

**ТЕРМИЗ ДАВЛАТ УНИВЕРСИТЕТИ**  
**ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ**  
**PhD. 03/31.03.2021.Fil.78.04 РАҚАМЛИ ИЛМИЙ КЕНГАШ**  

---

**ТЕРМИЗ ДАВЛАТ УНИВЕРСИТЕТИ**

**ХОЛИЁРОВ ЎРАЛ МЕНГЛИЕВИЧ**

**ЎЗБЕК ТИЛИ ТАЪЛИМИЙ КОРПУСИНИ ТУЗИШНИНГ**  
**ЛИНГВИСТИК АСОСЛАРИ**

**10.00.01 – Ўзбек тили**

**ФИЛОЛОГИЯ ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)**  
**ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ**

**Термиз – 2021**

**Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси  
автореферати мундарижаси**

**Оглавление автореферата диссертации доктора философии (PhD) по  
филологическим наукам**

**Contents of dissertation abstract of doctor of philosophy (PhD) on  
philological sciences**

**Холиёров Ўрал Менглиевич**

Ўзбек тили таълимий корпусини тузишнинг лингвистик асослари.....5

**Халияров Урал Менглиевич**

Лингвистические основы составления узбекского обучающего корпуса.....23

**Kholiyorov Ural Menglievich**

Linguistic basis of compiling uzbek educational corpus.....43

**Эълон қилинган ишлар рўйхати**

Список опубликованных работ

List of published works .....49

**ТЕРМИЗ ДАВЛАТ УНИВЕРСИТЕТИ**  
**ХУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ**  
**03/31.03.2021.Fil.78.04 РАҚАМЛИ ИЛМИЙ КЕНГАШ**  

---

**ТЕРМИЗ ДАВЛАТ УНИВЕРСИТЕТИ**

**ХОЛИЁРОВ ЎРАЛ МЕНГЛИЕВИЧ**

**ЎЗБЕК ТИЛИ ТАЪЛИМИЙ КОРПУСИНИ ТУЗИШНИНГ**  
**ЛИНГВИСТИК АСОСЛАРИ**

**10.00.01 – Ўзбек тили**

**ФИЛОЛОГИЯ ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)**  
**ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ**

**Термиз – 2021**

**Фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида B2020.4.PhD/Fil1490 рақам билан рўйхатга олинган.**

Диссертация Термиз давлат университетида бажарилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме)) Термиз давлат университети веб-саҳифасининг [www.tersu.uz](http://www.tersu.uz) ҳамда “Ziyonet” ахборот-таълим порталининг [www.ziyonet.uz](http://www.ziyonet.uz) манзилларига жойлаштирилган.

<b>Илмий раҳбар:</b>	<b>Менглиев Бахтиёр Ражабович</b> филология фанлари доктори, профессор
<b>Расмий оппонентлар:</b>	<b>Шахабитдинова Шохида Хошимовна</b> филология фанлари доктори, профессор <b>Худойберганова Дурдона Сидиковна</b> филология фанлари доктори, профессор
<b>Етакчи ташкилот:</b>	<b>Самарқанд давлат университети</b>

Диссертация ҳимояси Термиз давлат университети ҳузуридаги илмий даражалар берувчи PhD.03/31.03.2021.Fil.78.04 рақамли Илмий кенгашнинг 2021 йил «03» 08 соат 10:00 даги мажлисида бўлиб ўтади. (Манзил: 190111, Термиз ш., Баркамол авлод кўчаси, 43-уй. Тел.: (876) 221-74-55; факс: (876) 221-71-17; e-mail: [termizdu@umail.uz](mailto:termizdu@umail.uz)).

Диссертация билан Термиз давлат университети Ахборот-ресурс марказида танишиш мумкин (3 рақам билан рўйхатга олинган). (Манзил: 190111, Термиз ш., Баркамол авлод кўчаси, 43-уй. Тел.: (876) 221-74-55)

Диссертация автореферати 2021 йил «26» 07 куни тарқатилди.  
(2021 йил «26» 07 даги 3 рақамли реестр баённомаси).



**Б.Умуркулов**  
Илмий даражалар берувчи илмий кенгаш  
раиси, ф.ф.д., доцент

**А.Эшмўминов**  
Илмий даражалар берувчи илмий кенгаш  
котиби, ф.ф.ф.д. (PhD)

**Ш.Махмараимова**  
Илмий даражалар берувчи илмий кенгаш  
қошидаги Илмий семинар раиси,  
ф.ф.д., доцент

## КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

**Диссертация мавзусининг долзарблиги ва зарурати.** Жаҳон тилшунослигида табиий тилни қайта ишлаш соҳалари кенгайиб, компьютер лингвистикаси ютуқларидан тилни тадқиқ этиш, лингводидактика, лексикография, таржимонлик борасида кенг фойдаланилмоқда. Жумладан, тилни ўрганишда тезаурус, электрон луғатлар, автоматик таржима, ахборот-қидирув тизимлари қаторида тил корпуслари ҳам асосий воситалардан бирига айланмоқда. Натижада тил бирликлари тўғрисидаги лингвистик маълумотларни махсус қидирув орқали санокли сонияларда қулай шаклда тақдим қилувчи, тил ёки тилнинг муайян қисмини минглаб матн кўринишида ўз ичига олган, тил имкониятларини ўргатишга мўлжалланган, лингводидактик хусусият касб этувчи таълимий корпуслар ҳам юзага келмоқда. Ҳозирги глобаллашув жараёнида дунё тамаддунининг виртуал оламга кўчиши, айрим тиллар қўлланишининг кенгайиши бошқа тилларнинг сақланиб қолишига хавф солмоқда, шунинг учун мазкур корпуслар тилнинг яшаб қолиши учун имкониятлардан бири сифатида аҳамиятлидир.

Дунё тилшунослигида XXI асрга келиб, корпус муаммоларини илмий-назарий жиҳатдан ўрганиш ҳаракати кучайди. Ҳозирда тил корпусларининг ижтимоий аҳамияти, тузиш тамойиллари, лингвистик теглар тизимини ишлаб чиқиш, автоматик разметкаш масалалари бўйича мақсадли тадқиқотлар олиб борилмоқда. Матн таҳлили жараёнига сарфланадиган вақтнинг тежалиши, тил бирликларининг нутқ таркибидаги хусусиятларини минглаб мисоллар билан кўрсата олиш корпуснинг асосий самарали имкониятлари сифатида эътироф этилмоқда. Шунинг учун жадал ривожланаётган компьютер лингвистикаси соҳасида таълимий корпус бирликларини танлаш, уларни лингвистик, экстралингвистик теглаш, таълимий корпус тузиш алгоритминини ишлаб чиқиш, таълимий корпус учун лингвистик таъминот яратиш ижтимоий заруратга айланмоқда.

Ўзбек компьютер лингвистикаси соҳасида ўзбек тилига замонавий ахборот технологиялари воситасида ишлов бериш борасидаги илмий изланишларга эътибор қаратилиши натижасида амалий ишлар сифатида миллий корпуснинг илк кўринишлари юзага келмоқда. “Ўзбек тилига оид барча илмий, назарий ва амалий маълумотларни ўзида жамлаган электрон кўринишдаги ўзбек тили миллий корпусини яратиш”<sup>1</sup> халқимиз ижтимоий ҳаётида ва халқаро миқёсдаги обрў-эътиборини оширишга улкан ҳисса қўшади. Шу боисдан ўзбек тилининг ахборот-коммуникация тизимида кенг қўлланишига эришиш, бунинг учун миллий-маданий меросимизни чуқур ўрганиш, она тили таълимида электрон ресурсларни кўпайтириш, ушбу ресурслардан таълим олувчиларнинг чегараланмаган ҳолда фойдаланишига эришиш, миллий-маънавий тарбияга хизмат қилувчи маданий мерос – она тили материалларини бир платформага тўплаш долзарб вазифага айланди.

---

<sup>1</sup> Ўзбекистон Республикаси Президентининг 2020 йил 20 октябрдаги “Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида”ги ПФ-6084-сон Фармони.  
URL: <https://lex.uz/docs/5058351>

Шунингдек, корпус лингвистикасининг истиқболли илмий йўналиши сифатида ўзбек тили таълимий корпусини тузиш, унинг лингвистик таъминотини яратиш сингари масалаларни замонавий илмий тамойиллар асосида тадқиқ этиш давлат тилини ўқитиш тизимини янада такомиллаштириш, унинг илм-фан тили сифатидаги нуфузини ошириш ҳамда республикамизнинг ижтимоий ривожланишини таъминлайдиган илмий-технологик захираларни шакллантиришга хизмат қилади.

Ўзбекистон Республикаси Президентининг 2017 йил 7 февралдаги ПФ-4947-сон “Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида”, 2019 йил 21 октябрдаги ПФ-5850-сон “Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан ошириш чора-тадбирлари тўғрисида”, 2020 йил 20 октябрдаги ПФ-6084-сон “Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида”ги фармонлари, 2019 йил 4 октябрдаги ПҚ-4479-сон “Ўзбекистон Республикасининг “Давлат тили ҳақида”ги Қонуни қабул қилинганлигининг ўттиз йиллигини кенг нишонлаш тўғрисида”ги қарори, 2020 йил 24 январдаги Олий Мажлисга Мурожаатномаси ҳамда мазкур фаолиятга тааллуқли меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишда мазкур тадқиқот маълум даражада хизмат қилади.

**Тадқиқотнинг Ўзбекистон Республикаси фан ва технологиялари ривожланишининг устувор йўналишларига мослиги.** Тадқиқот республика фан ва технологиялар ривожланишининг I. “Ахборотлашган жамият ва демократик давлатни ижтимоий, ҳуқуқий, иқтисодий, маданий, маънавий-маърифий ривожлантириш, инновацион иқтисодиётни ривожлантириш” устувор йўналишига мувофиқ бажарилган.

**Муаммонинг ўрганилганлик даражаси.** Жаҳон тилшунослигида корпус лингвистикаси янги соҳа бўлишига қарамай, соҳада мақсадли тадқиқотлар талай. Браун корпуси тузувчилари Н.Френсис, Г.Кучераларнинг корпус тузиш принциплари ҳақидаги фикрлари, рус тилшунослигида В.П.Захаров, А.Б.Кутузов, Е.В.Недошивина, В.В.Риков, В.Плунгянларнинг назарий ва амалий таклифларини корпус борасидаги дастлабки тадқиқотлар<sup>1</sup> сифатида баҳолаш мумкин. Бу мутахассислар корпус, унинг турлари, ўзига хос хусусияти, корпуснинг ижтимоий аҳамияти, корпус тузиш тамойиллари

---

<sup>1</sup> Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – М., 1967.; Синклер Дж. Предисловие к книге «Как использовать корпуса в преподавании иностранного языка» (Электронный ресурс). URL: [http://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4child/texts/sinclair\\_introduction.pdf](http://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4child/texts/sinclair_introduction.pdf)

масаласини ўрганишган бўлса<sup>1</sup>, таълимий корпуслар В.В.Мамонтова, Д.В.Сичинавалар томонидан тадқиқ этилган<sup>2</sup>.

Туркий корпус лингвистикасида тил корпуслари яратиш тамойиллари атрофлича ўрганилган бўлса-да<sup>3</sup>, таълимий корпус тузиш махсус тадқиқ манбаи бўлмаган.

Ўзбек тилшунослигида компьютер лингвистикаси соҳасидаги тадқиқотлар сифатида А.Пўлатов, М.Айимбетов, Н.Абдурахмонова, М.Абжаловаларнинг ишларини қайд этиш мумкин<sup>4</sup>. Бу тадқиқотлар матнни инновацион ёндашув – компьютер лингвистикаси ютуқлари ёрдамида тадқиқ этишнинг замонавий усулларини тавсия этганлиги билан аҳамиятли.

Ўзбек тили таълимий корпусини яратиш муаммолари махсус тадқиқ предмети бўлмаган, аммо айрим изланишларда масаланинг муайян қирралари ҳақида фикр билдирилган. Чунончи, ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш<sup>5</sup>, ўзбек-инглиз тили машина таржимасининг лингвистик таъминоти<sup>6</sup>, ўзбек тили муаллифлик корпусини тузиш тамойиллари<sup>7</sup>, ўзбекча матнларни таҳрир ва таҳлил қилиш модулларини яратиш масаласи<sup>8</sup>, тил корпуси лингвистик базасини тузиш тамойиллари<sup>9</sup>,

---

<sup>1</sup> Кутузов А.Б. Корпусная лингвистика. Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс). URL: [lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf](http://lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf); Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург, 2006. 26 с.; Рыков В.В. Курс лекций по корпусной лингвистике. URL: <http://tykov-cl.narod.ru/c.html>; Плуноян В. Зачем мы делаем Национальный корпус русского языка? “Отечественные записки” 2005, №2. URL: [http://magazines.russ.ru/oz/2005/2/2005\\_2\\_20-pr.html](http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html)

<sup>2</sup> Мамонтова В.В. Корпус параллельных текстов и база данных для исследования переводческих соответствий: проблемы и процедуры формирования. – Альманах современной науки и образования Тамбов: Грамота, 2008. № 8 (15): в 2-х ч. Ч. II. – С. 117-118. URL: [www.gramota.net/materials/1/2008/8-2/50.html](http://www.gramota.net/materials/1/2008/8-2/50.html); Сичинава Д.В. Европейский перфект сквозь призму параллельного корпуса. – Acta Linguistica Petropolitana. Труды института лингвистических исследований. – 2016, № 2. – Т. 12. – С. 85-114.

<sup>3</sup> Сулейманов Д.Ш., Гильмуллин Р.А., Гатауллин Р.Р. Морфологический анализатор татарского языка на основе двухуровневой модели морфологии / Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 2. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 327 с.; Сулейманов Д.Ш., Гатаиатуллин А.Р. Модель татарской аффиксальной морфемы и ее реализация // Серия: Интеллект. Язык. Компьютер. – Вып.4. – Казань: Фэн. – 1996. – С. 113-127.

<sup>4</sup> Каримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Алфавитли луғат. Частотали луғат. Терс луғат. – Тошкент, 2007.; Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Филол. фан. док. дисс. ...автореф. – Тошкент, 2008.; Ўринбоева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010.; Жуманазарова Г.У. Фозил Йўлдош ўғли дostonлари тилининг лингвопоэтикаси: Филол. фан. док. дисс. ...автореф. – Тошкент, 2017.; Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида): Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. афгореф. – Тошкент, 2018.

<sup>5</sup> Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. – Тошкент, 2006.

<sup>6</sup> Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида): Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. афгореф. – Тошкент, 2018. – 52 б.

<sup>7</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Қарши, 2018. – 168 б.

<sup>8</sup> Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун): Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Фарғона, 2019. – 164 б.; Абжалова М. Матнларга автолингвистик ишлов бериш тизимлари // Шестая Международная конференция по компьютерной обработке тюркских языков “TurkLang-2018”. (Труды конференции) – Ташкент, 2018. – 320 с.

<sup>9</sup> Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Қарши, 2019. – 140 б.

Ўзбек тили атов бирликларини семантик теглашнинг лингвистик асослари<sup>1</sup>, ўзбек тили морфологик анализаторининг лингвистик таъминоти муаммолари<sup>2</sup> монографик планда ўрганилган. Шунингдек, ўзбек-инглиз параллел корпусини тузиш муаммолари<sup>3</sup>, ўзбек тилидаги отларни автоматик таҳлил қилиш<sup>4</sup>, сўз ясалишининг формал моделлари<sup>5</sup> ҳақида қатор мақолалар нашр этилган. Бу тадқиқотлар компьютер лингвистикаси муаммоларини тадқиқ этганлиги билан долзарб, аммо ўзбек тили таълимий корпуси кун тартибига махсус қўйилмаган. Диссертацияни тайёрлаш жараёнида юқорида санаб ўтилган илмий изланишлар атрофлича ўрганилди, муносабат билдирилди ва улардан тадқиқотда фойдаланилди.

**Диссертация тадқиқотининг диссертация бажарилган олий таълим муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги.** Тадқиқот Термиз давлат университети илмий-тадқиқот ишлари режаси таркибига кирувчи “Ўзбек тилшунослиги: бадиий нутқ таҳлили, антропоцентрик назария, этнолингвистика, корпус тилшунослиги” мавзусидаги тадқиқот доирасида бажарилган.

**Тадқиқотнинг мақсади** ўзбек тили таълимий корпусини тузишнинг назарий масалаларини белгилаш, ўзбек тили таълимий корпусининг лингвистик таъминотини яратиш асосларини очиқ беришдан иборат.

**Тадқиқотнинг вазифалари:**

таълимий корпусга умумий характеристика бериш, унинг таълим ресурси сифатидаги аҳамиятини ёритиш;

таълимий корпуснинг она тили таълими, ўзбек тилини хорижий тил сифатида ўқитиш (лингводидактика)даги афзалликларини аниқлаш;

ўзбек тили таълимий корпусига материал танлаш тамойилларини ишлаб чиқиш, таълимий корпус фрагменти учун материал тўплаш;

ўқув луғатларини таълимий корпусга тўплаш жараёнини аниқлаш ва уларнинг қидирув тизимини ишлаб чиқиш;

назарий хулосалар чиқариш учун таълимий корпус морфолексиконини шакллантириш;

ўзбек тили морфологик ва сўз ясалиши луғатининг лингвистик асосларини яратиш;

сўз маънолари изоҳига, сўзлараро шаклий муносабатларга доир маълумотлар базасининг таркиби ва тузилишини очиқ бериш.

---

<sup>1</sup>Akhmedova D.B., Mengliev B.R. Semantic Tag Categories in Corpus Linguistics: Experience and Examination International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-3S, October 2019. – P. 208-212.

