

КОРПУС ЛИНГВИСТИКАСИ АТАМАЛАРИНИНГ ҚИСҚАЧА ИЗОҲЛИ ЛУҒАТИ

Ушбу луғатда корпус лингвистикасининг энг асосий атамалари: корпус, унинг турлари, муаллифлик корпуси, ўзбек тили муаллифлик корпуси яратишнинг лингвистик асослари билан боғлиқ атамалар, жаҳон корпус лингвистикаси атамалари (тег, лемма, разметка, парсинг, таггинг, аннотация, корпус менежери, конкорданс каби) изоҳи, Интернет тармоғида мавжуд корпуслар ҳақида маълумотлар шарҳланган.

Мазкур луғат ижтимоий-гуманитар ва ахборот технологиялари соҳаси талабалари, магистрантлар, тилшунослар, тадқиқотчилар ҳамда соҳага қизиқувчилар оммасига мўлжалланган.



Хамроева Шахло Мирджоновна – Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёт университети докторанти. 2018 йилда “Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари” мавзусида филология фанлари бўйича фалсафа доктори (PhD) диссертациясини ёқлаган.



9 786200 613295

Globe
EDIT

Globe
EDIT



Шахло Хамроева

КОРПУС ЛИНГВИСТИКАСИ АТАМАЛАРИНИНГ ҚИСҚАЧА ИЗОҲЛИ ЛУҒАТИ

Терминологик луғат

Шахло Хамроева

КОРПУС ЛИНГВИСТИКАСИ АТАМАЛАРИНИНГ ҚИСҚАЧА
ИЗОҲЛИ ЛУҒАТИ

FOR AUTHOR USE ONLY

FOR AUTHOR USE ONLY

Шаҳло Хамроева

**КОРПУС ЛИНГВИСТИКАСИ
АТАМАЛАРИНИНГ ҚИСҚАЧА
ИЗОҲЛИ ЛУҒАТИ**

Терминологик луғат

FOR AUTHOR USE ONLY

GlobeEdit

Imprint

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this work is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: www.ingimage.com

Publisher:

GlobeEdit

is a trademark of

International Book Market Service Ltd., member of OmniScriptum Publishing Group

17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page

ISBN: 978-620-0-61329-5

Copyright © Шахло Хамроева

Copyright © 2020 International Book Market Service Ltd., member of OmniScriptum Publishing Group

FOR AUTHOR USE ONLY

Ш.М. ХАМРОЕВА

**КОРПУС ЛИНГВИСТИКАСИ
АТАМАЛАРИНИНГ
ҚИСҚАЧА ИЗОҲЛИ ЛУҒАТИ**

FOR AUTHOR USE ONLY

Ш.М. ХАМРОЕВА

**КОРПУС ЛИНГВИСТИКАСИ
АТАМАЛАРИНИНГ
ҚИСҚАЧА ИЗОҲЛИ ЛУҒАТИ**
(тўлдирилган кайта нашр)

Мазкур луғат Бухоро давлат университети ўқув-методик Кенгашининг 2018 йил 28 сентябрдаги навбатдан ташқари йиғилиши қарори билан нашрга тавсия қилинган.

Масъул муҳаррир:

Б.Менглиев – филология фанлари доктори, профессор

Тақризчилар:

М.Абузалова – филология фанлари доктори

С.Назарова – филология фанлари номзоди, доцент

Ушбу луғатда корпус лингвистикасининг энг асосий атамалари: корпус, унинг турлари, муаллифлик корпуси, ўзбек тили муаллифлик корпуси яратишнинг лингвистик асослари билан боғлиқ атамалар, жаҳон корпус лингвистикаси атамалари (тег, лемма, разметка, парсинг, таггинг, аннотация, корпус менежери, конкорданс каби) изоҳи, Интернет тармоғида мавжуд корпуслар ҳақида маълумотлар шарҳланган.

Мазкур луғат ижтимоий-гуманитар ва ахборот технологиялари соҳаси талабалари, магистрантлар, тилшунослар, тадқиқотчилар ҳамда соҳага қизиқувчилар оммасига мўлжалланган.

ТИЛ КОРПУСИ ВА УНИНГ АМАЛИЙ АҲАМИЯТИ

Корпус – маълум мақсадда йиғилган матнлар мажмуини ташкил этувчи тил бирликлари йиғиндиси, табиий тилдаги электрон шаклда сақланадиган ёзма ёки оғзаки, компьютерлаштирилган кидирув тизимига дастурий таъминот асосида жойлаштирилган on-line ёки off-line тизимда ишлайдиган матнлар жамланмаси. Лингвистик тадқиқот ва амалий топшириқлар ечими учун тил корпуслари замонавий тилшуносликнинг инкор этиб бўлмас иш қуроли. У оддий электрон кутубхонадан фарқланади. Электрон кутубхона халқнинг ижтимоий-сиёсий, маънавий, иқтисодий ҳаётини ақс эттирувчи бадиий ва публицистик асарларни нисбатан тўлиқ қамраб олишни мақсад қилган бўлади. Электрон кутубхона матнлари тилшунослик нуқтаи назаридан лексик, морфологик, грамматик, семантик белгилар асосида ишлов берилмаганлиги сабабли лингвистик тадқиқот учун ноқулайлик туғдиради. Чунки электрон кутубхона илмий тадқиқот материали базасини тайёрлаш мақсадида тузилмайди, балки миллий-маънавий меросни тўплашни мақсад қилган бўлади. Тил корпуси эса электрон кутубхонадан фарқли ўлароқ, тилни ўрганиш учун зарурий, фойдали ва қизиқарли матнларни тўплашни назарда тутди. Корпусни электрон кутубхонадан ажратиб турувчи биринчи омил ундаги матннинг хусусияти ва қўшимча маълумот билан бойитилганлиги ҳисобланади ҳамда бу белги корпуснинг алоҳида қисми – корпус бирликларига ёзилган изоҳни ташкил этади.

Фойдаланувчига бирор сўз керак бўлса, бунини оддий матн муҳаррири ҳам топиб беради. Лекин сўзнинг грамматик маъноси ва тузилишини “тушунадиган” матн билан ишлаш жуда кўп афзаллик ва қулайликка эга. Сўзшаклни кидириш керак бўлса, бундай дастурий таъминот, яъни корпус тадқиқотчи ёки фойдаланувчига жуда катта ёрдам бериши мумкин. Агар тадқиқотчи ўз иши учун мисолни топиш, уларни картотекага қўчириш (компьютер технологиялари ривожланишидан олдинги даврда)га ойлаб, баъзан йиллаб вақт сарфлаган бўлса, бугун дунё тил корпуслари ёрдамида санокли дақиқада юзлаб мисолни топишга, улар устида ишлаш имконига эга

бўлди. Махсус қидирув корпусдан маълумот олишга мўлжалланган бир қанча дастурдан иборат, статистик ахборот ва қидирув натижасини фойдаланувчига қулай шаклда тақдим этади. Тилда қандай жараён кечаётганлигини аниқ тасаввур қилиш учун корпус қамровини янада кенгайтириш, нафақат ёзма нутқ, балки оғзаки нутқ материалдан фойдаланиш ҳам мақсадга мувофиқ. Бундай корпус ёрдамида тил тараққиёти натижасида унда содир бўлган ва қутилаётган ўзгариш ҳақида аниқ хулоса чиқариш мумкин.

Корпус бўйича қидирув фойдаланувчига қуйидаги натижани беради:

- 1) белгиланган сўзнинг турли контекстдаги барча шаклдаги кўриниши;
- 2) тилнинг луғат таркибидаги ўрни ва вариантлари;
- 3) белгиланган сўз билан бирикиш имкониятига эга сўзлар рўйхати;
- 4) у ёки бу ёзувчининг айна сўздан фойдаланиш частотаси ёки статистикаси;
- 5) сўзнинг ўз ва кўчма маъноси;
- 6) лексик қўлланишининг яширин модели (имконияти);
- 7) тил тараққиётининг турли даврида қўлланиш ҳолати.

Тасниф белгиси	Корпус турлари
Нутқ турига кўра	Оғзаки, ёзма, аралаш
Матн тилига кўра	Русча, инглизча 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, ёпик
Мақсадига кўра	Тадқиқий, тавсифий
Изоҳланишига кўра	Қўшимча изоҳли, изоҳсиз
Ўзгарувчанлигига кўра	Ўзгарувчан, турғун
Структурасига кўра	Марказий, архив, маҳаллий
Изоҳ хусусиятига кўра	Морфологик, синтактик, семантик
Матн ҳажмига кўра	Тўлиқ матнли, лавҳа матнли
Хронологик жиҳатига кўра	Синхрон, диахрон
Умумийлигига кўра	Умумий, яқка муаллифли

Корпуснинг кенг қўламли бўлиши маълумотнинг ўзига хослигини кафолатлайди, тил ҳодисаларининг барча киррасини тўлиқ намоиш этишни таъминлайди. Турли типдаги маълумот тил корпусида ўзининг табиий контекстуал шаклида жойлашади, бу эса уни ҳар томонлама ва объектив ўрганиш учун асос бўлади. Бир марта тузилган ва тайёрланган ахборотлар массиви бир неча тадқиқотчи томонидан кўп марта, турли мақсадда ишлатилиши мумкин. Корпус тилни тадқиқ этиш (тилда сўзнинг ўзгариши, историзм, неологизм, маъно кенгайиши ва торайиши; янги фразеологизмларнинг пайдо бўлишини кузатиш), тил ўрганиш, луғат тузишда энг замонавий, кенг имкониятли дастурлаштирилган тизим. Корпуссиз бугунги кун назарий ва амалий филологиясини тасаввур этиш қийин. Тилшуносликка оид тадқиқотда факт билан иш кўриладиган ҳолларда материал йиғилиши, системага солиниши лозим. Бундай катта ҳажмли ишни бажаришда корпус тилшунос учун вақт ва меҳнатни тежайдиган иш курали. У техник жараёни тезлаштирувчи восита бўлиб қолмай, маълум тил замонавий шаклининг ахборот тизими бўлиб, қутилмаган саволга ҳам жавоб бера оладиган, тилшунослик олдига аввал қўйилмаган долзарб муаммони қўя оладиган тизим.

Мавжуд корпуслар таркибидаги матнларнинг нисбатига қарайдиган бўлсак, бадий адабиёт хиссаси 40% ни ташкил этишига гувоҳ бўламиз. Бунинг таркибига мемуар асарлар ҳам кириб кетадики, бу жанр тил хусусияти бадий ва публицистик услуб оралиғида бўлиб, жонли тилни ўрганиш учун анча қулай. Евпора тиллари корпусларида бадий адабиёт материали 20% ни ташкил этади. Масалан, замонавий ёзувчилар тил хусусиятини ўрганишга бағишланган 20дан ортиқ тадқиқот мавжуд бўлса-да, улар ҳали тўлалигича бу муаммони ўрганиб бўлди, дейиш қийин. Чунки алоҳида ёзувчи асарининг тил хусусиятидаги ўзгаришга ҳали тилдаги янги ҳодиса деб қараб бўлмайди. Ана шундай жараёнларни кузатиш ва тадқиқ этишнинг энг қулай воситаси тил корпуси. Шу сабабли матннинг мазмуни муҳим аҳамият касб этади. Корпус таркибига кирадиган матнлар алоҳида бир

муаллиф ёки бир неча ёзувчи асаридан олинган матн, маълум давр (бир неча ўнйиллик ёки юзйиллик)ни қамраб олган матн, белгиланган мавзудаги замонавий матн, тил ва жамиятнинг бугунги ҳолатини акс эттирувчи замонавий матнлардан иборат бўлиши мумкин.

