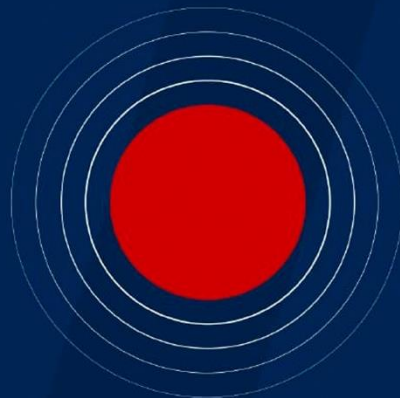


UZA | O'zbekiston
Milliy axborot
agentligi



**ЎЗБЕКИСТОН МИЛЛИЙ
АХБОРОТ АГЕНТЛИГИ – ЎЗА
ИЛМ-ФАН БЎЛИМИ**

(ЭЛЕКТРОН ЖУРНАЛ)

2021 ЙИЛ МАЙ

ТОШКЕНТ – 2021

О.Х.АБДУЛЛАЕВА. Тил корпусларини яратишда қўлланилувчи дастурлар таснифи-----	194
Ф.С.БАФОЕВ. Ҳерман Вамберининг «Сagataische sprachstudien» («Чигатой тили дарслиги») асаридаги ўзбек мақоллари хусусида-----	203
Д.Б.ОЛИМОВА. Аҳмад Яссавий “Меърожнома”сининг ирфоний моҳияти-----	210
Д.М.ЮЛДАШЕВА. Педагогическое наследие академика Л.В.Щербы и его актуализация в современной науке и образовании-----	219
С.С.ХОТАМОВ. Ёзувчи Хайриддин Султоннинг «Бобурийнома» асаридаги ўзига хос услубнинг таҳлилий жиҳатлари-----	228
Д.Н.САЪДУЛЛАЕВА. Ўзбек насрида рухоний образига муносабат масаласи-----	236
М.О.ҲАМИДУЛЛАЕВА. Алишер Навоий шеърлятида кўз, қош, киприк тимсоллари талқини-----	246

Тарих

С.С.РАУПОВ. Бухоро амирлигида маҳаллий ўзини-ўзи бошқарув-----	254
Ф.А.АДИЛОВ. Озарбайжон тарихчиси Зиё Мусо ўғли Буниётовнинг Ўзбекистон тарихини ўрганишга қўшган ҳиссаси-----	264
О.Р.РАШИДОВ. Ўзбекистонда большевиклар миллий сиёсатидаги тил ва ёзув масаласи-----	271
И.Х.ШОЕВ. Археология и перспективы изучения и использования археологических коллекций-----	281
Ф.А.КАЛАНДАРОВА. “Тарихи салотини манғития”да Бухоро амирлиги этник таркибининг ёритилиши-----	289

Фалсафа

Ш.Б.МУРТОЗАЕВ. Абу Ҳафс Умар Сухравардий ҳаёти ва илмий мероси-----	298
Ш.Ф.ДЖУРАЕВ, А.А.АЛАМОВ. Организация учебного процесса по предмету философия: краткий анализ-----	306

Техника

Х.Н.ЗАЙНИДИНОВ, Ш.А.АНАРОВА, Ж.С.ЖАББАРОВ. Ўхшашлик ёрдамида геометрик шаклларнинг фрактал ўлчовлари-----	317
Э.Т.ФАРМОНОВ, Т.АБДИЛЛАЕВ, Ф.Э.ФАРМОНОВА. Саксовул ва черкез уруғларини экадиган сеялка экиш аппаратининг параметрларини мақбуллаштириш-----	326
Ш.А.СУЛАЙМОНОВ. Кимёвий препаратлар ёрдамида куруқ пилла қобиғини чанг ва бошқа омиллардан сақлаш орқали сифатли хом ипак олиш усуллари-----	334