<sup>2</sup> Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Монография. (Электрон китоб) – GloeEdit, 2020. – 244 б. – 78 б.

<sup>3</sup> Karimov R., Mengliev B. Theoretical fundamentals of uzbek-english parallel corpus / Journal of critical reviews. ISSN- 2394-5125. – VOL 7, ISSUE 17, 2020. – P. 73-76.; Karimov R.A., Mengliev B.R. The Role of the Parallel Corpus in Linguistics, the Importance and the Possibilities of Interpretation International Journal of engineering and Advanced Technology (IJEAT). ISSN: 2249 – 8958, Volume-8, Issue-5S3 July 2019. – P. 388-391.

<sup>4</sup> Орхун М. Computational analysis of uzbek nouns / Шестая Международная конференция по компьютерной обработке тюркских языков “TurkLang-2018”. (Труды конференции) – Ташкент, 2018. – 320 с.

<sup>5</sup> Турсунов А. Вопросы словообразования в формальных моделях тюркских языков (на примере узбекского языка) // Шестая Международная конференция по компьютерной обработке тюркских языков “TurkLang-2018”. Труды конференции) – Ташкент, 2018. – 320 с.



**Тадқиқотнинг объекти** сифатида тил имкониятларини ўргатишга мўлжалланган, лингводидактик хусусият касб этувчи таълимий корпус танланган.

**Тадқиқотнинг предмети**ни ўзбек тили таълимий корпусининг лингвистик таъминоти масаласи ташкил этади.

**Тадқиқотнинг усуллари.** Тадқиқот мавзусини ёритишда таснифлаш, тавсифлаш, қиёслаш, статистик, моделлаштириш, корпус таҳлил усулларидан фойдаланилди.

**Тадқиқотнинг илмий янгилиги** қуйидагилардан иборат:

таълимий корпусга хос лингвистик, дидактик, инфорацион хусусиятлар жаҳон таълимий корпуслари функционал имкониятининг лингвистик таъминоти жиҳатидан фарқ қилиши, тилга автоматик ишлов бериш усуллари жиҳатидан ўхшаш эканлиги асосланган;

таълимий корпус материални автоматик қайта ишлаш тизимини йўлга қўйиш учун ўзбек тили морфолексикони ҳамда ўзбек тили морфологик базасини тузишнинг оптимал математик ва лингвистик усуллари аниқланган ҳамда морфолексикон ва морфологик база тузишнинг когнитив, лингво-статистик илмий-назарий асослари ишлаб чиқилган;

феъл шакллари луғатини тузиш тартиби, ўзак ва грамматик шаклларнинг корпусга жойлашувини тизимлаштиришнинг лингвистик, математик, мантикий, дастурий тамойиллари белгиланган ҳамда ўзбек тили таълимий корпусининг қидирув имкониятлари, интерфейс қурилиши, таълимий корпуснинг конкорданс ва электрон кутубхона сифатидаги таълимий, лингводидактик, ресурс имкониятлари очиб берилган;

таълимий корпус базасига ўқув луғатлари, сўз ясалиши, сўз маънолари, сўзлараро маъновий, шаклий муносабатларини ифодаловчи параметрларни кутубхона, конкорданс ва SQL қидируви усулида жойлаштириш тартиби ишлаб чиқилган.

**Тадқиқотнинг амалий натижалари** қуйидагилардан иборат:

таълимий корпус учун ўзбек тилини ўргатиш мақсадида ахлоқий-тарбиявий мазмундаги матнларни танлаш тамойиллари аниқланган, ўзбек тили таълимий корпусига материал тўпланган;

ўзбек тилида мавжуд ўқув луғатлари бир платформага тўпланган ҳамда қулай қидирув тизимига бўйсундирилган;

ўзбек тили таълимий корпуси лингвистик таъминоти – 37200 лексемадан иборат морфолексикони шакллантирилган; дастурий таъминот учун таклифлар тизими ишлаб чиқилган; ўзбек тили морфологик луғатини тузиш тартиблари асосида луғат яратилган.

**Тадқиқот натижаларининг ишончлилиги** ўрганилган материалнинг ўзбек тили табиатидан келиб чиққан ҳолда хулосалар қилишга ёрдам бериши, уларнинг асосли эканлиги, она тили ўқитиш методикасига назарий жиҳатдан мос келиши, методологик мукамаллиги, ўзбек тили таълимий корпусини тузиш тамойилларини ишлаб чиқишда амалда исботланган манбаларга таяниши билан изоҳланади.

**Тадқиқот натижаларининг илмий ва амалий аҳамияти.** Тадқиқот натижаларининг илмий аҳамияти ўзбек тили корпусларини яратишнинг асосларини ишлаб чиқишда, компьютер лингвистикаси йўналишида тадқиқотларни амалга оширишда илмий-назарий манба сифатида хизмат қилиши билан изоҳланади.

Тадқиқот натижаларининг амалий аҳамияти амалий филология бўлимларидан бири – корпус лингвистикасининг фан сифатида ўқитилиш жараёнида дастур, режалар тузиш, мавзуларни баён этишда манба, тил таълимида таълимий ресурсларни шакллантириш воситаси вазифасини ўташи билан изоҳланади.

**Тадқиқот натижаларининг жорий қилиниши.** ўзбек тили таълимий корпусини тузишнинг лингвистик асослари бўйича олинган илмий натижалар асосида:

таълимий корпус учун ахлоқий-тарбиявий мазмундаги матнларни танлаш тамойиллари, ўқув луғатлари, сўз маънолари изоҳи, сўзлараро маъновий, шаклий муносабатларга доир маълумотлар базасини шакллантириш, уларга мос қидирув тизимини яратиш, корпус базасига жойлаштириш ҳақидаги назарий маълумотлардан Ўзбекистон Республикаси Фанлар Академияси Қорақалпоғистон бўлими Қорақалпоқ гуманитар фанлар илмий-тадқиқот институти томонидан олиб борилган ФА-А1-Г007 «Қорақалпоқ нақл-мақоллари лингвистик тадқиқот объекти сифатида» мавзусидаги фундаментал илмий лойиҳасида фойдаланилган (ЎзР Фанлар академияси Қорақалпоғистон бўлимининг 2021 йил 22 апрелдаги 17.01/108-сон маълумотномаси). Натижада қорақалпоқ нақл-мақолларида тил бирликларининг морфологик шакллари ва позицияларини аниқлаштиришга, лингвистик база сифатида аҳамиятини очиб беришга эришилган;

ўзбек тили морфологик базасини тузиш орқали таълимий корпус морфолексиконини яратишнинг самарадор математик ва лингвистик усуллари, ўзбек тили лексемаларининг туркумга тегишлилик статистикасини аниқлаш, сўз ясалишининг лингвистик тадқиқ кўламини кенгайтириш; сўз ясалиши, сўз маънолари изоҳига доир маълумотлар базасининг таркиби ва тузилишига доир хулосалардан Ўзбекистон Республикаси Фанлар Академияси Қорақалпоғистон бўлими Қорақалпоқ гуманитар фанлар илмий-тадқиқот институтида бажарилган ФА-Ф1-Г003 “Ҳозирги қорақалпоқ тилида функционал сўз ясалиши” мавзусидаги илмий-амалий лойиҳасида фойдаланилган (ЎзР Фанлар академияси Қорақалпоғистон бўлимининг 2021 йил 22 апрелдаги 17.01/108-сон маълумотномаси). Натижада лойиҳа ишида ҳозирги қорақалпоқ тилида сўз ясалиш моделлари, сўз ясалиш асосларини таҳлил қилишда назарий умумлашмаларнинг мукаммаллашувига эришилган,

ўқув луғатлари, сўз ясалиши, сўз маънолари изоҳи, сўзлараро маъновий, шаклий муносабатларга доир маълумотлар базасининг таркиби ва тузилиши; уларни таълимий корпус базасига жойлаштириш ҳақидаги илмий-назарий муаммоларини ёритишдаги хулосалардан Олий ва ўрта махсус таълим вазирлиги томонидан дарслик сифатида тавсия этилган “Ҳозирги ўзбек тили”

дарслигининг “Лексик-семантик муносабатлар”, “Ўзбек тили лексикаси таснифи” деб номланган 17–18-параграфларида фойдаланилган (Ўзбекистон Республикаси Олий ва ўрта таълим вазирлигининг 2021 йил 24 майдаги 89-03-2167-сон маълумотномаси). Натижада сўз маънолари изоҳи, сўзлараро маъновий, шаклий муносабатларга доир мавзулар мазмунининг мукамаллашувига эришилган;

Ўзбек тили морфологик луғатини тузишнинг морфологик шакллари ва позицияларини аниқлаштириш, феъл шакллари луғатини тузиш тартиби, ўзак ва грамматик шаклларнинг корпусга жойлашувини тизимлаштириш тамойиллари ва имкониятлари, грамматик морфемаларнинг модели бўйича тавсиялардан ўзбек тили феъл шаклларининг морфологик луғатини яратишда фойдаланилган (O‘zbek tilining fe‘l shakllari lug‘ati (“kiymoq” fe‘li misolida) [Matn]: lug‘at. – Termiz: Termiz davlat universiteti nashr-matbaa, 2021. – 492 b. ISBN 978-9943-7243-3-4). Натижада морфологик шакллар базаси яратилиб, ўзбек тили миллий корпуси морфологик теглаш дастури ва морфологик анализаторини ишлаб чиқиш учун асосий лингвистик таъминот шакллантирилган;

Ўзбек тили таълимий корпусини тузиш зарурати, унга хос информацион хусусиятлар, лингвистик таъминоти, жаҳон таълимий корпусларидан фарқли ва ўхшаш жиҳатлари, лингводидактик аҳамиятига доир хулосалардан Сурхондарё вилоят телерадиокомпанияси адабий-бадиий, маънавий-маърифий дастурларининг сценарийсини тайёрлашда фойдаланилган (Сурхондарё вилоят телерадиокомпаниясининг 2021 йил 29 апрельдаги 01-01/54-сон маълумотномаси). Натижада юқори савияли баркамол авлодни вояга етиштириш, тил маданиятини шакллантириш ҳамда замонавий ахборот тизимида ўзбек тилининг имкониятларидан фойдаланиш ҳақидаги тарғибот ишлари мазмуни мукамаллашган.

**Тадқиқот натижаларнинг апробацияси.** Мазкур тадқиқот натижалари бўйича 5 та халқаро ва 3 та республика илмий-амалий анжуманларида маърузалар қилинган.

**Тадқиқот натижаларининг эълон қилинганлиги.** Диссертациянинг асосий мазмуни жами 15 илмий мақола (уларнинг 5 таси Ўзбекистон Республикаси ОАК тасарруф этган илмий журналлар ва 2 таси халқаро журналлар) ва тезисларда ўз ифодасини топган.

**Диссертациянинг тузилиши ва ҳажми.** Диссертация кириш, уч боб, умумий хулоса, фойдаланилган адабиётлар рўйхати ва иловадан иборат. Диссертациянинг умумий ҳажми 147 саҳифани ташкил этади.

## ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида тадқиқот мавзусининг долзарблиги ва зарурияти асосланган, тадқиқотнинг республика фан ва технологиялари ривожланишининг устивор йўналишларига мослиги, муаммонинг ўрганилганлик даражаси, диссертация бажарилган олий таълим муассасасининг илмий-тадқиқот ишлари билан боғлиқлиги, илмий

тадқиқотнинг мақсад ва вазифалари, илмий янгилиги, назарий ва амалий аҳамияти, объекти ва предмети, тадқиқот натижаларининг ишончлилиги, жорий қилиниши, апробацияси, тадқиқот натижаларининг эълон қилиниши, диссертациянинг тузилиши ва ҳажми ҳақида маълумотлар берилган.

Диссертациянинг биринчи боби **“Таълимий корпус лингводидактик база сифатида”** деб номланиб, унда ўзбек тили таълимий корпусининг умумий тавсифи, атамаси, корпуслар таснифидаги ўрни, лингводидактик (тил таълимидаги) аҳамияти, интерфейси, мазмунига қўйиладиган талаблар, қидирув менежери, лингвистик теглар тизими, таркиби, тузилиши, танланадиган матнлари, имкониятлари тавсифи таҳлил қилинган.

Таълимий корпус – материаллари тилни ўргатишга йўналтирилган, лингводидактик хусусиятга эга тил корпуси. Ўзбек тилининг таълимий корпуси ўзбек тили имкониятларини ўргатишга мўлжалланган, лингводидактик хусусият касб этган, электрон матнларни камраб олган, махсус сайт кўринишида амал қиладиган корпусдир. Ўзбек тили ёки ўзбек тилини чет тили сифатида ўрганувчилар эҳтиёжини қондириш учун тузиладиган ушбу корпус “таълимий корпус” атамаси сифатида қўлланиши мақсадга мувофиқ (инглиз ва рус тилларида: Learning Corpus, Educational Corpus, *обучающий корпус, учебный корпус*).

Инглиз ва рус тилларида таълимий корпуснинг хилма-хил намуналари яратилган. Масалан, инглиз тили сўзлашув тилини ўрганиш учун Ланкастер инглиз тили сўзлашув корпуси (Lancaster/IBM Spoken English Corpus<sup>1</sup>, Corpus of Spoken American English<sup>2</sup>), лингводидактик тадқиқотлар учун инглиз тили ўрганувчиларининг халқаро корпуси<sup>3</sup> шулар жумласидан. Ҳозирги вақтда миллионлаб сўзни камраб олган лингвистик корпуслар мавжуд: Британия Миллий корпуси (The British National Corpus, BNC 100 млн сўз), COBUILD Bank of English, (200 млн сўз)лар тилни ўрганиш учун жуда қулай воситалар сирасида олдинги ўринда туради. Шунингдек, оғзаки ва ёзма адабий тилни ифодалайдиган махсус корпуслар (масалан, CANCODE (Cambridge and Nottingham Corpus of Discourse in English), MICCASE (Michigan Corpus of Academic Spoken English (инглиз тили Оғзаки нутқ Мичиган корпуси) LIBELCASE (LIBEL Corpus of Spoken Academic English (Лимерик-Белфаст оғзаки академик нутқ корпуси) лингводидактиканинг бебаҳо қуруллари бўлиб улгурди<sup>4</sup>.

Рус тили миллий корпуси таркибида таълимий корпус (обучающий корпус)<sup>5</sup> мавжуд: у миллий корпуснинг ички корпуси сифатида ишлаб чиқилган, миллий корпуснинг таркибий қисми саналади. Унинг материали Миллий корпусдаги матнлардан ўқувчининг ёшига мослиги билан ажралиб

<sup>1</sup><https://www.routledge.com/A-Corpus-of-Formal-British-English-Speech-The-LancasterIBM-Spoken-English/Knowles-Assistant-Taylor-Research-Williams/p/book/9780582056398>

<sup>2</sup> <https://www.linguistics.ucsb.edu/research/santa-barbara-corpus>

<sup>3</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Қарши, 2018. – 168 б..

<sup>4</sup>Ахметова К.Ю. Корпусный подход в обучении иностранному языку. URL: <https://scipress.ru/philology/articles/korpusnyj-podkhod-v-obuchenii-inostrannomu-yazyku.html>

<sup>5</sup> <https://ruscorpora.ru/new/search-school.html>

туради. Шунингдек, Иркутск давлат тилшунослик университетида Таълимий мультимодал корпус (УМКО) тузиш давом этмоқда. Бу корпус материали рус, хитой тили ташувчиси, иккинчи тили сифатида ишлатувчи шахсларнинг маълум мавзу борасида олдиндан тайёрланмаган таълимий диалоглари видеоёзувидир.

Тил корпуслари турли тилларда олимлар томонидан тузилиши, мақсади, турғун, ўзгарувчанлиги ва бошқа жиҳатларига кўра бир-биридан фарқли бир қанча турларга ажратилади.<sup>1</sup> Ўзбек тилида махсус тадқиқот олиб борган Ш.Хамроева, А.Эшмўминовлар ҳам олимларнинг бу борадаги фикрини келтириб, корпус турлари таснифига алоҳида тўхталган.<sup>2</sup> Тадқиқотчи Ш.Хамроева тил корпуслари ҳақида батафсил маълумот келтиради ва уларни 13 хил нуқтаи назардан тасниф этиб, улар орасига таълимий ва миллий корпуслар киритилмаганлиги ҳақида маълумот беради.

Тасниф мезонларига кўра ҳар бир фарқланаётган, алоҳида кўрсатилаётган бирлик, албатта, тасниф ичида қарама-қарши турувчи бирликка эга бўлиши шарт. Шу жиҳатдан корпуслар йўналганлигига кўра академкорпуслар ва таълимий корпусларга бўлинади. Ҳозирги вақтда дунё тиллари бўйича тузилаётган миллий корпусларни академ корпуслар сифатида қараш мумкин.

Таълимий корпуслар йўналганлик нуқтаи назардан нутқ турига кўра: ёзма, оғзаки, аралаш; матн тилига кўра: русча, инглизча, ўзбекча ва ҳ.; параллеллигига кўра: бир тилли, икки тилли, кўп тилли; матннинг ихтисослашувига кўра: бадий, диалектал, суҳбат, терминологик, аралаш; жанрига кўра: адабий, фольклор, драматик, публицистик; кириш усулига кўра: эркин, пуллик, ёпиқ; мақсадига кўра: тадқиқий, тавсифий; изоҳланишига кўра: кўшимча изоҳли, изоҳсиз; ўзгарувчанлигига кўра: ўзгарувчан, турғун; изоҳ хусусиятига кўра: морфологик, синтактик, семантик ва ҳ. турларга бўлинади. Жаҳон тажрибасида таълимий корпуслар кўпроқ умумий, универсал табиатни намоён этади, лекин таълимий корпуслар ўрта мактаб, ўрта махсус касб-ҳунар коллежлари, лицей ўқувчилари, олий таълим талабалари ёки бошқа бирор мақсадда таҳсил олаётганлар учун алоҳида-алоҳида тузилиши мумкин.