Мавжуд миллий корпусларни кузатиш асосида корпуснинг тузилиши ва таркибини ўрганар эканмиз, корпус интерфейси ва қидирув тизими унинг энг асосий таркибий қисми, деган хулосага келамиз. Рус тили миллий корпуси (Национальный корпус русского языка (НКРЯ)) жойлашган www.ruscorpora.ru сайтининг биринчи саҳифасида корпус ва унинг тузувчилари ҳақида асосий маълумот, ўнг томондаги менюда исталган саҳифага ўтиш имконияти мавжуд. Меню тўрт қисмдан иборат бўлиб, бош саҳифа, сайтнинг қидирув ресурси, матн ҳақида унга бириктирилган қўшимча маълумот иловаси, корпус бирликларига изоҳ ёзиш принциплари, охириги блок корпус тузувчилари жамоаси, фойдаланилган дастур, матнларнинг муаллифлик ҳуқуқи ҳақида тўлиқ маълумотлар базасидан иборат.

Корпуснинг лексикографиядаги аҳамияти. Корпус катта ҳажмли луғатларни тузиш учун манба. Вақт ўтиши билан корпус турли лингвистик йўналишлар учун аҳамиятли бўлиши билан кучли информацион ресурсга айланди. Корпус асосида компьютер ёрдамида луғатлар аввалгига нисбатан тезлик билан тузилади ва қайта ишланади. Шу йўл билан иш бошланишидан тугаш жараёнигача (нашргача) тилни акс эттириб туради, луғат мақоласи «эскириш»га улгурмайди.

Лексикология соҳасидаги аҳамияти шундаки, сўзнинг қўлланиш даври ва частотасини аниқлашда ҳеч қандай восита корпусга тенглаша олмайди. Корпус лингвистикасининг кейинги тараққиёт босқичида статистик тадқиқ методи компьютер таржимаси, нутқни синтезлаш ва таниш, аниқлаш, орфографик текширув каби лингвистик амалларни бажаришда қўл кела бошлади. Рус ва Европа тилшунослигида тил бўйича қилинадиган барча тадқиқотларнинг корпусга асосланиши одатий ҳолга айланган ва ҳатто баъзи

таджикотларда мажбурий ҳам саналади. Яқин орада тил ўрганаётган ўқувчи ёки унинг бирор жиҳатини тадқиқ этаётган киши бугунги кунда луғатга бўлган эҳтиёжи корпус кўчиши шубҳасиздир.

Синтаксисни ўрганишдаги аҳамияти тилнинг жонли қурилишини, сўзларнинг ўзаро бирикиш имкониятини таҳлил қила олишида ҳисобланади. Анъанавий тилшуносликдаги мисолни олишда бадиий асар тилига таяниш тажрибасидан кўра корпусга таяниш мисолнинг бугунги кун учун ҳам ишонарлилигини таъминлайди.

Услубиятни ўрганишда матннинг услубий ўзига хослигини таҳлил этишда ҳам корпусга таяниш мумкин. Бу амалиётни бажаришда матнларнинг статистик ҳолати таҳлилини (матндаги жумланинг узунлиги, бир сўзнинг бошқа сўз билан бирикиш ҳолатининг доимий ёки ноодатийлиги) аниқлаштиради. Бундай усулда корпус ёрдамида ёзма нутқ билан бирга оғзаки нутқни ҳам ўрганиш мумкин бўлади. Масалан, расмий иш қоғозларида ҳужжат матнида *шартнома бўйича талаблар* шакли маъқулми ёки *шартнома талаблари* конструкцияси тўғрими? Корпусга бу тарзда берилган саволга санокли дақиқаларда қайси шакл кўпроқ қўлланишини аниқлаб беради. Шундай қилинув натижасида рус тили миллий корпусида *шартнома бўйича талаблар* конструкцияси 40 марта, *шартнома талаблари* шакли 3 марта қўлланганлиги тўғрисида статистик маълумот беради. Бу конструкциянинг қўлланиши албатта, рус тилига хос, эҳтимол, у ўзбек тилида бошқача натижа берса керак.

Корпуснинг таълимий аҳамияти. Корпус жуда муҳим бўлган ва фаол қўлланадиган соҳа бу лингводидактика бўлиб, она тили ва хорижий тилни ўрганишда бирдек аҳамиятли. Тилни ўргатишда луғат бойлигининг катталигини кўрсата олиш, сўзнинг қўлланиш имкониятини у ёки бу грамматик қурилиш орқали тушунтириш учун мисоллар массивини кўрсатишда корпус жуда қўл келади. Тил таълими учун муҳим бўлган мисолнинг доимий янгилиниб бориши, бунга акс эттириб туриш хусусияти ҳамда имконияти фақат корпусга хос. Ўқитувчи янги, ишонарли, чексиз

ҳамда хилма-хил мисолларни шу ердан топа олади, топшириқ ва машқларни белгилашда қийналмайди, бир неча дақиқада мавзу бўйича яп-янги мисоллардан иборат машқ тайёрлай олади. Корпусда мавжуд матнларни саралаш имкони мавжуд бўлиб, мисолни барча матнлардан эмас, балки тадқиқотчи учун кизиқарли ва керакли бўлганини ажратиш мумкин. Ҳатто корпус матнларини белгиланган давр (ҳатто аниқ йилигача), матннинг аниқ бир тури (масалан, реклама матни, иш ҳужжати ёки шунчаки бир неча муаллиф асарлари)ни танлаб олиш имконияти мавжуд. Турли мавзу ва жанрдаги матнлар билан корпуснинг доимий бойитилиши корпуснинг асосий хусусиятидан бири бўлиб, таълимда йўналтириб ўқитиш имкониятини очади. Академик лицей ва коллежда она тили барча йўналиш ва мутахассисликларда ўқитилади. Йўналишларнинг мавзуси бир хил бўлиши мумкин, лекин ҳар бир мутахассисликда мавзунини тушунтириш учун мос матнни танлашда корпусдаги имконият бирорта манбада йўқ. Аслида академик лицей ва касб-ҳунар коллежлари ДТСида мақсад ҳам эгаллаётган мутахассислиги бўйича ёзма ва оғзаки нутқини ривожлантириш деб белгиланган. Айни бир мавзу (мас., сўзнинг шакл ва маъно муносабатига кўра тури, атамашунослик, топоним, қўлланиш доираси чегараланган лексика ва ҳ.к.) турли йўналишда алоҳида, ихтисослаштирилган матнлар асосида ўрганилиши лозим. Шундагина тил таълими ўз мақсадига эришган бўлади. Тил таълимига бағишланган дарслик ва қўлланмаларда иқтисод, молия, сиёсат, ҳуқуқшунослик каби мавзудаги матнларга дуч келмаймиз, мисолни бадий асардан олишга одатланиб қолганмиз. Ваҳоланки, адабий тил бадий асар тили дегани эмас. Тил таълими ҳаётдан узилиб қолган: кундалик турмушда ўқувчи тузадиган, ишлатадиган ёки таҳлил этадиган матнлардан узилиб, машқ ва топшириқни жонли тилдан олмасликка кўникиб қолганмиз. Н.Р.Добрушина рус тили бўйича дарслик ва қўлланмаларни таҳлил қиларкан, шундай ёзади: “Типик рус тили дарслигида машқдаги мисолларнинг 60 фоизи XIX аср, 20 фоизи XX аср бадий адабиётига оид; 10 фоизи муаллиф томонидан тузилган; замонавий ОАВ материалдан эса битта ҳам мисол йўқ.

Албатта, дарсликда мумтоз наср намунаси берилиши табиий, ижобий ҳолат. Бугунги кун ўқувчиси ўзи учун анча нотаниш бўлган бу матнни ўқишни унутмаслиги учун ҳам керак. Агар биз таълимнинг ўқитаётганимиз шахсга аниқ қаратилишини ва таълим мақсадига эришишини, таълим ва ҳаёт ўртасида фавқулудда катта тафовут бўлмаслигини истасак, биз замонавий тилга (ҳозирги адабий тил) мурожаат этишимиз, дарсликка фақат бадиий адабиёт намунаси эмас, балки кундалик турмуш тарзи ва ҳаёт билан боғлиқ матн намуналарини ҳам киритишимиз керак”.

Ўқитувчи учун корпус тенгсиз хазина, ундан ҳар бир ўқув машғулотини учун серқирра ва мазмунли ўқув материали тайёрлаш мумкин. Юқорида айтиб ўтганимиздек, корпусда хилма-хил матнлар жамланган, ундан ўқитувчи исталган шаклда фойдалана олади. Масалан, ўқитувчи топшириқни тайёрлаш учун фақат бадиий адабиёт парчасидан (ёки нобадиий матндан: спорт, автобиография ёки шунчаки маълум ижодкор ижоди билан чегараланмоқчи) фойдаланмоқчи. Унда “Менинг корпусим” иловаси орқали ўзига керакли матнни ажратиб олиш имкони мавжуд, яъни мисоллар сирасини чегаралайди. Корпус турли жанр ва услубдаги матнларни камраб олганлиги сабабли ҳар қандай талабни кондириш имконига эга. Мисол учун, журналистика факультети талабалари учун ёзилган дарсликда публицистик матнлар анча эскирган. Ҳар куни бойитилиб борадиган корпусдан янги мисолларни олиш, уни талабага тақдим этиш ёки талабага ҳам шундай топшириқ бериш, албатта, таълимни бугунги ҳаётга яқинлаштиради. Кўпинча, қайта-қайта нашр этилган дарсликда мисол эскилигича қолиб кетади ва ахборот билан шиддатли янгиланиб бораётган бугунги кун учун мавзу эскиради, талаба билимни давр билан ҳамоҳанг ўзлаштирмайди. Улар таълим жараёнида сўзнинг нозик қиррасини ўрганиш, уни нутқда қўллаш қўникмасини шакллантиришда нимага таянади? XX аср боши ёки ўртасида ёзилган асар билан ишлаш жараёни унга ҳозирги адабий тил услубини шакллантириш учун етарли эмас. Бугун филолог ёки журналист бўлиб етишаётган мутахассис тилимизнинг барча нозик қирраларини ҳис этиши

учун фақат эски – 20-30 йил олдинги матн намунаси етарли эмас, бугун ёзилган барча соҳа, услуб ва жанрдаги асарлар билан ишлаши лозим, шундагина таълим ижтимоий буюртмани бажарган ҳисобланади. Қайси фанлар, унинг айнан қайси бўлимида корпусдан қандай фойдаланиш мумкинлигини аниқ мисолларда кўриб ўтамыз.

Тил корпусидан ўқитувчи, талаба ва мактаб ўқувчиси ҳам анча унумли фойдаланиши мумкин. Чунки фақат корпус орқали жуда осонлик билан кам ишлатиладиган сўз, ибора ва бирикмани топиш, қўлланиши ва ёзилиши (орфографияси)ни ўрганиш мумкин. Шунини таъкидлаб ўтиш жоизки, тил корпусида тил грамматика ва дарслик муаллифи тавсифланганидек эмас, балки жамиятда қандай яшаса, шундай акс этади. Бу эса ўқувчини ўзини ўраб турган муҳит – умумхалқ тили ва адабий тилни ўрганишда энг сермахсул восита бўлиб хизмат қилади.

Хорижий тил таълимида корпусдан фойдаланиш. Инглиз, немис, француз, рус тилларини хорижий тил сифатида ўқитиш масаласи методикада алоҳида тадқиқ этилмоқда. Айнан тил ўргатишни мақсад қилувчи корпуслар ҳам мавжуд бўлиб, “Учебный корпус русского языка”, “Learner corpus of english” шулар жумласидан. Хорижий тил вакиллари аудиторияси билан ишлаш жараёнида тил корпусининг аҳамияти бир неча мартага ортади. Тадқиқ предмети она тили бўлмаган (иккинчи ёки хорижий тил ҳисобланган) ўқитувчи ва ўқувчи учун ҳам корпус жуда муҳим ва қулай восита. Тил эгаси учун тилдаги турли қўлланишлар уни кундалик ҳаётда қамраб турса, хорижий тил эгаси учун бу ҳолат ноодатий бўлиб, у ўрганаётган ҳар бир тил категорияси билан бевосита муфассал, турли мисоллар билан кенг танишиши лозим. Таъкидлаш жоизки, илк рус тили корпуслари Россияда эмас, Европада рус тили тадқиқотчилари томонидан яратилган.

