

ЎЗБЕКИСТОН РЕСПУБЛИКАСИ
ОЛИЙ ВА ЎРТА МАХСУС ТАЪЛИМ ВАЗИРЛИГИ

ФАРҒОНА ДАВЛАТ УНИВЕРСИТЕТИ

**FarDU.
ILMIY
XABARLAR-**

1995 йилдан нашр этилади
Йилда 6 марта чиқади

4-2021

**НАУЧНЫЙ
ВЕСТНИК.
ФерГУ**

Издаётся с 1995 года
Выходит 6 раз в год

Муассис: Фарғона давлат университети.

«FarDU. ILMIY XABARLAR – НАУЧНЫЙ ВЕСТНИК. ФерГУ» журналі бир йилда олти марта чоп этилади.

Журнал филология, кимё ҳамда тарих фанлари бўйича Ўзбекистон Республикаси Олий аттестация комиссиясининг докторлик диссертациялари асосий илмий натижаларини чоп этиш тавсия этилган илмий нашрлар рўйхатига киритилган.

Журналдан мақола кўчириб босилганда, манба кўрсатилиши шарт.

Ўзбекистон Республикаси Президенти Администрацияси ҳузуридаги Ахборот ва оммавий коммуникациялар агентлиги томонидан 2020 йил 2 сентябрда 1109 рақами билан рўйхатга олинган.

Муқова дизайни ва оригинал макет ФарДУ таҳририят-нашриёт бўлимида тайёрланди.

Таҳрир ҳайъати

Бош муҳаррир
Масъул муҳаррир

ШЕРМУҲАММАДОВ Б.Ш.
ЎРИНОВ А.А.

ФАРМОҢОВ Ш. (Ўзбекистон)

БЕЗГУЛОВА О.С. (Россия)

РАШИДОВА С. (Ўзбекистон)

ВАЛИ САВАШ ЙЕЛЕК. (Туркия)

ЗАЙНОБИДДИНОВ С. (Ўзбекистон)

JEHAN SHANZADAN NAYYAR. (Япония)

LEEDONG WOOK. (ЖанубийКорея)

АЪЗАМОВ А. (Ўзбекистон)

КЛАУС ХАЙНСГЕН. (Германия)

БАХОДИРХОНОВ К. (Ўзбекистон)

ҒУЛОМОВ С.С. (Ўзбекистон)

БЕРДЫШЕВ А.С. (Қозоғистон)

КАРИМОВ Н.Ф. (Ўзбекистон)

ЧЕСТМИР ШТУКА. (Словакия)

ТОЖИБОЕВ К. (Ўзбекистон)

Таҳририят кенгаши

ҚОРАБОЕВ М. (Ўзбекистон)

ОТАЖОНОВ С. (Ўзбекистон)

ЎРИНОВ А.Қ. (Ўзбекистон)

РАСУЛОВ Р. (Ўзбекистон)

ОНАРҚУЛОВ К. (Ўзбекистон)

ҒАЗИЕВ Қ. (Ўзбекистон)

ЮЛДАШЕВ Г. (Ўзбекистон)

ХОМИДОВ Ғ. (Ўзбекистон)

АСҚАРОВ И. (Ўзбекистон)

ИБРАГИМОВ А. (Ўзбекистон)

ИСАҒАЛИЕВ М. (Ўзбекистон)

ҚЎЗИЕВ Р. (Ўзбекистон)

ХИКМАТОВ Ф. (Ўзбекистон)

АХМАДАЛИЕВ Ю. (Ўзбекистон)

СОЛИЖОНОВ Й. (Ўзбекистон)

МАМАЖОНОВ А. (Ўзбекистон)

ИСКАНДАРОВА Ш. (Ўзбекистон)

МЎМИНОВ С. (Ўзбекистон)

ЖЎРАЕВ Х. (Ўзбекистон)

КАСИМОВ А. (Ўзбекистон)

САБИРДИНОВ А. (Ўзбекистон)

ХОШИМОВА Н. (Ўзбекистон)

ҒОҒУРОВ А. (Ўзбекистон)

АДҲАМОВ М. (Ўзбекистон)

ХОНКЕЛДИЕВ Ш. (Ўзбекистон)

ЭГАМБЕРДИЕВА Т. (Ўзбекистон)

ИСОМИДДИНОВ М. (Ўзбекистон)

УСМОҢОВ Б. (Ўзбекистон)

АШИРОВ А. (Ўзбекистон)

МАМАТОВ М. (Ўзбекистон)

ХАКИМОВ Н. (Ўзбекистон)

БАРАТОВ М. (Ўзбекистон)

Муҳаррирлар: Ташматова Т.
Жўрабоева Г.

