

ЎЗБЕКИСТОН ДАВЛАТ ЖАҲОН ТИЛЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.03/30.12.2019.Fil/Ped.27.01 РАҚАМЛИ ИЛМИЙ КЕНГАШ

ЎЗБЕКИСТОН ДАВЛАТ ЖАҲОН ТИЛЛАРИ УНИВЕРСИТЕТИ

АТАБОЕВ НОЗИМЖОН БОБОЖОН ЎҒЛИ

ИНГЛИЗ ТИЛИ КОРПУСНИНГ ФУНКЦИОНАЛ ХУСУСИЯТЛАРИ
(СОСА мисолида)

10.00.04 – Европа, Америка ва
Австралия халқлари тили ва адабиёти

Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси
АВТОРЕФЕРАТИ

ТОШКЕНТ – 2020

Фалсафа доктори (PhD) диссертацияси автореферати мундарижаси

Оглавление автореферата диссертации доктора философии (PhD)

Contents of dissertation abstract of Doctor of Philosophy (PhD)

Атабоев Нозимжон Бобожон ўғли

Инглиз тили корпусининг функционал хусусиятлари

(СОСА мисолида)..... 3

Атабоев Нозимжон Бобожон угли

Функциональные характеристики английских корпусов

(на примере СОСА)..... 27

Ataboev Nozimjon Bobojon ugli

Functional Characteristics of the English Corpora

(on the example of СОСА)..... 51

Эълон қилинган ишлар рўйхати

Список опубликованных работ

List of published works..... 55

ЎЗБЕКИСТОН ДАВЛАТ ЖАҲОН ТИЛЛАРИ УНИВЕРСИТЕТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
DSc.03/30.12.2019.Fil/Ped.27.01 РАҚАМЛИ ИЛМИЙ КЕНГАШ

ЎЗБЕКИСТОН ДАВЛАТ ЖАҲОН ТИЛЛАРИ УНИВЕРСИТЕТИ

АТАБОЕВ НОЗИМЖОН БОБОЖОН ЎҒЛИ

ИНГЛИЗ ТИЛИ КОРПУСНИНГ ФУНКЦИОНАЛ ХУСУСИЯТЛАРИ
(СОСА мисолида)

10.00.04 – Европа, Америка ва
Австралия халқлари тили ва адабиёти

Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси
АВТОРЕФЕРАТИ

ТОШКЕНТ – 2020

Фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида В2019.1.PhD/Fil723 рақам билан рўйхатга олинган.

Докторлик диссертацияси Ўзбекистон давлат жаҳон тиллари университетида бажарилган.

Диссертация автореферати уч тилда (Ўзбек, рус ва инглиз (резюме)) Илмий кенгашнинг веб-саҳифасининг (www.uzswlu.uz) ва «ZiyoNet» Ахборот таълим порталида (www.ziynet.uz) жойлаштирилган.

Илмий раҳбар:

Тешабаева Дилфуза Мўминовна
филология фанлари доктори, профессор

Расмий оппонентлар:

Рахимов Ғанишер Худойкулович
филология фанлари доктори, доцент

Хамроева Шахло Мирджонова
филология фанлари бўйича фалсафа доктори (PhD)

Етакчи ташкилот:

Бухоро давлат университети

Диссертация химояси Ўзбекистон давлат жаҳон тиллари университети ҳузуридаги DSc.03/30.12.2019.Fil/Ped.27.01 рақамли Илмий кенгашнинг 2020 йил «11» ноябрь соат 14⁰⁰ даги мажлисида бўлиб ўтади. Манзил: 100138, Тошкент шаҳри, Учтепа тумани, Кичик ҳалка йўли кўчаси, 21 А уй. тел: (99871) 230-12-91. факс: (99871) 230-12-92; (99871) 230-12-91; факс: (99871) 230-12-91; email:uzswiu_info@mail.ru)

Диссертация билан Ўзбекистон давлат жаҳон тиллари университетининг Ахборот-ресурс марказида танишиш мумкин (12 рақами билан рўйхатга олинган). Манзил: 100138, Тошкент шаҳри, Учтепа тумани, Кичик ҳалка йўли кўчаси, 21 А уй. тел: (99871) 230-12-91. факс (99871) 230-12-92.

Диссертация автореферати 2020 йил «29» октябрь куни тарқатилди.

(2020 йил «29» октябрь да 12 рақамли реестр баённомаси).

Г.Х. Бокиева
Илмий даражалар берувчи Илмий кенгаш
раиси, филол.ф.д., профессор

Х.Б. Самигова
Илмий даражалар берувчи Илмий кенгаш
котиби, филол.ф.д., доцент

Ж.А. Якубов
Илмий даражалар берувчи Илмий кенгаш
кошидаги Илмий семинар раиси, филол.ф.д.,
профессор



КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

Тадқиқот мавзусининг долзарблиги ва зарурати. Жаҳон тилшунослигида ўтган асрнинг 60-йилларида лисоний корпусларни яратиш ва уларнинг татбиқи билан шуғулланувчи – корпус лингвистикасининг компьютер технологиялари билан ўзаро алоқаси жадаллашди. XIX асрнинг бошларида тилшунослик соҳаси корпус лингвистикасининг назарий тамойиллари шаклланиши натижасида илк намуна – Бровн корпуси яратилди. Натижада, инсониятнинг интеллектуал бойлигини электрон корпус асосида автоматик рақамлаштириш замонавий тилшуносликнинг бугунги кундаги долзарб вазифалари сирасига киритилди. Шу сабабли тил ёки тилнинг муайян қисмини ўзида ифода этган, ёзма ва оғзаки матнларни умумлаштирган электрон лисоний база – корпус таҳлил методларининг амалий ифодаси бўлган миллий корпуслар тил ривожини, лексик қатлам бойиши ва унинг бугунги кунини хусусидаги статистик таҳлилини аниқ сонларда такомиллаштириш долзарб аҳамият касб этади.

Дунёда лисоний тадқиқотлар сифати статистик таҳлилларга асосланар экан, соҳада тил корпусларини яратиш амалиётининг самарали натижаси матн таҳлили жараёнида сарфланадиган вақтнинг тежалишига имкон яратади. Лингвистик бирликларнинг хусусиятлари, уларнинг нутқ таркибидаги хоссаларига оид минглаб мисолларнинг конкордансда аниқланиши корпуснинг бирламчи хусусияти саналади. Шу жиҳатдан XXI асрнинг учинчи ўн йиллигига қадам қўяётган илм-фан компьютерларни амалиётда қўллаш заруратини қўяр экан, корпус лингвистикасининг илмий-назарий тадқиқи барча соҳаларда долзарблик касб этади.

Мамлакатимизда бугунги кунда кенг қўламли ислоҳотлар амалга оширилиб, давлатимизнинг дунё ҳамжамиятига интеграциялашуви чет тилларни мукамал эгаллаган мутахассис кадрларни тайёрлаш каби муҳим вазифаларни белгиламоқда. “Тараққиётга эришиш учун рақамли билимлар ва замонавий ахборот технологияларини эгаллашимиз зарур ва шарт. Бу бизга юксалишнинг энг қисқа йўлидан бориш имкониятини беради. Зеро, бугун дунёда барча соҳаларга ахборот технологиялари чуқур кириб бормоқда”¹. Шундай экан тадқиқот олиб борилаётган соҳа замонавий илм-фан ривожини сифатида қаралгани боис, уни мамлакатимиз тараққиёти йўлида салмоқли ўзлаштириш бугунги кун талабини ифода этади. Бунда корпус лингвистикасининг тарихий ривожланиш эволюцияси, корпус, уни таснифлаш, қўллаш принциплари таҳлили асосида инглиз тили мавжуд корпусларининг функционал хусусиятларининг назарий тадқиқини ишлаб чиқиш каби муҳим вазифалар белгиланди.

Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги ПФ-4947-сон «Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида»ги фармони, 2019 йил 8 октябрдаги «Ўзбекистон Республикаси олий таълим тизимини 2030 йилгача

¹ Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг Олий Мажлисга Мурожаатномаси. – President.uz сайти, 2020 йил 24 январь. <https://president.uz/uz/lists/view/3324> [мурожаат санаси: 17.02.2020]

ривожлантириш концепциясини тасдиқлаш тўғрисида»ги ПФ-5847-сон фармони, 2017 йил 11 августдаги 610-сон «Таълим муассасаларида чет тилларини ўқитишнинг сифатини янада такомиллаштириш чора-тадбирлари тўғрисида»ги, 2018 йил 5 июндаги ПҚ-3775-сон «Олий таълим муассасаларида таълим сифатини ошириш ва уларнинг мамлакатда амалга оширилаётган кенг қамровли ислохотларда фаол иштирокини таъминлаш бўйича қўшимча чора-тадбирлари тўғрисида»ги қарорлари, Ўзбекистон Республикаси Президентининг 2020 йилнинг 24 январидagi Олий Мажлис палаталарига Мурожаатномаси ҳамда мавзуга оид бошқа меърий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишда ушбу диссертация тадқиқоти муайян даражада хизмат қилади.

Тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига мослиги. Диссертация республика фан ва технологиялар тараққиётининг I. «Ахборотлашган жамият ва демократик давлатни ижтимоий, ҳуқуқий, иқтисодий, маданий, маънавий-маърифий ривожлантиришда инновацион ғоялар тизимини шакллантириш ва уларни амалга ошириш йўллари» устувор йўналишига мувофиқ бажарилган.

Муаммонинг ўрганилганлик даражаси. Жаҳон тилшунослигида Г. Лич биринчилардан бўлиб корпус лингвистикасининг пайдо бўлиши, тараққиёти ҳақида илмий изланиш олиб борди². Замонавий корпус лингвистикаси шаклланишига доир тадқиқотлар С. Хюстон, Д. Бибер, С. Грайс, М. Ҳаллидай ва В. Теуберт каби инглиз олимлари томонидан амалга оширилган³. Рус олимлари В. Баранов ва В. Захаров ишларида корпус лингвистикаси спецификасига, Н. Владимов, С. Вадяев, О. Горкина, В. Мамонтова ва А. Магаметова каби олимлар «корпус лингвистикаси», «лисоний корпус» тушунчаси ва уни таснифлаш муаммоларига атрофлича ёндашганлар⁴. Тожикистонлик олима Д. Джаъфарова тадқиқотида корпус лингвистикасининг компьютер лингвистикаси ва бошқа ўзаро ўхшаш соҳалар билан қиёсий таҳлили амалга оширилган⁵.

² Leech G. Corpora And Theories Of Linguistic Performance. In J.Svartvik (ed.). 1992. – 105–122 pp.

³ Hunston S. Corpora in Applied Linguistics: Cambridge University Press, 2002. – 234 p.; Halliday M.A., Teubert W., Yallop C. and Cermakova A. Lexicology and Corpus Linguistics, – London, 2004. – 184 p.; Gries S.Th. Introduction to S. Gries and A. Stefanowitsch (eds.). Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis. – Berlin: Heidelberg, New York: Mouton de Gruyter, 2006. – 1-18 p.; Douglas B., Ulla C. and Thomas A.U. Discourse On The Move. – Amsterdam, 2007. – 290 p.

⁴ Баранов А.Н. Компьютерная лингвистика // Введение в прикладную лингвистику: учебное пособие. – М.: Едиториал УРСС, 2003. – 320 с.; Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие / В.П. Захаров. – СПб, 2005. – 48с. – Режим доступа:<http://download.yandex.ru/class/zakharov/plan.pdf>; Владимов Н.В. Корпусный подход к решению переводческой проблем (на материале английского и русского языках). Дис...канд. филол. наук. – Москва: РГБ, 2005. – С 26.; Вадяев С.Е. Лингвистические принципы построения и использование корпуса текстов для Исследования официально-делового стиля современного немецкого языка(на материале электронного корпуса «DER») диссер... для кандидата филол. наук. – Самара, 2005. – 160 с.; Горина О.Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке. Дисс. ... канд. пед. наук. – Москва, 2014. – 321с.; Мамонтова В.В. Особенности перевода сложносоставных слов с английского языка на русский (на материале корпуса публицистических текстов). Дис. ... канд. фил.наук. – Ставрополь: 2008. – 139 с.; Магомедова А.Н. Корпусная лингвистика и контекстное разрешение лексической многозначности слов. Дис. ... канд. филол. наук – М.: РГБ, 2005. – 132 с.

⁵ Джаъфарова Д.Ф. Модели лингвистического анализа текстов таджикского языка (на материале газелей Хафиза) дисс. ... канд. фил. наук. – Душанбе, 2013. – 151 с.

Корпус таҳлилларидан луғатшуносликда фойдаланиш борасидаги ишлар А. Килгарриф, К. Изток, Ф. Карлсон ва И. Соженин тадқиқотларида кузатилган бўлса⁶, В. Рыков, А. Романов ва В. Плунгян каби олимлар «корпус» тушунчасини матн таҳлилларида қўллаш методини ишлаб чиқишган⁷.

Ўзбекистонда ҳам бугунги кунда корпус лингвистикаси методларини қўллаш, мавжуд корпуслардан фойдаланиш борасида кенг қўламли илмий тадқиқотлар олиб борилмоқда. Ж. Жумабоева инглиз тили корпусларидан, хусусан, BNC⁸ таркибидан семантик градацияларни топиш, Г. Эргашева BNC корпуси асосида гендерга оид концептуал метафораларни тадқиқ қилиш, Г. Собирова эса BNC корпусини тилларни ўқитишда қўллашга эътибор қаратган. Ш. Ҳамроева муаллифлик корпусини яратиш муаммосига ечим излаган⁹. Шунингдек, ҳозирда ўзбек тилининг миллий корпуси ҳамда ёш ва жисмоний фарқлиликни ифода этувчи қўллаб лисоний корпуслар яратиш бўйича катта ҳажмли лойиҳалар амалга оширилмоқда.

Тадқиқотнинг диссертация бажарилган олий таълим муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация тадқиқоти Ўзбекистон давлат жаҳон тиллари университети илмий-тадқиқот ишларининг “Таржима муаммолари” илмий мавзусининг таркибий қисми бўлиб, таржима назарияси ва амалиёти факультетининг илмий режаларига киритилган.

Тадқиқотнинг мақсади корпус лингвистикасининг тарихий ривожланиш эволюцияси, корпус, уни таснифлаш, қўллаш принциплари таҳлили асосида инглиз тили мавжуд корпусларининг функционал хусусиятларини СОСА корпуси мисолида очиқ беришдан иборат.

Тадқиқотнинг вазифалари:

корпус лингвистикасининг диахроник тадқиқи ва унинг специфик хусусиятларини, корпус терминининг турлича талқин ва тавсифларини изоҳлаш;

⁶ Kilgarriff A. and Kosem I. Corpus Tools For Lexicographers. – Trojina: Institute for Applied Slovene Studies, 2012. – 37 p. УРЛ: https://www.researchgate.net/publication/264884809_Corpus_tools_for_lexicographers; Karlsson F. Lexicography And Corpus Linguistics. – Helsinki: Department of General Linguistics University of Helsinki, 1992. УРЛ: <http://www.ling.helsinki.fi/~fkarlsson/INTJLLEX.pdf>; Kilgarriff A. Using Corpora [And The Web] As Data Sources For Dictionaries. – Helsinki: 2012. – 21 p. https://www.sketchengine.eu/wp-content/uploads/Using_corpora_2012.pdf; Саженин И.И. Корпусные методы в лексикографии: опыт создания модели Словарного корпуса Диссертация на соискание ученой степени кандидата филологических наук Новосибирск — 2013. 193 с.

⁷ Рыков В.В. Корпус текстов новый тип словесного единства Текст. / В.В. Рыков. – Электрон, изд. – Режим доступа к изд.: <http://rykov-cl.narod.ru/t31.html>; Романов А.С. Методика и программный комплекс для идентификации автора неизвестного текста: Автореф. дисс....канд. тех. наук. - Томск, 2010. 26 с.; Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. – Москва: №2., 2008. – С. 7-20.

⁸ BNC (BritishNationalCorpus) – Британия Миллий Корпуси: 100 миллион сўзли лисоний корпус.

⁹ Сабирова Г. Умумтаълим мактаблари ўқувчиларига инглиз тили олмошларини ахборот технологиялари асосида ўргатиш методикаси: Филол. фалс. д-ри (PhD) ...дисс. автореф. – Тошкент: 2017. – 41 б. URL: <http://uzswlu.uz/media/ba7be50d-84f1-c826-f935-6e7610cb6bf5.pdf>; Эргашева Г. Турли тизимдаги тилларда гендерга оид терминологиянинг шаклланишида лингвистик ва экстралингвистик омиллар: Фил.фан.д-ри.дисс. автореф. – Тошкент: 2018. – 61 б. URL: <http://demo.navoiy-uni.uz/sites/default/files/ziyonet.pdf>; Джумабаева Ж. Ўзбек ва инглиз тилларида лексик ҳамда стилистик градуонимия. докторлик диссертацияси автореферати. – Тошкент: 2016. – 82 б. URL: http://tashgiv.uz/wp-content/uploads/2017/01/avtoref_jumabayeva.pdf; Ҳамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Ф.б. фалс. д-ри (PhD) ... дисс. автореф. – Қарши: 2018. – 52 б. URL: <https://e-books.tsue.uz/pdf-ozbek-tili-mualliflik-korpusini-tuzishning-lingvistik-asoslari>

корпус яратиш ва уни қўллашга оид корпус лингвистикаси илмий-тадқиқот методологиясини такомиллаштириш;

экспериментал корпус яратиш орқали корпус лингвистикасининг тил хусусиятларини ёритиш;

корпус таҳлиллари натижасида корпус лингвистикасига оид частотали-изоҳли терминологик луғат яратиш;

мавжуд инглиз тили корпусларининг амалий аҳамиятини очиб бериш;

СОСА¹⁰ корпусининг яратилиш тарихи, унинг таркибий тузилиши, у тақдим этиши мумкин бўлган семантик, синтактик ва морфологик тадқиқотларга оид маълумотлар таҳлилинини келтириш;

СОСА корпусидан олинган маълумотларнинг турли тил соҳибларининг лингвомаданий тушунчаларни англашидаги аҳамиятини кўрсатиб бериш;

СОСА корпусининг луғат корпуси сифатида тақдим этиши мумкин бўлган лисоний статистик маълумотларнинг салмоғини аниқлаш.

Тадқиқотнинг объекти сифатида инглиз тили корпуслари, хусусан, замонавий америкача инглиз тили корпуси – СОСА танланган.

Тадқиқотнинг предметини СОСА корпуси мисолида инглиз тили корпусларининг функционал хусусиятлари ташкил этади.

Тадқиқотнинг усуллари. Тадқиқот мавзусини ёритишда контент таҳлил, таснифлаш, тавсифлаш, қиёслаш, статистик, когнитив ва корпус таҳлил методларидан фойдаланилган.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

корпус лингвистикаси тарихи, тараққиёт босқичлари, таснифлаш муаммоларининг статистик тадқиқи натижасида намунавийлик, мақсадга йўналтирилганлик ва қамровнинг чегараланганлиги каби корпус яратиш мезонлари мукамаллаштирилган; олти методологик тизимга биноан корпус яратиш ва уни амалда қўллаш механизми илмий-назарий далиллар асосида исботланган;

турли форматдаги матнлардан автоматик тарзда корпус яратувчи АнтКонк программаси ёрдамида амалга оширилган экспериментал корпус таҳлили асосида яратилган соҳавий частотали-изоҳли терминологик луғат орқали корпус лингвистикасининг терминологик аппарати бойитилган;

корпусларнинг репрезентативлик хусусиятига кўра корпуслараро (Британия Миллий корпуси – BNC ва Замонавий Америка инглиз тили корпуси СОСА корпуслари мисолида) метафоралар, коллокация (семантик бирикма)лар ва фразеологик бирикмаларнинг статистик таҳлили ёрдамида лингвокультурологик бирикларни қиёслаш имконияти исботланган;

илк бор луғат корпуси сифатида ўрганилган СОСА корпуси лексик бирикларнинг фонетик, орфографик, грамматик ва лексик-семантик хусусиятларини лексикография соҳаси тақдим этгани ҳамда инглиз тили луғатларини такомиллаштиришдаги аҳамияти далилланган.

Тадқиқотнинг амалий натижалари қуйидагилардан иборат:

¹⁰ СОСА (Corpus Of Contemporary American English) – *Замонавий Америка инглиз тили корпуси*: Йиллик ҳисобда 20 миллион ҳажмдаги матнларни ўз ичига олиб борадиган жанрлараро мувозанатли корпус

корпус лингвистикаси методлар тизими такомиллаштирилган;
корпус лингвистикасига оид частотали-таржима-изоҳли терминлар лугати ишлаб чиқилган;

корпуслараро лингвомаданий концептларининг намоён бўлиши очиб берилган;

СОСА корпусининг лексикография соҳасига ижобий татбиқи илмий-амалий жиҳатдан таҳлил этилган.