Корпуснинг бошқа соҳалардаги аҳамияти. Тил корпусига энг кўп эҳтиёж сезувчи мутахассис бу матнни автоматик қайта ишлаш (мас., таржима дастури) ва турли қидирув тизимлари билан ишлайдиган дастурчи. Чунки бу мутахассис табиий тил билан иш кўради ҳамда ушбу тилда ёзилган барча

матнлар структурасини (табиий ва жонли тилда, академ грамматика ва дарсликдаги мисолларга таяниб эмас!) мукамалроқ “тушуниши”, хис этиши лозим.

Тил корпусига эҳтиёж сезадиган мутахассислардан яна бири кундалик иш жараёнида ёзма ва оғзаки нутқ жозибасига тез-тез мурожаат этувчилар: газета ва журнал муҳаррири, журналист, радио ва телевидение ходимлари. Чунки бу мутахассис маълум бир сўз, ибора ёки конструкциянинг қўлланилиш ҳолати, даражаси, ким, қачон илк марта бу конструкцияни қўллаганлиги, қандай услуб учун хосланишини билишга грамматика билан шуғулланувчи олимдан кўра кўпроқ эҳтиёж сезади. Корпусдан ташқари ҳеч қандай ахборот банки бу каби саволларга зудлик билан жавоб бериши мумкин эмас. Шунинг учун корпус лингвистикаси тадқиқотчилари корпуслар журналист, мухбир, муҳаррир, ўқитувчи ҳамда дастурчи учун махсус яратила бошланган, деган хулосага ҳам келишади. Биз бугунгача дарслик, академ грамматика ва луғатга қандай таянган бўлсак, замонавий соҳа вакиллари корпусга шу қадар эҳтиёж сезишади, ундан фойдаланишади.

Жаҳон тилларининг жуда кўпчилиги мукамаллик даражаси ва матнни илмий қайта ишлаш имконияти билан фарқ қилувчи ўз миллий корпусига эга. Инглиз тилида яратилган Браун корпуси, Ланкастер-Осло/Берген (ЛОБ) корпуси, Лондон-Лунд корпуси, Лексикографик тадқиқотлар учун Америка мерос корпуси (American Heritage Intermediate), Ланкастер инглиз тили сўзлашув корпуси (Lancaster/IBM Spoken English Corpus, Corpus of Spoken American English, etc), диахроник корпус саналган Инглиз матнларининг Хельсинки корпуси, лингводидактик тадқиқотлар учун Инглиз тили ўрганувчиларининг халқаро корпуси, инглиз тилидаги корпусларнинг энг сўнгги авлоди сифатида Инглиз тили банки (The Cobuild Project / The Bank of English), Британия миллий корпуси, Халқаро инглиз тили корпуси (The International Corpus of English (ICE)), Америка миллий корпуси (American National Corpus) каби машҳур корпусларнинг мавжудлиги Миллий тил ва Давлат тили тараққиётида Миллий корпуснинг аҳамияти ва ўрнининг

қанчалық муҳимлигини кўрсатади. Дунё тили саналган инглиз, испан, хитой, араб, француз, рус, немис тилларининг миллий корпусидан ташқари поляк, поляк-украин, чех, словак, серб, хорват, босния, болгар, болгар-рус, македон, шотланд, идиш, нидерланд, нидерланд-француз, швед, дат, норвег, исланд, фарер, ўрта аср француз тилиси, испан, италян, португал, румин, литва, латиш, грек, шарқий армян, осетин, албан, хинд, цыган, хетт, фин, урал тиллари, эстон, вепс, венгер, удмурт, грузин, инглиз-грузин, лезгин, турк, татар, бошқирд, қрим-татар, қалмик, бурят, монгол, араб, иврит, амхар, япон, қадимий япон, баман, эсперанто тиллари корпуслар мавжуд. Дунё компьютер лингвистикасида Миллий тил корпусининг мавжудлигига тилнинг яшовчанлик ва компьютер тилига айланиш мезони сифатида қаралмоқда.

Юртимизда ўзбек тили корпуслари яратиш ишларини қоникарли деб бўлмайди. Бу соҳада олиб борилаётган илмий тадқиқот ишлари ўзбек корпус лингвистикасининг назарий асосларини ишлаб чиқиш, намуна сифатида бир-икки корпус лавҳаси яратиш жараёни босқичида турибди. Ўзбекистонда амалий тилшуносликнинг бир йўналиши саналган ўзбек корпус лингвистикани ривожлантириш уни илмий-назарий жиҳатдан тадқиқ этиш, компьютер ва корпус лингвистикаси маркази ва лабораторияси ташкил этиш ҳамда ушбу лабораторияга мутахассисларни бирлаштириш орқали амалга ошиши мумкин.

Корпуснинг ижтимоий аҳамияти кенг қамровлилиги, ундан лингвистик ва этно-психоллингвистик тадқиқотларда, она тили, адабиёт, хорижий тил таълимида, матнга автоматик ишлов бериш ва таржима дастурларини тузишда фойдаланиш мумкинлиги, корпуснинг тилшунос, таржимон, ўқитувчи, дастурчи, журналист, муҳаррир, умуман, қундалиқ фаолияти жараёнида “иш қуроли сўз бўлган” ҳар қандай киши учун замонавий ахборот воситасининг бир қўриниши эканлиги корпус яратишнинг долзарблигини белгилайди.

Масъул муҳаррир

А.С.ГРИБОЕДОВ ТИЛИ ЛУҒАТИ – улуғ ёзувчи асарлари тил хусусияти бўйича илк тўлиқ лексикографик тавсифий иш. Луғат иждодкорнинг нашр этилган барча асарлари лексик таркибини қамраб олган, асосий қисмини алифбо, частотали конкорданс ташкил этган. Конкорданс 12 000дан ортиқ луғат мақоласидан (150.000 сўз қўллаш ҳолати), луғат мақоласи эса лексема (бош сўз), унинг грамматик хусусияти (сўз туркуми, род, жонли/жонсизлик белгиси), барча матнлардаги миқдорий кўрсаткичи, зарурий ўринда қисқача изоҳдан иборат. Ҳар бир сўзшакл контекст билан берилиб, манбага ҳавола қилувчи манзилни (боб, бўлим, воқеа, ҳодиса ва б.) акс эттириб туради. Луғат конкорданс билан бирга “Горе от ума” асари матнига алифболи кўрсаткич, частотали луғат, насрий, шеърый нутқнинг частотали луғати, терс ҳамда грамматик луғат каби ёрдамчи луғат, кўрсаткичларни ҳам қамраб олган. Қўшимча луғатлар асосий луғатнинг турли тартибда жойлаштирилган (асосий шакл, частота, грамматик гуруҳ) таркибий қисми саналади. Улар фақат лексемалардан иборат, сўз қўллаш ҳолати киритилмаган. Бу луғат муаллиф тилини тадқиқ этувчи, рус адабиёти, адабий тили тарихи билан шуғулланувчи барча тилшуносларга мўлжалланган; ҳажми, кўлами, тузилиши жиҳатдан мукамал ишланганлиги сабабли кенг қамровли филологик тадқиқотлар: аниқ муаллиф ёки даврнинг тил хусусияти, турли асарлар тил хусусиятини киёсий таҳлил қилиш, маълум сўзларнинг рус тили тарихида қўлланиши, грамматик шаклини ўрганишда асқотади. (Қаранг: Поляков А.Е. Словарь языка А.С.Грибоедова. – М.: Языки славянской культуры, 2008.)

АГИОГРАФИК МАТНЛАР КОРПУСИ (СКАТ)¹ – тузиш жараёни анча мураккаб корпус. Санкт-Петербург XV–XVII аср агиографик матнлари корпуси (СКАТ)² – шундай корпус намунаси. СКАТ – Санкт-Петербург филология факультети математик лингвистика кафедраси томонидан тузилган қадимги рус агиографик адабий ёдгорликлари матни электрон

¹ <http://project.phil.pu.ru/skat>.

² <http://project.phil.pu.ru/skat>.

корпуси. Агиографик асарлар (Агиография (грек. агіос “муқаддас” ва γράφο “ёзаман”) – диний ва тарихий-черков муқаддас ёдгорликларини ўрганувчи мустақил фан соҳаси, диний турмуш тарзи лингвистик, диний-тарихий, тарихий, ижтимоий-маданий ва адабий нуқтаи назардан ўрганилади.) тили кўп жиҳатдан XV-XVII аср рус адабий тили характери, тақдирини белгилаб берган. Бу тилни акс эттиришнинг биринчи усули – ўша давр турмушига оид рус матнлари корпусини тузиш. Натижада, катта географик территорияда турли даврда асос солинган рус агиографик ёдгорликлари тили намоён бўлади. 2011 йилда корпус ҳажми 500.000 сўзни ташкил этган эди.

АМЕРИКА МИЛЛИЙ КОРПУСИ³ – “American National Corpus”нинг дастлабки кўриниши(парчаси); 2003 йилда ишлаб чиқилган. 100 миллион сўз қўллаш ҳолатини қамраб олиши режалаштирилган. Ҳозирча корпус ҳажми 11 миллионга етган, XML форматда; фойдаланиш пуллик.

АНАФОРИК РАЗМЕТКА – “ишора сўзлар”га ёзиладиган изох. Матнга ишлов беришда катта қийинчилик туғдирадиган туркум – олмош, чунки олмош матндаги қайси сўзга ишора қилишига қараб турли маънони билдиради. Ишора сўзнинг матндаги маъносини ажратиб олиш учун разметканинг алоҳида кўринишига эҳтиёж туғилади. Анафорик разметка шу хилдаги маълумотни кўшиш учун керак бўлади. Маъноси олмош билан ифодаланаётган сўз алоҳида кодланиб, кейинги ўринда шу сўзга ишора қилаётган олмош ёнига шу код бириктирилади. Натижада олмошнинг матндаги маъноси аниқланади ҳамда керакли тадқиқотларда корпуга асосланиш имконияти пайдо бўлади.

АРАЛАШ КОРПУС – тилнинг барча услубига тегишли бўлган матнлар акс этган миллий корпуслар. Замонавий катта ҳажмли корпуслар, одатда, аралаш – ёзма матн билан бирга оғзаки матнни ҳам қамраб олади. Ҳатто Британия миллий корпуси (BNC) матнларининг 10 фоизи оғзаки матнлар ёзуvidан иборат. Тил, асосан, оғзаки шаклда яшайди, унинг ёзма

³ <http://www.americannationalcorpus.org/>

шакли иккиламчи, шунинг учун оғзаки корпус жуда муҳим, ҳеч бўлмаганда, корпуснинг аралаш тури тил тадқиқи учун кўпроқ материал тақдим этади.

БИРИНЧИ АВЛОД КОРПУСЛАРИ – электрон шаклдаги разметкаланмаган корпус. Компьютернинг пайдо бўлиши ёзма (ноэлектрон) корпусларнинг йўқолишига олиб келмади. Ҳозир ҳам тилшуносликнинг бир қанча соҳаси учун картон картотека, сопол бўлақларига ёзилган материал билан ишлаш муҳим ҳисобланади. Айрим тил, матнлар тадқиқи учун мисол, фактни корпус типи ҳамда миқдорида ёйма шаклда ўрганиш муҳимлиги кузатилади. Корпус лингвистикасининг локомативи – инглиз тили. Бугунги кунда ҳеч ким инглиз тилидаги оддий матнларни йиғиш ҳақида ўйлаб ўтирмайди, чунки бу тилнинг барча варианты учун жуда катта ҳажмли, чуқур грамматик таҳлилга эга (аннотацияланган) корпуслар мавжуд. “Аннотацияланган корпус” давригача корпуслар аниқ мақсадли тўпланган, системалаштирилган матнлар йиғиндисидангина иборат эди. Бунга илк инглиз тили (Brown corpus, 1960 й.), рус тили (Упсальский корпус русских текстов, 1980 й.) корпуслари мисол бўла олади. Шу тариқа илк электрон корпуслар ўз аждодидан фақат маълумотни сақлаш шакли ҳамда сезиларли даражада ҳажмнинг кенгайиши билан фарқ қилди.

