ажратиш, номатний элементни ўчириш каби амалларни бажаришдан иборат;
- ностандарт (нолексик) элементни белгилаш, расмийлаштириш, махсус матний элементни (қисқартма асосида ёзилган ном (исм,

фамилия), бошқа алифбода ёзилган ўзлашма лексема, расмга берилган ном, изоҳ, зарварақ, адабиётлар рўйхати ва б.) бир хил мезон асосида қайта кўриб чиқиш. Албатта, бу амаллар автоматик равишда матн муҳаррири томонидан бажарилади” (4, 5).

Мутахассислар К. Ф. Мейер ва И. А. Мельчукларнинг технологик жараён босқичларини қўллаб-қувватлаган ҳолда қуйида “Ўзбек тили миллий корпуси”нинг технологик босқичларни таклиф этамиз:

**1. Матнни олдиндан қайта ишлаш босқичи.** Ушбу босқичда турли манбалардан олинган барча матнлар имловий тузатилади ва таҳрир қилинади. Шунингдек, матннинг библиографик ва экстралингвистик тавсифи тайёрланади.

Конверсия ва график таҳлил қилиш босқичи. Баъзи бир матнлар, шунингдек, дастлабки қайта ишлашнинг бир ёки бир неча босқичларини босиб ўтади, улар давомида турли хил кодлаш (керак бўлса), матн бўлмаган элементларни (расмлар, жадваллар) йўқ қилиш ёки ўзгартириш, матндан чизикларни олиб ташлаш, "чизикларнинг каттик учлари" ни таъминлаш. ягона ёзув тире ва бошқалар. Қоидага кўра, ушбу операциялар автоматик равишда амалга оширилади. Одатда, худди шу босқичда, матн таркибий қисмларига бўлинади.

**2. Матнни белгилаш босқичи.** Матнни белгилаш матнлар ва уларнинг таркибий қисмларига қўшимча маълумот (метадата) қўшишдан иборат. Корпус матнларининг мета-тавсифи иккала маълумот элементларини (библиографик маълумотлар, матннинг жанри ва услуб хусусиятларини тавсифловчи белгилар, муаллиф ҳақидаги маълумотлар) ва расмийларни (файл номи, кодлаш параметрлари, белгилаш тилининг версияси, иш босқичларини ижрочилари) ўз ичига олади. Ушбу маълумотлар одатда қўлда киритилади. Ҳужжатнинг таркибий белгилари (параграфлар, жумлалар, сўзларни танлаш) ва ҳақиқий лингвистик белгилар одатда автоматик равишда амалга оширилади.

**3. Кейинги босқичда автоматик маркировка натижаларини тўғирлаш амалга оширилади:** хатоларни тузатиш ва ажратиш (қўлда ёки ярим автоматик).

**Корпусга киришни таъминлаш.** Корпус дисплейи қуйидаги кўринишда: CD-ROMда тарқатилиши ва глобал тармоқ режимида мавжуд бўлиши мумкин. Фойдаланувчиларнинг турли тоифаларига турли хил ҳуқуқлар ва турли хил имкониятларга эга бўлади.

**4. Яқуний босқич** — тегли матнларни тез кўп қиррали кидириш ва статистик ишлов беришни таъминлайдиган ихтисослаштирилган лингвистик маълумот олиш тизимининг таркибига ўзгартириш (корпус менежери).

Албатта, ҳар бир ҳолатдаги босқичларнинг таркиби ва сони юқорида санаб ўтилганлардан фарқ қилиши ва реал технология анча мураккаб бўлиши ҳам мумкин.

Хулоса қилиб айтиш мумкинки, миллий корпусни яратиш — жуда катта ҳажмдаги материалларни таҳлил этишга мўлжалланган электрон шакллар ва кенг қўламдаги статистик материалларни йиғишга йўналтирилган ўзига хос машаққатли иш. Миллий корпус технологик жараёни қуйидагилардан иборатдир: танланган матнлар асосида лексема ва сўз шаклларининг такрорланиши луғатини яратиш; олинган такрорланиш луғатининг ҳар қандай бирлиги учун матнни кўриб чиқиш; графикли сўзни бўғинга ажратиш ва бўғинларнинг такрорланиш луғатини тузиш; сўз захираларини саралаш; бир

вақтнинг ўзида чекланмаган файлларни қайта ишлаш; ташқи белгиларга эга бўлган матнлар корпусларини яратиш; яратиладиган матнлар корпуслари ҳамда корпусга кирувчи алоҳида матнлар учун статистик маълумотларни ҳисоблаб чиқиш; дастлабки матнлар билан txt, doc и rtf форматда ишлаш, кодлаштиришни автоматик тарзда белгилаш.