Мусахҳиҳ: Шералиева Ж.

Таҳририят манзили:

150100, Фарғона шаҳри, Мураббийлар кўчаси, 19-уй.
Тел.: (0373) 244-44-57. Мобил тел.: (+99891) 670-74-60
Сайт: www.fdu.uz

Босишга рухсат этилди:

Қоғоз бичими: - 60×84 1/8

Босма табоғи:

Офсет босма: Офсет қоғози.

Адади: 50 нусха

Буюртма №

ФарДУ нусха кўпайтириш бўлимида чоп этилди.

Манзил: 150100, Фарғона ш., Мураббийлар кўчаси, 19-уй.

Фарғона,
2021.

М.Исамиддинов Қадимги Фарғона ва Суғд кулолчилигини ўрганишнинг айрим долзарб масалаларига доир	72
Б.Усмонов Фарғона водийси Темурийлар даври тарихининг жаҳон темурунослигида ўрганилиши	77
Лианг Юн, Н.Камбаров Қанғ маданияти ҳақида хитойлик олимларнинг фикрлари	81
О.Пуговкина Туркистонда 1917 йил февраль инқилоби: Туркистонда “собик” империя амалдорларига бўлган ижтимоий эҳтиёж ва уларнинг сиёсий фаоллиги	90
А.Ўткиров Хайрия ва ҳомийлик феномени илмий-назарий таҳлили	98
С.Ҳақимова XIX-XX аср боши тарихий манбаларида Ўзбекистон ҳудудидаги сув ҳавзаларига оид маълумотлар	103
С.Анаркулов Тошкент вилояти саноат шаҳарлари аҳолисининг этник тарихи ва таркиби хусусида.....	110

АДАБИЁТШУНОСЛИК

З.Қобилова, Н.Турсунова Татаббулар силсиласи.....	118
П.Туропова Услуб жилолари ёхуд адиб маҳорати	123
Д.Ғуломов Ретроспектив сюжет хусусида айрим мулоҳазалар	128
М.Маҳмудова Заҳириддин Муҳаммад Бобур ҳукмронлиги даврида сув иншоотлари ва боғдорчилик санъати.....	134

ТИЛШУНОСЛИК

Р.Сайфуллаева, Н.Ғайбуллаева Тиббий эвфемизмларнинг социал хосланиши	138
Т.Эназаров, С.Ғаниева Инглиз ва ўзбек тиллардаги аёллар исмларининг келиб чиқиши тарихидан	145
М.Мамажонов Антропонимларни дискурс релеванти сифатида тадқиқ этишнинг лингво-этномаданий асослари	150
Ж.Абдуллаев Тил қурилишининг асосий бирликлари академик Азим Ҳожиёв талқинида	155
Д.Собирова Ўзбек болалар шеърояти морфопоэтикаси	160
Д.Шокирова, Ш.Шокиров Инглиз тилида “property-мулк” концепти тадқиқи.....	164
Н.Абдурахмонова Ўзбек тили электрон корпуси учун матнлар репрезентативлиги масаласи	169
А.Мусоева Self-нинг Ўзбекистон таълим тизимида жорий этилиши билан боғлиқ муаммолар.....	174

ЎЗБЕК ТИЛИ ЭЛЕКТРОН КОРПУСИ УЧУН МАТНЛАР РЕПРЕЗЕНТАТИВЛИГИ
МАСАЛАСИTHE ISSUE OF REPRESENTATION OF TEXTS OF ELECTRONIC CORPUS OF THE
UZBEK LANGUAGEПРОБЛЕМА РЕПРЕЗЕНТАТИВНОСТИ ТЕКСТОВ ДЛЯ ЭЛЕКТРОННОГО КОРПУСА
УЗБЕКСКОГО ЯЗЫКААбдурахмонова Нилуфар Зайнобиддин қизи¹¹Абдурахмонова Нилуфар Зайнобиддин қизи – Тошкент давлат ўзбек тили ва адабиёти университети, доцент, PhD.