БОҒЛОВЧИ-ЛЕММА – боғловчи-лексема.

БРАУН КОРПУСИ (АҚШ) – 1 миллион сўз захирасига эга инглиз тили корпуси. Тўлиқ номи – Замонавий американча инглиз тилининг Браун университети стандарт корпуси (Brown University Standard Corpus of Present-Day American English), 1961-1964 йилларда тузилган. Корпус инглиз тилининг америка варианты ёзма матнларидан ташкил топган; 1 миллион сўз қўллаш мавжуд (бундай миқдор дастлабки корпусларнинг барчаси учун ҳажм мезони бўлган). Ўша даврда Ноам Хомский концепцияси ҳукмронлик қилганлиги сабабли Браун корпуси тузувчилари Нелсон Френсис, Генри Кусера учун корпус тузиш жараёни кўп қийинчилик туғдирган. Корпус 2000 сўзли 500 матндан иборат; қулай, пухта тузилишга эгаллиги ҳамда матн танлаш тамойилининг ҳар томонлама тўғри-лиги сабабли кейинчалик корпус

тадқиқотлари мезонлари яратилиши учун асос бўлган. Корпусдан тадқиқот учун эркин фойдаланиш анъанасини шу корпус бошлаб берган. Мазкур корпус асосида 1969 йилда “Америка мерос луғати”га (American Heritage Dictionary) асос солинган.

БРИТАНИЯ МИЛЛИЙ КОРПУСИ⁴ (British National Corpus (BNC)) – 100 миллион сўз қўллаш ҳолатидан иборат, инглиз тилини маълум жанрда эмас, балки тўлиқ қамраб олган. Бу корпуснинг ҳажмининг аниқлиги, чегараланганлиги билан Cobuild Projectдан фарқ қилади. 90 фоизини ёзма манба, 10 фоизини оғзаки матн ташкил этади. Корпусни ташкил этишда бир қанча ташкилотлар (Британия ҳукуматидан бошлаб) иштирок этган; мазкур иш 1995 йилда ниҳоясигига етган. Корпус 4124 матндан иборат, шундан 863таси оғзаки суҳбат, монологдан ёзиб олинган. Ҳар бир матн орфографик сегментланган, матндаги сўз автоматик тарзда туркумланган. Сўзни сегментлаш, классифика-циялаш Ланкастер университетида ишлаб чиқилган стохастик разметка дастури CLAWS асосида амалга оширилади. Барча матнлар энг оптимал усуллардан бири саналган SGML тили, TEI дастури ёрдамида разметкаланган. Корпусни тузишда матн танлашнинг янги усуллари қўлланилган, кўп даражали назорат тизимидан фойдаланилган.

ДИНАМИК КОРПУС – таркибининг тўлдирилиб борилишига кўра корпуснинг бир тури. Мақсади доимий равишда компьютер хотирасини матн билан тўлдириб, корпусни бойитиб бориш, асосий белгиси – янгиланиш, тўлдирилиш. Корпус лексикография ривож учун хизмат қилади, чунки тил доимий ҳаракатда, сўзлар янги-янги маънолар касб этиши, неологизм, архаизм, историзмларнинг пайдо бўлишини кузатиб бориш имконини беради. Динамик корпус катта даврнинг ёзма манбаларини акс эттириб, улар турли диахроник тадқиқотларни бажаришга мўлжалланган. Бунга Cobuild Project корпуси мисол бўла олади. Бундай корпуслар “очиқ корпус” деб аталади; уларнинг камчилиги шундаки, ҳар доим ҳам тилни бир хилда

⁴ <http://www.natcorp.ox.ac.uk/>

(адекват ҳолатда) акс эттирилмайди, матн танлашда аниқ мезонга амал қилинмайди.

ЁЗМА КОРПУС – оғзаки нутқ акс эттирилмаган (Браун корпуси, ЛОВ), фақат ёзма матнлардан иборат корпус.

ЁПИҚ КОРПУС – ўзига хос тор (хусусий) мақсадда тузилган, оммавий фойдаланишга мўлжалланмаган корпус.

ИДЕОГРАФИК РАЗМЕТКА – идеографик луғат ёки тезаурус яратишга асос бўлувчи разметка кўриниши. Идеографик луғатнинг аҳамияти корпусда идеографик разметкага эҳтиёж борлигини кўрсатади. Муаллиф асари мавзуларини тасниф қилиш натижасида бирликлар орасидаги мантиқий алоқадорлик асосида бирлашган сўзлар тўпламини акс эттирувчи идеографик луғат пайдо бўлади. А.П.Чехов насрий, драматик асарларининг мавзувий разметкаси хусусиятлари Е.В.Суровцева томонидан алоҳида тадқиқ этилган (Қаранг: Суровцева Е.В. Словарь тем писателя как разновидность идеографических словарей. На материале художественной прозы и драматургии А.П.Чехова. // Acta linguistica, Vol. 4 (2010), 2; Лапонина Л.В., Суровцева Е.В. Тематическая разметка в корпусе художественной прозы А.П.Чехова // Чехов и мировая культура: взгляд из XXI века. Тезисы докладов международной научной конференции (Москва, 29 января – 3 февраля 2010 года) / Сост. и ред. Р.Б.Ахметшин, М.О.Горячева, В.Б.Катаев (отв. ред.). М.: Издательство Московского университета, 2010.). Бундай мавзувий разметкани амалга ошириш учун унинг 583та насрий, 17та драматик асари матни танланган. Тахлилга олинган матнлар учун муаллифнинг мукамал асарлари тўплами (Чехов А.П. Полное собрание сочинений и писем. В 30 томах. М.: Наука, 1974-1983.) асос қилиб олинган. Идеографик разметка учун автоматик алгоритм – махсус дастур бўлмаганлиги сабабли разметка қўлда амалга оширилади. Чунки ҳар бир муаллиф асари мавзу доираси турлича бўлади. Аввал муаллифнинг барча асарлари матни таҳлил қилиниб, кейин умумий хулосага келинади. Мавзувий разметка матн мавзуси асосида лексик-семантик гуруҳлар шакллантиришни мақсад қилиб белгилайди. Матн танланиб, ундан маълум бир мавзу атрофида бирлашган сўзлар белгиланади. Сўзларни белгилашда матнни қайта айтиб бериш учун бутун матн мазмунини тўлиқ ифодалашга хизмат қилувчи сўз, иборалар ажратилади. Айнан шу матндан бошқа бир лексик-семантик

гурухга (ЛСГ) мансуб сўзлар алоҳида ранг билан белгиланади. Масалан, А.П.Чехов асарлари идеографик разметкаси учун ёзувчининг “Семиз ва орик” ҳикоясини объект сифатида танланган. Биринчи босқичда ҳикоядан *учрашув*, иккинчи босқичда *болалик*, *хотира* мазмуний гуруҳига кирувчи сўзлар алоҳида ранглар билан белгиланган, *дўстлик* мазмуний гуруҳига кирувчи сўзлар курсивда ажратилган. Шундай таҳлилдан кейин ЛСГ бир хил эмаслиги кўринган. Бу гуруҳлар *ҳолат*, *мавҳум тушунча*, *қаҳрамон тити* ёки *воқеа жойи* каби гуруҳлар бўлиши мумкин. Демак, битта ҳикояда бир неча ЛСГга кирувчи сўзлар гуруҳи аниқланган. Фақат битта ЛСГга кирувчи сўзлардан иборат асар (ҳикоя) кам учрайди. Масалан, “Баҳорни кутиш”да фақатгина *учрашув*, “Одамнинг қисқача анотомияси”да *илм*, “Парча”да *провинция* каби ЛСГлар мавжуд. Баъзи ҳикояларида эса ўнлаб ЛСГ аниқланган. Ҳикоячи, қаҳрамонда ҳаёл суриш, фикр юритиш, таҳлил қилишга мойиллик бўлган асарни мавзувий разметкаш осон кечади. Барча асарлар таҳлил қилингандан сўнг мавзулар иккига: макро, микромавзуга ажратилиб, баъзи мавзулар бир вақтнинг ўзиде икки макромавзуга тегишли бўлиб қолиши мумкин. Дейлик, *ҳайит* ЛСГи мазмунига кўра ҳам *дин*, ҳам *вақт* макромавзусига киритилиши мумкин. Корпусда бу типдаги разметканинг мавжудлиги рус тили (умуман олганда, барча тилни)ни хорижий тил сифатида ўқитишнинг самарадорлигини оширишда, корпус материалдан таълимий мақсадда фойдаланишда аҳамиятли. Чунки мавзулаштирилган матн хорижий тил таълими жараёнида ўқув материали тайёрлашда ахборот манбаи вазифасини бажаради; бу усулда разметкаланган корпус ёрдамида жаҳон муаллифлик картинасини кузатиш қулайлиги ҳам пайдо бўлади. Идеографик разметкали корпус ёрдамида бир муаллиф асарларини яхлит ҳолда ғоявий-бадий таҳлил қилишда, у ёки бу мавзунини қайси усул ва восита билан ифодалаганлигини ўрганишда ҳам асқотади. Масалан, А.Қаҳорнинг инсон, маънавият мавзуини кўтарилган барча асарларини бир призмадан ўтказиш орқали у ёки бу тил ифода воситаларини қўллаш маҳоратини ҳам ўрганиш мумкин. Шунингдек, стилистик, лингво-

поэтик, когнитив, прагматик, лингвомаданий, концептологик, дискурсив нуктаи назардан ўрганишда қўл келади.

ИККИНЧИ АВЛЮД КОРПУСЛАРИ – электрон шаклдаги разметкаланган корпус. Замонавий корпус у ёки бу тил ресурсини қай даражада ифодалашига, сифати, миқдорига қараб матнларнинг оддий электрон коллекцияси ёки чуқур аннотацияланган корпус каби турга бўлинади (Қаранг: История корпусной лингвистики. (Электрон ресурс) - <https://www.myfilology.ru/177/istoriya-korpusnoj-lingvistiki/>). Замонавий тилшунослик қўп ҳолларда матнни электрон шаклда ишлатиш эҳтиёжини сезади, бу эса автоматик равишда “корпуснинг биринчи авлоди”ни ташкил этади. Шунинг учун эски корпуслар “ўлмаиди”, истемолдан чиқиб кетмаиди, балки тўлдирилиб, чуқурлашиб, аннотацияга (турли лингвистик таҳлил) эга бўлиб бораверади. Ўтган асрнинг 90-йилларига келиб сақлаш, фойдаланиш технологияси 100 миллиондан ортиқ сўздан иборат корпусларни тузиш имконияти пайдо бўлди. Инглиз тили банки (The Cobuild Project/The Bank of English)ни тузиш 1980 йилда бошланган. Collins нашриёти янги луғат яратиш мақсадида корпусни тузишга киришган. 1990 йилда Бирмингем университети инглиз тили факультети, Collins нашриётининг бу борадаги саъй-ҳаракатларини “The Bank of English” номи остида бирлаштириш ҳаракати бошланган, натижада ушбу электрон маҳсулот мониторинг корпус сифатида шаклланган. Бу ҳақда лойиҳа муаллифи Жон Синклер: “Мониторинг корпус доимий ўзгаришда бўлган, аниқ ҳажмга эга бўлмаган, тилнинг кенг қўламли окимидан иборат корпусдир. Бу оқим лингвистик маълумотларни ажратиб оладиган филтрдан ўтади”, - деб ёзади. 1997 йилда қарийб 300 миллион сўз қўллаш ҳолати бўлса, 2005 йилда бу рақам 525 миллионга етди. Ҳар ойда корпусга 2,5 миллион янги бирикмалар қўшилади. Корпуснинг 25 фоизини оғзаки нутқ, 75 фоизини ёзма нутқ ташкил этади. Интернетда корпуснинг 56 миллион қўлланиш ҳолатидан иборат фрагменти⁵дан фойдаланиш мумкин.