#### Фойдаланилган адабиётлар

1. Абдурахмонова Н.З., Ҳайдаров М.Р. Ўзбек тилида WORDNET яратиш масалаларига доир // Ўзбекистонда хорижий тиллар (электрон илмий-методик журнал). № 4. 2019. — Б.19–28. — URL: <https://journal.fledu.uz/uz/issue/4-27-2019>
2. Абжалова М. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун). Филол. фан. бўйича фалсафа доктори (PhD)...дис. автореф. — Фарғона, 2019. — 45 б.
3. Атабоев Н.Б. Корпус лингвистикасининг асосий хусусиятлари // Ўзбекистонда хорижий тиллар (электрон илмий-методик журнал). — № 2 (25) / 2019. — Б. 37–45. — URL: <https://journal.fledu.uz/uz/issue/2-25-2019/>
4. Мельчук И.А. Порядок слов при автоматическом синтезе русского слова (предварительные сообщения) / Научно-техническая информация. 1985, № 12. — С.12–36.
5. Meyer C.F. English Corpus Linguistics. — Cambridge: Cambridge University Press, 2002. — URL: <https://pdfs.semanticscholar.org/c775/4bfab1d0f7>
6. Поляков А. Е. Технология подготовки информации в Национальном корпусе русского языка Текст. / Поляков А.Е. // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. — М., 2005. — 192 с.
7. Toirova G. Importance of Interface in Creating Corpus// International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277–3878, Volume 8, Issue 2S10, September 2019. — pp. 4591–4593.
8. Toirova G. The Role of Setting in Linguistic Modeling. International Multilingual Journal of Science and Technology. ISSN: 2528-9810 Vol. 4 Issue 9, September — 2019. — pp. 722–723.
9. Рыков В.В. Курс лекций по корпусной лингвистике. — URL: <http://rykov-cl.narod.ru/c.html>.
10. Савчук С.О. Корпус современной русской прессы: из опыта создания и использования // Труды Международной конференции «Корпусная лингвистика — 2011». — СПб.: Санкт-Петербургский государственный университет, 2011. — С. 149–154.
11. Эшмўминов А. Ўзбек тили миллий корпусининг синоним сўзлар базаси. Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. — Қарши, 2019. — 46 б.
12. Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Филол. фан. бўйича фалсафа доктори (PhD)...дис. автореф. — Қарши, 2018. — 45б.

#### References

1. Abdurakhmanova N.Z., Haydarov M.R., *Foreign Languages in Uzbekistan* (Electronic Scientific-Methodical Journal). No.4 (27), 2019, pp.19–28, available at: <https://journal.fledu.uz/>

2. Abzhalova M., *Uzbek tilidagi matnlarni tahrir va tahlil qiluvchi dasturning lingvistik modullari* (Linguistic modules of the program for editing and analyzing texts in the Uzbek language), Extended abstract of PhD's thesis, Fergana, 2019, 45 p.
3. Ataboev N.B., *Foreign Languages in Uzbekistan* (Electronic Scientific-Methodical Journal), No.2 (25) / 2019, pp. 37–45, available at: <https://journal.fledu.uz/uz/issue/2-25-2019/>
4. Mel'chuk I.A., *Nauchno-tehnicheskaya informatsiya*, No.12, 1985, pp. 12–36.
5. Meyer C.F. *English Corpus Linguistics*, Cambridge: Cambridge University Press, 2002, available at: <https://pdfs.semanticscholar.org/c775/4bfab1d0f7>
6. Polyakov A. E. *Tekhnologiya podgotovki informatsii v Natsional'nom korpuse russkogo yazyka Tekst* (Technology of information preparation in the National corpus of the Russian language Text), Polyakov A.E., *Natsional'nyi korpus russkogo yazyka: 2003–2005. Rezul'taty i perspektivy* (National corpus of Russian language: 2003-2005. Results and prospects), Moscow, 2005, 192 p.
7. Toirova G., *International Journal of Recent Technology and Engineering*, Vol. 8, 2019, pp. 4591–4593.
8. Toirova G., *International Multilingual Journal of Science and Technology*, Vol. 4, 2019, pp. 722–723
9. Rykov V.V., *Kurs leksii po korpusnoi lingvistike* (Course of lectures on corpus linguistics), available at: <http://rykov-cl.narod.ru/c.html>.
10. Savchuk S.O., Proceedings of the International Conference "Corpus linguistics-2011", Saint Petersburg State University, 2011, pp. 149–154
11. Eshmŷminov A., *Uzbek tili millii korpusining sinonim suzlar bazasi* (The National Body of the Uzbek Language is a Database of Synonyms), Extended abstract of PhD's thesis, Karshi, 2019, 46 p.
12. Hamroeva Sh., *Uzbek tili mualliflik korpusini tuzishning lingvistik asoslari* (Linguistic Foundations of the Creation of the Uzbek Language Authorship Corps), Extended abstract of PhD's thesis, Karshi, 2018, 45 p.