Аннотация

Мақолада ўзбек тили электрон корпуси (<http://uzbekcorpus.uz>) учун матнлар репрезентативлиги масаласи ўрганилган. Корпус учун матн танлашнинг статистик усули, ҳажми, миқдори каби омиллар корпус лойиҳалашнинг муҳим мезонларидан бири саналади. Унга кўра матн репрезентативлиги оғзаки ва ёзма матнлар учун турли ҳолатда бўлади. Ёзма матнларнинг графикаси, нормаллаштириш талаблари, матн элементларини компьютер таҳлилининг аниқлиги ва фойдаланиш имкониятининг қулайлиги учун мослаштириш, қолаверса, овозли матнларни транслитерациялаш ҳамда қидирув тизимида сегмент бирликларни унга мос тартибда белгилаш матн репрезентативлигининг асоси ҳисобланади.

Аннотация

Данная статья посвящена вопросу представления текста для электронного корпуса узбекского языка (<http://uzbekcorpus.uz>). Такие факторы, как статистический метод, размер, количество выделенного текста для корпуса являются одними из важных критериев дизайна корпуса. В соответствии с этим репрезентативность текста различна для устных и письменных текстов. Графика письменных текстов, требования стандартизации, адаптация текстовых элементов для точности и удобства использования компьютерного анализа, а также транслитерация аудиотекстов и присвоение сегментных единиц менеджеру корпуса в соответствующем порядке – основа репрезентативности текста.

Annotation

This article is devoted to the issue of text representation for the Uzbek language electronic corpus (<http://uzbekcorpus.uz>). Factors such as the statistical method, size, quantity of text selection for the corpus are one of the important criteria of the corpus design. According to this, the representativeness of the text is different for oral and written texts. Graphics of written texts, standardization requirements, adaptation of text elements for the accuracy and convenient of use of computer analysis, as well as transliteration of audio texts and assignment of segment units to the corpus manager in the appropriate order are the basis of text representativeness.

Таянч сўз ва иборалар: ўзбек тили корпуси, корпус менеджер, матн репрезентативлиги, лингвистик аннотациялаш.

Ключевые слова и выражения: корпус узбекского языка, менеджер корпуса, текстовое представление, лингвистическая аннотация.

Key words and expressions: Uzbek language corpus, corpus manager, text representation, linguistic annotation.

Корпус лингвистикаси электрон ҳисоблаш техникаси асосида сўнги ўн йилликда шаклланган мураккаб лингвистик соҳа ҳисобланади. У лингвистик корпусларни яратиш, маълумотларни қайта ишлаш усуллари ҳамда уларни яратиш ва қўллаш технологияларини ўрганади. Корпус бир нечта тилдаги электрон шаклдаги матнлар асосидаги қидирув тизими ҳисобланади [1, 125].

Дунё тажрибасида кўришимиз мумкинки, корпуснинг машина ўқий оладиган форматга келгунига қадар қатор

соҳаларда корпус тамойилларига асосланган тадқиқотлар олиб борилган. Манбаларда қайд этилишича, библиографик маълумотларни тўплаш, лексикографик масалаларда фойдаланиш, диалектларни ўрганиш, тил таълими ҳамда грамматикага доир изланишларда корпус талабларига ўхшаш меъёрларга амал қилинган. Хусусан, луғатларда матн фрагментларидан фойдаланиш, библиографик ва диалектик маълумотларни таснифлашда метамаълумотларни тўплаш ва саралаш, тилга ўқитишда матнларни

жанр ва услуби бўйича ўрганишда корпус яратилиш босқичларининг айрим элементларини кузатиш мумкин.

Адабиётларда корпус (text corpora, corpus), хужжатларнинг катта мажмуаси (large collection of documents) ёки маълумотлар базаси (data set) сифатида синонимик тушунчалар сифатида қўлланилган ҳолатлар ҳам учрайди. Бизнингча, бундай ёндашувга тўлиқ қўшилиб бўлмайди, гарчи барчасида матн бўлса ҳам, лекин корпуснинг имкониятлари икки турдаги матнлар мажмуасидан анча кенг ва фарқли.