⁵ <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>

ИЛЛЮСТРАТИВ КОРПУС – тадқиқот натижаси сифатида пайдо бўладиган корпус; унинг вазифаси фақат янги факт топиш эмас, балки у ёки бу тилга оид фактни аввалроқ бошқа усулда топилган мисолларни далиллашда лингвистик манба сифатида хизмат қилади. Иллюстратив корпуснинг типик намунаси маълум бир сўзнинг аниқ матндаги маъносини изоҳлашга мўлжалланган “Путеводитель по дискурсивным словам русского языка” корпусидир.

ИНКОР ҚИЛУВЧИ ФИЛЬТР (N-filter) – сўров билан мос келувчи сатр конкорданс рўйхатидан ўчирилади.

ИНФОРМАЦИОН СЎРОВ – маълум ахборотга бўлган эҳтиёжни сўз орқали ифодалаш усули (Қаранг: Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. работа с системой DDC.// Языковая инженерия: в поиске смыслов. (электрон ресурс: <https://docplayer.ru>). Сўров ўзининг предметлик белгиси, шаклий мундарижаси асосида таҳлил этилади, корпусда ишловчи дастурнинг сўров тили терминлар лугати асосида изоҳланади. Қидирув амалиёти корпусдаги ҳар бир элемент билан навбатма-навбат қиёслаш, уларнинг мослик даражасини аниқлаш билан яқунланади. Умуман олганда, сўров тили модели, асосан, қидирув элементи (информацион эҳтиёжни ифодаловчи атама), матн қидирув элементини нормаллаштирувчи восита, оператор йўналтирувчиси (конъюнкция, дизъюнкция, инкор), чизикли грамматика воситаси (оралиқ масофа оператори, позиция оператори) каби таркибий қисмларни ўз ичига олади. Қидирувнинг қўшимча шартлари сирасига корпуснинг алоҳида майдонларидаги қидируви (масалан, теглар ичида), қидирув майдонининг чегараланиши (аниқ бир муаллиф асари, ҳужжатни яратиш санаси ва х. бўйича), топилган натижаларни саралашга қўйиладиган талаб, қидирув натижасини ифодалаш шаклига қўйилган талаб, топилган натижаларнинг кўриниши, топилган ҳужжат миқдори каби омилларни киритиш мумкин. Масалан, нисбатан самарали, яхши ишлайдиган корпус менежери Bonito/Manatee1нинг қидирув тилидир. Лингвистик корпус билан ишлаш жараёнида кўп ишлатиладиган қидирув тили элементи ҳамда қидирув бўйича мисоллар изоҳини билиш керак. Bonito – менежер номи,

Manatee – корпус таг-тизимини бошқарувчи дастурий таъминот. Bonito корпус менежери корпус матни билан ишлашни тўлиқ қамраб олган дастурлар йиғиндиси; бу тизим икки қисмдан: сервер ҳамда график фойдаланиш интерфейси (GUI – graphical user interface)дан иборат. Ушбу интерфейс Чехиянинг Масарика университети информатика факультетида П.Рыхли ва NLPlab (Natural Language Processing Laboratory) гуруҳи томонидан тузилган. Тизим иши билан таништириш учун инглиз матнларининг SUSANNE (Surface and Underlying Structural Analysis of Natural English) корпуси⁶-дан фойдаланилган. Ушбу корпус Буюк Британиянинг Саксесс университетида яратилган, инглиз тилининг америка вариантыни ифодаловчи Браун корпусининг 130.000 сўзини ўзида мужассамлаштирган, SUSANNE схемасига мувофиқ разметкаланган. Bonito тизимининг асосий хусусияти сўров тилининг алоҳида атрибутлар кидирувига (сўзшакл, лемма, tag) эгаллиги, одатий ифодаларнинг қўлланиши, мантикий оператори, структурани ишлаб чиқиш воситасининг (gap чегараси ва б.) мавжудлиги, мураккаб сўровни тезлик билан ишлаб чиқа олиши билан характерланади. Конкорданс рўйхати ўзида фойдаланувчи сўрови тарихи, сўзшаклнинг морфологик хусусиятини кўриб чиқиш, леммани акс эттириши билан ажралиб туради. Конкорданс устида ишлаш жараёнида рўйхатни файлда сақлаб қолиш, рўйхатни босма шаклда чоп этиш, контекстдаги калит сўз асосида саралаш, контекстни чегараланмаган даражада кенгайтириш, филтрлаш (конкорданснинг кераксиз қисмини ўчириш), “такрор”ни ўчириш каби амалларни бажариш мумкин. Частотали тақсимот нуктаи назаридан корпус ёки контекстдаги сўз ёхуд бошқа атрибутлар частотаси, чегараланмаган даражада гурухлашни ҳам амалга ошира олади. Бундан ташқари, кодлаштириш танлови ҳам мавжуд, фойдаланувчи хусусий корпусини яратиши мумкин. Шунингдек, тегларни ихтиёрий танлай олиш, бошқа тилни қўшиш имконияти мавжудлиги каби хусусиятга ҳам эга. Корпус менежерининг асосий вазифаси фойдаланувчи сўровига жавоб

⁶ <http://www.grsampson.net>.

беришдан иборат экан, сўров ва уни амалга оширувчи буйруқ аҳамиятли. Юқорида тилга олинган корпус менежеридида фойдаланувчи қидирув тизими тилига мос равишда қидирув ойнасига ўзининг шахсий саволини киритиши ёки шаблон сўровдан фойдаланиши мумкин.

КВИРК КОРПУСИ (Survey of English Usage) – 1000000 сўз қўллаш ҳолати мавжуд, даставвал ҳар бири 17 сатрли матндан иборат 4x6 дюм ҳажмли 1000000 карточкадан иборат бўлган. Ушбу корпус энг сўнгги ноэлектрон корпус, уни тузиш 25 йил давом этган ҳамда 1989 йилда ниҳоясига етган, чунки бу даврда компьютер технологияси илгарилаб кетган эди. Зудлик билан корпусни электрон шаклга келтириш зарурати туғилди. Мазкур корпус ҳозир Лондон университетида сақланади; ўша ерда ундан фойдаланиш мумкин (Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf).

КЕНГАЙТИРИЛГАН ҚИДИРУВ ОЙНАСИ – асарлар, жанрлар, қаҳрамонлар, асар ёзилган йил кесимида қидирувни амалга оширишга ёрдам берадувчи ойна. Бу ойнанинг ишлаши корпус бирликлари(матн)га ёзилган экстралингвистик аннотация билан боғлиқ, яъни бу ойналар фаол ишлаши учун корпуснинг ҳар бир матнига асар жанри, ёзилган йили, жойи каби маълумотлар ёзилиши лозим. “Конкорданс” ойнаси Абдулла Қаҳҳор асарлари частотали луғатини яратишга ёки корпусдан частотали луғат сифатида фойдаланишга имкон яратадиган таркибий қисм. Луғатчилик тажрибасида А.Қаҳҳор асарлари асосида бир неча луғатлар (Қаримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати.–Тошкент: ЎзМЭ, 2007.– 434 б.) тузилганлиги маълум. Корпуснинг дастурий таъминоти мавжуд бўлса, А.Қаҳҳор асарлари частотали, идеографик ва б. луғатларининг янги авлодини яратиш имкони туғилади.

КОНКОРДАНС – қидириладиган бирликнинг контекстдаги лексик боғланиш ҳолатида манбага илова қилинган контекстлар рўйхати.

КОНКОРДАНС ҚИДИРУВИ – корпус интерфейсидаги сўз кидируви ойнаси; корпусда мавжуд сўзлар частотасини кўрсатади, уларнинг қўлланиш контекстига ишора қилади.

КОРПУС АННОТАЦИЯСИ – турли лингвистик топширик/амални бажариш учун матнга лингвистик ва экстралингвистик қўшимча маълумот билан ишлов бериш. Матн аннотацияси – матн ва унинг компонентларига берилган махсус изоҳ (масалан, сўз туркуми ҳақида маълумот). Бу изоҳ матн разметкаси (белги, ишора) деб ҳам аталади. Мукамал аннотация тадқиқотчига қидирилаётган сўз, сўзшакл ва контрукцияни топишни осонлаштиради. Чунки оддий электрон матнда унинг характерини очувчи алоҳида маълумот (аннотация) мавжуд эмас. Бизга шунчаки сўз керак бўлса, буни одатий матн муҳаррири ҳам топиб бера олади. Лекин сўзнинг грамматик маъноси, тузилишини “тушунадиган” матн билан ишлаш жуда кўп афзаллик ва қулайликка эга бўлади. Фақат лексемани эмас, балки сўзшаклни қидириш керак бўлса, бундай дастурий таъминот (корпус) тадқиқотчи ёки фойдаланувчига жуда катта ёрдам бериши мумкин. Аннотациянинг мукамал, тўлиқроқ бўлиши корпус аҳамиятини белгиловчи асосий омилдир. Катта ҳажмли матн(айтайлик, 50-100 саҳифа)дан ўтган замонни билдирувчи грамматик шаклларни топиш топширигини қўямиз. Бундай вазифани оддий матн муҳаррири, табиийки, бажара олмайди. Берилган грамматик шаклни топиш фақат олдиндан ишлов берилган матндагина бажарилиши мумкин, акс ҳолда, қидирув фақат қўлда бажарилади. Бу эса фойдаланувчидан кўп меҳнат ҳамда вақтни талаб қилади (Қаранг: Кутузов А.Б. Курс «Корпусная лингвистика». – Москва, 2008. – 26 с.). Мукамал ишлов берилган, катта ҳажмли, турли жанрдаги матнларни қамраб олган корпус тадқиқотчи учун бебаҳо ёрдамчи ҳисобланади. Чунки тил материали ўрганилган ҳар қандай тадқиқот материалининг бойлиги, ишонарлилиги, аниқ ва мукамаллиги билан баҳоланади. Агар тадқиқотчи ўз тадқиқоти учун мисолни топиш, уларни картотекага кўчириш (компьютер технологиялари ривожланишидан олдинги даврдаги ҳолат)га ойлаб, баъзан йиллаб вақт

сарфлаган бўлса, бугун дунё тил корпуслари ёрдамида санокли дақиқада зарур деб ҳисобланган юзлаб мисолни топиш, улар устида ишлаш имкониятига эга бўлди. Тил корпусларида бугунги кунда аннотациянинг морфологик, синтактик, семантик, акцентик каби бир неча кўриниши мавжуд. Яқин орада корпус лингвистикаси мутахассислари томонидан сўз ясалиш аннотациясининг киритилиши ҳам кутилмоқда.

КОРПУС ИНТЕРФЕЙСИ – фойдаланувчини корпуснинг бутун мазмуни билан таништирувчи, мундарижа вазифасини бажарувчи муҳим таркибий қисм. Муаллифлик корпуси интерфейси муаллиф ижодий мероси жанрий таснифини бир қарашдаёқ тушуниш учун яхши имконият, корпуснинг бошқа базалардан фарқи ҳам шунда. Форс-тожик мумтоз алломалари, замонавий адиблари корпуси мажмуасини ташкил қилган бош корпуснинг интерфейси ўзига хос кўринишга эга, бош корпус таркибига алоҳида муаллифлар корпуси⁷ бирлаштирилган. Интерфейсда муаллифлик корпуси давр бўйича таснифланган, мумтоз, замонавий шоирлар корпусига ажратилган. Муаллифлик умумий корпусидан мумтоз адабиёт намоёндалари Рудакий, Фирдавсий, Саъдий Шерозий, Мавлоно Румий, Ҳофиз Шерозий ҳамда муосир адабиёт вакиллари Лойикали Шерали, Мўмин Қаноат, Ҳабибулло Файзуллолар корпуси жой олган. Ҳар бир муаллиф корпуси алоҳида интерфейсга эга. Интерфейсда муаллиф ижодий мероси таснифи ёки муаллиф ижодининг қандай қирраси ҳақидаги маълумотлар тизими тўғрисида ахборот кўринади.