ТИЛ КОРПУСЛАРИНИ ЯРАТИШДА ҚЎЛЛАНИЛУВЧИ ДАСТУРЛАР ТАСНИФИ

Оқила Холмўминовна АБДУЛЛАЕВА

таянч докторант

Алишер Навоий номидаги

Ўзбек тили ва адабиёти университети

Тошкент, Ўзбекистон

Аннотация

Корпус лингвистикаси соҳасида лингвистик маълумотларни бевосита таҳлил қилиб, тадқиқотчига турли хил натижаларни кўрсатиб бера оладиган корпус дастурларнинг яратилиши, дастурларнинг имкониятлари ва хусусиятларига кўра таснифланиб, турларга ажратилади. Ҳар бир таснифланган қисмнинг ютуқ ва камчиликлари муҳокама қилинади.

Таянч сўзлар: корпус, дастурий таъминот, лингвистик таъминот, аннотация, конкорданс.

КЛАССИФИКАЦИЯ ПРОГРАММ, ИСПОЛЬЗУЕМЫХ ПРИ СОЗДАНИИ ЯЗЫКОВЫХ КОРПУСОВ

Оқила Холмуминовна АБДУЛЛАЕВА

базовый докторант

Университет узбекского языка и

литературы имени Алишера Навои

Ташкент, Узбекистан

Аннотация

В данной статье анализируются лингвистические данные в сфере корпусной лингвистики, рассматривается создание корпусных программ, показывающих различные результаты для исследователя, программы классифицируются и разделяются на группы в зависимости от их возможностей и особенностей. Обсуждаются успехи и недостатки каждой квалифицированной части.

Ключевые слова: корпус, программное обеспечение, лингвистическое обеспечение, аннотация, конкорданс.

Бугунги кунда тил фақат назарий эмас, амалий ёндашувга эҳтиёжи бор соҳалардан бирига айланди. Табиий тил матнлари (ёзма ёки аудио) орқали такрорланмас ва бир-бирига ўхшамаган лингвистик хусусиятларини, нутқ қолипларини кузатиш ва таҳлилий хулосалар беришда корпус дастурларидан кўплаб тилларда фойдаланилмоқда. Дастурлар ва уларнинг дизайни компьютер технологиялари воситасида, махсус дастурий таъминот орқали

яратилади. Тезкор ва қулай компьютер технологияларининг ривожланиши, миллий тилларнинг кенг ҳажмли корпуслари яратилиши сабабли юзга яқин турли мақсаддаги дастурлар яратилди.

Корпус дастурлари нима учун керак? Биринчидан, корпусларни яратиш махсус дастурларсиз амалга ошмайди. Иккинчидан, корпуслардаги кенг қатламли лингвистик бирликларнинг таҳлилида дастурларга эҳтиёж мавжуд. Учинчидан, корпус имкониятларидан нафақат мутахассис филологлар, лингвистлар, луғатшунослар, шунингдек, тил ўргатувчи ўқитувчилар, ўрганувчи ўқувчилар ҳам фойдаланиши мумкин. Бунда эса турли мақсадларда қўлланилувчи корпус дастурлари зарур бўлади.

Корпуслардан корпус дастурларининг ажратилишининг ўзига хос томонлари борлиги тўғрисида Т.МакЭнерий, А.Лауренселар дарслик ва мақолалари орқали маълумотлар берган. Корпус дастурларининг таснифланиши илк Т.МакЭнерий ва А.Ҳардие томонидан баён қилинган. Дастурларнинг гуруҳланиши уларнинг яратилиш даврлари ва фаолиятига кўра гуруҳланган. Олим дастурларнинг тўрт авлодга ажратади [3; 80].