Таълимий корпус электрон таълим ресурсларининг бир неча имкониятини ўзида мужассам этади, бу жиҳати билан улардан устунлик, мукамаллик касб этади.

Электрон ўқув ресурсларини яратишдаги каби таълимий корпусни тузишда ҳам умумий, дидактик, услубий, психологик-педагогик, техник-технологик, эстетик, эргономик ва санитар-гигиеник талаблар қўйилади.

Таълимий корпуслар миллий корпуслардан ўқув жараёнига мослашганлиги; ҳажмининг чегараланганлиги; киритилаётган турли

---

<sup>1</sup> Корпусная лингвистика. Статья в интернет-энциклопедии Википедия. URL: <http://wikipedia.org/wiki/>; Захаров В.П., Богданова С.Ю. Корпусная лингвистика: Учебник для студентов направления «Лингвистика». 2-е изд., перераб. и дополн., – СПб.: СПбГУ. РИО. Филологический факультет. – Санкт-Петербург, 2013. – С. 16.

<sup>2</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Қарши, 2018. – Б. 46-50; Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Қарши, 2019. – Б. 52-53.

матнларнинг аниқ, содда, тушунарли бўлиши; меъёрлаштирилганлиги; ўқув луғатларнинг киритилганлиги; турли қўшимча материаллар (расмлар, аудио, мультимедия) билан бойитиш мумкинлиги билан фарқ қилади.

Таълимий корпуснинг лингводидактикадаги аҳамияти махсус тадқиқ этилган<sup>1</sup>. Корпусдан унумли фойдаланувчи соҳа – лингводидактика; у она тили, хорижий тилни ўрганишда жуда аҳамиятли. Тил ўқитишда луғат бойлигини тўлиқ кўрсата олиш, сўзнинг қўлланиш имкониятини у ёки бу грамматик конструкция орқали тушунтириш учун мисоллар массивини кўрсатишда корпус жуда қўл келади. Ўқитувчи янги, ишонарли, чексиз ва хилма-хил мисолларни шу ердан топа олади, топшириқ, машқларни белгилашда қийналмайди, бир неча дақиқада мавзу бўйича янги мисоллардан иборат топшириқни тайёрлай олади.

Хорижий тилни (жумладан, ўзбек тилини ҳам хорижий тил сифатида) ўқитиш методологияси мураккаб, турли тил материални тақдим этишнинг содда, тушунарли усуллари топишга интилиш, тажрибани энг самарали тарзда узатиш (етказиш) усуллари доимий равишда излашдан иборат. Бу усулнинг энг қулай томони – корпус, у кўп миқдордаги маълумотларни бошқариш воситаси (русча, *инструментарий*), турли хил йўл билан тилнинг хаотик хилма-хиллигини тартибга солади; тартибга солинган сўзлар тўпламини шакллантиради<sup>2</sup>.

Таълимий корпус усуллари нафақат лингвистик тадқиқотларда, балки она тили, хорижий тил, таржима назарияси ва амалиётини ўқитишда ҳам жуда самарали инновацион манбадир. Ушбу усул эмпирик адекватлик, ҳақиқийлик, мослашувчанлик; муайян вазифа, мақсадли гуруҳга мослашиш каби жиҳатларни бирлаштиради. Корпус усули талаба/ўқувчига мустақил ишлаш, ўқитишда “кашфиёт” усулини қўллашга имкон беради, бу талаба/ўқувчини тилнинг чексиз имкониятини кўпроқ ўрганишга ундайди.

Таълимий корпусни фойдаланувчига янада қулайлик яратиш мақсадида турли платформалар билан интеграция қилиш мумкин. Масалан, мустаҳкамловчи-тренинг, “тьютор” дастурлари, тадқиқотлар ўтказишга ёрдам берувчи моделлаштирувчи дастурлар, таълимий ўйинлар ва ҳоказо.

Таълимий корпус интернет тармоғи орқали масофадан ўқитиш усулини намоён этади, бунда таълим олувчиларни: мактаб ўқувчилари; профессионал таълим тизими ўқувчилари; академик лицей ўқувчилари; абитуриентлар; талабалар; иккинчи олий ёки қўшимча маълумот олиш, малака ошириш ва

---

<sup>1</sup> Синклер Дж. Предисловие к книге «Как использовать корпуса в преподавании иностранного языка» (Электронный ресурс). URL: [http://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4child/texts/sinclair\\_introduction.pdf](http://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4child/texts/sinclair_introduction.pdf); Марданшина Р.М. Национальный корпус в практике лингвистических исследований и преподавании языка. URL: [https://kpfu.ru/staff\\_files/F1318514280/Nacionalnnyj.korpus.v.praktike.lingvisticheskikh.issledovaniy.i.prepodavaniy.yazyka.Mardanshina.RM.pdf](https://kpfu.ru/staff_files/F1318514280/Nacionalnnyj.korpus.v.praktike.lingvisticheskikh.issledovaniy.i.prepodavaniy.yazyka.Mardanshina.RM.pdf); Добрушина Н.Р., Левинзон А.И. Национальный корпус русского языка в образовании // Cuadernos de Rusística Española, 2006. № 2; <https://cyberleninka.ru/article/n/lc-tehnologii-v-issledovanii-osvoeniya-yazyka>. Камшилова О.Н. LC-технологии в исследовании освоения языка. URL: <https://cyberleninka.ru/article/n/lc-tehnologii-v-issledovanii-osvoeniya-yazyka/viewer>; Хамроева Ш. Таълимда корпусдан фойдаланиш // “Тил ва адабиёт таълими” журнали. 2017 йил сентябрь, № 9. – Б. 49-50.

<sup>2</sup> Krieger Daniel. Corpus Linguistics: What It Is and How It Can Be Applied to Teaching Siebold University of Nagasaki [Text] / Daniel Krieger (Nagasaki, Japan), 2003. URL: <http://iteslj.org/Articles/Krieger-Corpus.html>

қайта тайёргарлик ўташ истагида бўлганлар; минтақавий ҳокимият ва бошқарув раҳбарлари; анъанавий таълим тизимининг имкониятлари чекланганлиги сабабли маълумот ола олмаган ёшлар; ўз маълумот мақомини замонавий талаблар даражасига кўтариш истагида бўлган фирма ва корхоналар ходимлари; иккинчи параллел маълумот олишни хоҳлаган тингловчилар; марказдан узоқда, кам ўзлаштирилган минтақалар аҳолиси; эркин кўчиб юриши чекланган шахслар; жисмоний нуқсонлари бўлган шахслар; ҳарбий хизматда бўлган шахслар ва бошқалардан иборат ижтимоий гуруҳларга мансуб шахслар ташкил қилиши мумкин.

Фойдаланувчини корпус билан таништирувчи энг асосий таркибий қисм интерфейс ҳисобланади ва у мукамал, тушунарли, ўзбек тили хусусиятлари, ўқувчи ёшига мос жиҳатларни қамраши лозим.

Ўзбек тили таълимий корпуси материали сифатида ҳозирги ўзбек адабий тилидаги оммабоп, ахлоқий-тарбиявий мазмундаги расмий, илмий-оммабоп, бадий ҳамда публицистик услубдаги матнлар пропорционал тарзда танланади. Бугунги кунда ўзбек тили таълимий корпусини ишлаб чиқиш бўйича қуйидаги амалий натижаларга эришилди: ўзбек тили таълимий корпуси учун ахлоқий-тарбиявий мазмундаги матнлар танланган, корпусга жойлаштирилган; ўзбек тилида мавжуд ўқув луғатлари бир платформага тўпланган; қулай қидирув тизимига бўйсундирилган; ўзбек тили таълимий корпуси лингвистик таъминоти – 37200 лексемадан иборат морфолексикон шакллантирилган. Юқоридаги санаб ўтилган манбалар таълимий корпус учун бошланғич материаллар сифатида киритилган.

Таълимий корпусдан лексик бирлик бўйича қидирув қуйидаги кўрсаткичлар асосида амалга оширилади: 1) ёзилиши; 2) бўғинларга бўлиниши; 3) қайси имло қоидаси асосида ёзилганлиги; 4) имловий вариантдоши; 5) талаффузи ва урғуси; 6) товуш ўзгариши; 7) маъно(изоҳ)лари, кўп маънолилиқда маъно кўчиш усуллари; 8) синоним(лар)и; 9) антоним(лар)и; 10) токсонимик уяси, уядош(лар)и; 11) меронимик уяси, уядош(лар)и; 12) даражаланиш қатори; 13) шаклдош(омонимлар)и; 14) талаффуздош(пароним)и; 15) этимологияси; 16) ўз / ўзлашма қатламга мансублиги; 17) қўлланишига кўра: чегараланмаган / чегараланган: шевага хос/атама/арго ва жаргонлар; 18) даврига кўра: замонавий/эскирган/янги лексик бирлик; 19) диалектал (шева) варианты; 20) туб/ясамалиги; 21) таркиб (морфемалар)и; 22) тузилиш тури: содда/мураккаб: қўшма, жуфт, такрор; 23) туркумий белгиси турлари билан; 24) олд/орқа бирикувчилари; 25) услуби; 26) частотаси; 27) мисоллар.

Диссертациянинг иккинчи боби “**Таълимий корпус учун морфологик маълумотлар базаси**” деб номланган. Бобнинг “*Таълимий корпус учун морфолексиконни шакллантириши*” номли биринчи бўлимида табиий тилнинг грамматик луғати, морфолексикони, унинг хусусиятлари, табиий тилни қайта ишлаш имкониятлари ҳамда ўзбек тили морфолексикони бўйича бажарилган ишлар хусусида фикр юритилган.

Грамматик луғат – маълум бир тилдаги сўзларнинг туркумий белгиси, барча грамматик шакллари қамраб олувчи луғат. Морфолексикон – маълум

бир тил лексикасини қамраб оладиган, лексемага унинг туркумга мансублиги, грамматик хусусияти ҳақидаги маълумот бириктириладиган, табиий тилга ишлов бериш жараёнига мослаштириладиган маълумотлар базаси. Бундай маълумотлар базаси электрон кўринишда шакллантирилиб, табиий тилни қайта ишлашнинг турли воситалари (автоматик таржима, лингвистик корпус, морфологик, семантик, графематик анализатор)нинг лингвистик таъминоти сифатида амал қилади.

Ушбу бўлимда морфолексикон тузиш назарияси, тажрибасини ўрганиш асосида миллий корпус учун ўзбек тили морфолексиконини тузиш амалиёти тавсифланган.

Ўзбек тили морфолексиконини 4 босқичга бўлиб ишлаб чиқиш мақсадга мувофиқ: 1) ўзбек тилининг изоҳли луғатидаги 32500 лексемани маълумотлар базасига киритиш; 2) қўшма сўзлар луғати, имло луғати асосида қўшма сўзларни киритиш; 3) ўзбек тилининг изоҳли фразеологик луғатида берилган ибораларни, маталларни лексиконга киритиш; ҳамда туркумга мансублиги ҳақидаги ахборотни бириктириш; 4) ўзбек тилининг изоҳли луғатида берилган таркибли номларни саралаш ва базага қўшиш, турли соҳага оид атамаларни, атоқли отларни киритиш. Юқоридаги луғатлардаги ҳар бир лексик бирликка туркумга мансублик ва ўша туркум турлари ҳақидаги ахборотни бириктирилади. Натижада анча мукамал ўзбек тили морфолексикони вужудга келади.

Ўзбек тили морфолексиконининг дастлабки базасини шакллантириш жараёнида юқоридаги босқичларнинг 2 таси амалга оширилган ва қуйидагилар аниқланган: ўзбек тилида сўз туркумлари қуйидаги нисбатга эга: 22599 от, 7172 сифат, 3949 феъл, 1856 равиш, 800 тақлид сўз, 244 ундов сўз, 182 олмош, 132 модал сўз, 108 полифункционал сўз, 106 кўмакчи, 52 боғловчи, 45 юклама, 30 сон сўз туркумига мансуб лексема мавжуд; лексемаларнинг жами 37282 тани ташкил этади.

Бобнинг “*Ўзбек тили морфологик луғатини шакллантириш хусусида*” деб номланган иккинчи бўлимида морфологик шакллар луғатининг таълимий корпус учун лингвистик база сифатида аҳамияти, ўзбек тили морфологик тизимининг ўзига хос хусусиятлари, морфологик шакллар луғатини тузиш амалиёти изоҳланган.

Ривожланган мамлакатлар тилларининг (масалан, инглиз ва рус тили) ҳар бирида кўплаб грамматик луғатлар яратилган. Бироқ ўзбек тили грамматик луғатининг бирор тури ҳозиргача яратилмаган. Бунга сабаб ўзбек тили морфологик тизимининг жуда мураккаблиги ва морфологик шаклларнинг миқдори жуда кўплигида бўлса керак. Шунинг учун тилимизда таржима, орфографик текширув, морфологик автоматик анализатор ва бошқа компьютер дастурлари яратишда қийинчиликларга дуч келинмоқда. Анъанавий карточка усулида бундай далилларни тўплаш ва тартиблаш учун бир кишининг умри етмайди.

Ўзбек тили феъл шаклларининг морфологик луғатини тузишда ўзбек тилида мавжуд грамматик шакллар ва уларнинг барча позициясини аниқлаш, ҳар бир шаклнинг грамматик таркиби, грамматик морфемаларнинг моделини



кўрсатиш мақсад қилиб олинади. Мақсадни амалга оширишда амалий тадқиқ учун феъл шаклларининг морфологик луғатини туздик. Ўзбек тили феъл шакллари кўплиги боис фақат битта феъл (*киймоқ*) танланди.

Ўзбек тилидаги феъл шаклларининг миқдори ҳақидаги фикрлар ва амалий ишланмалар фақат битта манбада кузатилди ва унга кўра ўзбек тилида битта феълнинг тахминан 100 минг грамматик шакли бўлади. Амалий ишлар бу миқдорнинг келтириб чиқариш йўли унчалик ўзини оқламаслиги, миқдори ҳам тўғри эмаслиги, барча феъллар учун бир миқдорни белгилаш хато эканлигини исботлади.

Ўзбек тилининг бир феъл (*киймоқ*) мисолидаги феъл шакллари луғати нисбат, замон, майл, шахс-сон, бўлишли-бўлишсизлик, вазифа шаклларининг аранжировкасидан 30 минг атрофидаги сўзшаклни ҳосил қилди. Юқоридаги шаклларга кўмакчи феъллар бириктирилганида биргаликда 300 мингдан ортиқ шаклни олган бир феъл шакллари ҳосил қилинди. Амалий ишланмалар ўзбек тилида ҳар хил феълнинг грамматик шакллари 4-5 мингдан 300 мингдан ошишини кўрсатди (юкламалар билан биргаликда яна баробарларга кўпаяди). Мазкур грамматик луғат нафақат таълимий корпуснинг лингвистик маълумотлар омбори, балки келажакда, ўзбек тили миллий корпуси морфологик теглаш дастурини, морфологик анализаторини ишлаб чиқишда асосий лингвистик таъминот бўлиб хизмат қилади.

Бобнинг *“Таълимий корпус учун сўз ясалиши луғатини шакллантириши асослари”* деб номланган учинчи бўлимида ўқув луғатчилиги ва сўз ясалиш луғатлари тажрибаси, ўзбек тили таълимий корпуси учун сўз ясалиши луғати лингвистик таъминоти, тузилиши ва структураси хусусида фикр юритилади.

Ўзбекистонда ўқув луғатчилиги соҳасини намунали даражада ривожланган деб баҳолай олмасак-да, ўқув луғатчилигининг лингвистик асослари, тарихи ва истиқболлари каби масалаларнинг монографик планда тадқиқ этилганлиги<sup>1</sup>, бир қанча ўқув луғатининг нашр этилгани ўзбек таълимий лексикографиясининг бир жойда тўхтаб қолмаганидан далолат беради. Ушбу луғатлар ўзбек тили таълимий корпусини тузишда асосий лингвистик таъминотлардан бири бўлиб хизмат қилади.

2008 йилда муаллифлар жамоаси томонидан тузилган ўзбек тилининг сўз ясалиши ўқув луғати<sup>2</sup> таълимий корпус сўз ясалиш базаси учун лингвистик-лексикографик манба вазифасини ўтайди. Таълимий корпус учун сўз ясалиш луғатини шакллантиришда қуйидаги омилларга амал қилиш лозим: биринчидан, сўз ясалиш луғати вокабуласини ҳамда сўз ясалиш уясини аниқлаш; иккинчидан, деривацион луғат тузиш методикасига амал қилиш; учинчидан, онлайн маҳсулот эканлигини инобатга олиб, анъанавий луғатдан фаркли равишда ҳажмга чегара қўймаслик; тўртинчидан, луғат мақоласи учун бир хил параметр белгилаш; бешинчидан, бош сўз, ясалиш асоси ҳамда ясовчи воситани кўрсатиш; олтинчидан, ясалиш парадигмасини кўрсатиш учун

<sup>1</sup> Баҳриддинова Б. Ўзбекистонда ўқув луғатчилиги: лингвистик асослари, тарихи ва истиқболлари: Филология фанлари доктори диссертацияси. – Самарқанд, 2020. – 256 б.