Тадқиқот натижаларининг ишончлилиги хулосаларнинг ишончлилиги, муаммонинг аниқ баён қилиниши, ишда қўлланиладиган ёндашувлар ва усулларнинг самарадорлиги билан белгиланади.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Тадқиқотнинг илмий аҳамияти изланиш натижаларидан лисоний корпуслар яратиш ва уни мукаммаллаштиришнинг назарий асосларини ишлаб чиқишда, корпус лингвистикаси йўналишида тадқиқотлар олиб боришда илмий-назарий манба сифатида фойдаланиш мумкинлиги билан белгиланади.

Тадқиқотнинг амалий аҳамияти бугунги кунда илмий изланиш методологияси сифатида қаралаётган бўлса-да, яқин келажакда фан сифатида тараққий этиши кутилаётган корпус лингвистикасини ўқитиш жараёнида дастур, режалар тузиш ҳамда мавзуларни баён этишда назарий манба вазифасини ўташи, махсус экспериментал корпусларни яратиш ва мавжуд корпусларни қўллаш бўйича амалий намуна бўла олиши билан изоҳланади.

Тадқиқот натижаларининг жорий қилиниши. Инглиз тили корпусининг функционал хусусиятларини СОСА корпуси мисолида очиб бериш бўйича ишлаб чиқилган амалий таклиф ва тадқиқот натижалари асосида:

жаҳон ва ўзбек тилшунослигида корпус лингвистикаси концепциясининг тараққиёт босқичлари, тажриба мактабларини яратиш, таснифлаш мезонлари ва тамойилларининг илмий-назарий тадқиқи орқали корпусларга оид умумий талабларнинг шакллантирилишига оид таҳлиллардан давлат илмий-техник дастурлари доирасида бажарилган А-1-61 рақамли «Кўриш имконияти чекланган талабалар учун тилшунослик фанлари бўйича аудиокитоб таъминотини яратиш» номли амалий грант лойиҳаси асосида «2014-2016 йилларда нашр этилган тилшунослик фанларига оид мақолаларнинг қисқача мазмуни» («Ўзбек тили ва адабиёти» ҳамда «Филология масалалари» журналлари асосида») номли аудио маълумотлар тўпламини тайёрлашда фойдаланилган (Олий ва ўрта махсус таълим вазирлигининг 2020 йил 12 июндаги № 89-03-2032 - сонли маълумотномаси). Натижада, корпус яратиш, уларни амалда қўллашнинг тизимли методларини такомиллаштиришга доир муаммолар ўз ечимини топган;

эмпирик таҳлил асосида корпус лингвистикасининг яхлит тизим сифатида тадқиқ этилиши ва АнтКонк программаси ёрдамида амалга оширилган экспериментал корпус таҳлиллари натижасида яратилган соҳавий частотали-изоҳли терминологик лугат орқали корпус лингвистикасининг терминологик тизими такомиллаштирилганига оид илмий хулоса ва таклифлардан Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетининг «Ўзбек ва инглиз тилларининг электрон (инсон

қиёфаси, феъл-атвори, табиати ва миллий тимсоллар тасвири) поэтик луғатини яратиш» номли инновацион лойиҳасида фойдаланилган (Олий ва ўрта махсус таълим вазирлигининг 2020 йил 12 июндаги № 89-03-2032 - сонли маълумотномаси). Илмий натижаларнинг қўлланилиши соҳага оид терминларнинг изоҳлаш тамойилларини мукамаллаштирган;

корпусларнинг репрезентативлик хусусиятига кўра, корпуслараро (BNC ва СОСА корпуслари мисолида) метафоралар, коллокация (семантик бирикма)лар ва фразеологик бирикмаларнинг статистик таҳлили ёрдамида лингвокультурологик бирликларни қиёслашга доир амалий таклиф ва тавсиялардан ОТ-Ф1-18 – «Оммавий лисоний маданиятни шакллантириш методлари ва методологиясини ишлаб чиқиш» номли фундаментал лойиҳа доирасида амалга татбиқ этилган (Олий ва ўрта махсус таълим вазирлигининг 2020 йил 12 июндаги № 89-03-2032 - сонли маълумотномаси). Айни натижаларнинг қўлланилиши тил призмаси орқали юзага чиқадиган лингвомаданий концептларни корпус таҳлил орқали чуқурроқ англаш имконини берган;

СОСА корпуси илк бор луғат корпуси сифатида тадқиқ этилиб, лексикография соҳасига тақдим эта олган барча статистик ва амалий натижалар асосида инглиз тили луғатларини такомиллаштириш хусусиятлари аниқланганлиги асосида такомиллаштирилган илмий хулосалар Испаниянинг Барселона шаҳрида жойлашган Помпеу Фабра университети «Таржима ва тилшунослик» кафедраси профессор – ўқитувчилари ҳамда GEDIT (Grup d'Estudis del discurs I de la Traduccio') тадқиқот гуруҳи вакиллари иштирокида ўтказилган илмий семинарда муҳокамадан ўтди ва Университетнинг магистр талабалари диссертацияларида қўлланди (Помпеу Фабра университети «Таржима ва коммуникация» факультетининг 2020 йил 24 январдаги маълумотномаси). Натижада, қўлланилган методдан магистрлик ва докторлик диссертацияларини ёзишда фойдаланиш мумкин.

Тадқиқот натижаларнинг апробацияси. Тадқиқот натижалари бўйича 2 та халқаро ва 10 та республика илмий-амалий анжуманларда қилинган маърузаларда жамоатчилик муҳокамасидан ўтказилган.

Тадқиқот натижаларининг эълон қилинганлиги. Диссертация мавзуси бўйича жами 15 та илмий иш чоп этилган, жумладан, 1 та частотализохли терминологик луғат, Ўзбекистон Республикаси Олий аттестация комиссиясининг докторлик диссертациялари асосий илмий натижалари чоп этиш учун тавсия қилинган илмий нашрларда 14 та мақола, улардан 8 таси республика ва 6 таси хорижий журналда нашр этилган.

Диссертациянинг тузилиши ва ҳажми. Диссертация таркиби кириш, учта боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловадан иборат бўлиб, умумий ҳажми 189 саҳифани ташкил этади.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида илмий тадқиқотнинг мавзуси ва унинг долзарблиги асосланган бўлиб, тадқиқот ишининг мақсади ва вазифалари, объекти ва

предмети аниқланган, илмий ишнинг фан ва технологияларни ривожлантиришнинг муҳим йўналишларига мослиги кўрсатиб ўтилган. Шу билан бирга, диссертациянинг илмий янгилиги, амалий натижалари ва уларнинг ишончлилиги, ишнинг назарий ва амалий аҳамияти, эришилган натижаларнинг амалиётга жорий этилиши, илмий нашрларда эълон қилинганлиги, ишнинг тузилиши борасидаги маълумотлар келтирилган.

Диссертациянинг **“Жаҳон ва ўзбек тилшунослигида корпус лингвистикаси концепциясининг тараққиёт босқичлари”** номли биринчи бобида Корпус лингвистикаси(КЛ)нинг диахроник тадқиқи, КЛ спецификаси ҳамда «корпус» тушунчасини таснифлаш муаммосига оид назарий таҳлиллар асосида дунё ва мамлакат миқёсидаги тадқиқотлар таҳлилга тортилган.

КЛ – лисоний корпусларни яратиш ва уларнинг татбиқи билан шуғулланувчи тилшунослик соҳасидир¹¹. Ўз олдида лисоний корпусларни яратишни мақсад қилган КЛ ҳақида чуқурроқ тушунчага эга бўлиш учун унинг ривожланиш тарихига назар ташлаш ўринли. Бу атама XIX асрнинг 60-йилларида яратилган корпуслар амалиётини ривожлантириш билан боғлиқ бўлиб, 1980 йилдан буён компьютер технологиялари асосида амалга оширилмоқда¹².

КЛнинг тараққиёт босқичлари корпуснинг назарий тамойиллари билан узвий боғлиқ. КЛга оид тадқиқотлар тилшуносликда ва тилларни ўқитишни ташкил қилишда XVIII асрдан бошланган бўлса-да, матнларнинг илк электрон лисоний корпуси ўтган асрнинг 60-йилларида пайдо бўлган. 1961 йилда бошланган лойиҳа натижасида 1964 йилда Бровн университети (АҚШ) олимлари компьютерда (Бровн Корпуси) катта ҳажмдаги матнларни яратишган. У. Френсис ва Г. Кучера Америка нашрининг инглиз тилидаги беш юз икки минг насрий босма нусхасини яратди. Инглиз насрининг ўн бешта энг катта жанрларига тегишли матнлар Америка Қўшма Штатлари босма корпуси материаллари ҳамда унинг асосий статистик қайта ишланиши асосида юзага келган. 1961 йилда корпусдан олинган илк натижалар частота ва алфа-частотали луғат кўринишида чоп этилди. Х. Лингнинг¹³ фикрича, бундай ҳажмли корпуслар фақатгина фонетиканинг ўзига хос томонларини ўрганишда қўлланилган.

Н. Чомский¹⁴нинг фикрича, тилшунос тил моделини яратишда компетенция (competence)га асосланиши шарт, унинг ифода ҳолати (performance) бундан мустасно. У ўзининг юқоридаги назариясини 1950-1960 йилларда эълон қилди, яъни корпуснинг тилшуносликка татбиқини расман рад этди. Ана шу йилгача бўлган давр КЛ тарихида пре-Чомский номини олди. Ўз раддиясига асос қилиб, инглиз тилидаги *perform* феълининг қўлланилиши хусусидаги фикрини келтиради ва Бровн

¹¹ Корпусная лингвистика как раздел языкознания URL: <https://myfilology.ru/177/korpusnaya-lingvistika-kak-razdel-yazykoznaniya/>

¹² https://ru.wikipedia.org/wiki/Корпусная_лингвистика

¹³ Ling X. Cihui Yu.he Jisuan Yu. [Lexical Semantics And Computational Linguistics]. BeijingLingua Publishing house, 1999. – P. 240.

¹⁴ Chomsky N. Current issues in linguistic theory // J.A. Fodor and J.J. Katz. The Structure Of Language. Engelwood Cliffs, New Jersey: 1964, – P. 50–118.

корпусининг имкониятларига нисбатлаб, корпуслар қамровининг паст эканлигини кўрсатади. Бугунги кунда жуда катта ҳажмли корпусларнинг мавжудлиги бу сўз бирикмаси қаторини узоқ давом эттириш имконини беради. Масалан, биз СОСА корпусида кидирувга берганимизда, *perform ability, perform functions, perform tests, perform operation, perform surgery, perform dance* каби кўплаб сўз бирикмаларини аниқладик. Чомский танқиди замонавий инглиз тили корпуслар татбиқига нисбатан қаратилмаганлигига амин бўлдик.

КЛ корпусининг умумий принциплари ва тил хусусиятлари (матнлар корпуси)ни ишлаб чиқиш билан шуғулланар экан, лисоний корпус муайян меъёрлар ва стандартларга мувофиқ тўпланган ва махсус қидирув тизими билан таъминланган матнлар тўплами сифатида қаралади. Баъзан корпус – тўпланган ҳамда сараланган тил, жанр, муаллиф омилига кўра тартибланган матнларни яратиш даврини бирлаштирган тўплам ҳам дейилади.

П.Ю. Довнар ва А.В. Воронцов матн корпусларини яратишнинг мақсадга мувофиқлигини таъкидлайди ва қуйидаги хусусиятларни ажратиб кўрсатади: 1) ҳақиқий маънода лингвистик маълумотларни тақдим этиш; 2) етарлича катта маълумотни тақдим этиш (катта корпус ҳажми билан); 3) матннинг грамматик ва лексик-грамматик таҳлилини ва ҳ.к. сингдириш каби турли хил лисоний вазифаларни ҳал қилиш учун бир марта яратилган корпусдан қайта фойдаланиш имконияти ва бошқ.¹⁵

Дунё тилларининг аксарияти ўз миллий корпусларига эга. Умумий корпусларга мисол сифатида Британия Миллий корпуси (BNC)ни келтириш мумкин. Славян тилларининг корпуслари орасида Прагадаги Чарлз университетига ташкил этилган Чехия Миллий корпуси мавжуд. Миллий корпуслар немис, хитой, француз ва бошқа тилларда ҳам мавжуд. Рус тили учун бундай мавқега эга корпус Россия Миллий корпуси (НКР) ҳисобланади. Афсуски, ўзбек тили ўз миллий корпусига ҳали эга эмас.

Ўзаро яқин лисоний изланиш соҳаларининг муносабати хусусидаги изланишларимиз КЛ спецификаси ҳақида қуйидаги хулосаларга келиш имконини берди (1- расмга қаранг):

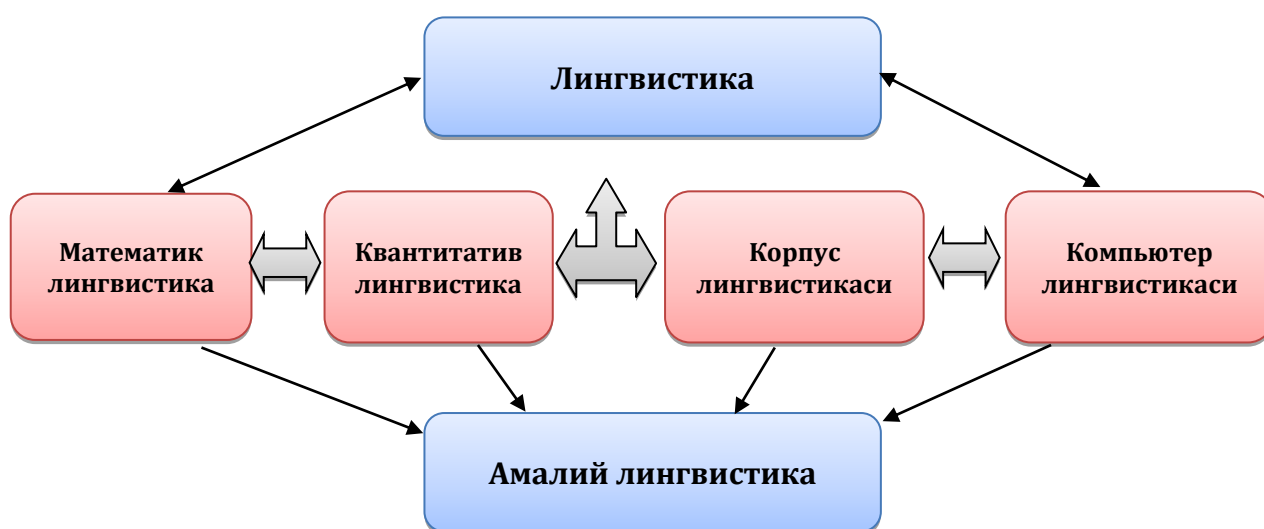
1. Квантитатив лингвистика ва корпус лингвистикаси тилшунослик соҳалари бўлиб, тўғридан-тўғри унинг муаммоларини ечишга йўналтирилган тармоқдир.

2. Математик лингвистика ва компьютер лингвистикаси тилшунослик билан ўзаро алоқадор тармоқлар сифатида унинг муаммоларини ечиш учун математик ва техник таъминот ишлаб чиқишни мақсад қилади.

3. Амалий лингвистика – Фарбда «applied linguistics» номи билан тарқалган соҳа сифатида компьютер қурилмаларидан фойдаланишдан манфаатдор барча соҳаларни қамраб олади ва юқоридаги тўрт соҳа билан ўзаро бир томонлама алоқага киришади.

¹⁵ Довнар П.Ю., Воронцов А.В. Лингвистический процессор китайского языка. Особенности разработки (рус.)// Международный конгресс по информатике: информационные системы и технологии: материалы международного научного конгресса 31 окт. — Минск: БГУ: БГУ, 2011.

4. КЛни лингвистиканинг мустақил соҳаси сифатида қараш асослидир. Бироқ КЛ ўз ривожда математик лингвистикадан лисоний ҳисоб-китобларга керакли формулаларни, квантитатив лингвистикадан миқдорий кўрсаткичларни келтириб чиқариш ва улардан хулоса чиқариш учун асосли алгоритмларни олса, компьютер лингвистикасидан автоматлаштирилган компьютер программалари ва қурилмаларини олади ҳамда ўз ютуқларини лисоний изланишларга қаратган ҳолда натижаларини амалий лингвистика билан тил ўқитиш жараёнида қўллаш учун бўлишади.



1- расм. Тилшунослик ва унга алоқадор соҳалараро муносабат шакллари

Бизнингча, КЛга тилшуносликнинг семантика, синтаксис ва шу каби бўлимлари сифатида қараш ноўрин, уни изоҳ ва тасвирлашни талаб этмайдиган методология дейиш мумкин. КЛга лингвистик тадқиқотлар олиб боришга қаратилган методология сифатида қараш Ch.F. Meyer, P. Baker, D. Biber, Y. Kawaguchi, N. Chanyoo, J. Flowerdew, M. Mahlberg, L.C. Steuber, A. Renouf, A. Hardie, R. Love, S. Hunston, P.E. Rayson, M. Fuster, B. Clavel¹⁶ каби кўп сонли олимлар ишларида кузатилди. Бу методологияни тилшуносликнинг биргина соҳаси билан чегараланиб қолмайдиган, балки ҳар қандай лисоний изланишларда қўллаш имконияти мавжуд илмий-тадқиқот методларига эга методология сифатида қараш ўринли.

¹⁶ Meyer Ch. English Corpus Linguistics: An Introduction – Boston: University of Massachusetts at Boston. 2004. – 168 p.; Baker P. Sociolinguistics and Corpus Linguistics: – Edinburgh University Press. 2010. – 189 p.; Biber D., Ulla C. and Upton T.A. Discourse on the Move: – Amsterdam: 2007. – 290 p.; Kawaguchi Yu. and others Corpus-Based Perspectives in Linguistics. – Tokyo. 2007. – 442 p.; Chanyoo N. A Corpus-Based Study Of Connectors And Thematic Progression in The Academic Writing of Thai EFL students: – University of Pittsburgh: 2013. – 146 p.; Flowerdew J. and Mahlberg M. Lexical Cohesion and Corpus Linguistics Amsterdam Philadelphia: 2009. – 125 p.; Lucas C. Steuber Disordered Thought, Disordered Language: A Corpus-Based Description Of The Speech Of Individuals Undergoing Treatment For Schizophrenia. – Portland State University: 2011 – 73 p.; Renouf A. Corpus Linguistics: Past And Present. – Birmingham: University of Central England, – 13 p; Hardie A. and Love R. Corpus Linguistics: 2013. – 376 p.; Hunston S. Corpus Linguistics. University of Birmingham, 2006. – 234–248 pp.; Rayson P.E. Matrix: A Statistical Method And Software Tool For Linguistic Analysis Through Corpus Comparison, Lancaster University: 2002. – 194 p.; Fuster M. and Clavel B. Corpus Linguistics and its Applications in Higher Education University of Valencia: 2010. – 51-67 pp.

КЛ ривожланиш босқичининг илк даврида тадқиқот методи сифатида қаралган ва уни компьютер ва математик ҳисоб-китобларга қаратилган фан тармоқларининг қисми ёки воситаси сифатида қаралган бўлса, бугунги кунда ўзининг мақсад-вазифалари ва тадқиқот методларига эга методология бўлиб улгурган ҳамда кенг илмий жамоатчилик томонидан қабул қилинган. Келажакда эса тез суръатда ривожланиб бораётган КЛ барча соҳадаги муаммолар ечимини топишга кучи етадиган лингвистик илмий-тадқиқот парадигмаси ёки мустақил фан сифатида баҳолаш мумкин.

Фикримизча, ҳар қандай лисоний корпус ўз таркибидаги матний элементлар йиғилган тил ёки суб тилнинг лингвистик хусусиятларини очиб бера олиш имкониятига эга бўлган ҳажм ва туркумлардан шаклланган бўлиши шарт. Ушбу хусусият корпуснинг лингвистик тадқиқот амалиётларида қўлланилиши ва уни тилнинг «электрон ҳолати» сифатида қаралишига олиб келади. Бунга эришиш учун эса ҳар бир корпус тузувчи матнларни категориялар асосида йиғиши керак. СОСА таркибидаги оғзаки ва ёзма матнларнинг ҳар бири муайян категорияга ажратилган. Масалан, ёзма – газета материали – газета номи каби аниқ кўрсатилган, белгилаб олинган тартибда матнлар жараёнга тортилиши керак.