КОРПУС ЛИНГВИСТИКАСИ – тилшуносликнинг корпус, унинг турлари, тарихи, таракқиёти, бугунги ҳолати, корпус тузиш тамойилларини ўрганувчи бўлими. Корпус лингвистикаси компьютер лингвистикасининг жуда тез ривожланётган соҳаси; ушбу фан жаҳоннинг етакчи олий таълим муассасаларида ўқитилади. Бу соҳа предмети корпус яратиш назарияси ва амалиёти бўлса, фан сифатида корпуснинг ўзига хослиги, дастурлаш (ва ҳ.к.) масалалари ўқитилади. Корпус лингвистикаси компьютер лингвисти-

⁷ www.sit.tj.

касининг таркибий қисми; тил корпусини яратиш, компьютер технологияси ёрдамида улардан фойдаланиш-нинг умумий назарияси ва амалиёти билан шуғулланади.

КОРПУС МАТЕРИАЛИ – корпус тақибдаги матнлар йиғиндиси. Мавжуд корпуслар таркибидаги матнлар нисбатида бадиий адабиёт ҳиссаси 40 фоизини ташкил этади. Бу таркибга тил хусусияти бадиий, публицистик услуб оралиғида, жонли тилни ўрганиш учун анча қулай саналган мемуар асарлар ҳам киради. Европа тиллари корпусларида бадиий адабиёт материали 20 фоизни ташкил этади. Замонавий ёзувчилар асарларининг тил хусусиятини ўрганишга бағишланган 20 дан ортиқ тадқиқот мажуд бўлса-да, улар ҳали тўлалигича мазкур муаммони ўрганиб бўлди, деб бўлмайди. Чунки маълум бир ёзувчи асарининг тил хусусиятидаги ўзгаришга ҳали тилдаги янги ҳодиса деб қараб бўлмайди. Шундай жараёнларни кузатиш, тадқиқ этишнинг энг қулай воситаси тил корпусидир. Шу боисдан ҳам бу борада матн мазмуни муҳим аҳамият касб этади. Корпус маълумотининг ишонарлилигини таъминловчи омил – унинг репрезента-тивлиги. Бу хусусиятга корпус ҳамжмининг катталиги, матнларнинг турли жанрга мансублигини таъминлаш билан эришиб бўлади. Нуткий фаолият, ҳақиқатан ҳам, ранг-баранг, турли шаклларда (оғзаки, ёзма, босма ва ҳ.) намоён бўлади ҳамда нуткий вокеланишларнинг чексизли-гини таъминлайди. XX асрнинг 60-йилларида яратилган корпуслар биринчи турга мансуб, улар универсалликка интилган. Репрезентатив корпус учун танланадиган матнлар 15 турдаги жанр(регистр)ни қамраб олиши, улардан бтадан 80тагача элементар белгиларга қўра танланиши лозим. Матнлар, одатда, матбуот: репортаж; матбуот: бош мақола; матбуот: обзор; диний матнлар; кўникма, машғулот, хобби; илмий-оммабоп адабиёт; беллетристика, биография, эссе; турли (бошқарув ҳужжатлари, корхона, ташкилот, ишлаб чиқариш ҳисоботи), илмий асар; бадиий адабиёт; мистика, детектив; илмий проза; саргузашт, мемуар; ишқий романлар; ҳажвий асарлар каби жанрларда бўлади.

КОРПУС МЕНЕЖЕРИ – лингвистик корпуснинг ажралмас, энг асосий қисми, унинг қидирув тизими, у матн ва лисоний бирликларни бошқарувчи система. Корпус менежери – корпус маълумотлари устида ишлаш учун мўлжалланган махсус қидирув тизими, у статистик маълумот ҳамда қидирув натижасини фойдаланувчига қулай кўринишда кўрсатиб берувчи дастурий таъминот ҳисобланади. Корпус фойдаланувчиси учун турли имконият ва қулайлик яратадиган, корпусдан зарурий ахборотни олишга ёрдамлашадиган бир неча махсус дастурий восита – конкорданс (қидирув тизимининг нисбатан оддий кўриниши)лар йиғиндиси. Қидирув натижаси, одатда, конкорданс шаклида намоён бўлади. Корпус менежерининг қидирув имконияти аниқ сўзшакли, сўзнинг *лемма* (корпусда лексема шундай аталади) шакли, бирикмали бўлинган ва бўлинмаган синтагма, морфологик белгиларга асосланган жамланма шаклини қамраб олади. Корпус менежерига қўйиладиган энг асосий талаб матннинг калит сўзлари рўйхати, тўлиқ конкорданс рўйхатни ярата олиш; фақат сўзни эмас, балки сўз бирикма сўровига ҳам жавоб бера олиш; шаблон асосида (мураккаб сўров) қидирувни амалга ошириш; олинган натижа (чиқарилган рўйхат)ни бир неча мезон асосида саралай олиш; сўзшаклга берилган сўровни чекланмаган миқдордаги контекстда акс эттириш; корпуснинг алоҳида элементлари бўйича статистик маълумот бера олиш; корпус разметкасидан келиб чиққан ҳолда лемма, сўзшаклнинг морфологик хусусияти ҳамда метаахборот (библиографик, типологик)ни тўлиқ ифода қилиш; натижаларни сақлаш, чоп этиш; файл, корпуснинг чекланмаган ҳажми билан ишлай олиш; қидирувни тез амалга ошириш, натижаларни чиқариш; турли матн форматларини (txt, doc, rtf, html, xml ва б.) ўқий олиш, шу формат билан ишлаш; малакали ҳамда янги фойдаланувчи учун бирдек қулай бўлишидир (Қаранг: Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. работа с системой DDC.// Языковая инженерия: в поиске смыслов. (электрон ресурс: <https://docplayer.ru>). SARA, XAIRA (BNC), Manatee/Bonito, SQR, DDC каби корпус менежерлари нисбатан универсал саналади, корпус тузиш тажрибасида шу менежерлардан кўпроқ фойдаланилади. Корпус

маълумотини қайта ишлаш учун маълумотлар базасини бошқариш ёки қидирувнинг янги тизимини ишлаб чиқиш мумкин.

КОРПУС РАЗМЕТКАСИ – корпус матнларига бериладиган лингвистик ҳамда экстралингвистик изох.

КОРПУС ТУЗИШНИНГ ТЕХНОЛОГИК ЖАРАЁНИ – бир неча босқичларни қамраб олган жараён⁸: Белгиланган манбага мувофиқ ҳолда матннинг корпусга киришини таъминлаш (1). Матнни автоматик ўқишиш шаклида қайта ишлаш (2): корпусга киритиладиган электрон шаклдаги матн турли усул билан олинган бўлиши мумкин: қўлда терилган, сканерланган, муаллифлик нусхаси, ҳада, айирбошлаш, Интернет, нашриётлар томонидан корпус тузувчисига бериладиган оригинал-макетлар. Таҳлил, матнга дастлабки ишлов бериш (3) босқичида турли манбалардан қабул қилинган матнлар филологик текширув, таҳрирдан ўтади. Конверсиялаш, графематик таҳлил (4). Баъзи матнлар қайта кодлаштириш жараёни амалга ошадиган илк машина ишловидан қайта-қайта ўтади, номатний қисмлар (расм, жадвал) ўчирилади ёки ўзгартирилади. Матндаги бўғин кўчириш, чегаралар (MS-DOS матнларида) бекор қилинади, тире, бошқа белгилар бир хиллигига эришилади. Графематик таҳлил корпусга кирувчи матнни қисмга (сўз, боғловчи) ажратиш, номатний элементни ўчириш каби амалларни бажаришдан иборат (5). Ностандарт (нолексик) элементни белгилаш, расмийлаштириш, махсус матний элементни (қисқартма асосида ёзилган ном (исм, фамилия), бошқа алифбода ёзилган ўзлашма лексема, расмга берилган ном, изох, зарварақ, адабиётлар рўйхати ва б.) бир хил мезон асосида қайта кўриб чиқиш (6). Албатта, бу амаллар автоматик равишда матн муҳаррири томонидан бажарилади.

КОРПУС ФЙДАЛАНУВЧИСИ – корпусдан фойдаланадиган мутахассислар: тадқиқотчи, филолог, журналист, ўқитувчи, талаба, мактаб ўқувчиси, программист.

⁸ <http://www.uni-koeln.de/phil-fak/englisch/bald/corpora>.
http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/history.html.

КОРПУС ФРАГМЕНТИ – яратилиш боскичида бўлган, тугалланмаган, аммо маълум кисмидан фойдаланиш мумкин бўлган корпус кисми.

КОРПУС ҲАЖМИ – корпус матнларининг XML тилидаги ҳажми, Мбда ҳисобланади.

КОРПУСНИ ЛОЙИҲАЛАШ – уни тузиш боскичи, кейинчалик такомиллаштириш йўлларини камраб олади. Корпус тушунчаси, тилшунос учун анъанавий картотекаларнинг янги шакли; улар XX асрга келиб компьютерлаштирилди, ундан оммавий фойдаланиш имконияти пайдо бўлди. Картотекаларнинг корпусга айланишида, албатта, Интернет тармоғи салмоқли аҳамият касб этди. Натижада, турли лингвистик тадқиқотлар олиб бориш имконини берувчи катта ҳажмли матнларнинг умумистеъмол варианти пайдо бўлди. Бу борада луғат, грамматикалар учун асос вазифасини ўтайдиган тил материалининг кўлами ҳамда баланси масаласи кун тартибига чиқиб, хусусан, миллий корпуслар яратиш жараёнида кўндаланг турди. Корпуснинг репрезентативлик масаласи матнлар етарлилиги, хилма-хиллиги билан ҳал этилди. Корпуснинг жанрий-мавзувий тузилиши кўриб чиқиладиганда корпус матни сифатида қандай бирликнинг олинishi муаммосига алоҳида эътибор қаратиш лозим (Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. –Иркутск: ИГЛУ, 2011.). Масалан, газеталардаги кичик реклама матни алоҳида матн сифатида қараладими ёки уларни бир матнга бирлаштириш лозими? Газета мақоласи матн саналадими ёки газетанинг битта сонини яхлит матн сифатида баҳолаш керакми? Ҳар бир шеър битта матнми ёки шеърый тўпламни яхлит ҳолда киритиш керакми? Бир-бирига жавоб тарзида ёзилган, моҳиятан бир мавзу муҳокама қилинган нашр этилган мактублар битта матнми ёки алоҳида корпус бирлиги сифатида ёндашиш лозими? Бу саволларга тузувчи корпуснинг тури, кейинчалик бажарадиган вазифасидан келиб чиқиб жавоб беради. Миллий корпус ёки махсус корпус эканлигига қараб корпус бирлиги белгиланади. Корпусни лойиҳалаштириш жараёнининг муҳим жиҳати – хроноло-гия масаласи. Масалан, тилнинг

замонавий корпуси деганда нима тушунилиши лозим? Турли жанрларда корпуснинг хронологик чегараси турлича бўлиши табиий. Корпус кенг омма фойдаланиши ҳамда хилма-хил топшириқлар бажарилиши учун (жумладан, бошқа графика асосида рус тилида ёзилган матнларни ўрганиш учун) тузилади. Корпусда матннинг бошланғич шаклидан қандай қисм олиниб, нималар чиқариб ташланиши – яна бир эътиборга молик масала. Матн таркибида мавжуд бўлган расмлар тил материалига тегишли бўлмаганлиги учун корпус таркибига кирган матндан чиқариб ташлаш, жадвалларни корпусга мослаб қайта ишлаш ҳам муҳим. Улар матннинг мазмунини ифодалашда аҳамиятли, лекин корпус таркибида қолдирилса, разметкашда қийинчилик туғдиради. Цитата, кўчирма гаплар, ўзлашма бирлик (атама)лар, ўлчов бирликлари ҳам алоҳида эътибор талаб қилади. Санаб ўтилган масалалар лойиҳалаштириш босқичида маълум принцип асосида ҳал этилса, айрими корпус тузиш жараёни, корпусдан фойдаланишда ҳал этилади. Шу билан бирга, корпусни ишга туширишдан олдин фойдаланувчи билан қайтар алоқани ҳам назарда тутиш лозим.