Дунё корпус лингвистикасига эътибор қаратсак, корпусларнинг турли белгиларига кўра фарқланишини унинг лингвистик билимлар базасида таркиб топган манбалар ҳам белгилаши мумкинлигининг гувоҳи бўламиз.

Корпусни кимлар ва қандай қилиб яратдилар, деган савол барчани қизиқтиради. Э.Мелконян бу борада корпусни илмий-тадқиқот гуруҳлари ёки ташкилотлар, шунингдек, ушбу соҳага алоқадор муатахассис корпус яратиши мумкинлиги ҳақида фикрларни илгари сурган [2.72].

Корпусни яхлит шаклда яратиш ва ундан фойдаланиш учун қуйидагилар талаб этилади [3.125].: 1) етарлича катта ҳажмга эга бўлиши; 2) турли типдаги маълумотларни ўзининг табиий контекстдаги шаклида корпусда берилиши; 3) бир марта тайёрланган ва яратилган маълумотлардан кўп мартаба фойдаланиш имкониятининг мавжудлиги.

Корпуснинг матн контенти ва унинг структурасини лойиҳалаш аудитория фойдаланувчиларининг мақсадига мослаштирилади. Шу ўринда тилнинг имконияти ҳисобга олинса, энг катта ҳажмдаги корпус ҳам тилни мукамал даражада ҳар жиҳатдан ўрганиш учун асос бўлмаслиги мумкин.

Ўзбек тили корпусини яратишда лингвистик ресурслар сифатида биз қуйидагилардан фойдаландик: web саҳифалари, луғатлар (бир тилли ва икки тилли), ўзбек тилининг морфологик базаси, бадиий адабиётлар. Корпус структурасини шакллантириш учун мувозанатлашган корпус тамойилларига асосланildi. Яъни олинган матнларнинг жанри, яратилган

даври, мавзуси ва услуби пропорционал тарзда тенг тақсимланди. Лингвистик ресурс сифатида олинган манбаларнинг оригинал ҳолатини сақлаган ҳолда ўзбекча матнларнинг ҳам кириллча, ҳам лотин ёзувидаги шакллари базага киритилди. Шу тариқа икки хил графемадан фойдаланилгани боис, луғатларни ҳам айнан икки ёзувда киритишга муваффақ бўлинди.

Корпуснинг асосий лингвистик базаси оғзаки ва ёзма матнлар ҳисобланади. Корпуснинг ёзма матнлари ҳақида ишимизнинг аввалги қисмларида маълумотлар келтириб ўтдик. Корпус учун оғзаки матн намуналарининг маълумотлар базасига киритилиши учун муайян оғзаки нутқ жараёнида ёзиб олинган аудио форматдаги матнлар транскрипция ёки транслитерация қилинади. Шунингдек, корпуснинг мультимодал тури ҳам мавжуд бўлиб, унда нафақат аудио, балки мулоқот жараёни акс этган видеолавҳа ҳам берилади. Бунда кишиларнинг нутқ жараёнида вербал ва невербал коммуникацион воситаларидан (кўз, қош, юз ва бошқа) нечоғлик нутқ жараёнида қўллаши акс этган видеолавҳалар орқали тил ўрганувчига маълумот беришга хизмат қилади.

Корпуснинг репрезентативлиги матнларнинг миқдор жиҳатдан етарли эканлиги ва жанр жиҳатдан хилма-хиллиги билан белгиланади. В.П.Захаров ва С.И.Богданованинг фикрича [4.16]., корпуснинг жанрий-мавзувий тузилиши корпус матни учун қандай манбалар танлашига эътибор қаратиш лозим.