1)Биринчи авлод дастурлари: Конкорданс генератори, Дискон дастурларини ўз ичига олади. Бу дастурлар кейинги дастурлар учун пойдевор вазифасини бажарса-да, бироқ фаолияти чегараланган, фақат чекланган миқдордаги белгиларнигина таниган. Кўпроқ матндаги сўзлар миқдорини ва рўйхатини кузатиш билан чегараланган. 2) Иккинчи авлод дастурлари: Оксфорд конкорданс программаси, Лонгман Мини-конкордансери, Микроконкорд каби дастурлари бўлиб, бу дастурларни шахсий компьютерларга ўрнатиш ва кичик таҳлиллардан тил ўқитишда фойдаланиш мумкин бўлган. 3) Учинчи авлод дастурлари: ВордСмит дастурлари, АнтКонк дастурлари бўлиб, фойдаланилиши ва турли хил функцияларга эгаллиги билан юқори даражада сифатли ва кенг имкониятли дастурлар сифатида тан олиш мумкин. 4)Тўртинчи авлод дастурлари: Corpus.byu.edu, SketchEngine дастурлари бўлиб, бу дастурлар ишлаш тезлиги ва йирик ҳажмдаги матнларни қамраб олганлиги билан ажралиб туради. Охирги ва замонавий, тезкор қидирув

тизимига эга дастурлар ҳисобланишса ҳам, лекин фойдаланувчига баъзи чекловлари мавжуд. Масалан, SketchEngine дастури муаллифлик ҳуқуқини тўлиқ ҳимоя қилган ҳолда, бу дастур ҳақида тўлиқ маълумот олиш қийинчилик туғдиради, вебга асосланган дастур бўлгани учун фойдаланувчи шахсий компьютерига юклаб ололмайди, фойдаланиш учун эса ойлик тўловни амалга ошириши керак. Corpus.byu.edu дастурини юклаш ва фойдаланиш бепул бўлса-да, қидирув тизими муайян миқдордан ошганда тўлов сўралади. Юқоридаги тўрт авлод дастурларининг кўпчилиги инглиз тили ёки америка-инглиз тилини ўз тадқиқот марказига олган ёки муайян тил корпусига боғлаб қўйилганки, тадқиқотчи исталган тил қидирувида ишлата олмади. Аммо бу авлод дастурлари ичида А.Лауренсе томонидан яратилган корпус дастурларини очиклиги ва корпусда осонроқ фойдаланиш мумкинлиги жиҳатидан юқори баҳолаш мумкин. АнтКонк дастури тил танламаслиги билан ҳам муҳимдир, яъни турли тил тадқиқотчилари дастурни юклаб олиб, тил ва тил хусусиятларини кузатиши ва статистик ва таҳлилий хулосалар бера олиши мумкин. Бу дастурни ўзбек тили мисолида ҳам қўлладик. А.Лауренсе 2013 йилда эълон қилган мақоласи орқали янги авлод дастури яратилиши, унинг тузилиши ва фаолияти ҳақида маълумот берган. Демак бешинчи авлод дастурининг номи АнтВебконк деб номланиб, уч бутунликни қамраб олади. Бу Model-View-Controller [2], яъни модел – корпус маълумотлар базаси, View – браузер интерфейси, контроллер – текширувчи (масалан конкорданс таҳлили ва ҳ.)дан иборат. Дастур ҳақида берилган маълумотларга кўра дастурлаш иши ва тилини чуқур билмаган исталган тадқиқотчи дастурни тушуниши ва фойдаланиши мумкин. Дастур эълон қилиниши кутилмоқда.

Лексикография бўйича Оксфорд қўлланмасида Изток Косем корпусларда сўровни амалга оширишда, айниқса луғатшунослар учун пухта тайёрланган корпус дастурларига эҳтиёж юқори эканлигини таъкидлайди. И.Косем дастурларни таснифлашда фойдаланилиш ҳолатига кўра 3 та типда гуруҳлайди: мустақил дастурлар, компьютерга асосланган дастурлар, онлайн