<sup>2</sup> Mengliyev B., Bahriddinova B., Xoliyorov O'. va boshq. O'zbek tilining so'z yasaliishi o'quv lug'ati. – Toshkent: Yangi asr avlodi, 2007. – 176 b.

ясалиш уясига ўзаро ҳавола қиладиган индекслар қўйиш; еттинчидан, бир туркумдан бошқа туркумнинг ясалиши ҳақида ахборот берувчи маълумот (тег) киритиш.

Сўз ясалиш луғати мақоласи тузилиши ҳамда электрон луғатнинг имкониятларини инобатга олган ҳолда, корпуснинг сўз ясалиш луғати таркибини шакллантириш бўйича қуйидаги тавсиялар ишлаб чиқилди: 1) ўзбек тили таълимий корпуси учун сўз ясалиш луғатига асос сифатида 2009 йилда нашр этилган сўз ясалиш ўқув луғати танлаб олинади; 2) сўз ясалиш таркиби: ясалиш асоси, ясовчи восита ажратиб кўрсатилади; 3) ҳар бир сўз ясалиш уясига ясалиш типи ҳақида ахборот берилади; 4) нотипик сўз ясалиш ҳолатлари ҳам изоҳланади; шундай хусусияти ажратиб кўрсатилади; 5) сўз бирикмасидан ясалиш ҳолати: ясалиш асоси, ясовчи восита ва ясалма кўрсатилади; 6) қоғоз луғатда, асосан, аффиксация усули билан ясалиш келтирилган. Корпус имкониятидан келиб чиқиб корпус базасидан маълум ўзакдан синтактик усул билан ясалган сўз ҳамда унинг контекстига ҳавола берилади.

Диссертациянинг учинчи боби **“Таълимий корпуснинг сўз маънолари таҳлилига доир маълумотлар базасини тайёрлаш”** деб аталади. Биринчи бўлим *“Сўз маънолари изоҳига доир маълумотлар базасини шакллантириши масалалари”* деб номланади.

Ўзбек корпус лингвистикасида миллий корпус учун бирликларни семантик теглаш, маънолар ҳақида тег (изоҳ) бириктириш, семантик маълумотлар базасини шакллантириш бўйича бир неча тадқиқот амалга оширилган. Жумладан, Ш.Хамроева семантик тегларни ишлаб чиқишнинг умумий тамойиллари, А.Эшмўминов миллий корпуснинг синоним сўзлар базасини шакллантириш, Д.Ахмедова ўзбек тили миллий корпусида сўзлараро лексик-семантик муносабатнинг ифодаланиши масаласини ўрганган<sup>1</sup>.

Ўзбек тили таълимий корпуси қидирув ойнасига исталган сўз киритилганда, дастлаб шу сўз ишлатилган контекстлар рўйхати кўринади. Унда қидирилган сўз алоҳида ранг билан ажратилган бўлади. Фойдаланувчида ихтиёрий контекстни ажратиб таҳлил қилиш имкони бўлади. Контекстдаги маълум бир сўзга сичқончанинг ўнг тугмаси босилса, гиперҳавола орқали шу сўзнинг маънолари акс этган алоҳида ойна очилади. Гиперҳавола сўз маъносини таълимий корпус лингвистик базасида мавжуд луғат таркибидан қилдириб, топиб, натижа ойнасига жойлайди.

Сўзнинг изоҳларини беришда ўзбек тилининг изоҳли луғати<sup>2</sup>га асосланиш – энг оптимал йўл, аммо изоҳли луғатдан фарқли равишда корпусдаги изоҳни соддалаштиришга алоҳида эътибор берилади. Ўзбек тили

---

<sup>1</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD)...дисс. – Бухоро, 2018. – 168 б.; Эшмўминов А.А. Ўзбек тили миллий корпуси синоним сўзлар базасини яратиш асослари: Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси. – Қарши, 2019. – 146 б.; Ахмедова Д.Б. Атов бирликларини ўзбек тили корпуслари учун лексик-семантик теглашнинг лингвистик асос ва моделлари: Филология фанлари бўйича фалсафа доктори (PhD) диссертацияси. – Бухоро, 2020. – 138 б.

<sup>2</sup> Ўзбек тилининг изоҳли луғати. 5 томлик. – Тошкент: Ўзбекистон Миллий энциклопедияси, 2000-2006.

ўқув изоҳли луғатининг ҳозиргача мавжуд эмаслиги таълимий корпусда сўз изоҳини беришда қийинчилик келтириб чиқаради.

Мавжуд луғатдан таълимий корпус базасини шакллантиришда қуйидагича фойдаланиш мумкин, масалан,

**БАҚАР**<sup>1</sup> [а. بقر – буқа, ҳўкиз] **айн.** сигир йили **қ.** сигир. *Ундан ёшини сўраганларида, у – муш, бақар, паланг, харгўш, наҳанг дея, эскича йил аздариб... жавоб беради. Ғ.Расул, Адолат. – Афанди, бузоқни арабча нима дейди? – Сигир бўлгунча сабр қилади, ундан кейин бақар дейди. «Латифалар».*

Ушбу изоҳдан сўзнинг маъноси билан бирга ўз/ўзлашма қатламга мансублик ҳамда этимологик маъноси маълум бўлади. Кўриняптики, *бақар* сўзи араб тилидан ўзлашган; араб тилида *буқа, ҳўкиз* деган маънони билдиради. Шу асосда таълимий корпус базасининг бир устунига сўзнинг қайси тилдан ўзлашганлиги, бир устунига асл тилда қандай маънони билдириши, бир устунга ўзбек тилида айнан нимани англатиши ёзилиши лозим. Масалан:

Сўз	Ўз/ўзлашганлиги	Асл тилдаги маъноси (этимологияси)	Ўзбек тилидаги маъноси
БАҚАР	арабча	буқа, ҳўкиз	сигир йили

Изоҳдан кўриняптики, ушбу сўзга “айн.” белгиси билан бошқа бир сўзга ҳавола қилиш ўрни мавжуд: **айн.** сигир йили **қ.** сигир. Шу сабабли маълумотлар базасига яна бир устун киритамиз. Бу устунга, агар мавжуд бўлса, сўз ҳавола этилган бирликни киритамиз ҳамда бу бирликларни ўзаро гиперҳавола билан боғлаймиз:

Сўз	Ўз/ўзлашганлиги	Асл тилдаги маъноси (этимологияси)	Ўзбек тилидаги маъноси	Ҳавола
БАҚАР	арабча	буқа, ҳўкиз	сигир йили	сигир

Мазкур имконияти билан корпус луғатдан фарқланади: сўз саҳифа оралаб қидирилмайди, битта тугма(гиперҳавола)ни босиш билан исталган сўзнинг изоҳи топилади.

Сўз изоҳида келтирилган иллюстратив мисолларни корпус маълумотлар базасига киритишга эҳтиёж йўқ. Биринчидан, изоҳли луғатдаги мисоллар эскиради, иккинчидан, корпус базасидаги минглаб матнлардан иллюстратив мисол танлаш имконияти мавжуд бўлади.

Луғатда полисемантик сўзларнинг маънолари ажратилган. Демак, маълумотлар базасида унинг бу хусусияти алоҳида кўрсатилиши лозим. Шунинг учун сўз изоҳи берилган устун битта бўлмайди: изоҳ<sub>1</sub>, изоҳ<sub>2</sub>, изоҳ<sub>3</sub> шаклида изоҳлар киритилади. Ушбу маълумотлар ҳамда битта алгоритм билан маълумотлар базасида полисемантик ва моносемантик сўзларни ажратиш олишга эришиш мумкин.

Бобнинг “Сўзлараро маъновий муносабатларга доир маълумотлар базасини яратиш” деб номланган бўлимида сўзлараро маъновий муносабатнинг полисемия, синонимия, антонимия, гипонимия, градуонимия,

<sup>1</sup> Ўзбек тилининг изоҳли луғати. 1-том. – Тошкент: Ўзбекистон Миллий энциклопедияси, 2000. – Б. 176.

партонимия, иерархонимия каби турларининг таълимий корпусда лингвистик базасини яратиш тавсифи изоҳланган, тавсиялар ишлаб чиқилган.

Ўзбек тилида яратилган антоним ва синонимларларнинг ўқув луғатларини маълумотлар базасига боғлаш, шу луғатлардан гиперҳавола асосида қидирувни йўлга қўйиш дастурчи учун қийинчилик туғдирмайди. Энг асосийси, хом ашё мавжуд, фақат уни базага жойлаштириш, қидирувга бўйсундириш масаласи ечилиши лозим. Бунинг учун ушбу луғат матнлари бир неча техник қайта ишловдан ўтказилиши лозим.

Ўзбек тилида гипероним ва гипоним орасида семантик, логик-семантик алоқа мавжуд. Бундай лингвистик базани қуйидагича шакллантириш мумкин: ўзбек тилининг умумий лексик базаси олинади; умумий база бир устунга қўйилгандан кейин ёнидан яна бир устун қўшилади; умумий лексик баъза гипоним сифати белгиланиб, ёнидаги устунга гипероними киритилади; дастурчи ёнма ён устунлардаги бирликларни бир-бирига гиперҳавола орқали бириктиради. Натижада барча сўзлар бир-бирига гипоним-гипероним муносабатда боғланиб қолади.

Гипоним	Гипероним
жўжа	товуқ
терақ	дарахт
хўроз	товуқ
тол	дарахт
гилос	дарахт

Баъзан бир лексема маънолари кенглигидан бир нечта гиперонимга боғланиши мумкин. Масалан, *олма*, *гилос*, *шафтоли*, *ўрик* каби лексемалар *дарахт* гипероними билан бирга *мева* гипероними билан ҳам бирикади. Бундай ҳолларда устунлар ёнига яна устунлар қўшиб кетавериш мумкин ва гипероним<sup>1</sup>, гипероним<sup>2</sup> атамалари билан номласа бўлади.

Ўзбек тили таълимий корпуси даражаланиш қатори маълумотлар базасини шакллантириш учун лингвистик таъминот етарли эмас, чунки ўзбек тилида бундай луғат яратилмаган. Ушбу маълумотлар базаси монографик тадқиқот натижалари<sup>1</sup>, ўқув даражаланиш луғати материалларидан фойдаланиб шакллантирилади.

Ўзбек тилида даражаланиш қатори маълумотлар базасини шакллантириш учун сўзлар рўйхати, даражаланиш – маънонинг кучайиб / камайиб боришига қараб бирликлар киритилади. Қидирув натижасига эса маълумотлар базасидаги устунлар эмас, балки қатор маълумоти чиқарилади: базада нечта сўз бўлса, шунча қатор шаклланади. Қидирув ойнасига база таркибига кирган сўзлар орасидан қайси сўз киритилса ҳам, ўша сўз даражаланиш қаторида кўрсатилади.

Ўзбек тилида бутун-бўлак муносабатини акс эттирувчи маълумотлар базасини шакллантириш учун от туркумига мансуб сўзлар базаси ажратиб олиниб, бутун/қисмни билдирувчи сўз эканлиги ҳақида ахборот киритилиши

<sup>1</sup> Орифжонов Ш. Ўзбек тилида луғавий градуонимия: Филол. фан. номз. дисс. ...автореф. – Тошкент, 1996. – 21 б.

лозим. Базани ажратмай, морфолексиконнинг ўзига битта устун қўшиш кифоя қилади. Аммо бу усул гиперонимик уяни кўрсатиш учун етарли эмас: отлар лексиконида бутун/қисмни ифодаловчи сўз эканлиги ҳақидаги ахборотни киритиш бу сўзнинг қайси уяга тегишли эканлиги ҳақида хулоса беролмайди. Таълимий корпусда бундай маълумотлар базасининг мавжудлиги ўзбек тили миллий корпуснинг мереологик қидирув тизимини ишлаб чиқиш учун асос бўлади.

Бобнинг *“Сўзлараро шаклий муносабатларга доир маълумотлар базасининг таркиби ва тузилиши”* деб номланган бўлимида таълимий корпус учун омоним, пароним базасини яратиш хусусида фикр юритилади.

Тадқиқотнинг амалий ишланмаси сифатида тузилган ўзбек тили таълимий корпусининг омоним сўзлари базаси 2384 лексемани ташкил этди. Омоним сўзлар маълумотлар базасини шакллантиришда ҳажми чегараламаслик тамойилига асосланиш; ўзбек тилида қайси манба бўлишидан қатъи назар, омоним сўзларни тўплаш, уларни базага киритиш ҳамда базани босқичма-босқич тўлдириб бориш функциясини ишга тушириш; тузилаётган корпус ўзбек тили имкониятларини тавсифлашга йўналтирилганлиги туфайли (оддий луғатдан фарқли ўлароқ) лексеманинг русча эквивалентини бермаслик; маълумотлар базасининг ҳажми катталиги сабабли, изоҳларни қисқартирмаслик тамойилларига амал қилиш лозим.

Ўзбек тили таълимий корпуси пароним сўзлар базасини автоматик /яримавтоматик усулда шакллантиришнинг имкони деярли мавжуд эмас: ягона йўл лингвистик таъминот тайёрлаш. Ҳатто нашр этилган луғатдан тўғридан-тўғри импорт қилиш учун ҳам ягона стандарт кузатилмади. Шу сабабдан луғат мақоласи тузилиши асосида жадвал шакллантирилади.

## ХУЛОСА

1. Ўзбек тилининг таълимий корпуси – фойдаланувчиларга ўзбек тилини она тили ёки чет тили сифатида ўргатишга мўлжалланган, электрон матнларни қамраб олган, махсус сайт кўринишида амал қиладиган корпус.

2. Корпусларни йўналганлигига кўра академикорпуслар ва таълимий корпусларга бўлиш мумкин. Таълимий корпуслар миллий корпуслардан ўқув жараёнига мослашганлиги; ҳажмининг чегараланганлиги; киритилаётган матнларнинг аниқ, содда, тушунарли бўлиши; меъёрлаштирилганлиги; ўқув луғатларнинг киритилганлиги; турли қўшимча материаллар (расмлар, аудио, мультимедия) билан бойитиш имконияти билан фарқ қилади

3. Ўзбек тили таълимий корпуси материали сифатида ҳозирги ўзбек адабий тилидаги оммабоп, ахлоқий-тарбиявий мазмундаги расмий, илмий-оммабоп, бадиий ҳамда публицистик услубдаги матнлар пропорционал тарзда танланади. Матнлар таълимий характерга эга бўлган лингвистик маълумотлар билан батафсил разметкаланади. Булар сирасида морфологик разметка алоҳида ўрин тутаяди. Бунга грамматик луғат асос бўлиб хизмат қилади. Грамматик луғат – маълум бир тилдаги сўзларнинг туркумий белгиси, барча грамматик шакллари қамраб олувчи луғат. Морфолексикон – маълум бир

тил лексик бирликларини камраб оладиган, унинг туркумга мансублиги, грамматик хусусияти ҳақидаги маълумот бириктириладиган, табиий тилга ишлов бериш жараёнига мослаштириладиган электрон маълумотлар базаси бўлиб, табиий тилни қайта ишлашнинг турли воситалари (автоматик таржима, лингвистик корпус, морфологик, семантик, графематик анализатор)нинг лингвистик таъминоти сифатида амал қилади.

4. Ўзбек тили морфолексикони шакллантириш мураккаб бўлганлиги сабабли 4 та босқичда амалга оширилади. Ўзбек тили морфолексиконининг дастлабки базасини шакллантириш жараёнида босқичларнинг 2 таси амалга оширилган ва қуйидагилар аниқланди: ўзбек тилида сўз туркумлари нисбатлари: 22599 от, 7172 сифат, 3949 феъл, 1856 равиш, 800 тақлид сўз, 244 ундов сўз, 182 олмош, 132 модал сўз, 108 полифункционал сўз, 106 кўмакчи, 52 боғловчи, 45 юклама, 30 сон сўз туркумига мансуб лексема мавжуд; лексемаларнинг жами 37282 тани ташкил этади.

5. Ўзбек тили таълимий корпус морфологик базасини шакллантиришда амалий тадбиқ учун феъл шакллариининг морфологик луғати тузилди. Ўзбек тилининг бир феъл (*киймоқ*) мисолидаги феъл шакллари луғати нисбат, замон, майл, шахс-сон, бўлишли-бўлишсизлик, вазифа шакллариининг аранжировкасидан 30 минг атрофидаги сўзшаклини ўз ичига олди. Юқоридаги шаклларга кўмакчи феъллар бириктирилганида биргаликда 300 мингдан ортиқ бир феъл шакллари ҳосил қилинди. Ўзбек тилида ҳар хил феълнинг грамматик шакллари 4-5 мингдан 300 минггача боради. Юкламалар билан биргаликда яна баробарларга ошади.

6. Омоним ва полифункционал сўзни морфолексикон базасидаги туркумга мансублик ҳақидаги ахборот билан аниқлаб бўлмайди. Бу каби бирликларнинг маъноси матнда реаллашгани учун омоним, полифункционал сўзни автоматик аниқлаш алгоритми махсус тадқиқни талаб этади. Антонимлар базасини шакллантиришда антонимик жуфтнинг иккала аъзосига ҳам изоҳли луғатга гиперҳавола тарзида ишлов бериш тавсия этилади. Синонимлар базасини шакллантиришда маълумотни синонимлар луғатидан маълумотлар омборига қўлда киритиш, қидирув тизимини ишга тушириш учун сўзларнинг ўзаро ҳаволага эгаллигини ҳал этиш талаб этилади.