Корпус – корпус тузувчининг прагматик мақсадидан келиб чиқиб қатъий тамойиллар асосида шакллантириладиган ва таснифланадиган, репрезентативлик мезонига жавоб бера оладиган, сараланган, етарли ҳажмдаги лисоний матнлар (оғзаки ва ёзма) тўплами бўлиб, ўзида қайта ишлаш, теглаш каби принципларни мужассамлаштирган, асосан компьютер ёрдамида жараёнларни автоматлаштирадиган мукамал тизимли қидирув, ҳаволали натижа ва эмпирик таҳлилга қулай миқдор ҳамда изчилликка эга базадир.

Тадқиқот жараёнида корпуслар таснифини махсус принципларга асосланган ҳолда келтириш, уларни маълум маънода қўллаш соҳаларига кўра туркумлаш, шартли равишда, якуний шаклга келтирилди ва мавжуд терминлардан фойдаланилди. Лисоний корпусларни қўллаш мумкин бўлган соҳалар қаторига қуйидагиларни киритдик: А) таржима амалиёти ва чоғиштирма лингвистика: *махсус/ умумий корпус; қиёсий корпус; бир/кўп тилли корпус; оғзаки/ёзма корпус; Параллел корпус ва ҳ.к.* Б) лексикография: *махсус/умумий корпус; ҳаволали корпус; оғзаки/ёзма корпус; қўш корпус; бир/кўп тилли корпус ва ҳ.к.* В) адабиёт назариясида: *параллел корпус; муаллифлик корпуси; адабий корпус ва ҳ.к.* Г) лингводидактика ва тилларни ўқитиш: *махсус/ умумий корпус; ўрганувчи/ педагог корпуси; кузатув корпуси; оғзаки/ёзма корпус ва ҳ.к.* Д) этимология ва тил тарихи тадқиқи: *махсус/ умумий корпус; тарихий/замонавий; кузатув корпуси; ҳудудий корпус ва ҳ.к.* Бу рўйхатни корпус таҳлилларга таянувчи барча тил хусусидаги назарий ва амалий тадқиқот соҳалари билан давом эттириш мумкин.

Диссертациянинг **“Корпуслар татбиқига оид амалий таҳлиллар: натижа ва ечим”** деб номланган иккинчи бобида КЛ методологияси – корпус таҳлил методлари, махсус корпус ЕСОСЛ асосида корпус

лингвистикасининг тил хусусиятлари ва инглиз тили корпуслари билан ишлаш ва улар тақдим этадиган маълумотлар аҳамияти тадқиқ қилинган.

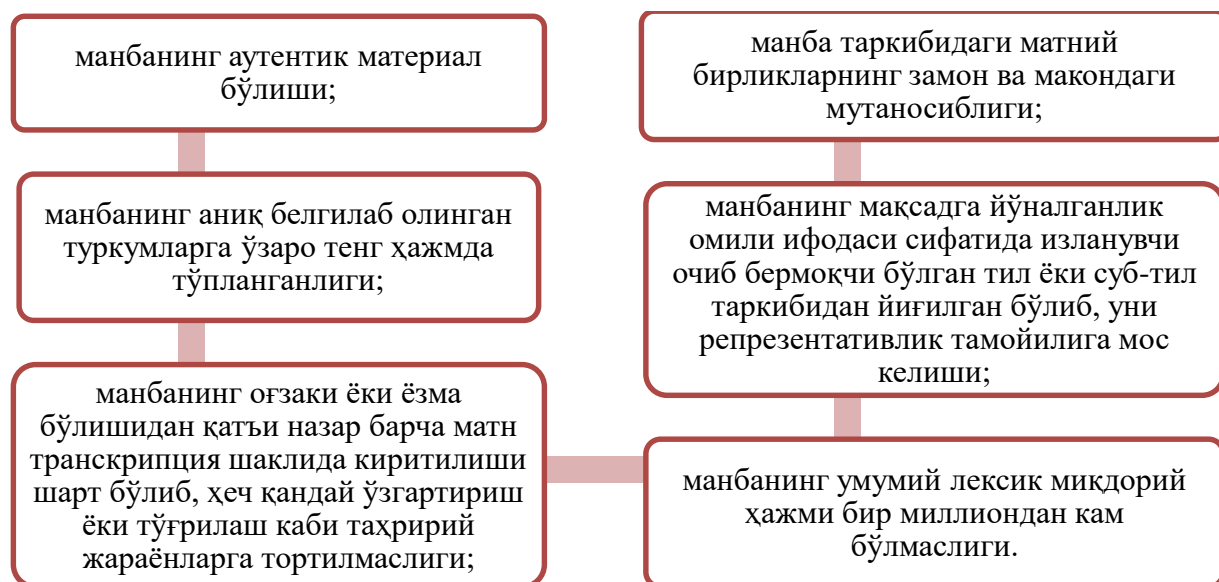
М.А. Большакова корпусга асосланган тадқиқотида КЛни қуйидаги босқичларга ажратиб кўрсатган¹⁷: *мақсадни белгилаш, ҳавола бериш, яратиш, қидирув ва тушунтириш босқичлари.*

Д.Ф. Джаъфарова¹⁸ КЛда корпус қуйидаги талабларга жавоб бериши кераклигини таъкидлайди: 1) намунавийлик ва репрезентативлик; 2) чегараланган ҳажм; 3) маълумотнинг автоматлаштирилганлиги.

Корпус тамойиллари асосида корпуслар яратиш ҳамда улардан амалий тадқиқотларда қўллашга қаратилган КЛ методологиясини олти та методдан иборат эмпирик тизимга ажратишни лозим топдик:

1) **йўналтирувчи метод.** Бунда корпуснинг *мақсадга йўналтирилганлигига* асосланиб, тузувчи яратаётган лисоний корпусини қандай тадқиқотларга мўлжалланганлигини аниқ белгилаб олиш шарт. Албатта, мақсадли корпуслар таснифига кўра, корпуснинг қайси турига эҳтиёж мавжудлиги ва унинг қандай хусусиятларга эга бўлишига нисбатан муҳим лисоний белгилар аниқланади.

2) **манба билан ишлаш ёки базавий метод.** Бу метод матнларни йиғиш ва йиғилган матнларни саралаш билан ишлашга қаратилади. Бунда корпус таркибига кирувчи матний манбаларга талаблар ишлаб чиқилади ва ҳар бир корпус тузувчи ушбу олти талаб асосида матнларни йиғиши тавсия этилади (2 - расмга қаранг). Бундан ташқари, ушбу босқич ўз ичига матнларни бир муаллиф ижоди, бир жанрга мансублик, таркибий ва соҳавий ўхшашликларга кўра туркумларга бўлиш жараёнини ўз ичига қамраб олади.



2-расм. Корпус матнларига қўйиладиган мезон талаблари

¹⁷ Большакова М.А. Лингвистический контент и программная реализация интеллектуального немецко -русского отраслевого словаря Диссер...канд.фил. наук. – Москва, 2013. – С 56-57.

¹⁸ Джаъфарова Д.Ф. Модели лингвистического анализа текстов таджикского языка (на материале газелей Хафиза) дисс. ... кандидата филологических наук. – Душанбе, 2013. – С 35

3) **метадата кўшиши методларидан** таркиб топади. Бу гуруҳга киритилган методлар корпус таркибига киритиш учун сараланган матнларга ҳаволавий характерга эга маълумотларни кўшишни ўз олдига мақсад қилади. Маълумот базасининг аниқ кўрсатилиши жуда содда метод ҳисобланади ва бу маълумот туркумга кўра (газета материаллари, оғзаки ва ҳ.к.) қидирув имконини беради ҳамда конкорданс қаторида келтирилган мисолда ҳавола тарзида улаб қўйилади. Бу босқичда теглаш методи муҳим бўлиб, бугунги кунда аксарият ҳолларда компьютерлар ёрдамида амалга оширилади.

Тадқиқот жараёнида ҳар бир синтактик қурилма ҳам тегланиши мумкин. Масалан, *I am* структураси PPSS+BEM (PPSS – non-third person nominative personal pronoun ва BEM – am form of Be) (1-жадвалга қаранг).

1-жадвал

Феъл формаларининг берилишини теглаш

| | | | | |
|-------------|-----------|-----------|------------|-----------|
| Sing /VB | Be /VB | Do /VB | Have /VB | Go /VB |
| Sings /VBZ | Is/VBZ | Does /VBZ | Has /VBZ | Goes /VBZ |
| Sang/VBD | Was/VBD | Did/VBD | Had/VBD | Went/VBD |
| Singing/VBG | Being/VBG | Doing/VBG | Having/VBG | Going/VBG |
| Sung /VBN | Been /VBN | Done /VBN | Had /VBN | Gone /VBN |

Бунда: VB – *verb* – феъл сўзи учун келган бўлса, унинг таркибига кўшилган Z учинчи шахс бирликда ҳозирги оддий замон кўшимчасидир, D эса ўтган замон шаклида феъл формасини билдирса, G герундий формасини ифодалайди ва N феълнинг сифатдош икки шаклини акс эттиради.

Тег-тизим эса ушбу индивидуал тегларнинг якуний систематик йиғиндиси каби шаклланади. *Penn Treebank* корпуси учун шакллантирилган тег-тизим таркибига жами 48 та теглаш кодлари санаб ўтилган бўлиб, улар сирасига, хусусан, қуйидагилар киритилган: CC – *coordinating conjunction*, CD – *cardinal number*, DT – *determiner*, FW – *foreign word*, JJ – *adjective*, JJS – *adjective in superlative degree*, JJR – *adjective in comparative degree*, NN – *noun*, NNS – *plural noun*, NNP – *proper noun*, RB – *adverb*, RBR – *adverb in comparative degree*, RBS – *adverb in superlative degree* ва бошқалар.

4) **автоматлаштириши методи** (корпусларни тўлиқ электрон шаклга ўтказишга қаратилган). Бу босқичдаги метод барча йиғилган, сараланган, ҳаволали метадатага бириктирилган ва текшириб чиқилган матнларни электрон шаклдаги корпус кўринишига келтириш ҳамда унинг дизайни билан ишлайди. Шунингдек, корпусларда конкордансларнинг шаклланиши, қидирувга берилган сўзнинг ва у бирикиб келган сўзларнинг рангидан тортиб, корпус таҳлил асосида эришилган натижаларни график кўринишларда ифода эта олиши каби кенг имкониятларни ўз ичига олади.

5) **баҳолаш методи**. Бу метод орқали «корпус» тушунчаси тўғрисида билдирилган тамойилларнинг ҳар бирига жавоб бера олиш имкониятини

аниқлаш орқали унинг мукамаллиги белгиланади. Бунда корпуснинг ҳар бир хусусияти баҳоланиши шарт.

б) *частотали қидирув-анализ методи*. Бунда сўзларнинг корпусда қўлланилиш частотасини кўриш орқали хулосалар чиқариш имконини берувчи метод сифатида қараш мумкин. Умуман олганда, барча корпус таҳлилига асосланган таҳлилларда частоталарнинг аҳамияти юқори бўлиб, улар билан ишлаш ҳам аниқ фарқлашни талаб этади. Чунки корпус таркибига кирувчи сўзларнинг қўлланилиш частоталари улар хусусида кенг қамровли хулосаларни чиқаришга имкон яратади. Т.С. Грайс¹⁹ уларни уч турга бўлади: бирламчи частоталар; меъёрий частоталар; логорифмланган частоталар.

Тадқиқотимизда частотали қидирув-анализ методи орқали корпус маълумотларини ўзаро қиёслашни мақсадга мувофиқ деб топдик:

Бирламчи частоталар (Raw frequencies) – қидирувга берилган сўзнинг корпусдаги қўлланганлик частотаси бўлиб, у аниқ сонларда берилади. Масалан, СОСАда *give* феъли 202515 марта қўлланилган, бу унинг бирламчи частотасидир (БЧ).

Меъёрий частоталар (Normalized frequencies) – бу корпуслар кўрсатган аниқ сонли частоталарни 10 нинг даражалари (1000, 1000000 ва ҳ.к.) га нисбатлашдир. Масалан, юқорида келтирилган *give* феълининг МЧсини топиш учун, СОСАнинг умумий миқдорини (УМ), яъни 570 миллионга нисбатан олиб ҳисобланади. Бунда сўзнинг миллионга нисбатланган меъёрий частотаси (МЧ) ини топиш учун унинг БЧ сини 570га бўлиш керак. Қуйидаги формулани тақдим этиш ўринли:

$$\text{МЧ(мил/1)} = \frac{\text{БЧ}}{\text{УМ/1000000}}.$$

Бундан келиб чиқадики, СОСАда *give* феълининг МЧ(мил/1) и тақрибан 355,2 га тенглигини кўриш мумкин. Бу бир тилли корпуслараро ёки турли тилли корпуслараро қиёслашда муҳимдир.

Логорифмланган частоталар (Logged frequencies) – аниқ танлаб олинган асосли ($e = 2,7182818$) логорифмга солинган бирламчи частотадан чиқарилган даража. Яъни логорифм асос ва даража муносабатини кўрсатади. $\log_b x = y$ формуласига қарайдиган бўлсак, $b=10$ бўлса, $x=100$ бўлса, $y=2$ чиқади. Соддароқ айтганда, 10 нинг нечанчи даражаси 100 бўлади саволига жавоб 2 экани тушунилади. Логорифмланган частота(ЛЧ)ни чиқаришда юқоридаги формулада $b=2,7182818$ (асос), $x=202515$ (бирламчи частота) ва бундан логорифм олинса, даража $y=12,218569364782$ эканлиги кўринади ва бу *give* феълининг СОСАдаги қўлланишининг логорифмланган частотасидир.

Бизнингча, ушбу қиёслаш жараёни уч хил кўринишда бўлиши мумкин: *ички қиёслаш; бир тилли корпуслараро қиёслаш; турли тилли корпуслараро қиёслаш.*

¹⁹ Stefan Th. Gries Quantitative designs and statistical techniques/ English Corpus Linguistics edited by Douglas Biber and Randi Reppen. Cambridge, United Kingdom. 2015. – 623 p. 50-71 pp. (52)

Корпус яратиш ва ундан амалий тадқиқотларда қўллашга йўналтирилган КЛ методологиясини олти методлар гуруҳига бўлиб тузишни таклиф қилдик. Айнан шу тизимли методлар узвийлиги орқали КЛнинг тил хусусиятларини ўрганишга қаратилган экспериментал корпус яратишни асосли деб топиш мумкин. Бунга кўра, яратиш режалаштирилган корпус шу жараёнга мос шакллантирилади.

Аввало, корпус яратишдан мақсад белгилаб олингандан сўнг кейинги босқичда ЕСОСЛ корпуси таркибига кирувчи матнлар билан ишланди. Манба билан ишлаш ёки базавий методлар орқали, аввало, корпус матнлари йиғилди. Корпус тамойилларига кўра, ишлаб чиқилган корпус матнларига қўйиладиган мезон талаблари асосида манба қуйидагича ишлаб чиқилди: 1) йиғилган база аутентик материал бўлиб, уларнинг барчаси инглиз тилида ёзилган ҳамда тил соҳиби томонидан таҳрир қилинган илмий маълумотлардир; 2) корпус манбасига танлаб олинган материалларнинг барчаси КЛга оид илмий материаллар; 3) манба аниқ ажратиб олинган туркумларга ўзаро тенг ҳажмда тўпланган (диссертация, китоб ва мақола каби); 4) корпус таркибига кирувчи барча материаллар КЛнинг тил хусусиятларини ўзида акс эттириш имконига эга; 5) манба тўлиқ ёзма нутқ намуналаридан ташкил топган; 6) ЕСОСЛга кирадиган материалларнинг умумий лексик миқдорий ҳажми бир миллиондан кўплиги аниқланди. Шу тарзда матнларни йиғиш ва саралаш ишлари олиб борилди ва бу босқичнинг асосий методларидан бири бўлган белгилаш (mark-up) методининг автоматик тарзда амалга оширилиши кўзда тутилди. Матнлар тури ва соҳаси бир хил бўлганлиги сабабли экстра-матний (extra-textual) маълумотларни қўшиш салмоқли аҳамият касб этмади.

Corpus терминининг грамматик кўплик формаси ўз эквивалентига эга. Бунда *corpora* ва *corpuses* каби шакллар бир хилда муомалада экани айтиб ўтилган. ЕСОСЛда берилган қидирув натижалари ҳар иккала грамматик шаклнинг ҳам мавжуд эканини кўрсатди, лекин *corpora* 13829 мартали БЧни кўрсатган бўлса, *corpuses* 53 марта қўлланилгани далилланди. Бундан инглиз тилида *corpus* терминининг грамматик кўплик шакли сифатида *corpora* кенгроқ қўлланилиши аниқ бўлди. Лекин ўзбек тилида ҳар иккаласи маъно жиҳатидан кўпликни англатгани боис бу шакл *корпуслар* тартибида берилди.

Тадқиқот натижасида жами 220 та термин ажратиб олинди ва уларнинг частотаси ЕСОСЛ конкорданс натижалари асосида киритилди. Кичик луғат дизайни “Ўзбек тилининг изоҳли луғати” форматига мос бўлган яқин шаклда юзага келди. Бунда ҳар бир терминнинг инглиз тилида ва унинг ўзбек тилида тавсия этилган муқобили берилди ҳамда унинг изоҳи келтирилди. Ҳар бир изоҳ олинган манба муаллифи ЕСОСЛ интерфейси берган имконият орқали тезкор топилиб, унга ҳавола (McEnergy, 2006) тарзида берилди. Биз кичик луғатнинг аҳамиятини КЛ соҳаси билан фойдаланишни бошлаган тадқиқотчиларга инглиз тилидаги терминларнинг ўзбек тилидаги муқобили ва изоҳи орқали тушуниш имконинигина бериб қолмай, у айнан, кенг қўлланиладиган терминлардан бошлаб ўрганишни таклиф этишида деб

топдик. Бу ўз ўрнида ўрганувчига соҳани босқичма-босқич кашф этиб боришда, нотаниш терминларни англаш учун замин яратади. Масалан, *корпус* тушунчасини аниқ фаҳмламасдан, унда қўлланиладиган кодлаш тизими бўлмиш *теглаш* атамасини мияда шакллантириб бўлмайди. Шу асосида яратилган ECOLL корпусидан олинган таҳлил натижаларидан фойдаланган ҳолда кичик частотали-изоҳли ва алфавитли-частотали-изоҳли луғат яратилди ва бунда сўзлар мос равишда частотасига ва алфавит тартибига кўра тартибланди.

Шунингдек, ушбу бобда корпус таҳлилларини тилшуносликнинг кўплаб тармоқларида қўллаш ижобий натижаларга сабаб бўлиши асосланди. КЛ ривожда инглиз тилида яратилган ва яратилаётган лисоний корпусларнинг ўз тарихий аҳамияти мавжудлиги назарда тутилган ҳолда инглиз тили лисоний корпусларини тақдим этувчи WWW кенгайтмали электрон манбалар ва улар тақдим этадиган маълумотлар ишда ўз аксини топган. Асосан, учта база кўриб чиқилган: *Sketch engine*, *CoRD*, ва *English-corpora.org*. ҳамда ҳар бир база таркибига кирувчи корпуслар хусусида қисқача тўхталиб ўтилган ва улардан амалий фойдаланишга кириш учун махсус сайтлар ҳаволалари ҳам келтирилган. *English Web 2015 (enTenTen15)*, *BNC* ва *Google Books n-gram* каби инглиз тили корпуслари тақдим этиши мумкин бўлган кенг қидирув ва конкорданс натижалари кўриб чиқилган. Бунда ҳар бир корпуснинг имкониятларини кўрсатиб берувчи оддий ва мураккаб қидирувлар амалга оширилган.

Учинчи боб **“СОСА корпусининг имкониятлари ва ўзига хос хусусиятлари”** деб номланади. Унда СОСАнинг функционал корпус тамойилларга мослиги, унинг афзалликлари, *СОСА* корпус таҳлили натижалари лингвوماданий бирликларнинг куантитатив ифодаси, *СОСА* лексикографияда лисоний маълумотлар тақдим этувчи луғат корпуси таҳлил этилган. Тадқиқот объекти ҳисобланган лисоний корпуснинг юқоридаги талабларни амалга ошириш имкониятларига эътибор берилди. Шунга кўра, замонавий инглиз тили корпусларидан бўлмиш иккинчи авлод вакили – *СОСА* танлаб олинди.