Мутахассислар корпус тузишнинг технологик жараёнида қуйидаги босқичларни ажратишади [5.15].: 1) белгиланган манбага мувофиқ ҳолда матннинг корпусга киришини таъминлаш; 2) матнни автоматик ўқилиш шаклида қайта ишлаш. Корпусга киритиладиган электрон шаклидаги матн турли усул билан олинган бўлиши мумкин: қўлда киритилган, сканерланган, муаллифлик нусхаси, Интернет, нашриётлар томонидан корпус тузувчисига бериладиган оригинал намуналар; 3) таҳлил ва матнга дастлабки ишлов бериш. Ушбу босқичда турли манбалардан қабул қилинган матнлар филологик текширув ва таҳрирдан ўтади; 4)

ТИЛШУНОСЛИК

конверсиялаш ва графематик таҳлил. Дасталаб баъзи матнлар учун қайта кодлаштириш жараёни амалга ошадиган машина ишловидан қайта-қайта ўтади, матн бўлмаган қисмлар (расм, жадвал) ўчирилади ёки ўзгартирилади. Матндаги бўғин кўчириш, чегаралар (MS-DOS матнларида) бекор қилинади, тире ва бошқа белгиларнинг бир хиллигига эришилади. Графематик таҳлил корпусга кирувчи матнни қисм (сўз ва боғловчи)га ажратиш, номатний элементни ўчириш каби амалларни бажаришдан иборат; 5) ностандарт (нолексик) элементни белгилаш ва расмийлаштириш, махсус матний элементни (қисқартма асосида ёзилган ном (исм, фамилия), бошқа алифбода ёзилган ўзлашма лексема, расмга берилган ном, изоҳ, сарлавҳа, адабиётлар рўйхати ва бошқалар) бир хил мезон асосида қайта кўриб чиқислади. Албатта, бу амаллар автоматик равишда матн муҳаррири томонидан бажарилади; 6) корпусни лойиҳалаштиришнинг кейинги босқичида манба сараланади. Корпуснинг аҳамияти унинг бир тилдаги кенг кўламли матнларни бир жойга йиғилиб, тартиб берилганлигида эмас, шу сабабли уни тузишда бир неча мезон асосида иш кўрилишидадир.

Даставвал корпус материални саралашда корпуснинг асосий бирлиги нимадан иборатлиги, унинг ҳажми қандай бўлиши (унда қанча сўз бўлгани маъқул), ёзма матн қайси манбага асосланиши ва қанча миқдорда бўлиши, унга кирувчи матн тилнинг қайси соҳасига тегишли бўлиши каби масалаларга ечим топилади. Ушбу масала илк маротаба 1965-1980-йилларда Р.Г.Пиатровский ва унинг шогирдлари томонидан самарали тарзда муҳокама қилиниб, ҳал этилган эди. Улар частотали луғат ва лингвостатистик тадқиқот ўтказиш учун матн танлаш тамойилларини тузиб чиқишган. Бу тадқиқот эса корпус манбаларини саралашда кейинчалик муҳим омил бўлиб хизмат қилган. Бу муаммо бошқа изланувчиларнинг ҳам қизиқишига сабаб бўлган. Масалан, Л.Н.Засорина таҳрири остидаги частотали луғат сўзбошисида ҳам бунга тўхталган. Ўшанда илк бор матн танлашнинг статистик усули, ҳажми, миқдори каби омиллар санаб ўтилган. Корпуснинг асосий бирлиги сўз шакл, ўзак (негиз, лемма) ва гапни ташкил этади. Тузиладиган корпус ҳажми эса

корпуснинг мақсадидан келиб чиқиб белгиланади. Агар корпус кўшимчаларни тадқиқ қилиш учун тузилган бўлса, унда ҳажмнинг йирик бўлиши талаб этилмайди ёхуд, аксинча, синтактик структуралар таҳлили назарда тутилса, тилнинг мавжуд имкониятларини очиб бериш мақсадида корпус катта ҳажмда бўлиши лозим. С.А.Шаровнинг фикрича[6], саралаш жараёнида қайси жанрга оид матн (наср, драма, шеърят, илмий матн, газета, журнал материали ва ҳ.к.) танлаш, матннинг қайси даврни (замонавий, 10 йиллик, 50 йиллик ва мумтоз матни) қамраб олиши, матн фақат адабий тилда бўлиши ёхуд бошқа манбалар ҳам кириши каби масалалар ҳам муҳим аҳамият касб этади. Жаҳон тилшунослиги тажрибасида бундай ҳолатларда корпус яратувчилари сўровнома ўтказишади, анкета методини қўллашади ва натижаларни соҳа мутахассислари билан муҳокама қилишади.