7. Таълимда корпусдан фойдаланиш (1) тилни ўрганувчи, унинг ташувчиси бўлган киши нутқини қиёслаш; (2) адабий оғзаки/ёзма тилни чоғиштириш; (3) тил услубларини китоб, дарслик, нашр материали бўйича қиёслаш, фарқини аниқлаш; (4) гапларда сўз тартибини аниқлаш каби жараёнлар учун замонавий инструментарий вазифасини бажаради.

**НАУЧНЫЙ СОВЕТ PhD. 31.03.2021. FIL. 78.04 ПО ПРИСУЖДЕНИЮ  
УЧЁНЫХ СТЕПЕНЕЙ ПРИ ТЕРМЕЗСКОМ ГОСУДАРСТВЕННОМ  
УНИВЕРСИТЕТЕ**

---

**ТЕРМЕЗСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**ХАЛИЯРОВ УРАЛ МЕНГЛИЕВИЧ**

**ЛИНГВИСТИЧЕСКИЕ ОСНОВЫ СОСТАВЛЕНИЯ УЗБЕКСКОГО  
ОБУЧАЮЩЕГО КОРПУСА**

**10.00.01 – Узбекский язык**

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ  
на соискание ученой степени доктора философии (PhD)**

**Термез – 2021**

Тема диссертации доктора философии (Phd) по филологическим наукам зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан за номером B2020.4.PhD / Fil1490.

Диссертация выполнена в Термезском государственном университете.  
Автореферат диссертации доступен на трех языках (узбекский, русский, английский (резюме)) размещен на сайте Научного совета (www.tertsu.uz) и информационно-образовательном портале «ZiyoNet» (www.ziynet.uz).

**Научный руководитель:** Менглиев Бахтиёр Раджабович  
доктор филологических наук, профессор

**Официальные оппоненты:** Шахабитдинова Шохида Хашимовна  
доктор филологических наук, профессор  
Худойберганова Дурдона Сидиковна  
доктор филологических наук, профессор

**Ведущая организация:** Самаркандский государственный университет

Защита диссертации состоится «03» 08 2021 года в 10:00 часов на заседании научного совета PhD.03 / 31.03.2021.Phil.78.04 при Термезском государственном университете. (Адрес: 190111, г. Термез, ул. Баркамол авлод, 43. Тел: (876) 221-74-55; факс: (876) 221-71-17; e-mail: termizdu@umail.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Термезского государственного университета (зарегистрирован под номером 3). Адрес: 190111, г. Термез, ул. Баркамол авлод, 43. Тел: (876) 221-74-55.

Автореферат диссертации разослан «26» 07 2021 года.  
(Реестр протокола рассылки № 3 от 2021 года «26» 07).



*[Signature]*  
**Б.Умуркулов**  
Председатель Научного совета  
по присуждению ученых степеней,  
доктор филол. наук, доцент

*[Signature]*  
**А.Эшмунинов**  
Ученый секретарь Научного совета  
по присуждению ученых степеней  
доктор философии (Phd)  
по филологическим наукам

*[Signature]*  
**Ш.Махмараимова**  
Председатель научного семинара при  
Научном совете по присуждению  
ученых степеней, доктор  
филологических наук, доцент



## **ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))**

**Актуальность и востребованность темы диссертации.** В мировой лингвистике расширилась область обработки естественного языка, а достижения компьютерной лингвистики получили широкое распространение в изучении языка, лингводидактики, лексикографии, перевода. В частности, наряду с тезаурусом, электронными словарями, автоматическим переводом, информационно-поисковыми системами, языковой корпус стал одним из основных инструментов в изучении языка. В результате этого сформировались обучающие корпуса, которые предоставляют лингвистическую информацию о языковых единицах удобным способом за считанные секунды с помощью специального поиска, содержат тысячи текстов на определенной части языка и предназначены для обучения языковым возможностям. В текущем процессе глобализации, переход мировой цивилизации в виртуальный мир, расширение использования одних языков угрожает выживанию других языков, поэтому эти корпуса важны как одна из возможностей для языкового выживания.

В мире лингвистики в XXI веке нарастает движение за научно-теоретическое изучение корпусных проблем. Целенаправленное исследование проводилось по социальной значимости языковых корпусов, принципам построения, разработке системы языковых тегов, автоматической разметке. Экономия времени, затрачиваемого на процесс анализа текста, возможность продемонстрировать особенности языковых единиц в структуре речи на тысячах примерах, признана основными эффективными возможностями корпуса. Поэтому в быстро развивающейся области компьютерной лингвистики выбор единиц обучающего корпуса, их лингвистическая, экстралингвистическая маркировка, разработка алгоритмов создания обучающего корпуса, создание лингвистической поддержки обучающего корпуса стало социальной необходимостью.

В результате сосредоточения внимания на научных исследованиях в области компьютерной лингвистики Узбекистана, по обработке узбекского языка с использованием современных информационных технологий первые выступления национального корпуса стали появляться как практические работы. Действительно, «создание национального корпуса узбекского языка в электронном виде, включающего в себя всю научную, теоретическую и практическую информацию об узбекском языке»<sup>1</sup> внесет большой вклад в повышение престижа нашего народа в общественной и международной жизни. Таким образом, для достижения широкого использования узбекского языка в системе информации и коммуникации, для углубленного изучения нашего национального и культурного наследия, для увеличения электронных ресурсов в обучении на родном языке, для достижения неограниченного использования этих ресурсов учащимися, для популяризации национального и духовного воспитания, сбор материалов родного языка в единую платформу

---

<sup>1</sup> Ўзбекистон Республикаси Президентининг 2020 йил 20 октябрдаги “Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида”ги ПФ-6084-сон Фармони.  
URL: <https://lex.uz/docs/5058351>

стал актуальной задачей. Также изучение таких вопросов, как формирование узбекского языкового образовательного корпуса, создание его лингвистического обеспечения на основе современных научных принципов, дальнейшее совершенствование системы обучения государственному языку, повышение его авторитета как научного языка и формирование технологических ресурсов для социального развития республики.

Данная диссертация в определенной степени служит выполнению задач, намеченных в Постановлениях Президента Республики Узбекистан № УП-4947 от 7 февраля 2017 года «О Стратегии действий по дальнейшему развитию Республики Узбекистан», № УП-5850 от 21 октября 2019 года «О мерах по кардинальному повышению престижа и статуса узбекского языка как государственного», № УП-6084 от 20 октября 2020 г. «О мерах по дальнейшему развитию узбекского языка и совершенствованию языковой политики в нашей стране», № ПП-4479 от 4 октября 2019 года «О широком праздновании тридцатой годовщины принятия Закона Республики Узбекистан «О государственном языке», обращение Президента Республики Узбекистан Олий Мажлису от 24 января 2020 года.

**Соответствие исследования приоритетным направлениям развития науки и технологий Республики Узбекистан.** Исследование проводилось в соответствии с приоритетом развития науки и технологии республики I.т «Социальное, правовое, экономическое, культурное, духовное и образовательное развитие информированного общества и демократического государства, развитие инновационной экономики».

**Степень изученности проблемы.** Хотя корпусная лингвистика – это новая область мировой лингвистики, в этой области проводится множество целевых исследований. Можно оценить взгляды основателей корпуса Брауна Н. Френсиса, Г. Кучера о принципах построения корпуса, теоретические и практические предложения русских лингвистов В.П.Захарова, А. Б. Кутузова, Е.В. Недошивина, В.В.Рикова, В.Плунгяна как предварительные исследования<sup>1</sup>. Если эти специалисты изучали корпус, его виды, особенности, социальную значимость корпуса, принципы построения корпуса<sup>2</sup>, обучающий корпус был изучен В.В. Мамонтовой и Д.В.Сичинава<sup>3</sup>.

---

<sup>1</sup> Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – М., 1967.; Синклер Дж. Предисловие к книге «Как использовать корпуса в преподавании иностранного языка» (Электронный ресурс). URL: [http://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4child/texts/sinclair\\_introduction.pdf](http://studiorum-ruscorpora.ru/wp-content/themes/bootstrap-basic4child/texts/sinclair_introduction.pdf)

<sup>2</sup> Кутузов А.Б. Корпусная лингвистика. Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс). URL: [lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf](http://lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf); Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург, 2006. 26 с.; Рыков В.В. Курс лекций по корпусной лингвистике. URL: <http://rykov-cl.narod.ru/c.html>; Плунгян В. Зачем мы делаем Национальный корпус русского языка? “Отечественные записки” 2005, №2. URL: [http://magazines.russ.ru/oz/2005/2/2005\\_2\\_20-pr.html](http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html)

<sup>3</sup> Мамонтова В.В. Корпус параллельных текстов и база данных для исследования переводческих соответствий: проблемы и процедуры формирования. – Альманах современной науки и образования Тамбов: Грамота, 2008. № 8 (15): в 2-хч. Ч.II. – С. 117-118. URL: [www.gramota.net/materials/1/2008/8-2/50.html](http://www.gramota.net/materials/1/2008/8-2/50.html); Сичинава Д.В. Европейский перфект сквозь призму параллельного корпуса. – Acta Linguistica Petropolitana. Труды института лингвистических исследований. – 2016, № 2. – Т.12. – С. 85-114.

Хотя принципы формирования языкового корпуса были подробно изучены в лингвистике турецкого корпуса<sup>1</sup>, формирование корпуса не было источником специальных исследований.

В качестве исследований в области компьютерной лингвистики в узбекской лингвистике можно назвать работы А.Пулатова, М.Айимбетова, Н.Абдурахмановой, М.Абджаловой.<sup>2</sup> Эти исследования знаменательны тем, что в нем рекомендованы современные методы исследования текстов с использованием инновационного подхода - достижений компьютерной лингвистики.

Проблемы создания узбекского языкового обучающего корпуса не были предметом специальных исследований, но некоторые исследования позволили предположить определенные аспекты проблемы. В частности, создание лингвистического программного обеспечения для компьютерных программ на основе глаголов<sup>3</sup>, лингвистическое обеспечение узбекско-английского машинного перевода<sup>4</sup>, принципы создания узбекского авторского корпуса<sup>5</sup>, вопрос создания модулей для редактирования и анализа узбекских текстов<sup>6</sup>, принципы создания лингвистической базы языкового корпуса<sup>7</sup>, лингвистические основы семантического приравнивания существительных узбекского языка<sup>8</sup>, проблемы лингвистического обеспечения морфологического анализатора узбекского языка<sup>9</sup> изучаются в морфологическом плане. Также были опубликованы ряд статей по проблемам

---

<sup>1</sup> Сулейманов Д.Ш., Гильмуллин Р.А., Гатауллин Р.Р. Морфологический анализатор татарского языка на основе двухуровневой модели морфологии / Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 2. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 327 с.; Сулейманов Д.Ш., Гатаиатуллин А.Р. Модель татарской аффиксальной морфемы и ее реализация // Серия: Интеллект. Язык. Компьютер. – Вып.4. – Казань: Фэн. – 1996. – С. 113-127.

<sup>2</sup> Каримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Алфавитли луғат. Частотали луғат. Терс луғат. – Тошкент, 2007.; Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Филол. фан. док. дисс. ...автореф. – Тошкент, 2008.; Ўринбоева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010.; Жуманазарова Г.У. Фозил Йўлдош ўғли дostonлари тилининг лингвопоэтикаси: Филол. фан. док. дисс. ...автореф. – Тошкент, 2017.; Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида): Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. афтореф. – Тошкент, 2018.

<sup>3</sup> Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. – Тошкент, 2006.

<sup>4</sup> Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида): Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. афтореф. – Тошкент, 2018. – 52 б.

<sup>5</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Қарши, 2018. – 168 б.

<sup>6</sup> Абжалова М.А. Ўзбек тилидаги матнларни таҳрир ва таҳлил қилувчи дастурнинг лингвистик модуллари (Расмий ва илмий услубдаги матнлар таҳрири дастури учун): Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Фарғона, 2019. – 164 б.; Абжалова М. Матнларга автолингвистик ишлов бериш тизимлари // Шестая Международная конференция по компьютерной обработке тюркских языков “TurkLang-2018”. (Труды конференции) – Ташкент, 2018. – 320 с.

<sup>7</sup> Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол. фан. бўйича фалсафа доктори (PhD) ...дисс. – Қарши, 2019. – 140 б.

<sup>8</sup> Akhmedova D.B., Mengliev B.R. Semantic Tag Categories in Corpus Linguistics: Experience and Examination International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-3S, October 2019. – P. 208-212.

<sup>9</sup> Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Монография. (Электрон китоб) – GloeEdit, 2020. – 244 б. – 78 б.

создания узбекско-английского параллельного корпуса<sup>1</sup>, автоматического анализа существительных в узбекском языке<sup>2</sup>, формальных моделей словообразования<sup>3</sup>. Эти исследования актуальны тем, что исследуют проблемы компьютерной лингвистики, но обучающего корпуса узбекского языка – не особо стоит в повестке дня. В процессе подготовки диссертации были детально изучены вышеперечисленные научные исследования, высказана позиция и использована в исследовании.

**Связь темы диссертационного исследования с планами научно-исследовательских работ высшего образовательного учреждения, в котором была завершена диссертация.** Исследование было в соответствии с планом исследований Термезского государственного университета по теме «Узбекское языкознание: анализ художественной речи, антропоцентрическая теория, этнолингвистика, корпусная лингвистика» и в рамках практического проекта AM-FZ-201908172 Ташкентского государственного университета узбекского языка и литературы им. Алишера Навои.

**Цель исследования.** Целью исследования является выявление теоретических вопросов формирования обучающего корпуса узбекского языка, а также научное освещение основ создания лингвистического обеспечения обучающего корпуса узбекского языка.

**Задачи исследования:**

дать общую характеристику обучающего корпуса, выделить его важность как образовательного ресурса;

выявление преимуществ обучающего корпуса в преподавании родного языка, обучении узбекскому языку как иностранному (лингводидактика);

разработка принципов отбора материала для узбекского языкового обучающего корпуса, сбор материала для фрагмента обучающего корпуса;

определение процесса сбора учебников в обучающем корпусе и разработать их поисковую систему;

формирование морфолексикона обучающего корпуса для теоретических выводов;

создание лингвистических основ морфологического и словообразовательного словаря узбекского языка;

описание строение и структуры базы данных по интерпретации значений слов, межсловных формальных отношений.

**Объектом исследования** является обучающий корпус, предназначенный для обучения языковым умениям, имеющий лингводидактический характер.

---

<sup>1</sup> Karimov R., Mengliev B. Theoretical fundamentals of uzbek-english parallel corpus / Journal of critical reviews. ISSN- 2394-5125. – VOL 7, ISSUE 17, 2020. – P. 73-76.; Karimov R.A., Mengliev B.R. The Role of the Parallel Corpus in Linguistics, the Importance and the Possibilities of Interpretation International Journal of engineering and Advanced Technology (IJEAT). ISSN: 2249 – 8958, Volume-8, Issue-5S3 July 2019. – P. 388-391.

<sup>2</sup> Орхун М. Computational analysis of Uzbek nouns / Шестая Международная конференция по компьютерной обработке тюркских языков “TurkLang-2018”. (Труды конференции) – Ташкент, 2018. – 320 с.

<sup>3</sup> Турсунов А. Вопросы словообразования в формальных моделях тюркских языков (на примере узбекского языка) // Шестая Международная конференция по компьютерной обработке тюркских языков “TurkLang-2018”. Труды конференции) – Ташкент, 2018. – 320 с.

**Предметом исследования** является вопрос лингвистического обеспечения узбекского языкового образовательного корпуса.

**Методы исследования.** Для освещения темы исследования использовались методы классификации, описания, сравнения, статистики, моделирования, корпусного анализа.

**Научная новизна исследования:**

лингвистические, дидактические, информационные особенности образовательного корпуса основаны на том, что функциональные возможности мирового образовательного корпуса различаются по лингвистическому обеспечению, доказано сходство по методам автоматической обработки языка;

определены оптимальные математические и лингвистические методы формирования морфолексикона узбекского языка и морфологической база узбекского языка для идентификации системы автоматической обработки учебно-строительного материала и были сформированы когнитивные, лингвистико-статистические, научно-теоретические основы формирования морфолексикона и морфологической базы;

отмечены порядок составления словаря глагольных форм, лингвистические, математические, логические, программные принципы систематизации основных и грамматических форм в корпусе, также открыты возможности поискового образовательного корпуса узбекского языка, строение интерфейса, конкорданс в качестве электронной библиотеки;

сформированы порядок размещения в базе учебного корпуса словарей, словообразования, значений слов, межсловной семантики, формальных отношений в виде библиотеки, конкорданса и поисковой системы SQL.

**Практические результаты исследования:**

определены принципы отбора текстов морально-воспитательного содержания для целей обучения узбекскому языку для обучающего корпуса, собран материал для обучающего корпуса по узбекскому языку;

существующие учебные словари на узбекском языке сгруппированы на одной платформе и имеют удобную систему поиска;

лингвистическое обеспечение обучающего корпуса узбекского языка – сформированный морфолексикон, состоящий из 37200 лексем;

разработана система предложений по программному обеспечению;

создан словарь на основе порядка составления морфологического словаря узбекского языка.

**Достоверность результатов исследования** объясняется тем, что изученный материал помогает сделать выводы, исходя из характера узбекского языка, его природе, теоретической совместимости с методами обучения родному языку, методического мастерства, опоры на проверенные источники в разработке принципов обучения узбекскому языку.

**Научная и практическая значимость результатов исследования.** Исследование служит научно-теоретическим источником при разработке основ корпуса узбекского языка, при выполнении исследований в области компьютерной лингвистики.