дастурлар. Мустақил дастурларни корпусга боғланмаган дастурлар сифатида баҳолайди, фойдаланувчининг компьютерига сақланган бўлади. WordСмит ва МоноКонк Про дастурларини шу гуруҳга киритади. Мустақил дастурлар маълум миқдордаги чекланган корпус маълумотларини қайта ишлайди, шу сабабли бу каби дастурлар замонавий катта ҳажмли корпусларда ўзининг фойдали томонини камайтириши мумкин. Онлайн дастурлар ҳам мустақил ва корпусга боғланган онлайн корпус дастурлари сифатида таснифлаган [1; 77-79]. Онлайн дастурлар сирасига кейинги йилларда энг машҳур ва такомиллаштирилиб борилаётган Sketch Engine дастурини киритиш мумкин. Мазкур дастур кўшимча қулайликлар билан тўлдирилиб бориши билан ҳам аҳамиятлидир. Лекин исталган тадқиқотчи бу дастурга тез ва осон киролмайди, дастурнинг талаб ва чекловлари мавжуд.

Тил корпусларини яратишда ва уларда лингвистик амалларни ижро этишда қўлланилувчи дастурларни синфлашда кузатув, эксперимент ва таққослаш методларини бевосита қўллаган ҳолда таҳлил қилинса мақсадга мувофиқ деб ўйлаймиз, чунки бир вақтда ва бир хил фаолиятни амалга оширувчи дастурларнинг ўхшаш ва фарқли томонларини изоҳлаш, бевосита кузатув ва тажрибани талаб қилади. Дастурларнинг амалда бажарадиган вазифасига ва қўлланилиш ҳолатига кўра гуруҳларга ажратишга ҳаракат қилиб кўрамиз.

Аввало, қўлланилиш ҳолатига кўра корпус дастурлари фойдаланувчининг интерфейсига ўрнатилувчи ёки онлайн ҳолатда фойдаланиладиган корпус дастурлари: 1) Windows, Mac, Linux операцион тизимларига ўрнатиладиган корпус дастурлари: аКонКорде, АННИС, АнтКонк, АНВИЛ, КорпКит, СегментАнт, ТагАнт, ТекстСТАТ каби корпус дастурлари. Масалан, АнтКонк конкорданс дастури Windows учун, Mac учун ҳам, Linux операцион тизимлари учун алоҳида версияларни ишлаб чиққан ва дастурни вебсаҳифа орқали ҳам ёки дастур мавжуд бошқа компьютерлардан ҳам юклаб олиш мумкин. 2) фақат онлайн ҳолатда фойдаланиш мумкин бўлган дастурлар бугунги кунда кўпчиликни ташкил қилади ва бу

дастурларнинг имконият доираси кенглиги туфайли мутахассис корпус имкониятларини тўлароқ ўргана олади: АМАЛГАМ, БНСВеб, КЛик, ХеиделГрам Веб-Басед дастурлари, СкетчЭнжийн, Кортекст Менежер каби дастурлар онлайн ишлайди. Баъзи онлайн корпус дастурлари хусусий тил корпусларининг ўзида жойлашган ва ўша корпус учун махсус яратилган. Масалан, АНК2го дастури Америка миллий корпуси (АМС) учун, БНСВеб Британия миллий корпуси (ВМС) учун яратилган. Аниқ бир тилга боғланган дастурлардан фойдаланиш ҳар доим ҳам эркин ва қулай бўлмайди, исталган тилдаги турли жанрдаги матнларнинг тил феноменини кузатиш имконсиз. Яна бир жиҳати онлайн ишлайдиган корпус дастурлари кенг ҳажмли ва кўп қаватли мураккаб матнларни қамраб олиб, нутқ қисмларини ажратиши (Part-of-speech tagging) ёки токенлаш каби лингвистик таҳлилларни секундлар ичида бажариши мумкин.