Марк Девизэс раҳбарлигидаги лойиҳа²⁰нинг самарали натижаси сифатида юзага келган 450+ миллион сўзли корпус, *СОСА – Corpus of contemporary American English* (Замонавий америка инглиз тили корпуси) америкача инглиз тилининг *ягона* йирик асосли (balanced) корпусидир. Уни ҳар ой 40 мингдан ортиқ индивидуал фойдаланувчилар ишлатади ва бу уни ҳозирги пайтда энг кўп ишлатиладиган онлайн корпусга айлантиради. Дизайни туфайли, бу лисоний корпус инглиз тилидаги мавжуд ўзгаришларни кўриб чиқиш учун ишлатилиши мумкин бўлган ягона йирик корпусдир. Яъни ҳозирда қарийб ўттиз йиллик тил маълумотларини ўзида жамлаганлиги боис унда шу вақт оралиғидаги тилнинг лексик қатламидаги ўзгаришларни кузатиб бориш мумкин, лойиҳа ҳали яқунланмаганлиги туфайли базага

²⁰[Mark Davies](http://davies-linguistics.byu.edu/personal/), Professor of Linguistics, [Brigham Young University](http://davies-linguistics.byu.edu/personal/). Available at URL:<http://davies-linguistics.byu.edu/personal/>

маълумотлар қўшиш давом этмоқда. *COCA* корпусига кириш English-Corpora.org сайти орқали бўлиб, унда мавжуд корпуслардан бепул фойдаланиш учун рўйхатдан ўтилади. Бу эса уни бугунги кунда энг кўп кириладиган сайтга айлантирган. Масалан, 2019 йил 1 октябрдан 31 октябргача бўлган муддатда сайтдан фойдаланувчилар сони 130 минг нафарга етган. Бир ойда юз мингдан ошган фойдаланувчиларни ўзига жалб этаётган ушбу сайт таркибидаги энг яхши ишланган лисоний корпус – *COCA*дир.

Бугунги кунга келиб эса, 2020 йилнинг март ойида сўнгги бор киритилган база ҳақидаги маълумотларга кўра, *COCA*нинг ҳажми 485202 матнлардан ташкил топган 1 миллиард 1 миллиондан ортиқ сўзни ўз ичига олади ва бу бўлиниш ҳар йил ҳисобига 30 миллион киритилиши натижасида 1990-2020 йиллар оралиғида шакллантирилган. Бу йиллар оралиғида корпус материаллари оғзаки (*spoken*), бадий (*fiction*), машҳур журналлар (*popular magazines*), газета (*newspapers*), илмий журналлар (*academic journals*), онлайн блоглар (*Blogs*), веб саҳифалар (*Web pages*), телевизор ва кино экраности ёзувлари (*TV/Movies subtitles*) каби саккиз соҳага ажратилиб, доимий равишда маълумотлар қўшиб борилган²¹:

*COCA*нинг функционал хусусиятларини очиқ бериш мақсадида унинг семантик, синтактик, фразеологик ва морфологик тадқиқотлардаги аҳамияти ўрганилган. Мисол тариқасида, *сабаб бўлмоқ* маъносида келадиган *cause* ва *result in* феълларининг семантик хусусиятларини таҳлил қилиш мақсадида *COCA* таркибида бирикиб келадиган сўзлар қидирувга берилди. Бунда *cause* ва *result in* феъллари билан кўп ишлатиладиган тўлдирувчи вазифасида келган отларни қидириш учун корпуснинг тегланганлиги қўл келди. Конкорданс жадвалида *cause* феъли 71078 марта ва *result in* феъли эса 84001 марта қўлланилгани маълум бўлди. *Cause+nn** ва *result in+nn** тартибида берилган қидирувлар натижалари шуни кўрсатдики, мос равишда 4394 ва 2907 ўринда ўзидан кейин от сўз туркумидаги тўлдирувчини олади. *Cause* феълининг тўлдирувчиларига эътибор берилганда, уларнинг деярли барчаси салбий маънони ифодаловчи отлар экани аниқланди. Масалан, *problems, cancer, trouble, harm, damage, disease, pain, illness, injury, stress, infection, discomfort* ва бошқалар. *Result in* феъли эса *death, loss, pressure, discrimination, rape, war, trauma* каби салбий маънони ифодаловчи отлар билан бир қаторда, *changes, chance, negotiations, improvements, power, quality* ва бошқалар каби ижобий тушунчалар билан ҳам бирикиб келиши аниқланди.

КЛ тарихида лисоний корпуслардан фойдаланиб, фразалар ва алоҳида сўзлар тадқиқини илк бор қўллаган олим Ж. Синклайр²² иккита турли фразаларни таҳлил қилади. Улар қуйидагилардир: *naked eye* ва *true feeling*. Ишида келтирилган ибораларнинг қўлланилишига оид мисолларни *BNC*

²¹ Corpus of Contemporary American English//English-corpora.org <https://www.english-corpora.org/coca/>

²² John Sinclair Trust the text: language, corpus and discourse. London: Routledge – 2004. 30-36

корпусидан олади ва *naked eye* учун 148 та мисол топилган бўлса, *true feeling* учун бу миқдор 53 тани ташкил этган.

Айнан шу бирикмалар *СОСА*дан қидирилганда бу сонлар сезиларли даражада ошгани кузатилди. Яъни мос равишда 654 ва 175 миқдоридаги натижаларга эришилди. Бу эса, ўз навбатида қидирувга берилган фразеологик бирикларни кенгроқ муҳокама этишга йўл очади.

Инглиз тилида сифат ясовчи *-ous* ва мавҳум от ясовчи *-ness* кўшимчалари мавжуд. Улар бир сўз таркибида келиши мумкинми, агар бирга кела олса, қандай тартибда жойлашади, деган саволларга жавоб *СОСА*да ўтказилган қидирув натижалари асосида берилди. Қидирув натижаларига кўра, **ness+ous* тартибидаги морфемалар бирикиши корпусда мавжуд эмас, бу орқали тилда ҳам йўқ, деган хулосага келдик. **ous+ness* тартибига мос келувчи натижалар 112 ҳолатда кузатилди, бу 10 дан ортиқ сўз-турларнинг умумий қўлланилган сонидир. Масалан, *consciousness, seriousness, nervousness, righteousness, graciousness, dangerousness* ва бошқалар.

Тадқиқотда корпус таҳлилларини линвокультурологик бирикларнинг қўлланилишига доир миқдорий мисолларни тақдим этадиган тадқиқот методлари сифатида қўллаш ўз-ўзидан ҳар томонлама ҳозирги замон талабларига мос келиши ўз исботини топди. Зеро, В.А.Маслова таъкидлаганидек, тил маданият билан ўзаро боғлиқ бўлиб, унинг таркибида тарқалади ва ривож топади, шунингдек, маданиятни ўзида акслантиради²³. С. Грайс ва А. Степановичларнинг “Когнитив лингвистикада корпора” номли монографиясида ҳам икки соҳа муштарақликда бир-бирини тўлдиради деб таъкидланган²⁴. Шу ўринда линвокультурологияни ҳам КЛ билан узвийлаштириш ўринли бўлади деб ҳисоблаймиз.

Барча танланган лисоний корпусларнинг ўхшаш дезайни, жанрлар тартиби, ва конкордансда маълумотларни бериш омиллари КЛ ва линвокультурология ўртасидаги алоқани таъминлашга хизмат қилиши аниқланди. Ушбу корпусларни бир-биридан шаклан фарқловчи хусусият – бу ҳажмдир. Бунда *BNC* ҳажм жиҳатидан *СОСА*даги сўзлар миқдори билан солиштирганда қарийб олти марта кичик миқдорга эга. Бундай ҳолат тадқиқотдаги қиёслаш жараёнига таъсир қилмайди, чунки частоталар БЧда эмас, балки МЧда берилади.

Корпуслардан олинган илк қидирув натижалари коллокацион бирикмаларни аниқлашга қаратилди. “СИФАТ + оила (_adj*+family)” тартибида берилган қидирув натижаларининг энг кўп бирикиб келадиган 300 таси ажратиб олинди ва қуйида маълум категорияларга бўлиб чиқилган ҳолда берилди ва *BNC*дан олинган натижалар қуйидагиларга ишора қилувчи гуруҳлар таркибида берилиши мумкинлиги асосланди: а) *оила ҳажмига*: large (1,44), small (0,72), average (0,60), big (0,55), normal (0,37), little (0,26); б) *аъзоларнинг келиб чиқишига*: royal (7,14), local (0,59), traditional (0,40), Jewish (0,35), holy (0,35), noble (0,31), black (0,26), British (0,26), Christian

²³ Маслова В.А. Лингвокультурология: учебное пособие. Академия изд., – Москва, 2004. – 208 с.

²⁴ Gries S.Th. and Anatol Stefanowitsch (eds.). Corpora in Cognitive Linguistics. Corpus Based Approaches to Syntax and Lexis. Trends in Linguistics: Studies and Monographs. – New York: Mouton deGruyter, 2006.– 352 p.

(0,26), human (0,25), poor (0,25), wealthy (0,21), working-class (0,19), farming (0,15), middle-class (0,13), English (0,18), French (0,11), Indian (0,11), Asian (0,10); в) *оила аъзолари мавжудлик таркибига*: extended (1,61), nuclear (1,36), immediate (0,89), private (0,23), individual (0,15), natural (0,18), birth (0,14), one-parent (0,11), single-parent(0,10), two-child (0,9); г) *оила аъзоларининг ўзаро муносабатига*: happy (0,69), close (0,68), joint (0,25), loving (0,15), ordinary (0,15), respectable (0,10), lovely (0,10), friendly (0,09), close-knit (0,09), ideal (0,06); д) *оила аъзоларининг ёшига*: new (1,10), old (0,70), young (0,60), growing (0,25), modern (0,15), contemporary (0,06).

Энди, СОСАда худди шундай қидирувни амалга ошириш орқали қиёсий планда ўхшаш ва фарқли томонларни кузатдик. Корпус натижалари юқорида келтирилган тартибда куйидагиларни ифодалашига кўра таснифланди: а) *оила ҳажмини*: large (0,86), big (0,85), average (0,56), small (0,52), little (0,40), median (0,36), normal (0,24), stable (0,10); б) *аъзоларнинг келиб чиқишини*: royal (3,28), American (2,01), black (0,97), traditional (0,66), holy (0,55), dysfunctional (0,46), human (0,46), Kennedy (0,62), Bush (0,45), Jewish (0,44), poor (0,34), white (0,31), middle-class (0,30), wealthy (0,30), military (0,28), Christian (0,22), immigrant (0,19), African-American (0,14), working-class (0,13), suburban (0,10), religious (0,10), Muslim (0,09), farming (0,08), free (0,08); в) *оила аъзолари мавжудлик таркибини*: extended (3,37), immediate (1,08), nuclear (0,91), host (0,21), adoptive (0,18), private (0,17), biological (0,12), two-parent (0,12), personal (0,09), single-parent (0,09), broken (0,08), surrogate (0,07); г) *оила аъзоларининг ўзаро муносабатини*: happy (0,61), close (0,12), great (0,41), loving (0,40), perfect (0,24), wonderful (0,23), close-knit (0,23), joint (0,20), typical (0,18), supportive (0,15), individual (0,15), devoted (0,10), ideal (0,09), tight-knit (0,07), lovely (0,06), beloved (0,06); д) *оила аъзоларининг ёшини*: new (1,19), old (0,76), modern (0,58), young (0,49), growing (0,25), younger (0,07).

Юқорида келтирилган мисоллар иккала корпусдан олинган натижалар асосида шакллантирилган бўлиб, таснифланган гуруҳларда келтирилган сифатлар “оила” тушунчасини изоҳлаб келади. Қиёсий планда ўрганишга имкон бериши учун уларнинг миллиондан бир нисбатдаги МЧси қавс ичида берилди ва маълумотларда ўзаро ўхшашлик ва фарқли томонлари кўзга ташланади. Масалан, оила аъзоларининг ёшини билдирувчи сифатлар таркибига “new – янги” сўзини СОСА таркибидан топилган мисоллар сони, яъни БЧлари 680 та ва ВНСда қидирув натижалари эса 110 та ўринда кузатилганидан, демак, СОСАда бу сўз бирикмаси кенгроқ ишлатилган деган хулоса мутлақо нотўғри. Уларнинг биз ҳисоблаб чиққан МЧлари СОСА учун 1,19 бўлса, ВНС учун 1,10 ни кўрсатди. Энди, корпусларда ва улар орқали тил қатламларида *new family* бирикмаси ўзаро яқин (америкача инглиз тилидаги озроқ устунлик билан) частотада қўлланилишини англаш мумкин. Шунингдек, *ёш оила* тушунчасини “*young – ёш*” сўзи билан ифодалашларини кузатдик. Бироқ, бунда устунлик британияликлар (ВНС ва СОСАда мос равишда 0,60 ва 0,49 МЧ кўрсатилди) тил қатламида бўлди. Бир-биридан фарқланишни келтириб чиқарган сифатлар кўриб чиқилди ва уларнинг

кенроқ контекстдаги эмирик таҳлиллари асосида корпусга асосланган таҳлил натижалари у ифодалайдиган тил ҳамда тил соҳибининг лингвомаданий тушунчаларни англашда қўлловчи лингвистик бирликларнинг улкан базаси сифатида қаралиши мумкин, деган хулосага келинган.

КЛ ва лингвокултурология ўртасидаги муносабат ҳақида сўз кетар экан, албатта, танлаб олинган тушунчага оид маъно кўчишининг метафора каби стилистик қурилмаларининг ҳам ўрганилиши аҳамият касб этади. Бу ўринда СОСА форматида KWIC орқали *family* – *оила* сўзи қидирувга берилганда, бир қадар кенроқ куйидаги метафоралар топилди: *family is responsibility*, *family is a blessing*, *families are building blocks*, *my family is my life*, *family is a source*, *family is the best experience*, *family is a school*, *family is a domestic church* ва *family is home*. СОСА: *family loyalty* (оилавий садоқат), *family status* (оилавий аҳвол), *family honor* (оилавий ғурур), *family value* (оилавий қадр-қиммат), *family environment* (оилавий муҳит), *family concerns* (оилавий ташвишлар), *family secrets* (оилавий сирлар), *family circle* (оила доираси), *family gathering* (оилавий йиғилиш), *family planning* (оилани режалаштириш), *family breakdown* (оиланинг бузилиши), *family involvement* (оиланинг иштироки), *family love* (оилавий муҳаббат).

Мисоллар таҳлили шуни кўрсатадики, *BNC* дан олинган фикрлар, асосан, оилани шакллантириш, унинг ҳажми, жамиятдаги ўрни ва аҳамиятига қаратилган бўлиб, уларни умумлаштирувчи тушунча бу оила қуриш ва уни шакллантириш анъаналарига таяниш ҳамда жамиятнинг бўлаги сифатида қарашга асосланган. Бунда оила – бу қадриятларга асосан қурилган, жамиятда ўрни ва ролига эга яқин инсонлардан шаклланган кичик қобик сифатида қаралади. СОСАдан олинган маълумотлар эса оиладаги муҳит, ички ўзаро муносабатларнинг ижобий бўлиши, меҳр ва инсон омилини қадрлаш каби ўзаро яқин қарашларни кўрсатди.

Тадқиқотда замонавий инглиз тилининг мавжуд бир тилли электрон луғатларининг фойдаланувчига тақдим этиши мумкин бўлган имкониятлари таҳлилга тортилди. Бунда ихтиёрий сўз танланди ва унинг инглиз тили луғатларидан қидириш натижасида олинган натижалар таҳлили ижобий эмаслиги аниқланди. Танланган '*fabric*' ихтиёрий сўзи СОСА таркибидан қидирилганда олиниши мумкин бўлган маълумотлар луғатлар тақдим эта оладиган ахборотлардан кескин ижобий томонга фарқланиши аниқланди.

ХУЛОСАЛАР

1. Илк тарихий ривожланиш босқичи XIX асрга бориб тақаладиган КЛ ўз олдида лисоний корпусларни йиғиш, қайта ишлаш ва улар устида лингвистик тадқиқотлар ўтказиш методларини мужассамлаштирувчи илмий тадқиқот методологияси эканлиги аниқланди. Шунингдек, жаҳон тилшунослигида соҳанинг бугунги ривожини ижобий ва жадал тус олмоқда. Хусусан, Ўзбекистонда бугунги кунда КЛга оид тадқиқотлар кўлами юқори даражада олиб бориляпти. КЛ диахроникаси Бровн корпусининг

яратилишидан бугунги кун электрон корпуслари тараққиётигача бўлган тараққиёт босқичлари даврий, назарий ва таснифий кўриб чиқилди.

2. КЛнинг мустақил тараққиётида компьютер лингвистикаси, математик лингвистика ва бошқа тилшунослик соҳалари асос бўлган. КЛ – ўз ривожланиш босқичининг илк даврида тадқиқот методи сифатида қаралгани ва уни компьютер ва математик ҳисоб-китобларга қаратилган фан тармоқларининг қисми ёки воситаси сифатида қаралгани, бугунги кунда ўзининг мақсад, вазифалари ва тадқиқот методларига эга методология бўлиб улгургани ҳамда кенг илмий жамоатчилик томонидан қабул қилингани тадқиқот мобайнида аниқ бўлди. Келажакда эса тез суръатда ривожланиб бораётган КЛни барча соҳадаги муаммоларга ечим топа оладиган лингвистик илмий-тадқиқот парадигмаси ва ёки мустақил фан сифатида барча хусусиятларга эга бўлишини юқори даражали илмий фараз каби баҳолаш ўринли.

3. Лисоний корпусларни шакллантиришда корпус матнларига қўйиладиган мезон талаблари сифатида қуйидагиларни белгилаш мумкин: 1) аутентик материал; 2) матн бирликларининг замон ва макондаги мутаносиблиги; 3) аниқ белгиланган туркумларга ўзаро тенг ҳажмда тўпланганлиги; 4) мақсадга йўналганлик омили ифодаси сифатида изланувчи очиб бермоқчи бўлган тил ёки суб-тил таркибидан йиғилган бўлиб, унинг репрезентативлик тамойилига мос келиши; 5) оғзаки ёки ёзма бўлишидан қатъи назар барча матн транскрипция шаклида киритилиши шарт, бунда ҳеч қандай ўзгартириш ёки тўғрилаш каби таҳририй жараёнларга тортилмаслиги; 6) умумий лексик миқдорий ҳажми бир миллиондан кам бўлмаслиги. Ушбу мезон талаблари корпус тамойилларидан келиб чиққан ҳолда яратилади.

4. Корпус – корпус тузувчининг прагматик мақсадидан келиб чиқиб, қатъий тамойиллар асосида шакллантириладиган ва таснифланадиган, репрезентативлик мезони талабига жавоб бера оладиган сараланган етарли ҳажмдаги лисоний матнлар (оғзаки ва ёзма) тўплами бўлиб, ўзида қайта ишлаш, теглаш каби принципларни мужассамлаштирган асосан компьютер ёрдамида жараёнлари автоматлаштирилган мукамал тизимли қидирув, ҳаволали натижа ва эмпирик таҳлилга қулай етарли миқдор ҳамда изчилликка эга базадир.

5. Корпусларни таснифлаш ҳам КЛ олдида турган характерли муаммолардан биридир. Таснифлашга нисбатан ёндашувлар турлича бўлиб, улар асосан корпуснинг ҳажм, шакл, дизайн, тил ва шу каби хусусиятларига кўра олиб борилганлигини кўриш мумкин. Корпусларни фойдаланувчининг прагматик мақсади нуқтаи назаридан таснифлаш тамойилига кўра, мавжуд корпус терминларидан фойдаланган ҳолда лисоний корпусларни қўллаш мумкин бўлган соҳалар каторига кирувчи: а) таржима амалиёти ва чоғиштира лингвистика; б) лексикография; в) адабиёт назарияси; г) лингводидактика ва тилларни ўқитиш; д) этимология ва тил тарихи тадқиқи каби тармоқларни киритиш мумкин. Зеро, корпусларнинг якуний стандарт шаклидаги таснифланишига нисбий тушунча сифатида қараш

ўринли, чунки ҳар қандай корпус яратилганда ўзи билан янги мақсад, янги дизайн ва янги таснифий турни фанга киритади.

6. Ишда КЛ лингвистик тадқиқотлар методологияси сифатида қаралиб, унинг тадқиқот методларининг изчилликка асосланган, ўз ичига аниқ жараёнларни олувчи олти босқичли методлар гуруҳи: йўналтирувчи методлар, манба билан ишлаш ёки базавий методлар, метамаълумот қўшиш методлари, автоматлаштириш методлари, баҳолаш методлари ва частотали кидирув-таҳлил методлари такомиллаштирилди.

7. Корпус таҳлиллари натижаларини қиёслаш жараёнини уч кўринишда амалга ошириш мумкин: 1) ички қиёслаш – бир корпуснинг ўзидан олинган бир неча кидирув натижаларини ўзаро таққослаш тушунилади. 2) бир тилли корпуслараро қиёслаш – бир тилни акс эттирувчи бир неча корпуслардан олинган, асосан, аниқ белгиланган лисоний бирлик(лар)нинг кидирув натижаларини ўзаро таққослаш демакдир. 3) турли тилли корпуслараро қиёслаш – ҳар хил тил асосида яратилган бир хил таснифга эга корпуслар ёки параллел (асосан, таржима амалиёти намуналаридан ташкил топган) корпусларда танланган лисоний бирлик ва унинг аниқ ёки эквивалент таржимаси белгиланган ҳолда кидирув натижаларининг таққосланишидир.