КЎМАКЧИ-ЛЕММА – кўмакчи-лексема.

КЎП ТИЛЛИ КОРПУС – икки ёки уч тили корпус. Корпус мутахассиси, асосан, таржимонларни доим бир неча тилли корпус яратиш қизиқтириб келган. Илк авлод корпуси давридаёқ инглиз, фин, француз, немис, грек, норвег, испан, швед ва бошқа тиллар учун икки тилли корпус пайдо бўла бошлади. Бундай корпус bitexts деб ҳам аталади. Корпусни икки, уч, тўрт, ундан ортиқ тилли қилишга ҳеч қандай тўсик йўқ. Соҳанинг бошқа бир мутахассислари эса параллеллик нуқтаи назаридан корпусни бир, икки, кўп тилли (*Қаранг:* Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011) каби турларга бўлишади. Бир тилли корпусда тил варианты, шевалар бир-бирига қарама-қарши қўйилади. Масалан, инглиз тили она тили ва хорижий тил сифатида алоҳида ўрганилади. Янги технологиялар пайдо бўлгунча инглиз тили хорижий тил сифатида ўрганилмай қолганлиги ҳақида фикрлар мавжуд. Икки ҳамда кўп тилли корпус бир мавзу доирасида турли

тилда ёзилган матндан ташкил топади. Масалан, маълум илмий муаммо борасида бир неча давлатларда турли тилда ўтказилган конференция материалларини қамраб олиши мумкин. Бундай корпус, одатда, атама билан ишлайди ҳамда таржимон томонидан фойдаланилади. Икки ва кўп тилли корпуснинг яна бир кўринишида оригинал матн ҳамда таржима матни мавжуд бўлади. Корпуснинг ушбу тури қиёсий чоғиштира тадқиқот олиб боришда, таржима назарияси ҳамда компьютер таржимасини ўрганишда муҳим, фойдали манба вазифасини бажаради. Икки тилли корпуснинг 2 тури мавжуд (*Қаранг:* Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурсы) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf). 1. Мосланмаган корпус (“қиёсий” корпус) – матнни унинг таржимаси билан тўғрилайдиган, унинг таржимада қайси бирликка тўғри келишини кўрсатадиган корпус. Мослаш автоматик равишда ёки қўлда бажарилиши мумкин. Биринчи усул осон, лекин хатолари кўп. Масалан, таржима жараёнида содда гап қўшма гап шаклида берилиши мумкин. Бундай пайтда қайси қурилиш оригинал эканлигини аниқлаш қийин бўлади. Кўп тилли “мосланган” корпус намунаси сифатида Европа Иттифоқининг Acquis Communautaire маълумотлар базасини келтиришимиз мумкин. Параллел корпуснинг қиммати унинг ҳажми, тилларнинг миқдори билан белгиланади. Acquis Communautaire – дунёдаги энг катта параллел корпус. Унинг икки муҳим жиҳати корпусдан фойдаланишнинг бепуллиги, мальти-эстон, словен-фин каби кам учрайдиган тиллар жуфтлигининг мавжудлигида. Бундай корпусдан қуйидаги мақсадларда фойдаланиш мумкин: типик таржима усуллари, трансформацияни юзага келтириш (1); автоматик таржима тизими статистикасини ўрганиш (2); бир ва кўп тилли луғатлар яратиш (3); маълумотни сақлаш, узатиш дастурларини ўрганиш, баҳолаш (4); таржиманинг тўғрилигини автоматик тарзда текшириш (5); эквивалент танлаш имкони кенглиги орқали таржимон меҳнатини осонлаштириш (6). Иккинчи хилдаги корпус “таржима корпуслари” (translation corpora) айни бир фикрнинг турли тилдаги ифодасини ўрганишда муҳим. Икки тилли корпус

тилга оид бир қанча қузатув натижасини (таржима назарияси тадқиқи ёки таржима маҳсулини баҳолаши) бера олганлиги сабабли тилшунослик йўналишини танлаган талаба учун кичик тадқиқотлар бажаришнинг жуда фойдали воситаси ҳамдир.

ҚАХРАМОН НУТҚИ ҚИДИРУВИ – муаллифлик корпусларида асар қаҳрамонлари нутқи қидирувини юзага чиқарувчи тизим.

ЛАВҲАЛИ (ФРАГМЕНТ МАТНЛИ) КОРПУС – матн ҳажмига кўра корпус тури. Браун, Ланкастер-Осло-Берген корпусига қўйилган дастлабки талаб қатъий равишда ҳар бири 2000 сўз(сўз қўллаш ҳолати)дан иборат матнлардан тузилиши лозим бўлган. Табиийки, бу мезонга қатъий амал қилувчи корпус йўқ. Шунинг учун, бундай корпус фрагмент, яъни лавҳа матнли ҳисобланади.

ЛАНКАСТЕР-ОСЛО/БЕРГЕН (ЛОБ) КОРПУСИ (Lancaster-Oslo/Bergen (LOB) Corpus) – 1970-78 йилларда Ланкастер, Осло университети ҳамда Берген илмий маркази лойиҳаси⁹ асосида тузилган корпус. Инглиз тилининг Британия варианты асосида 1 миллион сўз қўллаш ҳолатидан ташкил топган, тузилиш жиҳатдан Браун корпусига ўхшайди. Паст частотали бирликларни (улар тилда жуда катта фоизни ташкил этади) тадқиқ қилишда 1 миллион сўз етарли эмас. Шундай мукамал корпус мавжудлиги сабабли Браун, ЛОБ корпусларига асосланган жуда кўп сифатли, қизиқарли тадқиқотлар амалга оширилган.

ЛЕММА – корпус матнлардаги лексаманинг корпус лингвистикасида номланиши, лексема.

ЛЕММАТИЗАЦИЯ – сўзшаклнинг лемма шаклини ажратиб кўрсатиш. Сўзшаклни разметкалашнинг биринчи босқичи –леммалаш, яъни сўзшаклнинг лексема шаклини келтириш. Флексив тилларни разметкалашда энг қийин босқич – леммалаш (лемматизация), яъни сўзнинг лексема шаклини сўзшаклга тег сифатида бириктириш. Чунки флексив тилларда сўзшаклдаги грамматик маъно сўз ўзагига қоришиб кетган бўлади. Флексив

⁹ <http://khnt.hit.uib.no/icame/manuals/lobman/>

тиллардан фаркли ўларок, агглютинатив тилда леммалаштириш жараёни анча осон. Сўзшаклнинг грамматик шаклсиз қисми (ўзак ёки негиз) леммага тенг. Разметкада лемма <*> белгиси ичида берилади. Барча сўз туркумларида леммалаш шу асосда, яъни “сўзнинг ўзак-негиз қисми леммага тенг” тамойилига асосланса, феъл туркумида феъл-лемма II шахс буйруқ-истак майли шаклида берилади. Луғат мақолаларида феъл ҳаракат номи шаклида берилади: <бормоқ>. Аммо бу шакл корпус учун мос эмас, чунки корпусдаги матнда сўзнинг <бормоқ> шакли эмас, <бор> шакли қидирилади. Шу сабабли феъл-лемма **ўқитди** <ўқи>, **бўлмади** <бўл>, **кўрсатди** <кўр>, **олди** <ол> шаклида берилади. Феъл-иборанинг лемма шаклини “Фразеологик луғат”даги каби *-моқ* қўшимчаси билан бериш тўғрироқ, чунки *кўзи тиниб* каби гап қурилишли иборани <кўзи тин> шаклида леммалашда ҳеч қандай маъно йўқ. Фойдаланувчи ибора қидирганда морфологик эмас, балки семантик разметка натижасидан фойдаланади. Шунга асосланиб, иборанинг **кўзи тиниб** <кўзи тинмоқ>, **боши айланадиган** <боши айланмоқ> тарзида разметкаланиши мақсадга мувофиқ. Бошқа барча сўз туркумлари <*> белгиси ичида ўзак-негиз шаклида лексема ҳолида ёзилади: **сахарга** <сахар>. Албатта, сўз туркумларида ўзакнинг соддалашиш масаласи ҳам бор. Соддалашиш жараёнида бўлган ўзак шу ҳолида лемма шаклига келтирилади. От сўз туркумидаги луғавий шакл леммага киритилмайди, чунки бу шакл лексик маънога таъсир қилмайди, разметкада **қизча** <қиз> кўринишида акс этади. Айрим сўзлар бундан мустасно. Чунончи, *боғча, шолча, кўрпача* сўзлари шу ҳолида леммага тенг, чунки улардаги *-ча* кичрайтириш шакли эмас. Разметкада жараёнида ҳар бир сўзшаклга 5 тадан 10 тагача, баъзан ундан ҳам кўпроқ морфологик тег (изох) ёзиш талаб қилинади. Тегларни эса лингвистик моделлаштириш мақсадга мувофиқ, чунки лингвистик моделда морфологик тег шартли қисқартма шаклини олади. Ҳар бир сўз туркумини разметкадаш учун махсус лингвистик модел шакллари ишлаб чиқилади.

ЛИНГВИСТИК РАЗМЕТКА – матн таркибидаги сўз ёки гапга ёзилган лингвистик изох. Унинг ҳам ўз навбатида морфологик, синтактик, семантик, просодик разметка каби бир қанча кўринишлари мавжуд.

ЛОНДОН-ЛУНД КОРПУСИ¹⁰ – (London-Lund Corpus (LLC) инглиз тилининг оғзаки нутқ корпуси, 1975 йилда тузилган. Орфографик транскрипция, фонетик, просодик (урғу ва урғусиз, узун ҳамда қисқа бўғинларни талаффуз қилишга оид) разметкаси мавжуд, 500000 сўз қўллаш ҳолатини камраб олган. Корпус Лондон университети коллежида дастлаб қоғоз вариантда тайёрланиб, кейинчалик швед шаҳарларидан бири Лунд тилшунослари томонидан электрон кўринишга келтирилган.

МАНБАНИ САРАЛАШ – корпус тузишдаги асосий технологик жараёнлардан бири. Корпуснинг аҳамияти унинг бир тилдаги кенг қўламли матнларни бир жойга йиғиб, тартиб берилганида эмас, шу сабабли уни тузишда бир неча мезон асосида иш кўрилади. Корпус материални саралашда корпуснинг асосий бирлиги нимадан иборатлиги, унинг ҳажми қандай бўлиши (унда қанча сўз бўлгани маъқул), ёзма матн қайси манбага асосланиши, қанча миқдорда бўлиши, унга кирувчи матн тилнинг қайси соҳасига тегишли бўлиши каби масалаларга ечим топилади. Ушбу саволнинг илк жавоби 1965-80 йиллар оралиғида Р.Г.Пиатровский ва унинг шогирдлари томонидан берилган. Улар частотали луғат, лингвостатистик тадқиқот ўтказиш учун матн танлаш тамойилларини тузиб чиқишган. Бу муаммо Л.Н.Засорина таҳрири остидаги частотали луғат (Частотный словарь русского языка. (Под ред. Л.Н.Засорина.)– М.: Русский язык, 1977.– 936 с.) сўз бошисида ҳам кўтарилган. Ўшанда илк маротаба матн танлашнинг статистик усули, ҳажми, миқдори каби омиллар санаб ўтилган. Корпуснинг асосий бирликлари сўзшакл, ўзак (негиз, лемма) ва гап. Тузиладиган корпус ҳажми корпуснинг мақсадидан келиб чиқиб белгиланади. Агар у ҳарф, ҳарфий бирикма, товуш, дифтонгларни тадқиқ этишни мақсад қилган бўлса, у қадар катта бўлиши шарт эмас. Матннинг лексик бирликлари, морфологик ҳодиса, синтактик,

¹⁰ <http://www.ucl.ac.uk/english-usage>

услубий хослигини тадқиқ этиш мақсадида тузилса, катта ҳажм талаб этилади (Қаранг: Шаров С.А. Представительный корпус русского языка в контексте мирового опыта. (Электрон ресурс. <https://lamb.viniti.ru>). Саралаш жараёнида қайси жанрга оид матнни (наср, драма, шеърят, илмий матн, газета, журнал материали ва ҳ.) танлаш, матннинг қайси даврни (замонавий, 10 йиллик, 50 йиллик ва мумтоз матн) қамраб олиши, матн фақат адабий тилда бўлиши ёхуд бошқа манбалар ҳам кириши каби масалалар ҳам муҳим аҳамият касб этади. Корпус тузувчиси бу жараёнда, албатта, тилшунос, лингвостатистика мутахассиси ёки анкета методига мурожаат этади. Корпус тузиш жараёнида муаллиф ўз тажрибасига таянган ҳолда корпуснинг умумий ҳажми, матннинг нашр вақти, матн сони, элементар танлов ҳажми, танланадиган жанр хили ва турини асосий омил ҳисоблайди. Сўронома усули “Америка мерос корпуси” (“The American Heritage Intermediate Corpus”) тузувчилари томонидан қўлланган. 5 миллион сўзшакл ҳамда инглиз тилида 22 турдаги болалар, ўсмирлар жанрига оид матнлар киритилган. АҚШнинг 221 та мактабига қандай матнни танлаш мақсадга мувофиқлигини аниқловчи сўронома юборилган. Сўронома натижаси ўрганилгач, 19 000 номдаги китоблар рўйхати тузилган. Бу асосда ҳар бири 500 сўзшаклдан иборат 1045 матн танлаб олинган.