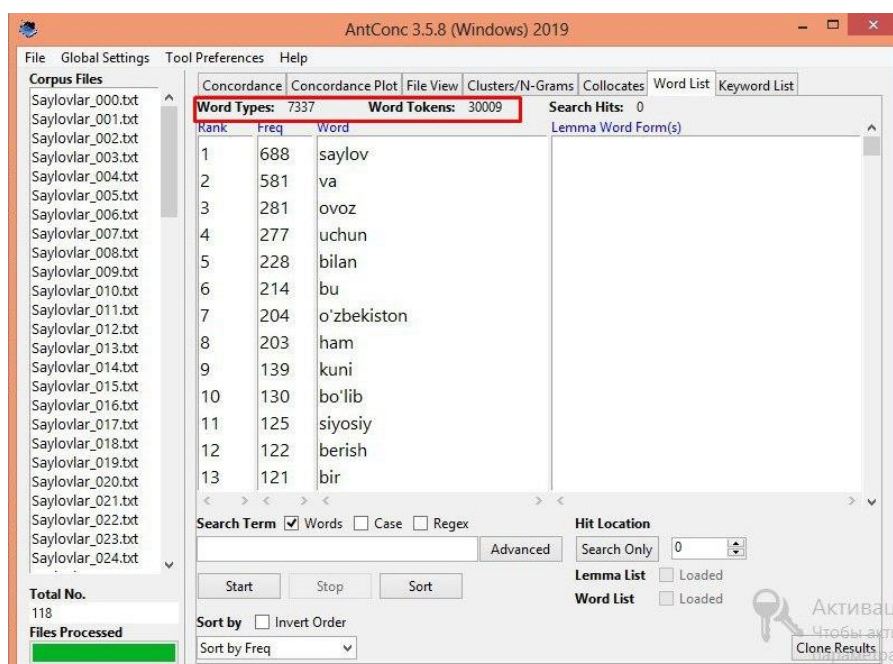
Корпус дастурларини бажарадиган вазифасига кўра: матн йиғувчи, изоҳловчи (аннотацион), конкорданс, статистик таҳлил қилувчи, тегловчи, грамматик таҳлил қилувчи (парсинг), токенловчи, семантик таҳлил қилувчи, аудио матнларни таҳлил қилувчи ва аралаш таҳлилни (mixed) амалга оширувчи дастурлар каби гуруҳларга ажратдик.

1. Матн йиғувчи дастурлар: корпус матнлар тўпламини фойдаланувчи компьютеридаги маълумотлар базасидан ёки интернетдан аниқ бир адрес бўйича ёки умуман турли жанрлар, соҳалар юзасидан тўплаш мумкин. Интернетдан матнларни йиғиб берувчи дастур WebBootCat дастури, СкетчЭнжийн воситасида эса компьютер хотирасида мавжуд ёки бирор манбани скрин қилиб корпус маълумотлар тўпламини яратиш мумкин.

2. Изоҳловчи дастурлар: @аннотате, АМАЛГАМ, Атомик, Дехтер, ПАЛинка дастурлари фақат аннотацияловчи дастур ҳисобланади. Мазкур дастурлар ҳам корпусдаги қўйилаётган масаладан келиб чиқиб изоҳлайди. Масалан, оғзаки корпусларда синтактик таҳлилдан кўра нутқни таниш, просодик хусусиятларини изоҳлаш муҳим ҳисобланса, худди шунингдек корпусда семантик таҳлил кўзланган бўлса семантик аннотацияни ўз ичига