Практическая значимость исследования объясняется тем, что одна из отраслей прикладной филологии – корпусная лингвистика как наука, в процессе преподавания как программа, план, источник в описании служит инструментом формирования образовательных ресурсов в языковом образовании.

#### **Внедрение результатов исследования:**

На основе научных результатов, полученных на лингвистической основе формирования обучающего корпуса узбекского языка:

теоретическая информация о принципах отбора текстов этического и образовательного содержания для обучающего корпуса, учебников, словарей, формирование базы данных межсловных духовных, формальных связей, создание соответствующей поисковой системы, размещение в базе данных корпуса использовано в фундаментальном научном проекте FA-A1-G007 «Каракалпакские пословицы как объект лингвистического исследования» (Каракалпакский филиал Академии наук Республики Узбекистан в 2021 г. № 17.01 / 108 от 22 апреля) Каракалпакским филиалом Академии наук Республики Узбекистан Каракалпакским научно-исследовательским гуманитарным институтом. В результате удалось определить морфологические формы и положение языковых единиц в каракалпакских пословицах, выявить их значение как лингвистической основы;

эффективные математические и лингвистические методы создания морфолексикона образовательного корпуса через формирование морфологической базы узбекского языка, определение статистики принадлежности к категории узбекских лексем, расширение сферы лингвистических исследований словообразования; словообразование, выводы по содержанию и структуре базы данных по толкованию значений слов использовано в научно-практическом проекте FA-F1-G003 «Функциональное словообразование в современном каракалпакском языке», реализуемом в Каракалпакском научно-исследовательском гуманитарном институте Каракалпакского филиала Академии наук Республики Узбекистан (Справка № 17.01 / 108 от 22 апреля 2021 г). В результате проектной работы созданы модели словообразования в современном каракалпакском языке, усовершенствованы теоретические обобщения при анализе основ словообразования.

структура и состав базы данных по учебным словарям, словообразованию, толкованию значений слов, межсловным духовным, формальным отношениям; выводы о научно-теоретических задач при их размещении в обучающем корпусе использованы в пунктах 17-18 учебника «Современный узбекский язык» «Лексико-семантические отношения», «Классификация узбекской лексики», рекомендованных Министерством высшего и средне-специального образования. (Справка Министерства высшего и среднего образования Республики Узбекистан № 89-03-2167 от 24 мая 2021 года). В результате достигнута трактовка значений слов, улучшение содержания тем, относящихся к межсловным духовным, формальным отношениям;

определение морфологических форм и позиций морфологического словаря узбекского языка, порядок составления словаря глагольных форм, принципы и возможности систематизации основных и грамматических форм корпуса, рекомендации по образцу грамматических морфем, использованных при создании узбекского морфологического словаря (O‘zbek tilining fe’l shakllari lug‘ati (“kiymoq” fe’li misolida) [Matn]: lug‘at. – Termiz: Termiz davlat universiteti nashr-matbaa, 2021. – 492 b. ISBN 978-9943-7243-3-4). В результате была создана база данных морфологических форм и сформировано базовое лингвистическое обеспечение разработки программы морфологического приравнивания и морфологического анализатора Узбекского национального национального корпуса;

выводы о необходимости создания обучающего корпуса узбекского языка, его специфические информационные особенности, лингвистическое обеспечение, его отличие и сходства с мировым образовательным корпусом, и выводы об их лингводидактической значимости использовались при составлении сценариев литературных, художественных, духовных и образовательных программ Сурхандарьинского областного телерадиокомпании (Справка № 01-01/54 Сурхандарьинского областного телерадиокомпании от 29 апреля 2021 года). В результате было улучшено содержание пропагандистской работы по воспитанию высокого гармонично развитого поколения, формированию языковой культуры у молодежи и использованию узбекского языка в современной информационной системе.

**Апробация результатов исследования.** Результаты исследования были представлены на 5 международных и 3 национальных научных конференциях.

**Публикация результатов исследования.** Основное содержание диссертации отражено в 15 научных статьях (5 из них - научные журналы ВАК Республики Узбекистан и 2 международных журнала) и тезисах.

**Структура и объем диссертации.** Диссертация состоит из введения, трех глав, общего заключения, списка использованной литературы и приложения. Общий объем диссертации 147 страниц.

## **ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ**

Во введении диссертации даны сведения об актуальности и востребованности темы исследования, релевантности исследования приоритетам науки и техники, степени изученности проблемы, актуальности диссертации для научно-исследовательской работы вуза, научной новизне, теоретической и практической значимости, объекте и информации по теме, достоверности, реализации, апробации результатов исследования, публикация результатов исследования, структуре и объеме диссертации.

Первая глава диссертации озаглавлена «**Учебный корпус как лингводидактическая база**», в которой содержится общее описание узбекского обучающего корпуса, термины, место в классификации корпуса, лингводидактика (в языковом обучении), интерфейс, требования к содержанию, поисковый менеджер, система лингвистических тегов, анализируются состав и структура, выбранные тексты, описание вариантов.

Обучающий корпус – это языковой корпус с лингводидактическими свойствами, материалы которого ориентированы на обучение языку. Обучающий корпус узбекского языка представляет собой корпус, предназначенный для обучения возможностям узбекского языка, имеет лингводидактический характер, содержит электронные тексты, действует как специальный сайт. Желательно использовать термин «образовательный корпус» (на английском и русском: Learning Corpus, Educational Corpus, *обучающий корпус, учебный корпус*), который создан для удовлетворения потребностей изучающих узбекский или узбекский как иностранный язык. Созданы различные образцы учебного корпуса на английском и русском языках. Например, Корпус разговорного английского языка Ланкастер (Lancaster / IBM, Spoken English Corpus<sup>1</sup>, Corpus of Spoken American English<sup>2</sup>), из их числа международный корпус<sup>3</sup> изучающий английский язык для лингводидактических исследований. В настоящее время существует лингвистический корпус, охватывающий миллионы слов: Британский национальный корпус (The British National Corpus, BNC 100 миллионов слов), COBUILD Bank of English (200 миллионов слов) являются одними из самых удобных способов для изучения языка. Существуют также специальные корпорации, представляющие устный и письменный литературный язык (например, CANCODE (Cambridge and Nottingham Corpus of Discourse in English), MICCASE (Michigan Corpus of Academic Spoken English LIBELCASE) (Академический корпус разговорного английского языка LIBEL). Устный академический корпус Лимерика-Белфаста стал бесценным инструментом лингводидактики<sup>4</sup>.

В русском национальном корпусе есть образовательный корпус (*обучающий корпус*)<sup>5</sup>: он задуман как внутренний корпус национального корпуса и является составной частью национального корпуса. Его материал отличается от текстов в Национальном корпусе тем, что соответствует возрасту. Также в Иркутском государственном лингвистическом университете продолжается создание Образовательного мультимодального корпуса (УМКО). Этот корпусный материал представляет собой видео неподготовленных образовательных диалогов на определенную тему между людьми, которые являются носителями русского, китайского языков как второго языка.

Языковые корпуса разных языков делятся учеными на несколько различных типов в соответствии с их структурой, назначением,

---

<sup>1</sup><https://www.routledge.com/A-Corpus-of-Formal-British-English-Speech-The-LancasterIBM-Spoken-English/Knowles-Assistant-Taylor-Research-Williams/p/book/9780582056398>

<sup>2</sup> <https://www.linguistics.ucsb.edu/research/santa-barbara-corpus>

<sup>3</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – 168 б.

<sup>4</sup>Ахметова К.Ю. Корпусный подход в обучении иностранному языку // <https://scipress.ru/philology/articles/korpusnyj-podkhod-v-obuchenii-inostrannomu-yazyku.html>

<sup>5</sup> <https://ruscorpora.ru/new/search-school.html>



стабильностью, изменчивостью и другими аспектами<sup>1</sup>. Ш.Хамроева и А.Эшмуминов, проводившие специальное исследование на узбекском языке, также процитировали мнение ученых по этому поводу, сделав акцент на классификации видов корпуса<sup>2</sup>. Исследователь Ш.Хамроева дает подробную информацию о языковых корпусах и классифицирует их с 13 различных точек зрения, включая образовательный и национальный корпус.

Согласно критериям классификации, каждая различающаяся единица, показываемая отдельно, должна иметь противоречащую единицу в рамках классификации. В этом отношении корпус делится на академический корпус и обучающий корпус в зависимости от их направленности. Формируемый в настоящее время национальный корпус мировых языков можно рассматривать как академический корпус.

Обучающие корпуса по направленности и по типу речи делятся на: письменный, устный, смешанный; по языку текста: русский, английский, узбекский и др.; по параллелизму: одноязычный, двуязычный, многоязычный; по специализации текста: литературный, диалектный, разговорный, терминологический, смешанный; по жанрам: литературный, фольклорный, драматический, публицистический; по способу доступа: бесплатный, платный, закрытый; по назначению: исследовательский, описательный; по толкованию: с дополнительным комментарием, без комментария; по изменчивости: переменная, постоянная; по характеру описания: морфологическое, синтаксическое, семантическое и др.

Образовательный корпус включает в себя несколько возможностей электронных обучающих ресурсов, которые в этом отношении дают им преимущество, совершенство.

Общие, дидактические, методические, психолого-педагогические, технико-технологические, эстетические, эргономические и санитарно-гигиенические требования устанавливаются при создании обучающего корпуса, а также при создании электронных учебных ресурсов.

Адаптация воспитательного корпуса из национального корпуса к образовательному процессу; ограниченный размер; ясность, простота, понятность различных введенных текстов; стандартизованный; внедрение учебных словарей; отличается возможностью обогащения различными дополнительными материалами (картинками, аудио, мультимедиа).

Специально исследовалось значение обучающего корпуса в лингводидактике<sup>3</sup>. Область эффективного использования корпуса –

---

<sup>1</sup> Корпусная лингвистика. Статья в интернет-энциклопедии Википедия. <http://wikipedia.org/wiki/>; Захаров В.П., Богданова С.Ю. Корпусная лингвистика: Учебник для студентов направления «Лингвистика». 2-е изд., перераб. и дополн., – СПб.: СПбГУ. РИО. Филологический факультет. – Санкт-Петербург, 2013. – С. 16.

<sup>2</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – Б. 46-50; Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2019. – Б. 52-53.

<sup>3</sup> Синклер Д. Предисловие к книге “Как использовать корпуса в преподавании иностранного языка”/ Д.Синклер [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/corpora-infro.html>, свободный.; <https://kpfu.ru/.../Nacioanlnyj.korpus.v.praktike.lingvisticheskikh.issledovaniy.i.prepod.> // Марданшина Р.М. Национальный корпус в практике лингвистических исследований и преподавании языка.; Добрушина Н.Р., Левинзон А.И. Национальный корпус русского языка в образовании// Cuadernos de Rusística Española, 2006.

лингводидактика; это очень важно в изучении родного языка, иностранного языка. В обучении языку корпус очень полезен, так как показывает множество примеров, чтобы полностью продемонстрировать богатство словарного запаса, объяснить возможность использования слова с помощью той или иной грамматической конструкции. Здесь учитель может найти новые убедительные, бесконечные и разнообразные примеры, нетрудно сформулировать задачу, упражнения, за несколько минут он может подготовить задание, состоящее из новых примеров по теме.

Методика обучения иностранному языку (в том числе узбекскому как иностранному) сложна, мы стремимся найти простые, понятные способы подачи материала на разных языках, постоянно ищем наиболее эффективные способы передачи опыта. Самым удобным аспектом этого метода является корпус, который представляет собой средство управления большим объемом данных (русский, *инструментарий*), которые по-разному регулируют хаотическое разнообразие языков; образует упорядоченный набор слов<sup>1</sup>.

Методы образовательного корпуса являются очень эффективным инновационным дополнением не только в лингвистических исследованиях, но и в обучении родному языку, иностранному языку, теории и практике перевода. Этот метод эмпирической адекватности, валидности, гибкости; конкретная задача сочетает в себе такие аспекты, как адаптация к целевой группе. Корпусный метод позволяет студенту работать самостоятельно, использовать метод «открытия» в обучении, что побуждает студента/ученика больше узнавать о безграничных возможностях языка.

Обучающий корпус можно интегрировать с различными платформами, чтобы сделать его более удобным для пользователя. Например, обучение с подкреплением, обучающие программы, программы моделирования для помощи в исследованиях, образовательные игры и т. д.

Обучающий корпус демонстрирует методику дистанционного обучения через Интернет, в которую входят студенты: школьники; студенты системы профессионального образования; учащиеся академических лицеев; абитуриенты; студенты; желающие получить второе высшее или дополнительное образование, обучение и переподготовку; главы региональных властей и администраций; молодые люди, которые не могут получить доступ к образованию из-за ограниченных возможностей традиционной системы образования; сотрудники фирм и предприятий, желающие поднять свой информационный статус до уровня современных требований; слушатели, желающие получить вторую параллельную информацию; жители отдаленных, слаборазвитых территорий; лица с ограниченным свободным движением; лица с ограниченными физическими

---

– № 2; <https://cyberleninka.ru/article/n/lc-tehnologii-v-issledovanii-osvoeniya-yazyka>. О. Н. Камшилова . LC-технологии в исследовании освоения языка.; Хамроева Ш. Таълимда корпусдан фойдаланиш // “Тил ва адабиёт таълими” журналы. 2017 йил сентябрь, № 9. – Б. 49-50.

<sup>1</sup> Krieger Daniel. Corpus Linguistics: What It Is and How It Can Be Applied to Teaching Siebold University of Nagasaki [Text] / Daniel Krieger (Nagasaki, Japan), 2003. // <http://iteslj.org/Articles/Krieger-Corpus.html>

возможностями; лица, принадлежащие к социальным группам, состоящим из лиц, проходящих военную службу, и других лиц:

Основным компонентом, который знакомит пользователя с корпусом, является интерфейс, и он должен охватывать особенности совершенного, понятного узбекского языка, соответствующего возрасту читателя.

В качестве материала обучающего корпуса узбекского языка пропорционально отобраны тексты популярного, нравственного, образовательного, официального, научно-популярного, художественного и публицистического стиля современного узбекского литературного языка. До сегодняшнего дня достигнуты следующие практические результаты в развитии узбекско-языкового обучающего корпуса: отобраны, помещены в корпус тексты нравственно-воспитательного содержания для узбекского обучающего корпуса; существующие учебные словари на узбекском языке собраны на одной платформе; при условии удобной поисковой системы; сформировано лингвистическое обеспечение обучающего корпуса узбекского языка - морфолексикон, состоящий из 37 200 лексем. Перечисленные выше ресурсы включены в качестве исходных материалов для образовательного корпуса.

Поиск лексических единиц в обучающем корпусе осуществляется по следующим показателям: 1) письмо; 2) разделение на части; 3) на основании какой орфографической нормы; 4) вариант написания; 5) произношение и ударение; 6) изменение звука; 7) значения (объяснения), способы смысловой передачи неоднозначности; 8) синоним(ы); 9) антонимы; 10) токсонимическое гнездо, гнездо(а); 11) меронимное гнездо, гнездо(а); 12) рейтинговая серия; 13) омонимы; 14) произношение (паронимы); 15) этимология; 16) принадлежность к слою самоассимиляции; 17) по применению: неограниченное / ограниченное: диалектный / термин / сленг и жаргон; 18) по периодам: современная / устаревшая / новая лексическая единица; 19) диалектные варианты; 20) исходность /производность; 21) содержание (морфем); 22) тип структуры: простая / сложная: суставная, двойная, повторяющаяся; 23) по видам типовых признаков; 24) передние / задние соединения; 25) методы; 26) частота; 27) примеры.

Вторая глава диссертации называется «**Морфологическая база данных для образовательного корпуса**». В первом разделе главы, озаглавленном «*Формирование морфолексикона для обучающего корпуса*», рассматриваются грамматический словарь естественного языка, морфолексикон, его особенности, возможности обработки естественного языка и работа, проделанная над морфолексиконом узбекского языка. язык.

Грамматический словарь – это группа символов слов на определенном языке, словарь, охватывающий все грамматические формы. Морфолексикон – это база данных, содержащая лексику определенного языка, к которой прикрепляется информация о лексеме, ее принадлежности к категории, ее грамматических особенностях, адаптированная к процессу обработки естественного языка. Такая база данных формируется в электронном виде и выступает в качестве лингвистической поддержки различных средств

обработки естественного языка (автоматический перевод, лингвистический корпус, морфологический, семантический, графический анализатор).

В этом разделе описывается теория формирования морфолексикона, практика составления морфолексикона узбекского языка для национального корпуса на основе изучения опыта.

Морфологическую лексику узбекского языка целесообразно развивать в 4 этапа: 1) ввести в базу данных 32500 лексем из толкового словаря узбекского языка; 2) , введение словосочетаний на основе орфографического словаря и словаря составных слов; 3) включение в словарный запас словосочетаний, пословиц, приведенных в толковом фразеологическом словаре узбекского языка; а также прикрепление информации о принадлежности к категории; 4) сортировка и добавление в базу данных имен, приведенных в толковом словаре узбекского языка, ввод терминов, относящихся к разным сферам, известных имен. К каждой лексической единице в приведенных выше словарях прилагается информация о принадлежности к категории и типах этой категории. В результате получился более совершенный узбекский морфолексикон.

В процессе формирования исходной базы морфологической лексики узбекского языка было выполнено 2 вышеуказанных этапа и выявлено следующее: словесные группы в узбекском языке имеют следующее соотношение: 22 599 существительных, 7 172 прилагательных, 3949 глаголов, 1856 обстоятельств, 800 имитационных слов, 244 побудительных слов, 182 наречий, 132 модальных слов, 108 полифункциональных слов, 106 вспомогательных слов, 52 союзов, 45 предлогов, 30 числовых лексем; общее количество лексем составляет 37 282.