8. *СОСА*нинг функционал хусусиятлари лисоний бирликларнинг ўзаро семантик фарқланиши корпус таҳлил натижалари асосида аниқланади. Улар сирасига семантик бирикмаларни ўрганишга қаратилган изланишлар учун корпуснинг тутган ўрни киритилиши мумкин. Синтактик тадқиқотларда *СОСА*дан олинган мисоллар феъл шакллари ва уларнинг ўзаро грамматик муносабати ҳамда синтактик қурилмалар хусусида кенг маълумот манбаидир. *СОСА*да фразеологик бирликлар тадқиқи учун зарур *KWIC* тизими орқали бирикиши мумкин бўлган сўзлар ва келиши мумкин бўлган орфографик белгилар таҳлили имкониятлари мавжуд. Морфологик бирликлар тадқиқида *СОСА* амалий жиҳатдан морфологик бирликларнинг шаклан ва бирикма ҳолатида инглиз тилида мавжуд ёки мавжуд эмаслигини кўрсатади.

9. Замоनावий инглиз тили корпусларидан бўлмиш иккинчи авлод вакили *СОСА* таркибан лингвомаданий бирликларнинг тилдаги ифодасини аниқлашга доир миқдорий мисоллар манбаидир. Америка ва Британия инглиз тилларида лингвомаданий тушунча ва бирликларни чоғиштирма аспектда қиёслашни *BNC* ва *СОСА* таҳлиллари асосида олиб бориш асослидир. Корпус таҳлилларида муайян тушунча (концепт) танлаб олинади ва унга оид конкорданс мисоллари бир тилли корпуслараро қиёсий ўрганиб чиқилади. Маданий бирликни ифодалашга керак бўлган барча маълумотлар йиғилиб, икки маданиятда «*оила*» тушунчасини англашдаги ўзаро ўхшаш ва фарқли жиҳатлар аниқланди.

10. *СОСА*га луғатлар яратиш жараёнида луғат корпуси сифатида қараб, унинг мавжуд функциялари замонавий инглиз тили бир тилли изоҳли луғатларининг ички имкониятлари билан солиштирилди. Бунда корпуснинг фойдаланувчига: 1) сўзнинг изоҳи ва у билан бирикиб келадиган сўзлар

рўйхатини; 2) жанрлар асосида бўлиб берилган таҳлил натижаларини; 3) маълумотларни график кўринишларда тақдим этиши; 4) тил қатламидаги неологизмлар ва уларнинг пайдо бўлиш этимологиясининг тилдаги талқинини; 5) сўзнинг қўлланилишига оид кўп сонли мисолларни; 6) предлогли бирикмаларнинг тўғри бирикуви ҳамда лингвистик бирликларнинг синтактик, талаффуз ва пунктуацион характерларини ўрганишда аниқ статистик маълумотлар ва бошқаларни тақдим эта олиши аниқланди.

**НАУЧНЫЙ СОВЕТ DSc.03/30.12.2019.Fil/Ped.27.01
ПО ПРИСУЖДЕНИЮ УЧЕНЫХ СТЕПЕНЕЙ ПРИ УЗБЕКСКОМ
ГОСУДАРСТВЕННОМ УНИВЕРСИТЕТЕ МИРОВЫХ ЯЗЫКОВ**

**УЗБЕКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ МИРОВЫХ
ЯЗЫКОВ**

АТАБОЕВ НОЗИМЖОН БОБОЖОН УГЛИ

**ФУНКЦИОНАЛЬНЫЕ ХАРАКТЕРИСТИКИ АНГЛИЙСКИХ
КОРПУСОВ (на примере СОСА)**

**10.00.04 – Языки и литература народов
Европы, Америки и Австралии**

**АВТОРЕФЕРАТ
диссертации доктора философии (PhD) по филологическим наукам**

ТАШКЕНТ – 2020

Тема диссертации доктора философии (PhD) зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за В2019.1.PhD/Fil723.

Докторская диссертация выполнена в Узбекском государственном университете мировых языков.

Автореферат диссертации на трех языках (узбекский, русский, английский резюме)) размещен на веб-странице Научного совета по адресу (www.uzswlu.uz) и на информационно-образовательном портале «ZiyoNet» по адресу (www.ziyo.net).

Научный руководитель: Тешабаева Дилфуза Муминовна
доктор филологических наук, профессор

Официальные оппоненты: Рахимов Ганишер Худайкулович
доктор филологических наук, профессор

Хамроева Шахло Мирджоновна
доктор философии (PhD) по филологии

Ведущая организация: Бухарский Государственный Университет

Защита диссертации состоится «19» ноября 2020 года в 14⁰⁰ часов на заседании Научного совета DSc.03/30.12.2019.Fil/Ped.27.01 по присуждению ученых степеней при Узбекском государственном университете мировых языков (Адрес:100138, город Ташкент, Учтепинский район улица Кичик халка, дом 21А. Тел.: (+99871) 230-12-92; факс: (+99869) 227-07-61; e-mail: uzswlu_info@mail.ru)

С диссертацией можно ознакомиться в Информационно-ресурсном центре Узбекского государственного университета мировых языков (зарегистрирована № 12). (Адрес: 100138, город Ташкент, Учтепинский район улица Кичик халка, дом 21А. Тел.: (+99871) 230-12-92.)

Автореферат диссертации разослан «29» октября 2020 года.

(Реестр протокола рассылки № 12 от «29» октября 2020 года).


Г.Х. Бокиева
Председатель Научного совета по присуждению ученых степеней, д.ф.н., профессор


Х.Б. Самитова
Ученый секретарь Научного совета по присуждению ученых степеней, д.ф.н., доцент


Ж.А. Якубов
Председатель Научного семинара при Научном совете по присуждению ученых степеней, д.ф.н., профессор

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и востребованность темы исследования. В 60-е годы прошлого века в мировой лингвистике усилилось взаимодействие корпусной лингвистики, которая занимается созданием и внедрением лингвистического корпуса, с компьютерной техникой. В начале XIX века в результате формирования теоретических основ корпусной лингвистики в области лингвистики был создан первый образец – Корпус Брауна. В результате автоматическая оцифровка интеллектуального богатства человечества на основе электронного корпуса стала сегодня одной из актуальных задач современной лингвистики. Именно поэтому актуальным является совершенствование электронной языковой базы, отражающей язык или определенную часть языка, объединившей письменные и устные тексты – национальных корпусов, являющихся практическим отражением методов анализа, статистического анализа, демонстрирующего развитие языка, пополнение лексического богатства и его современное состояние.

В то время как качество лингвистических исследований в мире основано на статистическом анализе, эффективный результат практики создания языкового корпуса в сфере позволяет сэкономить время, затрачиваемое на анализ текста. Своеобразие лингвистических единиц, выявление в конкордансе тысяч примеров их особенностей в речи является первичной особенностью корпуса. В связи с этим, поскольку наука, вступая в третье десятилетие XXI века, ставит необходимость внедрения компьютеров на практике, научно-теоретические исследования корпусной лингвистики приобретают все большее значение во всех областях.

Сегодня в нашей стране проводятся масштабные реформы, и интеграция нашей страны в мировое сообщество определяет такие важные задачи, как подготовка специалистов, свободно владеющих иностранными языками. “Нам необходимо осваивать цифровые знания и современные информационные технологии, чтобы добиться развития. Это позволит нам выбрать кратчайший путь к восхождению. Ведь сегодня в мире информационные технологии проникают глубоко во все сферы”¹. Таким образом, поскольку область исследований рассматривается как развитие современной науки, ее значимое освоение на пути прогресса нашей страны отражает потребности сегодняшнего дня. При этом обозначены такие важные задачи, как разработка теоретического исследования функциональных свойств существующего корпуса английского языка на основе анализа эволюции исторического развития корпусной лингвистики, корпуса, его классификации, принципов применения.

Данная диссертация в определенной степени служит реализации задач, обозначенных в Указе Президента Республики Узбекистан №УП-4947 «О Стратегии действий по дальнейшему развитию Республики Узбекистан» от 7 февраля 2017 года, №УП-5847 «Об утверждении Концепции развития

¹ Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг Олий Мажлисга Мурожаатномаси. – President.uz сайти, 2020 йил 24 январь. <https://president.uz/uz/lists/view/3324> [мурожаат санаси: 17.02.2020]

системы высшего образования Республики Узбекистан до 2030 года» от 8 октября 2019 года, Постановлении Президента Республики Узбекистан №ПП-3775 «О дополнительных мерах по повышению качества образования в высших образовательных учреждениях и обеспечению их активного участия в осуществляемых в стране широкомасштабных реформах» от 5 июня 2018 года, Постановлении Кабинета Министров №610 «О мерах по дальнейшему совершенствованию качества преподавания иностранных языков в образовательных учреждениях» от 11 августа 2017 года, Послании Президента Республики Узбекистан палатам Олий Мажлиса от 24 января 2020 года, а также в других нормативно-правовых актах.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Диссертация выполнена в соответствии с I приоритетным направлением развития науки и технологий республики «Формирование системы инновационных идей и способов их реализации в социальном, правовом, экономическом, культурном, духовном и образовательном развитии информационного общества и демократического государства».

Степень изученности проблемы. В мировой лингвистике Г. Лич был одним из первых, кто провел научные исследования по вопросам возникновения и развития корпусной лингвистики². Исследования по формированию современной корпусной лингвистики отражены в исследованиях таких английских ученых, как С. Хьюстон, Д. Бибер, С. Грайс, М. Халлидай и В. Тауберт³. В работах русских ученых В. Баранова и В. Захарова, всесторонне освещена специфика корпусной лингвистики, в трудах Н. Владимова, С. Вадяева, О. Горкиной, В. Мамонтовой и А. Магаметовой – такие понятия, как “корпусная лингвистика” и проблемы её классификации⁴. В исследовании таджикской ученой Д. Джафаровой проведен сравнительный анализ корпусной лингвистики с компьютерной лингвистикой и другими схожими областями⁵.

² Leech G. Corpora And Theories Of Linguistic Performance. In J.Svartvik (ed.). 1992. – 105–122 pp.

³ Hunston S. Corpora in Applied Linguistics: Cambridge University Press, 2002. – 234 p.; Halliday M.A., Teubert W., Yallop C. and Cermakova A. Lexicology and Corpus Linguistics, – London, 2004. – 184 p.; Gries S.Th. Introduction to S. Gries and A. Stefanowitsch (eds.). Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis. – Berlin: Heidelberg, New York: Mouton de Gruyter, 2006. – 1-18 p.; Douglas B., Ulla C. and Thomas A.U. Discourse On The Move. – Amsterdam, 2007. – 290 p.

⁴ Баранов А.Н. Компьютерная лингвистика // Введение в прикладную лингвистику: учебное пособие. – М.: Едиториал УРСС, 2003. – 320 с.; Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие / В.П. Захаров. – СПб, 2005. – 48с. – Режим доступа:<http://download.yandex.ru/class/zakharov/plan.pdf>; Владимов Н.В. Корпусный подход к решению переводческих проблем (на материале английского и русского языках). Дис...канд. филол. наук. – Москва: РГБ, 2005. – С 26.; Вадяев С.Е. Лингвистические принципы построения и использование корпуса текстов для исследования официально-делового стиля современного немецкого языка (на материале электронного корпуса «DER») диссер... для кандидата филол. наук. – Самара, 2005. – 160 с.; Горина О.Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке. Дисс. ... канд. пед. наук. – Москва, 2014. – 321с.; Мамонтова В.В. Особенности перевода сложносоставных слов с английского языка на русский (на материале корпуса публицистических текстов). Дис. ... канд. фил.наук. – Ставрополь: 2008. – 139 с.; Магомедова А.Н. Корпусная лингвистика и контекстное разрешение лексической многозначности слов. Дис. ... канд. филол. наук – М.: РГБ, 2005. – 132 с.

⁵ Джафарова Д.Ф. Модели лингвистического анализа текстов таджикского языка (на материале газелей Хафиза) дисс. ... канд. фил. наук. – Душанбе, 2013. – 151 с.

Работы А. Килгаррифа, К. Изтока, Ф. Карлсона и И. Соженина посвящены применению корпусных анализов в лексикографии⁶, а В. Рыков, А. Романов и В. Плуноян разработали метод использования понятия “корпуса” в текстовых анализах⁷.

В Узбекистане сегодня проводятся масштабные научные исследования по применению методов корпусной лингвистики, использованию существующих корпусов. Ж. Жумабоева акцентирует внимание на нахождении семантической градации в англоязычных корпусах, в частности, BNC⁸, Г. Эргашевой исследованы гендерные концептуальные метафоры на основе BNC корпуса, Г. Собирова обращает внимание на применение BNC корпуса в преподавании языка. Ш. Хамроева ищет решения проблемы авторского корпуса⁹. Кроме того, в настоящее время реализуются фундаментальные проекты, посвященные национальному корпусу узбекского языка, лингвистическим корпусам, выражающим различие в плане возраста и физиологии.

Связь диссертационного исследования с планами научно-исследовательских работ высшего образовательного учреждения, где выполнена диссертация. Исследование диссертации является составной частью темы исследования Узбекского государственного университета мировых языков «Проблемы перевода» и входит в научные планы факультета теории и практики перевода.

Цель исследования – выявить функциональные особенности существующих корпусов английского языка на примере корпуса COCA на основе анализа эволюции исторического развития корпусной лингвистики, корпуса, его классификации, принципов применения.

Задачи исследования:

диахроническое изучение корпусной лингвистики и интерпретация ее особенностей, различных толкований и описаний термина корпус;

⁶ Kilgarriff A. and Kosem I. Corpus Tools For Lexicographers. – Trojina: Institute for Applied Slovene Studies, 2012. – 37 p. УРЛ: https://www.researchgate.net/publication/264884809_Corpus_tools_for_lexicographers; Karlsson F. Lexicography And Corpus Linguistics. – Helsinki: Department of General Linguistics University of Helsinki, 1992. УРЛ: <http://www.ling.helsinki.fi/~fkarlsson/INTJLLEX.pdf>; Kilgarriff A. Using Corpora [And The Web] As Data Sources For Dictionaries. – Helsinki: 2012. – 21 p. https://www.sketchengine.eu/wp-content/uploads/Using_corpora_2012.pdf; Саженин И.И. Корпусные методы в лексикографии: опыт создания модели Словарного корпуса. Диссертация на соискание ученой степени кандидата филологических наук. – Новосибирск – 2013. 193 с.

⁷ Рыков В.В. Корпус текстов – новый тип словесного единства Текст. / В.В. Рыков. – Электрон, изд. – Режим доступа к изд.: <http://rykov-cl.narod.ru/t31.html>.; Романов А.С. Методика и программный комплекс для идентификации автора неизвестного текста: Автореф. дисс....канд. тех. наук. - Томск, 2010. 26 с.; Плуноян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики// Русский язык в научном освещении.–Москва: №2., 2008. – С. 7-20.

⁸ BNC (British National Corpus) – Британия Миллий Корпуси: 100 миллион сўзли лисоний корпус.

⁹ Сабирова Г. Умумтаълим мактаблари ўқувчиларига инглиз тили олмошларини ахборот технологиялари асосида ўргатиш методикаси: Филол. фалс. д-ри (PhD) ...дисс. автореф. – Тошкент: 2017. – 41 б. URL: <http://uzswlu.uz/media/ba7be50d-84f1-c826-f935-6e7610cb6bf5.pdf>; Эргашева Г. Турли тизимдаги тилларда гендерга оид терминологиянинг шаклланишида лингвистик ва экстралингвистик омиллар: Фил.фан.д-ри.дисс. автореф. – Тошкент: 2018. – 61 б. URL: <http://demo.navoiy-uni.uz/sites/default/files/ziyonet.pdf>; Джумабаева Ж. Ўзбек ва инглиз тилларида лексик ҳамда стилистик градуонимия. докторлик диссертацияси автореферати. – Тошкент: 2016. – 82 б. URL: http://tashgiv.uz/wp-content/uploads/2017/01/avtoref_jumabayeva.pdf; Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Ф.б. фалс. д-ри (PhD) ... дисс. автореф. – Қарши: 2018. – 52 б. URL: <https://e-books.tsue.uz/pdf-ozbek-tili-mualliflik-korpusini-tuzishning-lingvistik-asoslari>

совершенствование методологии научного исследования корпусной лингвистики, связанной с созданием корпуса и его применением;

выделение лингвистических особенностей корпусной лингвистики путем создания экспериментального корпуса;

создание частотно-аннотированного терминологического словаря корпусной лингвистики по результатам корпусного анализа;

раскрытие практического значения существующих корпусов английского языка;

проведение анализа данных по истории создания корпуса СОСА¹⁰, его структурного строения, семантических, синтаксических и морфологических исследований, которые он может предоставить;

раскрытие важности информации от корпуса СОСА для понимания лингвокультурных концепций носителями разных языков;

определение веса лингвистических статистических сведений которые СОСА может предоставить в виде корпуса словарей.

Объектом исследования выбраны корпуса английского языка, в частности, корпус современного американского английского – СОСА.

Предмет исследования – функциональные особенности корпусов английского языка на примере корпуса СОСА.

Методы исследования. Для освещения темы исследования использовались методы контент-анализа, классификации, характеристики, сравнения, статистического, когнитивного и корпусного анализа.

Научная новизна исследования состоит в следующем:

в результате статистического изучения истории корпусной лингвистики, этапов развития, проблем классификации были улучшены критерии создания корпуса, такие как моделирование, целенаправленность и ограниченный объем; создание корпуса по шести методологическим системам и механизм его применения на практике доказан на основе научных и теоретических данных;

терминологический аппарат корпусной лингвистики дополнен отраслевым частотно-аннотированным терминологическим словарем, созданным на основе экспериментального анализа корпуса, выполненного с помощью программы AntKonk, которая автоматически создает корпус из текстов различных форматов;

доказана возможность сравнения лингвокультурологических единиц с использованием метафор, словосочетаний (семантических комбинаций) и статистического анализа фразеологических комбинаций в соответствии с репрезентативной природой корпусов (British National Corpus – BNC и Corpus Of Contemporary American English – СОСА);

обосновано, что корпус СОСА, сначала изученный как корпус словарей, обеспечивает фонетические, орфографические, грамматические и лексико-

¹⁰ СОСА (Corpus Of Contemporary American English) – *Замонавий Америка инглиз тили корпуси: Йиллик ҳисобда 20 миллион ҳажмдаги матнларни ўз ичига олиб борадиган жанрлараро мувозанатли корпус*

семантические свойства лексических единиц в области лексикографии, а также ее важность для улучшения английских словарей.

Практические результаты исследования следующие:

усовершенствована система методов корпусной лингвистики;
разработан частотно-аннотированный терминологический словарь корпусной лингвистики;

выявлено проявление межкорпусных лингвокультурных концепций;
с научной и практической точки зрения проанализировано положительное применение корпуса СОСА в области лексикографии.

Достоверность результатов исследования определяется обоснованностью выводов, четкой постановкой проблемы, эффективностью подходов и методов, используемых в работе.

Научная и практическая значимость результатов исследования. Научная значимость исследования определяется тем, что результаты исследования могут быть использованы в качестве научно-теоретического источника при разработке теоретических основ для создания и совершенствования лингвистического корпуса, проведения исследований в области корпусной лингвистики.

Хотя практическое значение исследования сегодня рассматривается как методология научных исследований, оно определяется тем, что будет служить теоретическим источником для программирования, планирования и представления тем в процессе обучения корпусной лингвистике, которая, как ожидается, в ближайшем будущем будет развиваться как научный предмет.