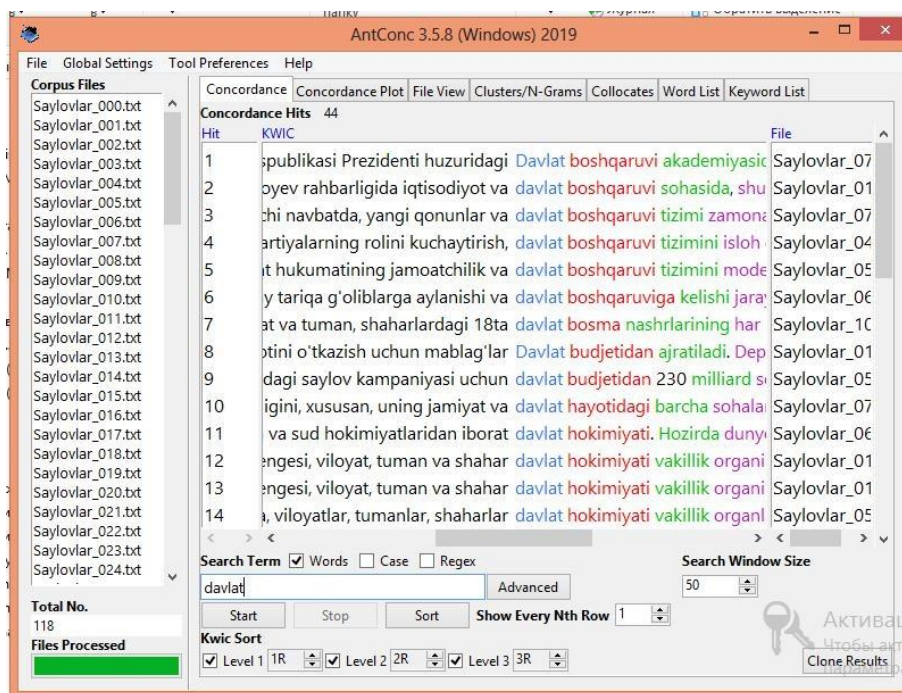
олади [4;75]. Аннотацияловчи дастурлар корпус ҳақида тўлиқ маълумот беради: фойдаланилган матн муаллифи, нашри, олинган манба (китоб, газета, веб-саҳифа), матн ҳажми ва ҳ. Шунингдек, кўшимча матндаги нутқ қисмлари, лемма ва токенлар ҳақида ҳам маълумот келтирилиши мумкин. Лекин Синклер, Лауренсе каби олимлар ҳар бир корпусни аннотациялаш матнларнинг бутунлигини йўқотишга сабаб бўлади деб ҳисоблайди, бошқа томондан эса корпус аннотациялари орқали ўша тилни чуқур билувчи тадқиқотчи матннинг хусусиятларини: замон масаласини, сўз ва гаплар қурилишини, жанр масаласини осонгина таҳлил қила олади.

3. Конкорданс дастурлари: АнтКонк, БНСВеб, Конкордансер, ПараКонк, Вордстатикс дастурлари. Корпус дастурлари ичида оддий, лекин муҳим дастурлардан бири ҳисобланади. Чунки бу дастурлар ҳар бир сўз ва сўз шаклларини контекстда неча марта учраши ҳолатини кўриш ва таҳлил қилиш имконини беради. Конкорданслар корпус лингвистикасида таҳлилнинг муҳим ҳисобланувчи икки шаклини, яъни сифат ва миқдорни аниқлаб беради [3;30]. Конкорданс дастурлари корпусдан маълум узунликдаги сўз, сўз шакли ёки фраземаларни қидиришга имкон беради. Ўзбек тилида хабар мазмунидаги “Сайловлар” номли матнлар корпуси АнтКонк конкорданс дастури орқали таҳлил қилинди.



1-расм. Сўзлар рўйхати

Албатта бу жуда кичик корпусни тажриба синов учун тайёрладик. 1-расмда тасвирланганидек 118 та матн мавжуд бўлиб, 7337 та сўз шакли ва 30009 та токенлар мавжудлиги кўрсатилган. Матнлар хабар сайтларидан олинганлиги сабабли энг кўп учраган сўзлар частотасида 688 марта сайлов сўзи учраганини дастур кўрсатиб турибди.



2-расм. Конкорданси тасвирланган.