Матнни корпусга тайёрлаш борасида матнларни нормаллаштириш босқичи жараёни муҳим саналади. Дунё тилларининг грамматик жиҳатдан ўзига хослиги корпус анализида, айниқса, токенизация жараёнида алоҳида методологик ёндашувни талаб қилади. В. Захаров барча тилларнинг токенизация жараёнида қуйидаги муаммолар учраши мумкинлиги ҳақида тўхталиб ўтади[7.16].:

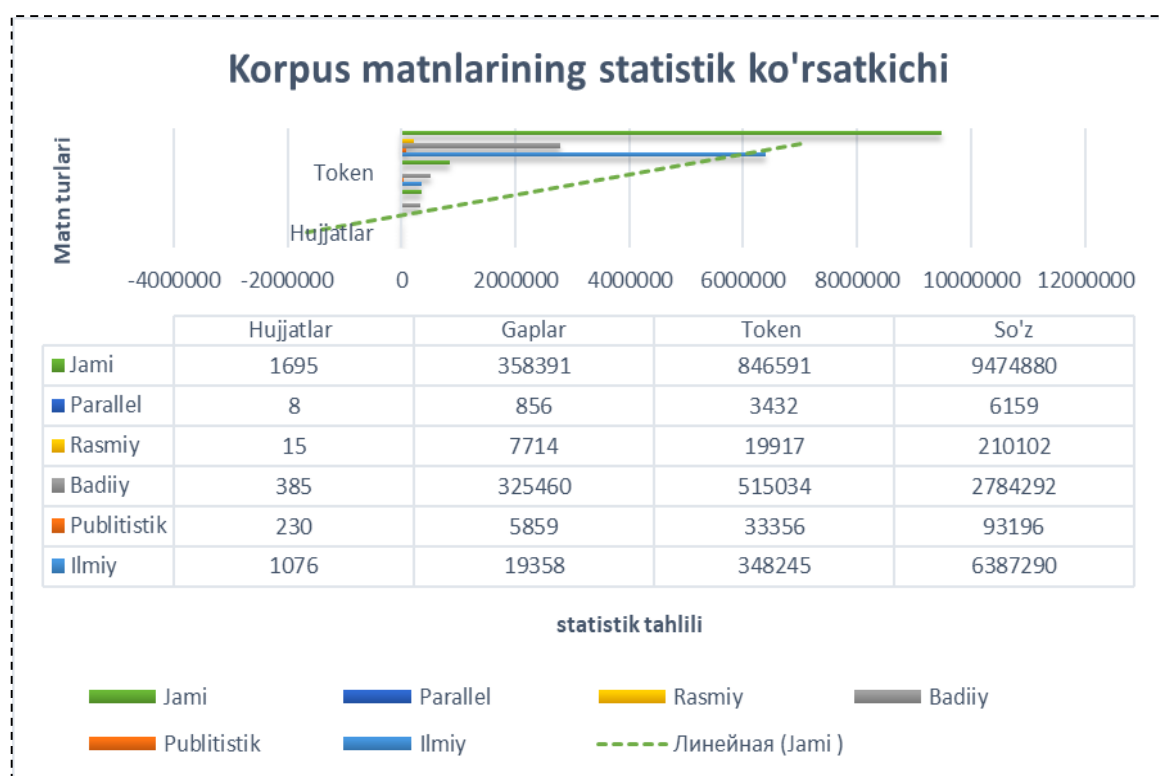
– Олим токенлаш жараёнида учрайдиган айрим муаммолар сифатида сўзларни нормаллаштириш босқичига тўхталиб ўтади. Таъкидланишича, матндаги барча сўзларни умумий қаторга келтириш керакми ёки турдош отлардан атоқли отларни фарқлаш учун биринчи бош ҳарфларни қолдириш керакми, бошқа бир томондан, бош ҳарфлар билан ёзилган сарлавҳаларни кичик ҳарфларга ўтказиш, сарлавҳалар ёки матндаги бўш жойларни олиб ташлаш, сўзларни бирлаштириш тўғри ечим бўла оладими? Ушбу матннинг нормаллаштириш усули муайян даражада сўзларни морфологик жиҳатдан аннотациялаш учун муҳим аҳамият касб этади. Чунки турдош отдан атоқли отга (ўткир –Ўткир, лола – Лола) ёки атоқли отдан турдош отга ўтган сўзлар (Рентген – рентген)да ушбу масалани муайян даражада ечиш имконияти бор. Матн бошида келган сўзларнинг ҳолатини мана шу усулда ҳал қилиш тўғри бўлади. Шунингдек, бир сўзда