Внедрение результатов исследования. На основе результатов исследования и разработанных практических рекомендаций по функциональным особенностям англоязычных корпусов (на примере СОСА):

результаты анализа формирования общих требований к корпусам посредством изучения этапов развития концепции корпусной лингвистики в мировой и узбекской лингвистике, формирования экспериментальных школ, научно-теоретического изучения критериев и принципов классификации применены при подготовке сборника аудиоданных «Сводка статей по лингвистике, опубликованных в 2014-2016 гг.» («По материалам журналов «Узбекский язык и литература» и «Вопросы филологии») в рамках практического грантового проекта А-1-61 «Создание аудиокниг по лингвистике для слабовидящих студентов», реализованного на основе государственных, научно-технических проектов (Справка Министерства высшего и среднего специального образования № 89-03-2032 от 12 июня 2020 г.). В результате решены проблемы создания корпуса, совершенствования систематических методов их применения на практике;

научные выводы и рекомендации по совершенствованию терминологической системы корпусной лингвистики путем изучения корпусной лингвистики в качестве целостной системы на основе эмпирического анализа и отраслевого частотно-аннотированного терминологического словаря, созданного в результате экспериментального анализа корпусов с помощью программы AntKonk применены в

инновационном проекте «Создание электронного поэтического словаря узбекского и английского языков (облик, характер, природа человека и изображение национальных символов)» Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои (Справка Министерства высшего и среднего специального образования № 89-03-2032 от 12 июня 2020 г.). Применение научных результатов усовершенствовало принципы интерпретации терминов в этой области;

практические предложения и рекомендации по сравнению лингвокультурных единиц с использованием статистического анализа межкорпусных (на примере корпуса BNC и COCA) метафор, словосочетаний (семантических сочетаний) и фразеологических сочетаний согласно репрезентативности корпуса реализованы в рамках фундаментального проекта ОТ-Ф1-18 – «Разработка методов и методологии формирования массовой языковой культуры» (Справка Министерства высшего и среднего специального образования № 89-03-2032 от 12 июня 2020 г.). Применение этих результатов позволило посредством анализа корпуса глубже понять лингвокультурные концепции, возникающие через призму языка;

улучшенные научные выводы, основанные на том факте, что корпус COCA впервые изучен как корпус словарей, и все статистические и практические результаты выявили особенности улучшения английских словарей, были обсуждены на научном семинаре с участием профессор-преподавателей кафедры «Перевод и лингвистика» Барселонского университета Помпеу Фабра и представителей исследовательской группы GEDIT (Grup d'Estudis del discurs I de la Traduccio') и использованы в диссертациях магистрантов Университета (Справка кафедры «Перевод и коммуникация» Университета Помпео Фабра, 24 января 2020 г.). Как следствие, данный метод может быть использован при написании кандидатских и докторских диссертаций.

Апробация результатов исследования. Результаты исследования прошли общественное обсуждение в рамках выступлений на 2 международных и 10 республиканских научно-практических конференциях.

Опубликованность результатов исследования. Всего по теме диссертации опубликовано 15 научных работ, в том числе 1 частотно-аннотированный терминологический словарь, 14 статей в научных изданиях, рекомендованных к публикации основных научных результатов докторских диссертаций ВАК Республики Узбекистан, в том числе 8 – в национальных и 6 - в зарубежных журналах.

Структура и объем диссертации. Диссертация состоит из введения, трех глав, заключения, списка использованной литературы и приложения, общий объем составляет 189 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обосновывается востребованность и актуальность научного исследования, определяются цели и задачи, объект и предмет

исследования, показано соответствие актуальность научной работы приоритетным направлениям развития науки и технологий. Также приводится информация о научной новизне диссертации, практических результатах и их достоверности, теоретической и практической значимости работы, внедрении результатов в практику, публикации в научных изданиях, структуре работы.

В первой главе диссертации «**Этапы развития концепции корпусной лингвистики в мировой и узбекской лингвистике**» проанализированы результаты диахронных исследований в корпусной лингвистике (КЛ), зарубежные и отечественные исследования, посвященные специфике КЛ и проблемам классификации понятия корпус.

КЛ – это раздел языкознания, который занимается созданием и применением лингвистических корпусов¹¹. Чтобы глубже понять КЛ, цель которой самостоятельно создавать лингвистические корпуса, следует взглянуть на историю его развития. Термин связан с развитием корпусной практики, созданной в 1960-х годах и применяемой с 1980 года на основе компьютерных технологий¹².

Этапы развития КЛ неразрывно связаны с теоретическими положениями корпуса. Хотя исследования КЛ в языкознании и организации преподавания языков начались в XVIII веке, первый электронный лингвистический корпус текстов появился в 60-х годах прошлого столетия. В результате проекта, начатого в 1961 году, в 1964 году ученые Брауновского университета (США) создали большие объемы текста на компьютере (Brown Corps). У. Фрэнсис и Г. Кучера создали пятьсот две тысячи прозаических экземпляров американского издания на английском языке. Тексты, относящиеся к пятнадцати основным жанрам английской прозы, были сформированы на основе материалов корпуса прессы США и их базовой статистической обработки. В 1961 году первые полученные из корпуса результаты были опубликованы в виде частотного и альфа-частотного словаря. По мнению Х. Линга¹³, такие объемные корпуса использовались только при изучении определенных аспектов фонетики.

Согласно Н. Хомскому, лингвист должен основывать свою языковую модель на компетенции (competence), за исключением производительности (performance). Он опубликовал свою теорию в 1950-1960 гг., то есть официально отклонил применение корпуса к лингвистике. Таким образом, период до этого года стал называться пре-Хомским в истории КЛ. Основываясь на своем опровержении, он дает свое мнение об использовании глагола *perform* в английском языке и показывает, что охват корпуса низок по сравнению с возможностями корпуса Брауна. В настоящее время существует множество объёмных корпусов, что даёт возможность

¹¹ Корпусная лингвистика как раздел языкознания URL: <https://myfilology.ru/177/korpusnaya-lingvistika-kak-razdel-yazykoznanija/>

¹² https://ru.wikipedia.org/wiki/Корпусная_лингвистика

¹³ Ling X. Cihui Yu.he Jisuan Yu. [Lexical Semantics And Computational Linguistics]. BeijingLingua Publishing house, 1999. – P. 240.

бесконечного продолжения *ряда словосочетаний*. Например, при осуществлении поиска корпуса COCA, *perform ability, perform functions, perform tests, perform operation, perform surgery, perform dance*, нами обнаружено множество словосочетаний. Мы удостоверились, что критика Хомского не касается современных корпусов английского языка.

В то время как КЛ занимается разработкой общих принципов и языковых характеристик (корпуса текстов) корпуса, лингвистический корпус рассматривается как набор текстов, собранных в соответствии с определенными нормами и стандартами и снабженных специальной поисковой системой. Иногда его называют корпусом – сборником, объединяющий период создания собранных и отсортированных текстов, упорядоченных по языку, жанру, авторскому фактору.

П.Ю. Довнар и А.В. Воронцов подчеркивают целесообразность создания корпусов текстов и выделяют следующие особенности: 1) представление языковой информации в истинном смысле; 2) предоставление достаточно больших данных (с большим размером корпуса); 3) возможность повторно использовать когда-то созданный корпус для решения различных лингвистических задач, таких как грамматический и лексико-грамматический анализ текста и др.¹⁴.

Большинство языков мира имеют свои национальные корпуса. В качестве примеров общих корпусов можно привести Британский Национальный корпус (BNC). Среди корпусов славянских языков стоит отметить Чешский национальный корпус, созданный в университете Чарльза в Праге. Национальные корпуса существуют в немецком, китайском, французском и других языках. В русском языке существует Российский Национальный корпус (РНК). К сожалению, узбекская лингвистика на данный момент не располагает национальным корпусом узбекского языка.

Наши наблюдения о соотношении взаимосвязанных исследовательских отраслей позволили сделать следующие умозаключения о специфике КЛ (см. рис. 1):

1. Квантитативная лингвистика и корпусная лингвистика – это области лингвистики, которые непосредственно ориентированы на решение ее проблем.

2. Математическая лингвистика и компьютерная лингвистика как взаимозависимые области лингвистики нацелены на разработку математического и технического программного обеспечения для решения своих задач.

3. Прикладная лингвистика. Как область, распространенная на Западе под названием «*applied linguistics*», она охватывает все области, заинтересованные в использовании компьютерных устройств, и в

¹⁴ Довнар П.Ю., Воронцов А.В. Лингвистический процессор китайского языка. Особенности разработки (рус.) // Международный конгресс по информатике: информационные системы и технологии: материалы Международного научного конгресса 31 окт. — Минск: БГУ: БГУ, 2011.

одностороннем порядке взаимодействует с вышеуказанными четырьмя областями.

4. КЛ разумно рассматривать как самостоятельную отрасль языкознания. Хотя КЛ в своем развитии переняла у математической лингвистики необходимые формулы для лингвистических расчетов, у квантитативной лингвистики – обоснованные алгоритмы для выявления и обобщения количественных показателей, у компьютерной лингвистики – автоматизированные компьютерные программы и устройства, она направляет свои достижения на языковые исследования, делится своими результатами для их применения в прикладной лингвистике и преподавании языка.

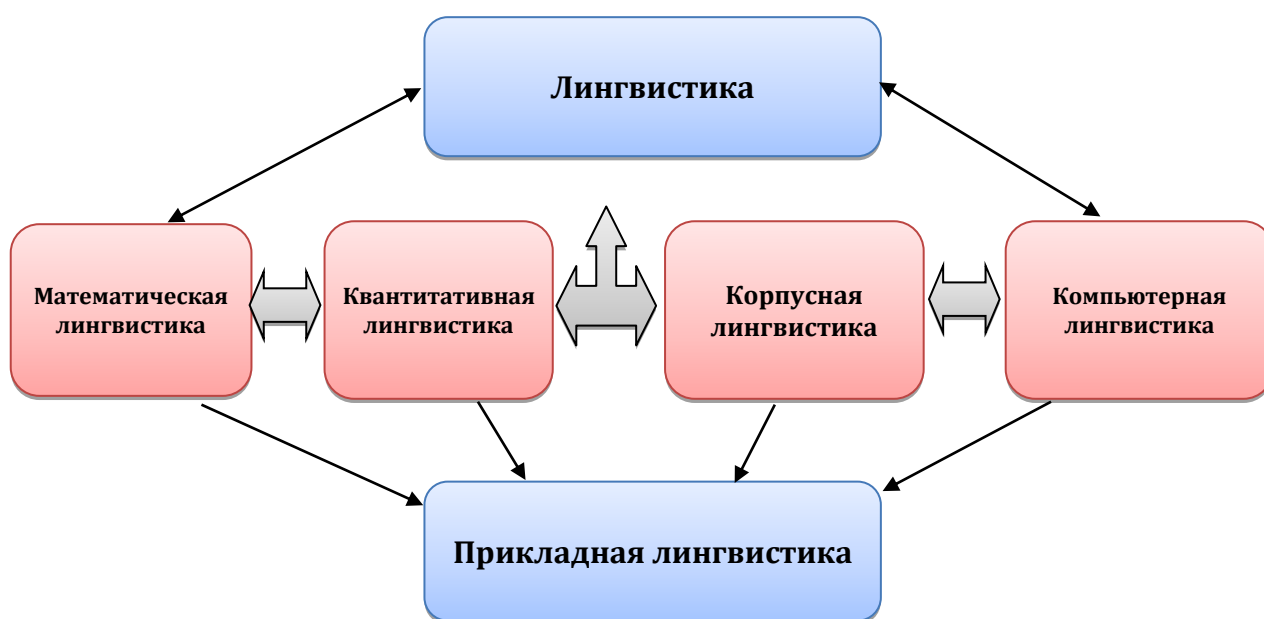


Рис. 1. Формы взаимосвязи языкознания и связанных с ним отраслей

На наш взгляд, неуместно рассматривать КЛ как семантику, синтаксис и аналогичный раздел лингвистики, она своеобразная методология, не требующая объяснения и описания. Рассмотрение КЛ как методологии лингвистических исследований наблюдается в работах множества ученых – Ch.F. Meyer, P. Baker, D. Biber, Y. Kawaguchi, N. Chanyoo, J. Flowerdew, M. Mahlberg, L.C. Steuber, A. Renouf, A. Hardie, R. Love, S. Hunston, P.E. Rayson, M. Fuster, B. Clavel¹⁵. Уместно рассматривать ее как

¹⁵ Meyer Ch. English Corpus Linguistics: An Introduction – Boston: University of Massachusetts at Boston. 2004. – 168 p.; Baker P. Sociolinguistics and Corpus Linguistics: – Edinburgh University Press. 2010. – 189 p.; Biber D., Ulla C. and Upton T.A. Discourse on the Move: – Amsterdam: 2007. – 290 p.; Kawaguchi Yu. and others Corpus-Based Perspectives in Linguistics. – Tokyo. 2007. – 442 p.; Chanyoo N. A Corpus-Based Study Of Connectors And Thematic Progression in The Academic Writing of Thai EFL students: – University of Pittsburgh: 2013. – 146 p.; Flowerdew J. and Mahlberg M. Lexical Cohesion and Corpus Linguistics Amsterdam Philadelphia: 2009. – 125 p.; Lucas C. Steuber Disordered Thought, Disordered Language: A Corpus-Based Description Of The Speech Of Individuals Undergoing Treatment For Schizophrenia. – Portland State University: 2011 – 73 p.; Renouf A. Corpus Linguistics: Past And Present. – Birmingham: University of Central England, – 13 p; Hardie A. and Love R. Corpus Linguistics: 2013. – 376 p.; Hunston S. Corpus Linguistics. University of Birmingham, 2006. – 234–248 pp.; Rayson P.E. Matrix: A Statistical Method And Software Tool For Linguistic Analysis Through Corpus Comparison, Lancaster University: 2002. – 194 p.; Fuster M. and Clavel B. Corpus Linguistics and its Applications in Higher Education University of Valencia: 2010. – 51-67 pp.

методологию, которая не ограничивается какой-либо одной областью лингвистики, но имеет исследовательские методы, которые могут быть применены в любом лингвистическом исследовании.

В то время как на ранних этапах своего развития КЛ рассматривался как метод исследования и как часть или инструмент научных сетей, ориентированных на компьютерные и математические вычисления, сегодня он стал методологией со своими собственными целями и методами исследования, широко принятой научным сообществом. Однако в будущем быстро развивающийся КЛ можно будет рассматривать как парадигму лингвистических исследований или как независимую науку, способную находить решения проблем во всех областях.

По нашему мнению, любой лингвистический корпус должен состоять из томов и категорий, в которых текстовые элементы в нем способны раскрыть лингвистические особенности собранного языка или подязыка. Эта функция делает корпус применимым в практике лингвистических исследований и делает его «электронным состоянием» языка. Для этого каждый компилятор корпуса должен компилировать тексты по категориям. Каждый из устных и письменных текстов, содержащихся в СОСА, разделен на определенную категорию. Например, письменный – газетный материал – тексты должны быть задействованы в процессе в установленном порядке, четко обозначенном, как название газеты.

Корпус – это набор отобранных достаточных лингвистических текстов (устных и письменных), которые могут быть сформированы и классифицированы на основе строгих принципов, основанных на прагматической цели составителя корпуса, соответствуют критериям репрезентативности, сочетают такие принципы, как обработка, теги, совершенная компьютерная автоматизация, он представляет собой базу данных с удобным объемом и согласованностью для систематического поиска, справочных результатов и эмпирического анализа.

В процессе исследования приводится классификации корпусов согласно определенным принципам, условная группировка по сферам применения, используются существующие термины. Нами представляется, что языковые корпуса можно применять в следующих отраслях: А) практика переводов и сопоставительная лингвистика: *специальный / общий корпус; сопоставительный корпус; одно-/многоязыковой корпус; устный/письменный корпус; параллельный корпус и др.* Б) лексикография: *специальный/общий; ссылочный корпус; устный/письменный корпус; двойной корпус; одно-/многоязычный корпус и т.д.* В) теория литературы: *параллельный корпус; авторский корпус; литературный корпус и т.д.* Г) лингводидактика и преподавание языков: *специальный корпус/общий корпус; обучающий/ педагогический корпус; наблюдающий корпус; устный/письменный и т.д.* Д) этимология и исследование истории языка: *специальный/общий корпус; исторический/современный; наблюдательный корпус; региональный корпус и т.д.* Этот перечень можно продолжить,

приводя теоретические и прикладные отрасли языкознания, опирающиеся на анализ корпусов.

Вторая глава, «**Прикладные анализы о внедрении корпусов: результаты и решения**», исследует методологию КЛ – методы корпусного анализа, лингвистические особенности корпусной лингвистики, основанной на специальном корпусе ESOCL, а также важность работы с корпусами английского языка и предоставляемой ими информации.

М.А. Большакова в своем корпусном исследовании разделила¹⁶ КЛ на следующие этапы: *этапы постановки цели, направления, создания, поиска и объяснения*.

Д.Ф. Джафарова подчеркивает¹⁷, что корпус в КЛ должен соответствовать следующим требованиям: 1) образцовость и представительность; 2) ограниченный объем; 3) автоматизация информации.

Мы сочли необходимым разделить методологию КЛ на эмпирическую систему из следующих шести методов, направленных на создание корпусов на основе принципов корпуса и их применение в прикладных исследованиях:

1) **реферальный метод**. При этом, исходя из целенаправленности корпуса, необходимо четко определить, для каких исследований предназначен лингвистический корпус, создаваемый компилятором. Конечно, согласно классификации целевых корпусов, важные языковые особенности определяются в зависимости от того, какой тип корпуса нужен и какие функции он имеет;

2) **работа с источниками** или **базовый метод**. Этот метод ориентирован на работу с коллекцией текстов и сортировку собранных текстов. Здесь разрабатываются требования к текстовым ресурсам, составляющим корпус, и рекомендуется, чтобы каждый компилятор корпуса компилировал тексты на основе этих шести требований (см. Рис. 2). Кроме того, этот этап включает в себя процесс разделения текстов на категории в соответствии с работами одного автора, принадлежностью к жанру, структурным и отраслевым сходствам;

3) **метадата** состоит из **методов объединения данных**. Методы, включенные в эту группу, направлены на добавление справочных данных к отобранному тексту для включения в корпус. Четкое представление базы данных – это очень простой метод, который позволяет выполнять поиск по категориям информации (газетные материалы, молва и т. д.) и связывать ее в качестве ссылки в примере, приведенном в строке согласования. На этом этапе важен метод тегирования, и сегодня он, в основном, выполняется с помощью компьютеров.

¹⁶ Большакова М.А. Лингвистический контент и программная реализация интеллектуального немецко -русского отраслевого словаря Диссер...канд.фил. наук. – Москва, 2013. – С 56-57.

¹⁷ Джафарова Д.Ф. Модели лингвистического анализа текстов таджикского языка (на материале газелей Хафиза) дисс. ... кандидата филологических наук. – Душанбе, 2013. – С 35

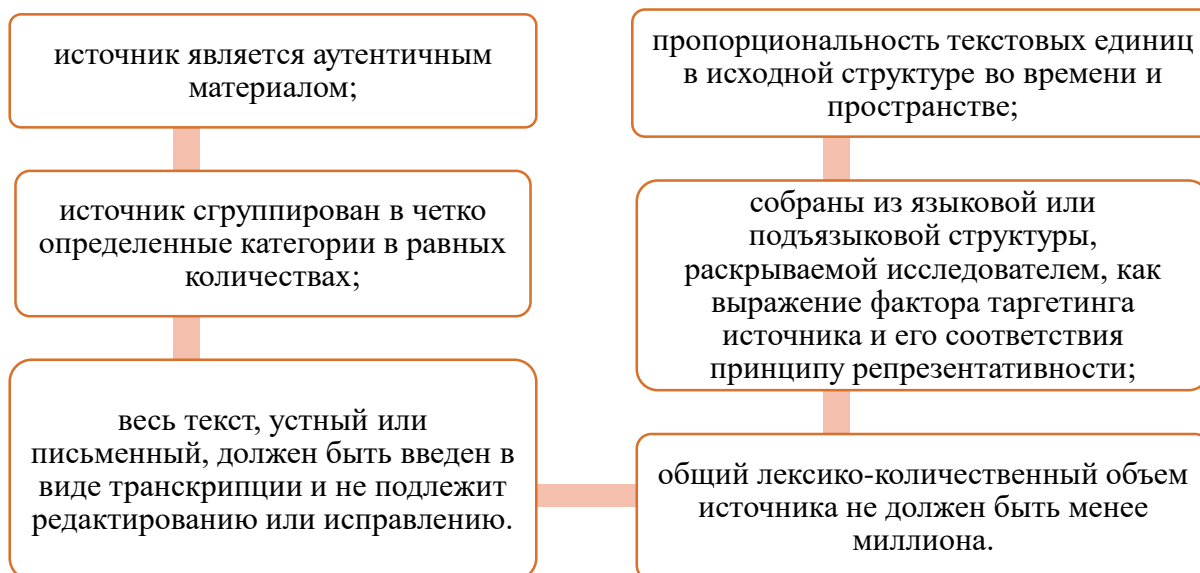


Рис 2. Требования критериев к текстам корпуса

Любой синтаксический прием также можно использовать в процессе исследования. Например, структура *I am* PPSS + BEM (PPSS – non-third person nominative personal pronoun в BEM – am form of Be) (см. Табл. 1).

Таблица 1

Тегирование приведения глагольных форм

| | | | | |
|-------------|-----------|-----------|------------|-----------|
| Sing /VB | Be /VB | Do /VB | Have /VB | Go /VB |
| Sings /VBZ | Is/VBZ | Does /VBZ | Has /VBZ | Goes /VBZ |
| Sang/VBD | Was/VBD | Did/VBD | Had/VBD | Went/VBD |
| Singing/VBG | Being/VBG | Doing/VBG | Having/VBG | Going/VBG |
| Sung /VBN | Been /VBN | Done /VBN | Had /VBN | Gone /VBN |

В этом случае, если VB – *verb* – заменяет глагол, третье лицо Z, добавленное к нему, является суффиксом настоящего простого времени в единственном числе, в то время как D представляет форму глагола в прошедшем времени, G представляет форму герундия, а N представляет две формы прилагательного глагола.