Во втором разделе главы «*О формировании морфологического словаря узбекского языка*» разъясняется значение словаря морфологических форм как лингвистической базы для обучающего корпуса, особенности морфологической системы узбекского языка, практика составления словаря морфологических форм. Создано множество грамматических словарей для каждого из языков развитых стран (например, английского и русского). Однако никакого грамматического словаря узбекского языка до сих пор не создано. Вероятно, это связано со сложностью морфологической системы узбекского языка и большим количеством морфологических форм. Поэтому в нашем языке возникают трудности с созданием программ перевода, проверки орфографии, морфологического автоматического анализатора и других компьютерных программ. В традиционном карточном методе сбора и сортировки таких доказательств не хватит и жизни человека.

При составлении морфологического словаря глагольных форм узбекского языка ставится задача определить грамматические формы, существующие в узбекском языке, и все их положения, показать грамматическое строение каждой формы, образец грамматических морфем. Мы составили морфологический словарь глагольных форм для практических исследований в реализации цели. Из-за большого количества узбекских глагольных форм был выбран только один глагол.

Идеи и практические разработки относительно количества глагольных форм в узбекском языке наблюдаются только в одном источнике, и, согласно ему, в узбекском языке существует около 100 тысяч грамматических форм одного глагола. Практическая работа доказала, что способ получения этого количества не настолько самооправдан, что количество неверно, что было бы ошибкой устанавливать количество для всех глаголов.

Словарь глагольных форм узбекского языка на примере глагола (*киймок*) образовал около 30 000 слов из расположения форм пропорции, времени, наклонения, лица-числа, неделимости, функции. Когда к вышеперечисленным формам были присоединены вспомогательные глаголы, образовалась единая глагольная форма, которая вместе приняла более 300 000 форм. Практические разработки показали, что грамматические формы различных глаголов в узбекском языке увеличиваются с 4-5 тысяч до 300 тысяч (вместе с нагрузками они снова умножаются). Этот грамматический словарь будет служить не только лингвистической базой данных обучающего корпуса, но и ключевым лингвистическим подспорьем в будущем при разработке программы морфологического тегирования, морфологического анализатора национального корпуса узбекского языка.

В третьем разделе главы, озаглавленном «*Основы словаря словообразования для обучающего корпуса*», обсуждается опыт создания учебной лексики и словарей словообразования, лингвистическое обеспечение, структура и состав словаря словообразования для образовательного корпуса узбекского языка.

Хотя область образовательной лексикографии в Узбекистане не считается образцовой, изучение таких вопросов в монографическом плане<sup>1</sup>, как лингвистические основы, история и перспективы образовательной лексикографии, издание нескольких учебников показывает, что узбекская образовательная лексикография не прекратилась. Эти словари служат одним из основных лингвистических ресурсов при формировании обучающего корпуса узбекского языка.

Учебник по узбекскому словообразованию, созданный коллективом авторов в 2008 году<sup>2</sup>, служит лингвистико-лексикографическим источником словообразовательной базы образовательного корпуса. При формировании словообразовательного словаря для образовательного корпуса необходимо учитывать следующие факторы: во-первых, определить лексику словообразовательного словаря и словообразовательного гнезда; во-вторых, следовать методике составления производного словаря; в-третьих, не устанавливать ограничения по размеру, в отличие от традиционного словаря, учитывая, что это онлайн-продукт; в-четвертых, установка того же параметра для словарной статьи; в-пятых, указать предисловие, основу конструкции и средства строительства; в-шестых, чтобы поместить индексы перекрестных

---

<sup>1</sup> Бахриддинова Б. Ўзбекистонда ўқув луғатчилиги: лингвистик асослари, тарихи ва истикболлари: филология фанлари доктори диссертацияси. – Самарқанд, 2020. – 256 б.

<sup>2</sup> Mengliyev B., Bahriddinova B., Xoliyurov O'. va boshq. O'zbek tilining so'z yasalishi o'quv lug'ati. – Toshkent: Yangi asr avlodi, 2007. – 176 b.

ссылок в строительное гнездо, чтобы показать строительную парадигму; в-седьмых, ввести информацию (теги), информирующую о создании другой категории из одной категории.

С учетом структуры словообразовательного словаря статьи и возможностей электронного словаря были разработаны следующие рекомендации по формированию корпуса словообразовательного словаря: 1) словообразовательный словарь, как основа корпуса узбекского образовательного языка, издан в 2009 году; 2) структура словообразования: выделены основы словообразования, средства словообразования; 3) информация о типе конструкции дана в каждом словообразовательном поле; 4) разъясняются случаи нетипичного словообразования; такая особенность выделяется; 5) состояние конструкции фразы: указаны основание конструкции, средства строительства и конструкции; 6) в книжном словаре, в основном сделанном методом аффиксации. В зависимости от возможностей корпуса даются синтаксически сформированное слово из определенного ядра из основания корпуса и ссылка на его контекст.

Третья глава диссертации называется **«Подготовка базы данных для анализа смыслового значения образовательного корпуса»**. Первый раздел озаглавлен *«Вопросы формирования базы данных по толкованию значений слов»*.

В узбекской корпусной лингвистике было проведено несколько исследований по семантическому уравниванию единиц для национального корпуса, прикреплению тегов (комментариев) о значениях и формированию семантической базы данных. В частности, Ш. Хамроева изучала общие принципы развития семантических тегов, А.Эшмуминов изучал формирование синонимичной словарной базы национального корпуса, Д. Ахмедова изучала выражение лексико-семантических отношений между словами в национальном корпусе узбекского языка<sup>1</sup>.

При введении любого слова в поле поиска узбекского учебного корпуса, появится список контекстов, в которых это слово было использовано. Слово, которое искали в нем, будет выделено отдельным цветом. Пользователь сможет выборочно анализировать контекст. Щелчок правой кнопкой мыши по определенному слову в контексте открывает отдельное окно с гиперссылкой, в которой отображаются значения этого слова. Гиперссылка делает значение слова из словаря доступным в лингвистической базе обучающего корпуса, находит его и помещает в окно результатов.

Использование узбекского толкового словаря<sup>2</sup> – лучший способ толкования слова, но, в отличие от толкового словаря, особое внимание уделяется упрощению корпуса. Отсутствие толкового словаря в узбекском языке затрудняет толкование слов в обучающем корпусе.

---

<sup>1</sup> Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD)...дис. – Бухоро, 2018. – 165 б.; Эшмўминов А.А. Ўзбек тили миллий корпуси синоним сўзлар базасини яратиш асослари: филология фанлари бўйича фалсафа доктори (PhD) диссертацияси. – Термиз, 2019. – 146 б.; Ахмедова Д.Б. Атов бирликларини ўзбек тили корпуслари учун лексик-семантик теглашнинг лингвистик асос ва моделлари: филология фанлари бўйича фалсафа доктори (PhD) диссертацияси. – Бухоро, 2020. – 138 б.

<sup>2</sup> Ўзбек тилининг изоҳли луғати. 5 томлик. – Тошкент: Ўзбекистон Миллий энциклопедияси, 2000-2006.

Существующий словарь может быть использован при формировании базы данных учебного корпуса следующим образом, например,

**БАҚАР**<sup>1</sup> [а. بقر – бык] **точ.** год коровы, корова. Когда его спрашивают, сколько ему лет, он отвечает: *муш, бақар, паланг, харғуш, наҳанг, эскича йил ағдариб..... Г.Расул, Адолат. Афанди, бузоқни арабча нима дейди? –Сигир бўлгунча сабр қилади, ундан кейин бақар дейди.* «Анекдоты».

Из этого объяснения, наряду со значением слова, становится ясным принадлежность и этимологическое значение слоя ассимиляции. Судя по всему, слово *бакар* было заимствовано из арабского языка; по-арабски означает *бык, буйвол*. Исходя из этого, в одном столбце базы данных образовательного корпуса следует записать, от какого языка произошло слово, в одном столбце - что оно означает на языке оригинала, в одном столбце - что именно оно означает на узбекском языке. Например:

Слово	Ассимиляция	Значение в языке оригинала (этимология)	Значение в узбекском языке
БАҚАР	по-арабски	бык, буйвол	год коровы

Из комментария видно, что со знаком «айн» к этому слову мы можем ссылаться на другое слово: **айн.** год коровы **қ.** корова. В связи с этим, мы добавляем еще один столбец в базу данных. В этом столбце, если таковой имеется, мы вводим единицу, к которой относится слово, и связываем эти единицы гиперссылкой:

Слово	Ассимиляция	Значение в языке оригинала (этимология)	Значение в узбекском языке	Ссылка
БАҚАР	по-арабски	бык, буйвол	год коровы	корова

Этой функцией корпус отличается от словаря: слово не ищется постранично, одним нажатием кнопки (гиперссылки) можно найти объяснение любого слова.

Нет необходимости включать иллюстративные примеры в глоссарий в базу данных корпуса. Во-первых, примеры в аннотированном словаре устареют, а во-вторых, появится возможность выбрать иллюстративный пример из тысяч текстов в базе данных корпуса.

Словарь различает значения многозначных слов. Это означает, что эта функция должна быть указана в базе данных. Следовательно, столбец, в котором интерпретируется слово, не будет одним: комментарии вводятся в виде комментарий<sup>1</sup>, комментарий<sup>2</sup>, комментарий<sup>3</sup>. С помощью этих данных, а также единого алгоритма можно различать многозначные и однозначные слова в базе данных.

В главе «Создание базы данных межсловесных духовных отношений» описаны типы межсловесных духовных отношений, такие как полисемия, синонимия, антонимия, гипонимия, градуонимия, партонимия, иерархия, а

<sup>1</sup> Ўзбек тилининг изоҳли луғати. 1-том. – Тошкент: Ўзбекистон Миллий энциклопедияси, 2000. – Б. 176.

также описывается создание лингвистической базы данных в образовательном корпусе.

Программисту несложно подключиться к базе словарей антонимов и синонимов, созданных на узбекском языке, найти эти словари по гиперссылкам. Самое главное, сырье есть, но необходимо решить вопрос о его размещении в базе данных с учетом поиска. Для этого тексты этих словарей должны пройти некоторую техническую обработку.

В узбекском языке существует смысловая, логико-семантическая связь между гиперонимом и гипонимом. Такая языковая база может быть сформирована следующим образом:

получается общая лексическая база узбекского языка; как только общая база помещается в один столбец, рядом с ним добавляется другой столбец; общая лексическая база определяется как гипоним, и гипероним вводится в следующем столбце;

программист соединяет блоки в соседних столбцах друг с другом с помощью гиперссылки. В результате все слова связаны в отношениях гипонима-гиперонима.

Гипоним	гипероним
Цыплёнок	курица
Бревно	дерево
Петух	курица
Ива	дерево
Вишня	дерево

Иногда одна лексема может быть связана с более чем одним гиперонимом в широком диапазоне значений. Например, такие лексеммы, как *яблоко, вишня, персик, абрикос*, сочетаются с гиперонимом *дерева*, а также с гиперонимом *фрукта*. В таких случаях вы можете добавить дополнительные столбцы рядом с столбцами и назвать их гиперонимом<sup>1</sup>, гиперонимом<sup>2</sup>.

Лингвистической поддержки недостаточно для формирования базы данных узбекского языкового корпуса, так как такой словарь в узбекском языке не создан. База данных сформирована с использованием результатов монографического исследования<sup>1</sup>, материалов учебно-степенного словаря.

Для формирования базы данных на узбекском языке список слов, иерархия – единицы добавляются по возрастанию / убыванию значения. Результатом поиска являются не столбцы в базе данных, а информация о строке: чем больше слов в базе данных, тем больше строк формируется. В поле поиска, какое бы слово ни было введено из базы данных, оно будет отображаться в строке ранжирования.

Чтобы сформировать базу данных, отражающую отношение целостности в узбекском языке, необходимо разделить базу данных слов, принадлежащих к категории существительных, и включить информацию о том, что это слово обозначает целое / часть. Не отделяя основание, достаточно

<sup>1</sup> Орифжонова Ш. Ўзбек тилида луғавий градуонимия: филол. фан. номз. дисс. автореф. – Тошкент, 1996. – 21 б.



добавить один столбик к самому морфолексикону. Но этого метода недостаточно для обозначения гиперонимической ячейки: ввод информации о том, что слово представляет собой целое / часть в лексиконе существительных, не дает заключения о том, к какому гнезду принадлежит это слово. Наличие такой базы данных в учебном корпусе станет основой для разработки мереологической поисковой системы национального корпуса узбекского языка.

В разделе главы, озаглавленном «*Структура и строение базы данных по межглагольным формальным отношениям*», обсуждается создание одноименной, паронимической базы данных для образовательного корпуса. База данных одноименных слов узбекского языкового образовательного корпуса, сформированная как практическое развитие исследования, состоит из 2384 лексем. На основе принципа неограниченного размера при формировании базы данных одноименных слов необходимо:

активировать функцию сбора омонимических слов, ввод их в базу данных и постепенное заполнение базы независимо от источника на узбекском языке;

не давать русского эквивалента лексемы (в отличие от простого словаря) в связи с тем, что корпус направлен на описание возможностей узбекского языка;

из-за большого размера базы данных следует соблюдать принцип не укорачивать комментарии.

В обучающем корпусе узбекского языка практически невозможно сформировать словарную базу-пароним автоматически / полуавтоматически: единственный способ - подготовить лингвистический материал. Не наблюдался даже единый стандарт для импорта напрямую из опубликованного словаря. По этой причине таблица формируется на основе структуры словарной статьи.

## **ЗАКЛЮЧЕНИЕ**

1. Обучающий корпус узбекского языка – корпус, предназначенный для обучения пользователей узбекскому языку как родному или иностранному, содержащее электронные тексты, работающее в виде специального сайта.

2. Корпус можно разделить на академический корпус и обучающий корпус по их направленности. Адаптация воспитательного корпуса из национального корпуса к образовательному процессу; ограниченный размер; вводимые тексты ясны, просты и понятны; стандартизованы; внедрение учебных словарей; отличается возможностью обогащения различными дополнительными материалами (картинками, аудио, мультимедиа)

3. Тексты официального, научно-популярного, художественного и публицистического стиля популярного, нравственного и просветительского содержания современного узбекского литературного языка выбираются пропорционально в качестве материала обучающего корпуса узбекского языка. Тексты подробно помечены лингвистической информацией образовательного характера. Особое место среди них занимает морфологическая разметка. Основой для этого служит грамматический

словарь. Грамматический словарь – это группа символов слов на определенном языке, словарь, охватывающий все грамматические формы. Морфолексикон – это электронная база данных, которая содержит лексические единицы определенного языка, информацию о его принадлежности, грамматические особенности, адаптированные к процессу обработки естественного языка, различные средства обработки естественного языка (автоматический перевод, лингвистический корпус, морфологический, семантический, грамматический анализатор), выступает в качестве лингвистической поддержки.

4. В связи со сложностью формирования морфолексикона узбекского языка оно проводится в 4 этапа. В процессе формирования исходной базы морфолексикона узбекского языка было проведено 2 этапа и выявлены: пропорции словосочетаний в узбекском языке: 22599 существительных, 7172 прилагательных, 3949 глаголов, 1856 обстоятельств, 800 имитационных слов, 244 побудительных слов, 182 наречия, 132 модальных слова, 108 полифункциональных слов, 106 вспомогательных слов, 52 союза, 45 предлогов, 30 числовых лексем; общее количество лексем - 37 282.

5. Морфологический словарь глагольных форм для практического применения при формировании морфологической основы обучающего корпуса узбекского языка. Словарь глагольных форм на примере глагола узбекского языка включает около 30 тысяч слов из расположения форм отношения, времени, наклонения, лица-числа, неделимости, функции. Когда к вышеперечисленным формам были присоединены вспомогательные глаголы, вместе образовалось более 300 000 одинарных глагольных форм. Грамматические формы различных глаголов в узбекском языке колеблются от 4-5 тысяч до 300 тысяч. Вместе с нагрузками он снова удваивается.

6. Омонимное и полифункциональное слово не может быть определено информацией о принадлежности к категории на основе морфолексикона. Поскольку значение таких единиц реализуется в тексте, алгоритм автоматического определения омонимического, полифункционального слова требует специального исследования. При формировании базы данных антонимов рекомендуется обрабатывать оба члена пары антонимов в гиперссылке на толковый словарь. Формирование базы данных синонимов требует ручного ввода информации из словаря синонимов в базу данных, решая перекрестные ссылки слов для запуска поисковой системы.

7. Использование корпуса в образовании (1) для сравнения речи человека, изучающего язык, с его носителем; (2) сравнение литературной устной / письменной речи; (3) сравнить языковые стили в книгах, учебниках, публикациях, выявить различия; (4) служит современным инструментом для таких процессов, как определение порядка слов в предложениях.

**SCIENTIFIC COUNCIL Ph.D. 31.03.2021. FIL. 78.04 ON AWARD OF  
SCIENTIFIC DEGREE DOCTOR OF SCIENCES  
AT TERMEZ STATE UNIVERSITY**

---

**TERMEZ STATE UNIVERSITY**

**KHOLIYAROV URAL MENGLIEVICH**

**LINGUISTIC BASIS OF COMPILING UZBEK  
EDUCATIONAL CORPUS**

**10.00.01 – Uzbek language**

**DISSERTATION ABSTRACT  
OF THE DOCTOR OF PHILOSOPHY (PHD) ON PHILOLOGICAL SCIENCES**

**Termez – 2021**

**The theme of dissertation for doctor of philosophy (PhD) on philological sciences was registered at the Supreme Attestation Commission of the Cabinet of Ministers of the Republic of Uzbekistan under № B2020.4.PhD/Fil1490.**

The dissertation has been prepared at Termez State University.