икки графеманинг (алифбонинг) қўлланилиши (call-centre)ни нормаллаштириш ёки тилда қандай мавжуд бўлса, шу ҳолда қолдириш керакми, деган масалага тўхталиб ўтади.

Назаримизда, бу каби тил ва нутқ бирликларини шу ҳолда қолдириш тўғри бўлмайди. Масалан, тилга янги ўзлашиб келаётган бирлик калькалаштирилган тақдирда ҳам тилнинг мавжуд лингвистик унсурларидан фойдаланиш керак. Акс ҳолда корпус менеджерининг қидирув тизимида муайян даражада муаммога дуч келиш мумкин. Бироқ ўзбек тилида ҳам шу каби сўзлар сақланиб қолган (бебсайт Википедия каби). В.Захаров ҳарф ва рақамдан иборат турли бирикмаларни ҳам муайян даражада нормаллаштириш лозимлигини таъкидлайди. Бу борада

элементлари сифатида киритилганда, дастурлар томонидан матннинг элементи сифатида қабул қилиниши мумкин. Масалан, лотин графикасида И, ИИ, ИИИ деб ёзиш дастур учун {ъиъ} тарзда ўқишга олиб келиши мумкин.

Корпус структураси турлича бўлиши мумкин. Тадқиқотчилар корпус соҳасида корпус контенти, яъни жанр мезонлари ва тартиби бўйича чуқур таҳлиллар олиб борганлар. Таҳлиллар натижасида [8.15] жанрларни мавзу жиҳатдан структурлашда айрим масалаларга тўختалади. Жумладан, улар илмий тўпламда берилган аннотацияларни алоҳида матн сифатида олиш керакми ёки матн билан бирлаштириш муҳимми, деган саволни ўртага ташлайди. Бизнингча, аннотацияни алоҳида матн сифатида олиш у даражада мантиқий эмас.



тилнинг маълумотлар базасидаги лингвистик ресурсга боғланиши мумкин бўлган моделлар асосида текшириладиган вақтда ушбу масалани алоҳида ўрганиш талаб қилинади. Масалан, ўзбек тилида 1-дан, 100 минглаб, 4-си каби бирликлар матнда учрайди. Ушбу вазиятда сон+[морфема] моделини алоҳида маълумотлар базасига киритиш лозим. Олим рим ёки араб рақамлари, лотин ёки кирилл ҳарфлари ёрдамида рўйхат

Чунки аннотация ҳам сарлавҳа остида мақолага бирлашади. Матн мазмуни жиҳатидан асосий матн шакли ва ғоясини такрорлайди. Терминологик жиҳатдан мақола матнидаги лексикондан умуман фарқ қилмайди. Демак, уни алоҳида қилиш шарт бўлмайди.

Кейинги ўринда газетада бериладиган у даражада катта бўлмаган эълон матнлари алоҳида матнлар сифатида берилиши керакми ёки матнларни

ТИЛШУНОСЛИК

бирлаштириш тўғри бўладими, деган асосли масалани кўтаради. Тадқиқотларимиздан шу нарса аён бўлдики, газета бериладиган ҳар бир матн алоҳида муаллифга тегишли. Шунингдек, бериладиган маълумотлар мавзу кўлами турлича. Гарчи айни даврий нашрда эълон қилинган мақолаларни ўзаро солиштириш учун ҳам уларни алоҳида бериш мақсадга мувофиқ бўлса-да. Бундан ташқари, тематик рефератив журналлар асосида корпус яратиладиган бўлса, аннотацияни алоҳида матн сифатида олиш керакми ёки тематик бўлимни аннотация билан бирга базага киритиш лозимлиги ҳақидаги фикрни ўртага ташлайди. Ўзбек тилининг электрон корпусини яратиш асносида юқоридаги талабларнинг барчаси инобатга олинди. Электрон корпус контентининг статик кўрсаткичи.