Система тегов формируется как окончательная систематическая сумма этих отдельных тегов. Система тегов для корпуса *Penn Treebank* содержит всего 48 кодов тегов, в том числе: CC – coordinating conjunction, CD – cardinal number, DT – determiner, FW – foreign word, JJ – adjective, JJS – adjective in superlative degree, JJR – adjective in comparative degree, NN – noun, NNS – plural noun, NNP – proper noun, RB – adverb, RBR – adverb in comparative degree, RBS – adverb in superlative degree и другие;

4) **метод автоматизации** (направлен на перевод корпусов в полностью электронный вид). На этом этапе метод работает со всеми собранными, отсортированными, связанными метаданными и проверенными текстами в виде электронного корпуса и его дизайна. Он также включает в себя

широкий спектр возможностей, таких как формирование соответствий в корпусе, цвет искомого слова и слов, которые оно объединяет, а также возможность графического представления результатов, полученных на основе анализа корпуса.

5) *метод оценки*. Благодаря этому методу его совершенство оценивается путем определения его способности реагировать на каждый из заявленных принципов концепции корпуса. В этом случае необходимо оценить каждую особенность корпуса;

6) *частотный поисково-анализирующий метод*. Его можно рассматривать как метод, позволяющий делать выводы, глядя на частоту использования слов в корпусе. В целом важность частот высока во всех основанных на корпусе анализах, и работа с ними также требует четкого различия. Потому что частота использования слов, входящих в корпус, позволяет нам делать исчерпывающие выводы. Т.С. Грайс¹⁸ подразделяет их на три типа: основные частоты; стандартные частоты; логарифмические частоты.

В настоящем исследовании мы сочли целесообразным сравнить данные корпуса с помощью метода частотного поиск-анализа:

Базовые частоты (Raw frequencies) – это частоты использования слова в корпусе при поиске, которое указывается в точных числах. Например, в СОСА глагол *give* использовался 202515 раз, что является ее основной частотой (ОЧ).

Нормативные частоты (Normalized frequencies) – это соотношение точного количества частот, указанных корпусом, к уровням 10 (1000, 1000000 и т.д.). Например, для нахождения НЧ приведенного выше глагола *give*, общая количества (ОК) СОСА рассчитывается относительно 570 миллионов. При этом для того, чтобы найти нормативную частоту слова (НЧ), соотношенную к миллиону, его ОЧ необходимо разделить на 570. Следует рассчитать следующую формулу:

$$\text{НЧ(мил/1)} = \frac{\text{ОЧ}}{\text{ОК}/1000000}.$$

Из этого следует, что в СОСА НЧ глагола *give* приблизительно равна (мил/1) 355,2. Это важно при межкорпусном сравнении одного языка или сравнении корпусов разных языков.

Логарифмические частоты (Logged frequencies) – это степень, возведенная из первичной конкретно отобранной и логарифмической частоты ($e = 2,7182818$). То есть логарифм указывает на соотношение основы и степени. Если обратиться к формуле $\log_b x = y$ и если $b=10$, $x=100$, то $y=2$. Иными словами, на вопрос, какая степень числа 10 равняется 100, ответ – 2. При выявлении логарифмической частоты (ЛЧ) в приведенной выше формуле $b=2,7182818$ (основа), $x=202515$ (первичная частота) и из этого

¹⁸ Stefan Th. Gries Quantitative Designs And Statistical Techniques/ English Corpus Linguistics edited by Douglas Biber and Randi Reppen. Cambridge, United Kingdom. 2015. – 623 p. 50-71 pp. (52)

берется логарифм, выявляется, что степень $y=12,218569364782$, это логарифмическая частота использования глагола *give* в СОСА.

На наш взгляд, данный процесс сравнения может быть в трех видах: *внутреннее сравнение, межкорпусное сравнение в рамках одного языка, сравнение разноязычных корпусов.*

Мы предложили разделить методологию КЛ на шесть групп методов, направленных на создание корпуса и применение его в прикладных исследованиях. Именно благодаря интеграции этих систематических методов может быть оправдано создание экспериментального корпуса, направленного на изучение лингвистических особенностей КЛ. Соответственно, планируемый к созданию корпус будет формироваться в соответствии с этим процессом.

Прежде всего после определения цели создания корпуса следующим шагом стала работа с текстами, составляющими корпус ЕСОСЛ. Сначала с помощью *работы с источниками или базового метода* собирали корпусные тексты. Исходя из критериальных требований к корпусным текстам, разработанным в соответствии с корпусными принципами, источник был разработан следующим образом:

1) собранная база данных является достоверным материалом, все это научная информация, написанная на английском языке и отредактированная носителем языка; 2) все материалы, выбранные в качестве источника дела, являются научными материалами, относящимися к КЛ; 3) источник сгруппирован по равным категориям (например, диссертации, книги и статьи); 4) все материалы, включенные в корпус, обладают способностью отражать лингвистические особенности КЛ; 5) источник состоит из полных образцов письменной речи; 6) выявлено, что общий лексико-количественный объем материалов, входящих в ЕСОСЛ, составляет более миллиона. Таким образом осуществлялся сбор и сортировка текстов, и одним из основных методов этого этапа был метод автоматической разметки. Добавление внетекстовой информации не сыграло существенной роли, потому что тип и площадь текстов были одинаковыми. Подобным образом проводятся работы по сбору и отбору текстов, основной метод данного этапа – метод маркировки (*mark-up*) осуществляется автоматически. В силу того, что одинаковы виды и отрасли текстов включение экстра-текстовых (*extra-textual*) данных не имеет важного значения.

Грамматическая форма множественного числа термина *Corpus* имеет свои эквиваленты. При этом, одинаково употребляются формы *corpora* и *corpuses*. Результаты поисков в ЕСОСЛ показывают, что существуют оба грамматических варианта, однако если *corpora* показала ОЧ в 13829 раз, то *corpuses* использовано 53 раз. Из этого следует, что в английском языке в качестве грамматической формы множественного числа термина *corpus* чаще используется термин *corpora*. Однако в узбекском языке обе формы передаются словом *корпуслар*, так как обе формы по значению выражают множественное число.

В результате исследований выделены 220 терминов и их частота включена на основе результатов ESOCL конкорданса. Дизайн небольшого словаря оформлен, соответственно, формату “Толкового словаря узбекского языка”. При этом приведено английское наименование термина и его эквивалент на узбекском языке, а также толкования терминов. Каждое толкование оперативно найдено благодаря возможностям, данным ESOCL интерфейса, приведена в форме ссылки к нему (McEnergy, 2006). Важность небольшого словаря, на наш взгляд, заключается в том, что исследователи, приступившие к изучению отрасли КЛ могут понять значение английских терминов с помощью их эквивалентов на узбекском языке и их толкований, также данный словарь предлагает изучение, начиная с наиболее широко распространенных терминов. Это, в свою очередь, помогает изучающему поэтапно открывать для себя данную отрасль, готовит базу для понимания незнакомых терминов. Например, не поняв сути понятия *корпус*, невозможно в сознании сформировать термин тегирование, являющимся системой кодирования. С использованием результатов анализов, полученных от корпуса ESOCL, созданы небольшие частотные-толковые и алфавитно-частотные-толковые словари, в них слова упорядочены соответственно частоте использования и алфавитному порядку.

Также, в данной главе обосновано, что применение анализов корпуса во многие отрасли языкознания дает положительные результаты. В развитии КЛ особую историческую роль играют языковые корпуса, созданные и создаваемые на английском языке, в работе также нашли свое отражение электронные источники с расширением WWW, представляющие языковые англоязычные корпуса и включенные в них данные. В основном, рассмотрены три базы: *Sketch engine*, *CoRD*, ва *English-corpora.org*, также приведены краткие сведения о корпусах, входящих в состав каждой базы, специальные сайты для их практического использования. Рассмотрены результаты широких поисков и конкорданса, которые могут предоставить англоязычные корпуса *English Web 2015 (enTenTen15)*, *BNC* и *Google Books n-gram*. При этом реализованы обычные и сложные поиски, демонстрирующие возможности каждого корпуса.

Третья глава называется «**Возможности корпуса СОСА и её специфические особенности**». В данной главе проанализированы соответствие СОСА принципам функционального корпуса, её приоритеты, результаты анализов *СОСА* корпуса, куантитативное выражение лингвокультурных единиц, словарный корпус, предоставляющий языковые данные в лексикографии СОСА. Учтены возможности реализации осуществления приведенных выше требований к языковому корпусу, являющегося объектом исследования. В силу этого выбран *СОСА*, являющийся вторым поколением современных англоязычных корпусов.

Корпус из более 450 миллионов слов, являющийся продуктивным результатом проекта под руководством Марка Девизеса¹⁹, – это корпус COCA – Corpus of contemporary American English (Современный американский английский корпус), единственный широкообоснованный (balanced) корпус американского английского языка. Его каждый месяц используют более 40 тысяч индивидуальных пользователей, что делает его наиболее часто используемым онлайн-корпусом. Благодаря своему дизайну этот языковой корпус является единственным крупным корпусом, который можно использовать для рассмотрения существующих изменений в английском языке. В настоящее время он охватывает данные тридцатилетнего периода, поэтому можно понаблюдать изменения в лексическом составе языка за данный отрезок времени, в силу того, что проект еще не завершен, продолжается дополнение базы данных. Вход в корпус COCA осуществляется через сайт English-Corpora.org, необходимо зарегистрироваться для бесплатного использования существующих корпусов. А это сделало его наиболее часто посещаемым сайтом на сегодняшний день. Например, на период от 1 октября до 31 октября 2019 года количество потребителей сайта достигло 130 тысяч. Наиболее хорошо разработанный лингвистический корпус данного сайта, ежемесячно привлекающего более ста тысяч пользователей – это COCA.

В настоящее время согласно последним данным, внесенным на дату март 2020 года, объём COCA составляет более 485202 текста, который охватывает более 1 миллиарда 1 миллиона слов, в течение 1990-2020 гг. каждый год включались более 30 миллиона. В течение данного времени материалы корпуса распределены на следующие 8 видов: устный (spoken), художественный (fiction), известные журналы (popular magazines), газеты (newspapers), научные журналы (academic journals), онлайн-блоги (Blogs), веб-страницы (Web pages) и записи телепередач и кино (TV/Movies subtitles), данные вносятся регулярно²⁰.

В целях раскрытия функциональных особенностей COCA изучена её значимость в семантических, синтаксических, фразеологических и морфологических исследованиях. В качестве примера, в целях анализа семантических особенностей глаголов *cause* и *result in* в значении *быть причиной*, задан поиск словосочетаний в COCA. При этом тегирование корпуса очень удобно для поиска существительных в функции дополнения, наиболее часто используемых с глаголами *cause* и *result in*. В таблице конкорданса глагол *cause* использован 71078 раз, а глагол *result in* 84001 раз. Результаты поисков, приведенных в порядке *Cause+nn** и *result in+nn**, соответственно 4394 и 2907 раз берет после себя дополнение части речи существительного. Если обратить внимание на дополнения глагола *cause*, выясняется, что почти все они выражены существительными отрицательного значения. Например, *problems, cancer, trouble, harm, damage, disease, pain, illness, injury, stress, infection, discomfort* и другие. А глагол *Result in*, наряду с

¹⁹ [Mark Davies](http://d Davies Linguistics), Professor of Linguistics, Brigham Young University. Available at URL:<http://d Davies Linguistics byu.edu/personal/>

²⁰ Corpus of Contemporary American English/English-corpora.org <https://www.english-corpora.org/coca/>

существительными отрицательного значения *death, loss, pressure, discrimination, rape, war, trauma*, может присоединяться также к существительным положительного значения *changes, chance, negotiations, improvements, power, quality*.

Ж.Синклайр²¹, впервые в истории КЛ исследовавший фразы и отдельные слова с применением лингвистических корпусов, анализирует две различные фразы. Это следующие фразы: *naked eye* и *true feeling*. Примеры использования проанализированных в работе фраз берет из корпуса *BNC* и если для *naked eye* находит 148 примеров, то для *true feeling* этот показатель равен 53.

При поиске именно этих словосочетаний в *COCA* выясняется, что эти цифры значительно увеличиваются. То есть получены результаты 654 и 175 раз соответственно. А это, в свою очередь, открывает путь к более широкому обсуждению искомых фразеологических единиц.

В английском языке существуют аффиксы образования прилагательных *-ous* и *-ness*, образующие абстрактные существительные. На такие вопросы, как: могут ли они быть в составе одного слова, если да, то в каком порядке они располагаются, можно получить ответ на основе результатов поиска в *COCA*. По данным поисков, в корпусе нет сочетание морфем **ness+ous*, на основе чего заключено отсутствие такого сочетания и в языке. Результаты, соответствующие порядку **ous+ness* равны 112, это общее количество использования более 10 слов. Например, *consciousness, seriousness, nervousness, righteousness, graciousness, dangerousness* и др.

Использование в исследовании корпусных анализов в качестве методов исследования, предоставляющих количественные результаты употребления лингвокультурологических единиц, всесторонне соответствуют современным требованиям. Ибо, как правильно отмечает В.А. Маслова, язык тесно связан с культурой, распространяется и развивается в его структуре, а также отражает в себе культуру²². В монографии С. Грайса и А. Степанович “Корпуса в когнитивной лингвистике” утверждается, что эти две отрасли взаимодополняют друг друга²³. При этом, нам представляется целесообразным соотнести лингвокультурологию с КЛ.

Установлено, что схожий дизайн всех выбранных лингвистических корпусов, жанровая структура и факторы приведения в конкордансе данных служат обеспечению связи КЛ и лингвокультурологии. Отличающий признак этих двух корпусов – это объём. При этом объём *BNC* по количеству включенных в него слов в шесть раз меньше *COCA*. В исследовании это не влияет на результаты сравнений, поскольку частоты даны не в ОЧ, а в НЧ.

Первые результаты поисков, полученных на базе корпусов, направлены на выявление коллокационных словосочетаний. Отобраны наиболее часто сочетаемые 300 соединений в порядке “ПРИЛАГАТЕЛЬНОЕ + семья

²¹ John Sinclair Trust the text: language, corpus and discourse. London: Routledge – 2004. 30-36

²² Маслова В.А. Лингвокультурология: учебное пособие. Академия изд., – Москва, 2004. – 208 с.

²³ Gries S.Th. and Anatol Stefanowitsch (eds.). Corpora in Cognitive Linguistics. Corpus Based Approaches to Syntax and Lexis. Trends in Linguistics: Studies and Monographs. – New York: Mouton deGruyter, 2006.– 352 p.

(*_adj*+family*)” и приведены, разбив на определенные категории, обосновано, что результаты, полученные на основе *BNC*, могут быть приведены в составе групп, указывающих на следующее: а) *размер семьи*: large (1,44), small (0,72), average (0,60), big (0,55), normal (0,37), little (0,26); б) *происхождение членов*: royal (7,14), local (0,59), traditional (0,40), Jewish (0,35), holy (0,35), noble (0,31), black (0,26), British (0,26), Christian (0,26), human (0,25), poor (0,25), wealthy (0,21), working-class (0,19), farming (0,15), middle-class (0,13), English (0,18), French (0,11), Indian (0,11), Asian (0,10); в) *структуру существования членов семьи*: extended (1,61), nuclear (1,36), immediate (0,89), private (0,23), individual (0,15), natural (0,18), birth (0,14), one-parent (0,11), single-parent (0,10), two-child (0,9); г) *взаимоотношение членов семьи*: happy (0,69), close (0,68), joint (0,25), loving (0,15), ordinary (0,15), respectable (0,10), lovely (0,10), friendly (0,09), close-knit (0,09), ideal (0,06); д) *возраст членов семьи*: new (1,10), old (0,70), young (0,60), growing (0,25), modern (0,15), contemporary (0,06).

Затем, с помощью реализации такого же поиска в *COCA* нами в сравнительном плане изучены общие и отличительные стороны. Результаты корпуса классифицированы в приведенном выше порядке для выражения следующих: а) *объем семьи*: large (0,86), big (0,85), average (0,56), small (0,52), little (0,40), median (0,36), normal (0,24), stable (0,10); б) *происхождение членов*: royal (3,28), American (2,01), black (0,97), traditional (0,66), holy (0,55), dysfunctional (0,46), human (0,46), Kennedy (0,62), Bush (0,45), Jewish (0,44), poor (0,34), white (0,31), middle-class (0,30), wealthy (0,30), military (0,28), Christian (0,22), immigrant (0,19), African-American (0,14), working-class (0,13), suburban (0,10), religious (0,10), Muslim (0,09), farming (0,08), free (0,08); в) *структуру существования членов семьи*: extended (3,37), immediate (1,08), nuclear (0,91), host (0,21), adoptive (0,18), private (0,17), biological (0,12), two-parent (0,12), personal (0,09), single-parent (0,09), broken (0,08), surrogate (0,07); г) *взаимоотношения членов семьи*: happy (0,61), close (0,12), great (0,41), loving (0,40), perfect (0,24), wonderful (0,23), close-knit (0,23), joint (0,20), typical (0,18), supportive (0,15), individual (0,15), devoted (0,10), ideal (0,09), tight-knit (0,07), lovely (0,06), beloved (0,06); д) *возраст членов семьи*: new (1,19), old (0,76), modern (0,58), young (0,49), growing (0,25), younger (0,07).

Вышеизложенные примеры оформлены на основе результатов, полученных на базе двух корпусов, приведенные качества в группах классификации толкуют понятие “семья”. Для возможностей сравнительного изучения их НЧ в соотношении 1 к миллиону приведена в скобках, бросается в глаза сходство и отличие в данных. Например, количество примеров в *COCA* со словом “new – новый”, выражающим возраст членов семьи, то есть ОЧ равна 680, а результаты поисков в *BNC* равны 110, но из этого не следует, что в *COCA* данное словосочетание используется чаще. Вычисленные нами НЧ для *COCA* составляет 1,19, а для *BNC* 1,10. Из этого следует, что словосочетание *new family* в корпусах и языковых пластах в относительно близких друг другу частотах (в американском английском языке частота

немного выше). Кроме того, установлено, что понятие *молодая семья* выражается словом “*young* – молодой”. Однако превалирование на стороне британцев (в *BNC* и *COCA* НЧ показали 0,60 и 0,49, соответственно). Рассмотрены прилагательные, отличающиеся друг от друга, на основе эмпирических анализов более широкого контекста заключено, что результаты анализа, основанные на корпусе, могут рассматриваться как огромная база лингвистических единиц, используемых для осмысления лингвокультурных понятий языка и носителей языка.

Рассуждая о связи КЛ и лингвокультурологии, несомненно, важно изучение стилистических конструкций типа метафоры, касающихся выбранного понятия. При этом при осуществлении поиска слова *family* – *семья* в формате *COCA* с помощью KWIC найдены более распространенные метафоры: *family is responsibility*, *family is a blessing*, *families are building blocks*, *my family is my life*, *family is a source*, *family is the best experience*, *family is a school*, *family is a domestic church* ва *family is home*. *COCA*: *family loyalty* (семейная вера), *family status* (семейное положение), *family honor* (семейная гордость), *family value* (семейная ценность), *family environment* (семейная атмосфера), *family concerns* (семейные хлопоты), *family secrets* (семейные секреты), *family circle* (семейный круг), *family gathering* (семейное собрание), *family planning* (планирование семьи), *family breakdown* (разрушение семьи), *family involvement* (участник семьи), *family love* (семейная любовь).

Анализ примеров показал, что рассуждения, полученные от *BNC*, в основном, направлены на формирование семьи, его состав, роль и значение в обществе, объединяющее их понятие это – опора на традиции создания и формирования семьи, рассмотренные в качестве части общества. При этом семья – это небольшое звено, построенное на базе ценностей, состоящее из близких друг другу людей, их места и роли в обществе. Данные, полученные от *COCA*, демонстрирует семейную атмосферу, положительность внутренних взаимоотношений, ценность фактора любви и человеческого фактора.

В исследовании проанализированы возможности современных англоязычных моноязычных электронных словарей в плане полезности для пользователей. При этом выбирается любое слово и выясняется, что анализ результатов, полученных по результатам поисков по англоязычным словарям, не дали положительных результатов. Установлено, что результаты, которые можно получить при поиске выбранного произвольного слова ‘*fabric*’ в структуре *COCA* кардинально отличаются в положительную сторону от информации, предоставляемой словарями.