The abstract of the dissertation is posted in three languages (Uzbek, Russian, English (resume)) is placed on the website of Scientific Council ([www.terstu.uz](http://www.terstu.uz)) and the website of «Ziyonet» Information and Educational portal ([www.ziyonet.uz](http://www.ziyonet.uz)).

**Scientific supervisor:** **Mengliev Bakhtiyor Rajabovich**  
Doctor of Philological Sciences, Professor

**Official opponents:** **Shakhobitdinova Shohida Hoshimovna**  
Doctor of Philological Sciences, Professor

**Khudoyberganova Durdona Siddikovna**  
Doctor of Philological Sciences, Professor

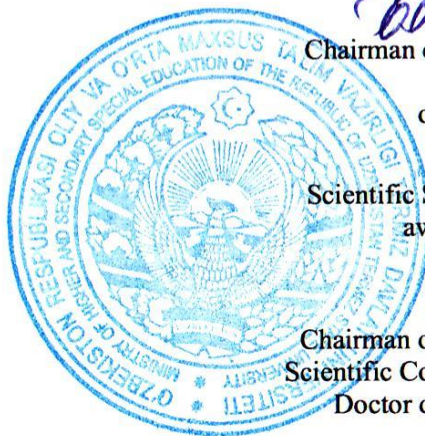
**Leading organization:** **Samarkand State University**


The defense of the dissertation will take place on “03” 08, 2021 at 10:00 at the session of Scientific Council PhD.03/31.03.2021.Fil.78.04 on award of scientific degree of Doctor of Sciences at Termez State University. Address: 190111, Termez, Barkamol avlod str. 43. Phone: (+99876) 221-74-55; Fax:(+99876) 221-71-17. E-mail: termizdu@umail.uz


The doctoral dissertation can be looked through at the Information Resource Center of the Termez State University (Registration № 3). (Address: 190111, Termez, Barkamol avlod str. 43. Phone: (+99876) 221-74-55


The abstract of the dissertation is distributed on «26» 07 2021.

(Mailing protocol № 3 dated «26» 07 2021.



  
**B. Umurkulov**  
Chairman of the Scientific Council awarding scientific degrees, Doctor of Philological Sciences, Professor

  
**A. Eshmuminov**  
Scientific Secretary of the Scientific Council awarding scientific degrees, PhD on philological sciences

  
**Sh. Mahmaraimova**  
Chairman of the Scientific Seminar under the Scientific Council awarding scientific degrees, Doctor of Philological Sciences, Professor

## INTRODUCTION (Abstract of PhD dissertation)

**Topicality and relevance of the dissertation theme.** In world linguistics, the field of natural language processing has expanded, and the achievements of computational linguistics have become widespread in the study of language, linguodidactics, lexicography, and translation. In particular, along with the thesaurus, electronic dictionaries, automatic translation, information retrieval systems, the language corpus has become one of the main tools in language learning. As a result, training corpuses have been formed that provide linguistic information about language units in a convenient way in a matter of seconds using a special search, contain thousands of texts in a specific part of the language and are designed to teach language capabilities. In the current process of globalization, the transition of world civilization to the virtual world, the expansion of the use of some languages threatens the survival of other languages, therefore these corpora are important as one of the possibilities for linguistic survival.

In the world of linguistics in the 21<sup>st</sup> century, there is a growing movement for the scientific and theoretical study of corpus problems. Purposeful research was carried out on the social significance of language corpora, principles of construction, development of a system of language tags, automatic markup. Saving time spent on the process of text analysis, the ability to demonstrate the features of language units in the structure of speech in thousands of examples, is recognized as the main effective capabilities of the corpus. Therefore, in the rapidly developing field of computational linguistics, the selection of units of the training corpus, their linguistic, extralinguistic labeling, the development of algorithms for creating the training corpus, the creation of linguistic support for the training corpus has become a social necessity.

As a result of focusing on scientific research in the field of computational linguistics in Uzbekistan, on the processing of the Uzbek language using modern information technologies, the first speeches of the national corpus began to appear as practical works. Indeed, «the creation of a national corpus of the Uzbek language in electronic form, which includes all scientific, theoretical and practical information about the Uzbek language»<sup>1</sup> will greatly contribute to increasing the prestige of our people in public and international life. Thus, in order to achieve widespread use of the Uzbek language in the information and communication system, for in-depth study of our national and cultural heritage, to increase electronic resources in teaching in the native language, to achieve unlimited use of these resources by students, to popularize national and spiritual education, collection materials of the native language into a single platform has become an urgent task. Also, the study of issues such as the formation of the Uzbek language educational corpus, the creation of its linguistic support based on modern scientific principles, further improvement of the system of teaching the state language, increasing its authority as a scientific

---

<sup>1</sup> Ўзбекистон Республикаси Президентининг 2020 йил 20 октябрдаги “Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида”ги ПФ-6084-сон Фармони.  
URL: <https://lex.uz/docs/5058351>

language and the formation of technological resources for the social development of the republic.

The given research work serves to a certain extent to perform the tasks specified in the Resolutions of the President of the Republic of Uzbekistan PF-4947 «About the Development Strategy of the Republic of Uzbekistan» dated February 7, 2017, PQ-5850 «On measures to radically increase the prestige and status of the Uzbek language as a state language» dated October 21, 2019, PQ-6084 «On measures to further develop the Uzbek language and improve language policy in our country» dated October 20, 2020, PP-4479 «On the wide celebration of the thirtieth anniversary of the adoption of the Law of the Republic of Uzbekistan» On the state language» dated October 4, 2019, address of the President of the Republic of Uzbekistan to the Oliy Majlis on January 24, 2020.

**The aim of research work** is to identify theoretical issues of the formation of the teaching corpus of the Uzbek language, as well as the scientific coverage of the foundations of the creation of linguistic support for the teaching corpus of the Uzbek language.

**The tasks of the research:**

to give a general description of the teaching corpus, highlight its importance as an educational resource;

identification of the advantages of the teaching corpus in teaching the native language, teaching Uzbek as a foreign language (linguodidactics);

development of principles for selecting material for the Uzbek language training corpus, collecting material for a fragment of the training corpus;

defining the process of collecting textbooks in the teaching corpus and developing their search system;

formation of the morpholexicon of the teaching corpus for theoretical conclusions; creation of linguistic foundations of the morphological and derivational dictionary of the Uzbek language;

description of the structure and structure of the database for the interpretation of the meanings of words, interword formal relations.

**The object of the research** is a teaching corpus designed for teaching language skills, which has a linguodidactic character.

**The scientific novelty of the research** is conditioned by:

It is justified the differentiation of educational corpus on the fact that the linguistic, didactic, informational features specific in terms of linguistic support from functional capabilities of the world teaching corpus and similarities in terms of methods of automatic language processing;

Optimal mathematical and linguistic methods of formation of the morpholexicon of the Uzbek language and the morphological base of the Uzbek language for identification of the system of automatic processing of educational building material are defined and cognitive, linguistic-statistical scientific-theoretical bases of formation of the morpholexicon and morphological base are developed;

It is defined the order of compiling a dictionary of verb forms, linguistic, mathematical, logical, programming principles of systematization of stem and

grammatical forms in placing the corpus, and it is revealed the search opportunities of Uzbek language educational corpus, interface construction, concordance of educational corpus and their educational, linguodidactic, resource opportunities as electronic library;

It is worked out the order of placement in the database parameters of educational corpus of dictionaries, word formation, word meanings, inter-word semantic, formal relations in the form of library, concordance and SQL search method.

**Implementation of the research results.** Based on the scientific results obtained on the linguistic basis of the formation of the teaching corpus of the Uzbek language: theoretical information on the principles of selection of texts of ethical and educational content for the teaching corpus, textbooks, dictionaries, the formation of a database of interword spiritual, formal connections, the creation of an appropriate search engine, placement in the corpus database was used in the fundamental scientific project FA-A1-G007 «Karakalpak Proverbs as an object of linguistic research» (Certificate No. 17.01 / 108 Karakalpak branch of the Academy of Sciences of the Republic of Uzbekistan issued on April 22, 2021,) by the Karakalpak branch of the Academy of Sciences of the Republic of Uzbekistan, Karakalpak research humanitarian institute. As a result, it was possible to determine the morphological forms and position of linguistic units in the Karakalpak proverbs, to reveal their meaning as a linguistic basis;

effective mathematical and linguistic methods for creating the morpholexicon of the educational corpus through the formation of the morphological base of the Uzbek language, determining the statistics of belonging to the category of Uzbek lexemes, expanding the scope of linguistic studies of word formation; word formation, conclusions on the content and structure of the database on the interpretation of word meanings were used in the scientific and practical project FA-F1-G003 «Functional word formation in the modern Karakalpak language», implemented at the Karakalpak Scientific Research Humanitarian Institute of the Karakalpak Branch of the Academy of Sciences of the Republic of Uzbekistan (Certificate No. 17.01 / 108 dated April 22, 2021). As a result of the design work, models of word formation in the modern Karakalpak language have been created, theoretical generalizations have been improved when analyzing the foundations of word formation.

the structure and composition of the database on educational dictionaries, word formation, interpretation of the meanings of words, inter-word spiritual, formal relations; conclusions about scientific and theoretical tasks when placed in the teaching building were used in paragraphs 17-18 of the textbook «Modern Uzbek language» «Lexical-semantic relations», «Classification of Uzbek vocabulary», recommended by the Ministry of Higher and Secondary Specialized Education. (Certificate of the Ministry of Higher and Secondary Education of the Republic of Uzbekistan No. 89-03-2167 dated May 24, 2021). As a result, an interpretation of the meanings of words was achieved, an improvement in the content of topics related to inter-verbal spiritual, formal relations; definition of morphological forms and positions of the morphological dictionary of the Uzbek language, the order of

compiling the dictionary of verb forms, principles and possibilities of systematizing the basic and grammatical forms of the corpus, recommendations on the model of grammatical morphemes used to create the Uzbek morphological dictionary (O'zbek tilining fe'l shakllari lug'ati ("Kiymoq" fe'li misolida) [Matn]: lug'at. – Termiz: Termiz davlat universiteti nashr-matbaa, 2021. - 492 b. ISBN 978-9943-7243-3-4). As a result, a database of morphological forms was created and a basic linguistic support was formed for the development of a morphological equalization program and a morphological analyzer of the Uzbek National Corpus;

conclusions about the need to create a teaching corpus of the Uzbek language, its specific informational features, linguistic support, its difference and similarities with the world educational corpus, and conclusions about their linguodidactic significance were used in drawing up scenarios of literary, artistic, spiritual and educational programs of the Surkhandarya regional television and radio company (Certificate No. 01-01 / 54 of the Surkhandarya regional television and radio company dated April 29, 2021). As a result, the content of propaganda work on the education of a high, harmoniously developed generation, the formation of a linguistic culture among young people and the use of the Uzbek language in the modern information system was improved.

**The structure and volume of the dissertation.** The dissertation consists of an introduction, three chapters with outcomes, general conclusions, the list of the used literature. The basic volume of the dissertation is 147 pages.



**ЭЪЛОН ҚИЛИНГАН ИЛМИЙ ИШЛАР РЎЙХАТИ**  
**СПИСОК ОПУБЛИКОВАННЫХ РАБОТ**  
**LIST OF PUBLICATION**

**I бўлим (I часть; I part)**

1. Холиёров Ў.М. Compiling a dictionary of morphological forms of the uzbek language // Berlin Studies Transnational Journal of Science and Humanities. Berlin. Volume 1, Issue 1.2, May 2021. – P. 22-28. <http://berlinstudies.de> (10.00.00; Европа мамлакатлари нашрлари № 10)

2. Холиёров Ў.М. Таълимий корпусда сўзлараро маъновий муносабатларга доир маълумотларни базасини яратиш // ЎЗМУ хабарлари. Ижтимоий-гуманитар фанлар туркуми. Илмий журнал. – Тошкент, 2021. № 1/3. – Б. 281-285. (10.00.00 № 15)

3. Холиёров Ў.М. Таълимий корпусда сўз маънолари изоҳига доир маълумотлар базасини шакллантириш // Тил ва адабиёт таълими. Илмий-методик журнал. – Тошкент, 2021. № 5. – Б. 46-48. (10.00.00 № 9)

4. Холиёров Ў.М. Formation of word formation dictionary for educational corpus // ACTUAL PROBLEMS OF MODERN SCIENCE, EDUCATION AND TRAINING IN THE REGION. ELECTRONIC JOURNAL. March, 2021-VI. ISSN 2181-9750. – P. 38-45. <http://khorezmscience.uz/web/archive/1617347811.pdf> (10.00.00 № 27)

5. Холиёров Ў.М. Таълимий корпус учун морфолексиконни шакллантириш // Сўз санъати халқаро журнали. – Тошкент, 2021. 4-жилд, 3-сон. – Б. 100-106. (10.00.00 № 31)

6. Холиёров Ў.М. Educational building and electronic dictionaries // 3rd International Multidisciplinary Scientific Conference on Ingenious Global Thoughts Hosted from Kuala Lumpur, Malaysia, May 31st 2021. – P. 67-68. <https://conferencepublication.com/index.php/aoc/article/view/1086/1166>

7. Холиёров Ў.М. Хориж лингводидактикасининг электрон таълим ресурсларидан фойдаланиш тажрибаси хусусида // Ўзбек тилини дунё миқёсида кенг тарғиб қилиш бўйича ҳамкорлик истиқболлари: Халқаро илмий-амалий анжуман материаллари. – Тошкент, 2020 йил 19-20 октябрь. – Б. 461-465.

8. Холиёров Ў.М. Таълимий корпуслар учун гипонимик, градуонимик ва партонимик муносабатлар билдирувчи бирликлар базасини шакллантириш // Компьютер лингвистикаси: муаммолар, ечим, истиқболлар. Республика I илмий-техникавий конференция. – Тошкент. Vol. 1. №. 01 (2021). – Б. 30-35. <http://compling.navoiy-uni.uz/index.php/conferences/article/view/46/42>

9. Холиёров Ў.М. Таълимий лингвистик корпус зарурати // Ўзбек тилшунослигининг долзарб масалалари: Онлайн илмий анжуман материаллари. – Термиз, 2020 йил 21 октябрь. – Б. 18-21.

## II бўлим (II часть; II part)

10. Холиёров Ў.М. The morpholexicon of the uzbek language as a source for the corpus // ACADEMICIA: An International Multidisciplinary Research Journal. – India. Vol. 11, Issue 4, April 2021. – P. 1904-1910. ISSN: 2249-7137, <https://saarj.com>.

11. Холиёров Ў.М. Акс маънонинг социал-фалсафий ва амалий лингвистик аспектига доир // ЎзМУ хабарлари. Ижтимоий-гуманитар фанлар туркуми. Илмий журнал. – Тошкент, 2018. №1/3. – Б. 492-495. (10.00.00 № 15)

12. Холиёров Ў.М. Educational building as an electronic resource // 2th International Multidisciplinary Scientific Conference on Ingenious Global Thoughts Hosted from Berlin, Germany. April 30th 2021. – P. 15-17. <https://conferencepublication.com/index.php/aoc/article/view/883/956>

13. Холиёров Ў.М. Тил таълимида электрон таълим ресурсларининг ўрни // Bildiri tam metinleri kitabi // Uluslararası Mersin Sempozyumu, KİTAP 4, Mersin / Türkiye. 22 – 24 Ekim, 2020, – Б. 156-170.

14. Менглиев Б.Р., Хамроева Ш.М, Холиёров Ў.М. Ўзбек тили таълимий корпуси: тузилиши, таркиби, имкониятлари // Ўзбек миллий ва таълимий корпусларини яратишнинг назарий ва амалий масалалари: Халқаро илмий-амалий конференция материаллари. 2021 йил 7 май. Электрон нашр / ebook. – Тошкент: ТошДЎТАУ, 2021. – 330 б. – Б. 122-125. <http://compling.navoiy-uni.uz>

15. Менглиев Б.Р., Холиёров Ў.М., Хамроева Ш.М. Ўзбек тили таълимий корпусининг қидирув тизими хусусиятлари // Ўзбек миллий ва таълимий корпусларини яратишнинг назарий ва амалий масалалари / Халқаро илмий-амалий конференция материаллари. 2021 йил 7 май. Электрон нашр / ebook. – Тошкент: ТошДЎТАУ, 2021. – 330 б. – Б. 126-135. <http://compling.navoiy-uni.uz>

16. Хамроева Ш.М., Холиёров Ў.М., Абдуалимова Г.Ҳ. Ўзбек тилининг феъл шакллари луғати (“киймоқ” феъли мисолида): луғат / – Термиз: Термиз давлат университети нашр-матбаа, 2021. – 492 б.

Авторефератнинг ўзбек, рус ва инглиз (резюме) тилларидаги нусхалари “ТерДУ нашр-матбаа маркази” таҳририятида таҳрирдан ўтказилди (23.07.2021 й.).

Босишга рухсат этилди 23.07.2021 й.  
Офсет босма қоғози. Қоғоз бичими 60 1/16. «Times» гарнитураси.  
Офсет босма усули. Ҳажми 3,3 босма табоқ.  
Адади 100. Буюртма – № 24.

“Термиз давлат университети нашр-матбаа маркази” босмахонасида чоп этилди.  
Манзил: Термиз шаҳри, Баркамол авлод кўчаси, 43-уй.