Лингвистик маълумотлар базасида ўзбек тилининг морфологик базаси орқали корпусда мавжуд сўзларнинг турли лингвистик маълумотларининг турли жиҳатдан статистикасини чиқариш имконияти туғилади. Шунингдек, корпуснинг қидирув тизими орқали ихтиёрий сўзнинг г-грам модели бўйича қўлланилиши натижасида сўзларнинг конкорданслари ҳосил бўлади.

Корпус технологияси матнларнинг репрезентативлиги лексикографияда лексик маълумотлар базасининг, яъни базага кирган лексиконни сўз туркумлари бўйича

теглаш, парсинг (синтактик таҳлил), сўз маънолари, семантик кўпмаънолилик каби НЛПнинг турли вазифаларини бажаришда калит сўз вазифасини бажаради; грамматика ва лексикон ўртасида турувчи бирлик сифатида кўп таркибли бирикмаларнинг назарий масалаларини ўрганишда йўналиш беради; корпус маълумотларидан кўп таркибли бирикмаларни идентификациялаш ва қидириш каби вазифаларни бажаришда муҳим аҳамият касб этади

Тадқиқотимиз натижаси сифатида <http://uzbekcorpus.uz/> да яратилган корпус тадқиқотларда ўрганилган мезонлар бўйича ҳавола қилинади.

Корпуснинг қидирув тизими сифатида сўзларнинг беҳато эканлигини тасдиқлаш ёки унинг хатолигини унга яқин, эҳтимолга яқин сўзлар базасидан қидириш корпус технологияси учун муҳим саналади. Джаро Винклер алгоритми ва ФСТ технологияси матнга жорий этиш орқали эҳтимолга асосланган метриканинг ўхшаш сатрларни ўзаро солиштириш ва энг яқин натижани ҳавола этиш имкониятининг яратилишида омил вазифасини бажаради.

Хулоса сифатида корпусни лойиҳалаштириш жараёнида материал (матн) танлаш, саралаш, уни техник жиҳатдан корпусга мослаштириш энг асосий босқич эканлигини қайд этиш жоиз

Адабиётлар:

1. Карпович С.Н. Русскоязычный корпус текстов scdm-ru для построения тематических моделей / Труды СПИИРАН. 2015. Вып. 2(39). ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн) www.proceedings.spiiras.nw.ru,
2. Мелконян Э. Соотношение корпусной лингвистики и типологии / Proceedings of the international conference «CORPUS LINGUISTICS–2019»
3. Карпович С.Н. Русскоязычный корпус текстов SCTM-RU для построения тематических моделей труды СПИИРАН. 2015. Вып. 2(39). ISSN 2078-9181 (печ.), ISSN 2078-9599 (онлайн) www.proceedings.spiiras.nw.ru,
4. Захаров В.П., Азарова И.В. и др. Моделирование в корпусной лингвистике: Специализированные корпусы русского языка, ИЗДАТЕЛЬСТВО САНКТ-ПЕТЕРБУРГСКОГО университета, 2019.
5. Kurs "Korpusnaya lingvistika"/ Kutuzov A.B. / Litsenziya Creative commons Attribution Share-Alike 3.0 Unported. - <http://www.ruscorpora.ru>.
6. Sharov S.A. Predstavitelnyy korpus russkogo yazyka v kontekste mirovogo opita. - <https://lamb.viniti.ru>.
7. Захаров В.П., Азарова И.В. и др. Моделирование в корпусной лингвистике: Специализированные корпусы русского языка, ИЗДАТЕЛЬСТВО САНКТ-ПЕТЕРБУРГСКОГО университета, 2019.
8. Захаров В.П., Азарова И.В. и др. Моделирование в корпусной лингвистике: Специализированные корпусы русского языка, ИЗДАТЕЛЬСТВО САНКТ-ПЕТЕРБУРГСКОГО университета, 2019.

(Тақризчи: Ш.Искандарова – филология фанлари доктори, профессор)