ЗАКЛЮЧЕНИЕ

1. Выяснилось, что КЛ, первый этап исторического развития которого восходит к XIX веку, представляет собой методологию научных исследований, сочетающую методы сбора, обработки и проведения исследований лингвистических корпусов. Кроме того, текущее развитие сферы в мировой лингвистике является позитивным и динамичным. В частности, объем исследований по КЛ в Узбекистане сегодня находится на

высоком уровне. Диахроника КЛ от создания корпуса Брауна до разработки электронного корпуса рассмотрены периодически, теоретически и классифицируемо.

2. Самостоятельное развитие КЛ было основано на компьютерной лингвистике, математической лингвистике и других областях лингвистики. Стало понятным, что на ранних этапах своего развития КЛ рассматривался как метод исследования и как часть или инструмент информатики и математической вычислительной науки, сегодня же он стал методологией со своими собственными целями, задачами и методами исследования. Однако в будущем уместно оценивать быстро развивающийся КЛ как научную гипотезу высокого уровня, обладающую всеми чертами парадигмы лингвистических исследований и/или независимой науки, которая может решать проблемы во всех областях.

3. Критерии требований к корпусу текстов при формировании лингвистического корпуса можно определить как: 1) аутентичный материал; 2) пропорциональность текстовых единиц во времени и пространстве; 3) объединение в четко определенные категории в равных количествах; 4) выражение целевого фактора, собранного с языка или подязыка исследователя, что соответствует принципу репрезентативности; 5) весь текст как устный, так и письменный, должен быть в форме транскрипции без каких-либо изменений и исправлений; 6) общее лексическое количество не должно быть меньше миллиона. Требования этого критерия основаны на корпусных принципах.

4. Корпус – это достаточно большой набор языковых текстов (устных и письменных), которые могут быть сформированы и классифицированы на основе строгих принципов, исходя из прагматических целей разработчика корпуса, отвечающих требованиям репрезентативности, совершенной автоматизации процессов с помощью компьютера. Он представляет собой базу данных с достаточным количеством и согласованностью, удобную для справочного результата и эмпирического анализа.

5. Классификация корпусов также является одной из характерных проблем, стоящих перед КЛ. Подходы к классификации различаются, и можно видеть, что они, в основном, осуществляются в зависимости от размера, формы, дизайна, языка и аналогичных характеристик корпуса. Согласно принципу классификации корпусов с точки зрения прагматической цели пользователя, лингвистический корпус с использованием существующих корпусных терминов является одной из областей, которые могут быть применены: а) в практике перевода и сравнительной лингвистике; б) в лексикографии; в) в теории литературы; г) в лингводидактике и обучении языкам; этимологии и исследовании истории языка и др. Действительно, классификацию корпусов в окончательной стандартной форме уместно рассматривать как относительную концепцию, потому что, когда создается любой корпус, он приносит с собой новую цель, новый дизайн и новый тип классификации.

6. В диссертации КЛ рассматривается как методология лингвистического исследования, как группа шестиэтапных методов, которые включают в себя определенные процессы, основанные на последовательности используемых методов исследования: методы направления, управление ресурсами или основные методы, методы добавления метаданных, методы автоматизации, усовершенствованы методы оценки и методы частотного поиска-анализа.

7. Процесс сравнения результатов корпусного анализа может осуществляться тремя способами: 1) Внутреннее сравнение – понимается как взаимное сравнение нескольких результатов поиска, полученных из самого корпуса. 2) Сравнение межъязыковых корпусов – означает взаимное сравнение результатов поиска четко определенной лингвистической единицы (единиц), полученных из нескольких корпусов, представляющих один и тот же язык. 3) Межкультурное сравнение разных языков – сравнение результатов поиска с идентификацией выбранной языковой единицы и ее точным или эквивалентным переводом в корпусах с той же классификацией или параллельно (в основном состоящих из примеров переводческой практики), созданных на основе разных языков.

8. Функциональные свойства *СОСА* определяются по результатам корпусного анализа взаимной семантической дифференциации языковых единиц. Это может включать роль корпуса для исследований, направленных на изучение семантических соединений. Примеры из *СОСА* в синтаксических исследованиях являются обширным источником информации о формах глаголов и их взаимосвязанных грамматических отношениях, а также о синтаксических приемах. *СОСА* имеет возможность анализировать слова и возможные орфографические символы, которые можно комбинировать с помощью системы KWIC, необходимой для изучения фразеологизмов. При изучении морфологических единиц *СОСА* практически указывает, присутствуют ли морфологические единицы в английском языке в форме и форме словосочетания.

9. *СОСА* – современный корпус английского языка второго поколения, является источником количественных примеров для определения лингвистического выражения лингвокультурных единиц. Сравнительный анализ лингвокультурных концепций и единиц в американском и британском английском уместен на основе анализа BNC и *СОСА*. При анализе корпуса выбирается конкретное понятие, и связанные с ним согласованные примеры изучаются сравнительно между одноязычными корпусами. Была собрана вся информация, необходимая для выражения культурного единства, и были выявлены сходства и различия в понимании концепции *семьи* в двух культурах.

10. Рассматривая *СОСА* как словарь в процессе создания словарей, его существующие функции сравнивались с внутренними возможностями современных англоязычных одноязычных словарей. Выяснилось, что в этом случае корпус предоставляет пользователю: 1) описание слова и список слов, связанных с ним; 2) результаты анализа по жанрам; 3) информацию в

графическом виде; 4) лингвистическую интерпретацию неологизмов в языковом слое и этимологию их возникновения; 5) большое количество примеров употребления слова; 6) правильное сочетание предлогов и умение давать точную статистику и т.д. – при изучении синтаксических, произношений и пунктуационных характеристик языковых единиц.

**SCIENTIFIC COUNCIL ON AWARDING
SCIENTIFIC DEGREES DSc.03/30.12.2019.Fil/Ped.27.01
AT UZBEKISTAN STATE UNIVERSITY OF WORLD LANGUAGES**

UZBEKISTAN STATE UNIVERSITY OF WORLD LANGUAGES

NOZIMJON BOBOJON UGLI ATABOEV

**FUNCTIONAL CHARACTERISTICS OF THE ENGLISH CORPORA
(on the example of COCA)**

**10.00.04– Languages and literature of the people of Europe,
America and Australia**

**ABSTRACT OF THE DISSERTATION
of doctor of philosophy (PhD) on philological sciences**

TASHKENT – 2020

The theme of the dissertation of Doctor of Philosophy (PhD) was registered at the Supreme Attestation Commission at the Cabinet of Ministers of the Republic of Uzbekistan under number B2019.1.PhD/Fil723

The doctoral dissertation has been carried out at the Uzbekistan State University of World Languages.

The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (abstract)) on the Scientific Council website (www.uzswlu.uz) and on the website "ZiyoNet" information and educational portal (www.ziynet.uz).

Scientific supervisor: **Teshabaeva Dilfuza Muminovna**
Doctor of Philological Sciences, Professor

Official opponents: **Rakhimov Ganisher Khudaykulovich**
Doctor of Philological Sciences, Professor

Khamroeva Shahlo Mirdjonovna
Doctor of Philosophy (PhD) on philological sciences

Leading organization: **Bukhara State University**


The defense will take place on «11th of November 2020 at 14⁰⁰ at the meeting of the Scientific Council DSc.03/30.12.2019.Fil/Ped.27.01 at Uzbekistan State University of World Languages (Address: 100138, Tashkent city, Uchtepa district, Kichik halqa yoli street, 21A Tel: (99871) 230-12-91, fax: (99871) 230-12-92, e-mail: uzswlu_info@mail.ru).

The dissertation can be reviewed at the Information Resource Center at Uzbekistan State University of World Languages (Address: 100138, Tashkent city, Uchtepa district, Kichik halqa yoli street, 21A Tel:(99871) 230-12-91, fax: (99871) 230-12-92.

The abstract of the dissertation was distributed on «29th of October 2020 y.
(Registry record No. 12 dated «29th of October 2020.)


G.Kh. Bakieva
Chairperson of the Scientific Council on awarding scientific degrees, Doctor of Philological Sciences, Professor


Kh.B. Samigova
Scientific Secretary of the Scientific Council on awarding scientific degrees, Doctor of Philological Sciences, Associate Professor


J.A. Yakubov
Chairperson of the Scientific Seminar of the Scientific Council on awarding scientific degrees, Doctor of Philological Sciences, Professor

INTRODUCTION (abstract of the PhD dissertation)

The aim of the research work is revealing the functional characteristics of the existing corpus of English on the basis of the analysis of the evolution of the historical development of CL, the corpus, its classification, the principles of application.

The object of the research work: the English language corpora, in particular the modern American English corpus COCA, were selected as the object of the study.

Scientific novelty of the research work:

on the basis of the history, stages of development of corpus linguistics, the problem of classification and the results of statistical approaches, the corpus creating criteria as representativeness, target-oriented and limited scope of corpus data have been improved; six-stage the methodology of creating corpora and their application in practice has been proved on the basis of scientific and theoretical evidence;

the terminological system of corpus linguistics has been enriched through the newly formed specific frequency-definitional dictionary based on the results of experimental corpus created by means of the AntConc program for various text-format data analysis;

the principle of comparative analysis of linguocultural concepts by means of analysis of metaphors, collocations and phraseological units taken from the results of corpora (on the example of British National Corpus – BNC and Corpus Of Contemporary American English – COCA corpus) has been developed based on the representativeness of a corpus;

the COCA corpus was first considered as a dictionary corpus, and all the statistical and practical results on the characteristics of lexical units such as phonetic, orthographic, grammatical and semantic it can provide to the field of lexicography have been demonstrated in comparison with the data that existing English dictionaries are able to supply.

Implementation of the research results. On the basis of research results and practical outcomes worked out to reveal the functional characteristics of the English corpus on the example of COCA:

the stages of development of the concept of corpus linguistics in world and Uzbek linguistics, the methodology of creating corpora, their application in practice are proved on the basis of scientific and theoretical evidence. The results of the work have been included in the project of the creation of an audiobook on linguistics for visually impaired students, numbered A-1-61, carried out under state scientific and technical programs on the basis of scientific and technical research on corpus linguistics and its current stages of development on the basis of a practical grant project “Summary of articles on linguistics published in 2014-2016 (from “The Uzbek language and literature” and “Philological Issues”)” (Reference № 89-03-2032 of The Ministry of Higher and Secondary Specialized Education as of June 12, 2020). As a result, the theoretical problems of creating corpora and the development of the systematic methods of corpus application have been solved;

the results of the theoretical analysis and the formation of a school of experience, the elaboration of general requirements for a language corpus, the creation of corpus for educational and scientific purposes, the development of methods, classification criteria and principles of application of corpus texts in various fields of linguistics. The results were used under the innovative project at Tashkent State University of the Uzbek Language and Literature of the Ministry of Higher and Secondary Specialized Education of the Republic of Uzbekistan for “Creating an electronic poetic dictionary (human image, verb, nature and national symbols) of the Uzbek and English languages” (Reference № 89-03-2032 of The Ministry of Higher and Secondary Specialized Education as of June 12, 2020). Consequently, the usage of the scientific results accomplished the principles of defining the terms related to the sphere;

the language features of CL were analyzed by creating an experimental (one-time) corpus by collecting materials (texts) devoted to CL using the AntConc program, and the principle of comparative analysis of linguocultural concepts based on the representativeness of corpora (in the example of BNC and COCA). The results, practical proposals and recommendations were implemented by Andijan State University under the fundamental project numbered OT-F1-18 of the Ministry of Higher and Secondary Specialized Education of the Republic of Uzbekistan – “Development of methods and methodology for the formation of mass linguistic culture” (Reference № 89-03-2032 of The Ministry of Higher and Secondary Specialized Education as of June 12, 2020). As a consequence, the implementation of these results enabled to comprehend the linguocultural concepts, which can be cognited through the prism of a language, better;

for the first time, the COCA corpus was considered to be a dictionary corpus, and all the statistical and practical results that could be provided to the field of lexicography were presented in comparison with the information provided by existing English dictionaries. Improved scientific conclusions based on empirical analysis were discussed at a scientific seminar with the participation of professors of the Department of Translation and Linguistics of the University of Pompeu Fabra in Barcelona, Spain and representatives of the research group GEDIT (Grup d'Estudis del discurs I de la Traducció) and the results obtained were implemented in the MA dissertations (Reference of translation and Communication faculty of Pompeu Fabra University as of January 24, 2020). Thus, the used method was able to be used in writing masters' and doctoral dissertations.

The structure and scope of the dissertation. The thesis consists of an introduction, three chapters, conclusions, a list of used literature and appendix. The total volume is 189 pages of the text.

ЭЪЛОН ҚИЛИНГАН ИЛМИЙ ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ НАУЧНЫХ РАБОТ
LIST OF PUBLISHED SCIENTIFIC WORKS

I бўлим (I часть; I part)

1. Ataboev N.B. Chomsky's Critics to Corpus and the Problem of Expert Speakers' Linguistic Intuition // Филология масалалари. – Тошкент. (4)/2015. – Б. 108-111. (10.00.00. №18)
2. Ataboev N.B. The Main Theoretical Problems of Corpus Linguistics // Ўзбекистонда хорижий тиллар илмий-методик электрон журнал. – Тошкент. №6(14)/2016. – Б. 46-51. (10.00.00. №17)
3. Ataboev N.B. Conceptual Classification of Corpora by N.S. Dash // Филология масалалари. – Тошкент. (2)/2016. – Б. 120-124. (10.00.00. №18)
4. Ataboev N.B. The Procedure of Corpus Compiling (on the material of the lexeme) // Филология масалалари. – Тошкент. (3)/2017. – Б. 68-74. (10.00.00. №18)
5. Атабоев Н.Б. Корпус тилшунослиги шаклланишининг диахроник тадқиқи // Бердақ номидаги Қорақалпоқ давлат университетининг Ахборот-номаси. – Нукус. №4/ 2018. – Б. 147-150. (10.00.00. №12)
6. Ataboev N.B. Compiling Dictionaries by Using Corpus Analysis and its Advantages // International Journal of Progressive Sciences and Technologies (IJPSAT). – Spain. Vol. 9 № / July 2, 2018. – P. 206-212. (№ 23; SJIF: 3,9)
7. Атабоев Н.Б. Корпус лингвистикасининг асосий хусусиятлари // Ўзбекистонда хорижий тиллар илмий-методик электрон журнал. – Тошкент. №2(25)/2019. – Б. 37-44. (10.00.00. №17)
8. Атабоев Н.Б. Корпус ва компьютер лингвистикасининг фарқли жиҳатлари // Самарқанд давлат чет тиллари институти Хорижий филология тил-адабиёт-таълим Илмий-услубий журнал. – Самарқанд. №2(71)/2019. – Б. 109-112. (10.00.00. №10)
9. Ataboev N.B. Корпус таҳлили натижалари лингвомаданий бирликларнинг куантитатив ифодаси сифатида // Филология масалалари. – Тошкент. (2)/2019. – Б. 97-103. (10.00.00. №18)
10. Ataboev N.B. Application of Corpus Analysis in Teaching Foreign Languages // International Journal of Progressive Sciences and Technologies (IJPSAT). – Spain. Vol. 14 № 2 /May 2019. – P. 351-355. (№ 23; SJIF: 3,9).
11. Ataboev N.B. Problematic Issues of Corpus Analysis and its Shortcomings // International Scientific Journal Theoretical & Applied Science. – USA. Issue: 10, Vol. 78. / 14.10.2019. – P. 170-173. (№ 5; GIF: 0.564, № 2; JIF: 1.500, № 23; SJIF: 5.667)
12. Ataboev N.B. ICT in Linguistic Studies: Application of Electronic Language Corpus and Corpus-based Analysis // International journal of TEST Engineering and Management. – India. Vol.81/ November-December, 2019. – P. 4170-4176. (№ 3; Scopus: SJR 2019: 0.101)

13. Ataboev N.B. Corpus-Based Research on the Language Features of Corpus Linguistics: in the Example Of ECOCL // International Journal of Scientific & Technology Research. – India. Volume 9, Issue 02 / February 2020. – P. 4176-4179 (№ 3; Scopus: SJR 2019: 0,2)

14. Ataboev N.B. The Structure and Components of British National Corpus. // Шарқ ва ғарб тили, маданияти ва адабиёти ёшлар нигоҳида мавзусидаги талабалар республика илмий-амалий конференцияси мақолалар тўплами. –Тошкент. 2018. – Б. 36-37.

15. Ataboev N.B. Practical Application of Corpus Analyses in Translation and its Advances // Functional Aspects of Intercultural Communication. Translation And Interpreting Issues// Proceedings of international conference – 2019 Volume 6, – Москва: РУДН, 2019. – С. 708-718.

16. Атабоев Н.Б. Корпус таҳлиллари асосида ўзбек бадий асарларининг частотали луғатларини ишлаб чиқиш // МАТЕРИАЛЫ VI Международной научно-практической конференции «GLOBAL SCIENCE AND INNOVATIONS 2019: CENTRAL ASIA» – Нурсултан: 2019. VII ТОМ. – С. 111-115.

17. Атабоев Н.Б. Лисоний корпуслар таснифига назарий ёндашувлар. // Замонавий тилшунослик ва лингводидактиканинг коммуникатив аспекти Республика илмий-амалий анжумани материаллари. – Самарқанд. 7 май 2019. – Б. 112-114.

II бўлим (II часть; II part)

18. Ataboev N.B. Special Terminology in Using Corpora // «Бугунги кун ёшларининг тил ва адабиётни ўрганишдаги интеллектуал салоҳияти» мавзусидаги магистр ва иқтидорли талабаларнинг илмий-амалий анжуман материаллари. – Тошкент. 2017. – Б. 3-4.

19. Ataboev N.B. Structure of ‘COCA’ (Corpus Of Contemporary American English) and Simple Queries on it // Наука, образование и культура. – Россия. № 10 (44) / 2019. – С. 32-34.

20. Ataboev N.B. Practical Use of Corpus Analyses in Teaching Foreign Languages // Тошкент вилояти Чирчиқ давлат педагогика институти «Барқарор ривожланишда узлуксиз таълим: муаммо ва ечимлар» халқаро илмий-амалий анжуман илмий ишлар ТЎПЛАМИ. 1 Том. 21-24 май, 2019 й. – Б. 158-159.

21. Атабоев Н.Б. Инглиз тили корпус таҳлиллари асосида “она” концептининг ифодаланишини ўрганиш (BNC мисолида) // Best practices in distance education in learning foreign languages// International conference. Tashkent. 14-15 May, 2019. – Б. 28-30.

22. Ataboev N.B. Corpus Linguistics: Dysfunctions of Corpora // Trends in Teaching Foreign Languages: Achievements, Challenges and Solutions// international conference proceeding. – Tashkent. 16-17 October, 2019. – P. 172-173.

23. Атабоев Н.Б. Корпус таҳлилларининг таржимон фаолиятидаги ўрни // Translation, information, communication – political bridge // proceedings international scientific-practical conference. – Samarkand. July 7, 2019. – P. 259-260.

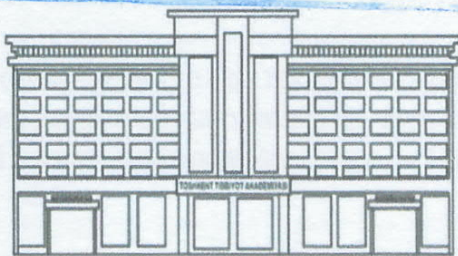
24. Атабоев Н.Б. Корпус таҳлиллари асосида лингвомаданий бирликларнинг қиёсланиши. // Intercultural communication through the prism of tourism in Uzbekistan: Experience, current issues and perspectives. – Samarkand. June 27-28, 2019. – P. 142-143.

25. Ataboev N.B. Peculiar Features of BNC and COCA // The role of innovation in the development of modern philology// international scientific conference. – Toshkent. 2019. – P. 61-64.

26. Атабоев Н.Б. Cord инглиз тили корпуслари тўғрисида маълумот берувчи база сифатида. // ИСТИҚЛОЛ ВА ТИЛ илмий даврий тўплами. – Тошкент. 2020. – Б. 133-136.

Автореферат «Жамият ва бошқарув» журнали
таҳририятида таҳрирдан ўтказилди

"AKADEMIYA" nashriyoti
30SISHGA RUXSAT ETTILADI
Muharrir _____
Muharririyat mud. _____
Bosh rassom _____
Texn.muhar. _____
Bosh muharrir _____
" " 20 yil



MUHARRIRIYAT VA NASHRIYOT BO'LIMI

Разрешено к печати: 24 октября 2020 года
Объем – 3,12 уч. изд. л. Тираж – 100. Формат 60x84. 1/16. Гарнитура «Times New Roman»
Заказ № 0730 -2020. Отпечатано РИО ТМА
100109. Ул. Фароби 2, тел: (998 71)214-90-64, e-mail: rio-tma@mail.ru