Юқоридаги 2-расмда контекстда сўз ва сўз бирикмалари тасвирланган. Айнан сўз иштирок этган контекст орқали тасвирланган. Бу жараён корпус лингвистикасида KWIC, яъни “контекстдаги калит сўз” (key word in context) деб номланади [3;35]

1. Статистик таҳлил қилувчи дастурлар ҳам конкорданс дастурларига ўхшайди. Масалан WordSmith дастури, йирик ҳажмли кутубхоналардаги маълумотлар статистикасини аниқловчи UCS Toolkit дастури ёки икки корпус ҳажми ва сўзлар частотасини текширадиган Log-Likelihood ва Effect-Size Calculator дастурлари мазкур гуруҳга мансуб. Оддий тадқиқотчи учун мураккаблик туғдирадиган корпус деб ҳисобладик.

2. Тегловчи ёки нутқ қисмларини ажратувчи дастурлар мураккаблиги ва ҳар бир тилда ўзига хослиги билан бошқа дастурлардан ажралиб туради. Гапдаги ҳар бир бўлак тегланганда ҳам, яъни кодланганда ҳам бу тег бошқа бир матнда гап мазмуни ва структурасидан келиб чиқиб ўзгариши мумкин.

Бундан ташқари корпусларда қўлланилувчи дастурлар ўз олдига қўйилган сўровига кўра грамматик таҳлилни, семантик таҳлилни, аудио матнлар таҳлилини кузатишда фойдаланилувчи ва аралаш кўп қисмли дастурлар ҳолатида таснифлаш мумкин.

Корпус лингвистикасига оид тадқиқотлар ва манбаларда дастурий таъминотни таснифлашда ҳар хилликни кузатиш мумкин. Масалан, Хумболдт университетида корпус лингвистикаси ва морфологиясига оид вебсаҳифасида дастурий таъминотни онлайн дастурлар, АПИ дастурлаш интерфейси ва фреймворклар, корпус яратувчи дастурлар, аннотацияловчи дастурлар, тегловчилар, корпусни таҳлил қилиш воситалари ва бошқалар каби туркумлаган [5].

Шунингдек дастурларни исталган фойдаланувчи учун очиқ ёки ёпик, пулли (муайян муддатга ёки чекловсиз) ёки текин каби турларга ҳам ажратиш мумкин.

Хулоса қилиб айтганда, тил корпуслари имкониятларини кўрсатувчи дастурий таъминотни ўз ички хусусиятларига кўра яна ҳам чуқурроқ таснифлаш мумкин, аммо бу дегани ҳар қандай турдаги дастурдан корпусларда фойдаланиш мумкин дегани эмас. Одатда ҳар бир миллий тил корпуси учун алоҳида сайтлар яратилиб, унинг ичида таҳлил ва хулосаларни таъминлаб берувчи дастурлар ўрнатилади ёки энг оддийи фойдаланувчи интернетдаги текин ва ўрнатиш қулай бўлган дастурлардан фойдаланишни афзал билади. Назаримизда, корпус дастурлари диққат марказида аниқ бир тил турмаса, тадқиқотчи учун кириш ва фойдаланиш йўриқномаси тўлиқ бўлса, фойдаланувчининг дастурлаш тилини чуқур билишини талаб этмаса ўша дастур энг кўп фойдаланувчига эга ҳисобланади. Чунки бу каби

дастурлар нафақат илмий тадқиқот ишларида, шунингдек тил ўқитиш ва умуман таълим тизимининг турли босқичларида қўлланилиши мумкин.

Фойдаланилган адабиётлар:

1. Kosem I. The Oxford handbook of Lexicography. – United Kingdom, 2016. – 787.
2. Laurency A. A critical look at software tools in corpus linguistics. Linguistic Research, 2013. <https://www.researchgate.net/publication/267631312>
3. McEnery T., Hardie A. Corpus linguistics: Method, theory and practice. Cambridge: Cambridge University Press, 2012. – 312.
4. McEnery T., Xiao R., Tono Y. Corpus-based language studies: an advanced resource book. – Routledge, 2006. – 408.
5. Интернет манба: <https://www.linguistik.hu-berlin.de/en/institut>