МАТНЛАР КОРПУСИ – стандартланган, бир хир шаклга келтирилган; матнлари лингвистик воқеликнинг маълум қисмини акс эттиришга мўлжалланган.

МАТННИНГ АВТОМАТИК РАЗМЕТКАСИ – парсинг ва таггинг дастурлари томонидан матнни автоматик разметкалаш. Катта ҳажмли корпусни қўлда разметкалаш узок муддатли, қимматли меҳнатни талаб этади. Ўтган асрнинг 70-йилларидаёқ бу вазифани сунъий интеллектга юклатиш бўйича бир неча лойиҳалар ишлаб чиқилди (Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf). Ўшанда TAGGIT дастури Браун корпусидаги сўзларнинг 77 фоизини сўз туркумига тўғри ажратган эди, қолганини эса 10 йил давомида қўлда бажаришга тўғри келганди. Лекин 80-йилларга келиб,

CLAWS (Constituent Likelihood Automatic Word-tagging System) дастури Браун корпуси сўзларини 95 фоиз тўғри таҳлил қилди. Бугунги кунга келиб Европа тиллари учун сўз туркуми автоматик разметкаси (word-class tagging), гап бўлаклари автоматик разметкаси (parsing) ишлаб чиқилган. Бу ишнинг натижаси автоматик таржима ва интернет-қидирув тизимининг ишлашида намён бўлади. Матнга автоматик ишлов бериш устида иш олиб бораётган олимлар (сайт <http://www.aot.ru>), асосан, назарий лингвистикани замонавий ахборот технологияларида қўллаш устида иш олиб боришмоқда. Бу борада рус, инглиз ва немис тилларидаги матнни таҳлил қилишнинг графем (сўз чегарасини аниқлаш), морфологик (сўз туркумини аниқлаш), синтактик (гап бўлакларини аниқлаш) ҳамда семантик (сўзлараро семантик муносабат) модулларини яратишган. Корпусни разметкалаш (аннотациялаш) дастурлаштирилган йўллар билан амалга оширилади. Бунда, аввало, вақтни тежаш, меҳнатни камайтириш назарда тутилса, иккинчидин, матнга автоматик ишлов бериш муаммосига ечим топилади. Ҳозирча анафорик, просодик разметка қийин, разметка фақат қўлда бажариляпти, кейинчалик дастурлаштирилади. Морфологик, синтактик разметка эса теггер, парсинг ёрдамида амалга оширилса-да, бу дастурларнинг ҳам аксарияти автоматик разметкадан кейинги тузатишни талаб қилади. Морфологик омонимия (қўпроқ флексив тилларга хос), синтактик қўпмаънолилик ҳолатида дастур хулосанинг бир неча кўринишини таклиф қилади, тадқиқотчи эса кераклисини танлайди. Янги авлод корпуслари ҳажмининг фавқулодда катталашгани мутахассислар олдига разметканинг тўлиқ автоматлаштирилган турига ўтиш, янги, мукамал теггер ва парсинг яратиш вазифасини қўяди. Автоматик морфологик таҳлил (теггер) ёрдамида ҳар бир лексик бирликка (сўз туркуми, лемма, граммема гуруҳи) алоҳида грамматик характерис-тика (шахс-сон, келишик, бошқа грамматик категория) берилади. Масалан, Браун корпусида сўзнинг частотасини аниқлаш осон. Фақат бу сўзшаклнинг (корпус тилида *токен*) частотаси бўлади. Лексеманинг частотасини аниқлаш учун эса ҳар бир сўзга унинг леммаси бириктирилган

бўлиши керак. Корпусни автоматик разметкалашнинг оддий усули сўзнинг лексик категорияси кўрсатилган ҳажман катта электрон луғатни разметкаланган корпус билан бирлаштиришдир. Шунда электрон луғатдаги изоҳ (грамматик категория тавсифи) разметкаланмаган корпусдаги сўзга тег сифатида ўзлаштирилади. Масалан, корпус ва электрон луғатда *ахборот, сийёсат* сўзлари мавжуд бўлса, луғатдаги “от” теги автоматик тарзда корпусга кўчади. Лекин бу усул билан ҳам корпусни тўлиқ разметкалашнинг имкони йўқ. Чунки баъзи сўз ва бирикмалар бир вақтнинг ўзида бир неча категорияга мансуб бўлиши мумкин. Бу ҳолат морфологик кўпмаънолилиқ (*ambiguity*) муаммоси билан боғлиқ. *Олма, математик, этик, сурма, сузма, бўлмоқ, қўллар, боғлар* каби сўзшакллар бирдан ортиқ грамматик категорияга тегишли, шу сабабли бундай сўзларни фақат электрон луғат ёрдамида разметкалаб бўлмайди. Контекстда сўзшакл фақат битта категорияга тегишли бўлиб қолади, шунинг учун разметканинг яна ҳам мукамалроқ кўриниши: морфологик разметка учун синтактик разметка, синтактик разметка учун семантик разметка қилиш корпусни тўлиқ ва тўғри разметкалашга олиб келади. Лингвистик разметканинг барча (морфологик, синтактик, семантик, анафорик, просодик) турлари куйидаги тамойиллар асосида амалга оширилади: разметка схемасини тавсифлаш (асослаш) (1); умумий лингвистик тушунчалар тизимини аниқлаш (2); фойдаланувчи учун маълум бўлган таҳлил схемасини шакллантириш (4); разметка схемасининг назарий анъанавийлигига эришиш (5); халқаро андозаларга амал қилиш (6).

МАҲСУС МАТНЛАР КОРПУСИ – балансга келтирилган корпус, у ҳажман катта бўлмаган, алоҳида тадқиқий топшириққа хизмат қилувчи, маълум мақсадни кўзлаган, тузувчи режасига мувофиқ тайёрланган корпус (Захаров В.П. Корпусная лингвистика в России. CrossLingua’2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.). Санкт-Петербург инглиз тилини ўрганувчи (SPbEFLLC) мактаб ўқувчиларининг таълимий матнлари корпуси шундай корпуслар сирасига киради. А.И.Герцен номидаги педагогика университети амалий тилшунослик кафедраси томонидан яратилган бу корпусни тузишдан асосий

максад рус ўқувчисида энг кўп кизиқиш уйғотган инглиз матнларининг ўзига хослигини ўрганиш бўлган. Корпус учун материал вазифасини ўтаган бундай матнларнинг асл нусхаси Санкт-Петербург мактабларида 2007 йил ноябрь-декабрь ойларида йиғилган. Ушбу матнларнинг муаллифи 911та синфнинг олдиндан махсус тестдан ўтказилган инглиз тилини билиш даражаси ўрта (intermediate (26%)), ўртадан юқори (upper-intermediate (74%)) сифатида баҳоланган 78 нафар ўқувчи бўлди. Корпус ҳажми 50.000 сўз миқдориди. Корпус базасига асосланган тадқиқот шуни кўрсатдики, инглиз тили хорижий тил сифатида (English as a Foreign Language) эгалланган шахс нутқида содда матн ва гап тузилиши қўлланилади. Бу EFL (English as a Foreign Language) грамматикаси инглиз адабий тили грамматикаси билан айнан бир хил структурага эга эмаслигини кўрсатади. Махсус корпусга бўлган эҳтиёж миллий корпусдан қолишмайди. Хоҳлаган махсус корпус аниқ бир соҳа ёки ёндош соҳани (кемасозлик, металлургия, экология, навигация ва ҳ.) ўрганиш учун қўлланиши табиий. Чунки бу корпус мутахассисга атамани конкрет профессионал қуршовда, у ёки бу муаллиф шу атама остида нимани назарда тутганлигини кузатиш имконини беради. Махсус корпуслар сирасида Рус тилининг Регенбург диахрон корпуси (қадимги рус матнлари), Қадимги Рус ёзма ёдгорликлари корпуси (қўлёзма китоблар, солнома, қайин пўстлоғига ёзилган қадимги рус ёдгорликлари), “Слово о полку Игореве” параллел корпуси, “Манускрипт” рус электрон мерос корпусларини санаш мумкин.

МАХСУС ОҒЗАКИ КОРПУС – Лондон Лунд корпуси (London Lund Corpus (LLC, 1975)), Ланкастер (Lancaster/IBM Spoken English Corpus (SEC 1992)) каби корпуслар кўриниши. Кейингиси 52600 сўз ҳажмида, аудиоёзуви билан бирга CD шаклида фойдаланувчиларга тақдим этилган. Урғу, интонация, тўхтам жиҳатидан тўлиқ разметкаланган, аммо ижтимоий мавқеи, респондентнинг шаклланиши ҳақида маълумотлар бўлмаганлиги корпуснинг социолингвистикада қўлланишини черағалайди (Қарағ: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурсы) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf).

МИЛЛИЙ КОРПУС – жанридан қатъи назар тилнинг миллий бойликлари жамланган электрон қидирув тизими.

МОДАЛ-ЛЕММА – модал сўз туркумига мансуб лексема.

МОРФОЛОГИК РАЗМЕТКА – синтактик ва семантик разметкага асос бўлувчи, инглиз тилида part-of-speech tagging деб аталувчи, сўзларни туркумга ажратувчи аҳамиятли разметка. Разметканинг бу тури тег ёрдамида амалга оширилиб, тегнинг матнда мавжудлик даражаси, кўлами корпуснинг хусусиятидан келиб чиқиб, ҳар хил бўлади. Тег қанча кўп бўлса, корпуснинг лингвистик амалларни бажариш имконияти шунча кенг бўлади. Лекин кейинги авлод корпуслари ҳажмининг катталиги сабабли тегни соддалаштириш йўли қулай деб топилди. Кодировканинг соддалаштирилган тизими ортиқча хатоликларнинг олдини олади, морфологик кўп маънолилиқни келтириб чиқармайди, бир неча миллион сўзни қамраб олувчи катта массивли матнларнинг разметкаланишини тезлаштиради. Морфологик разметка тизимига *сўзшакл*, *лемма* ва *тег* киради. Сўзшакл – танланган матндаги морфологик бирлик. Сўзшаклни разметкаланишнинг биринчи босқичи уни леммалаш, яъни сўзшаклнинг лексема шаклини келтириш. Флектив тилларни разметкалашда энг қийин босқич бу леммалаш (лемматизация), яъни сўзнинг лексема шаклини сўзшаклга тег сифатида бириктириш ҳисобланади. Чунки флектив тилларда сўзшаклдаги грамматик маъно сўз ўзагига қоришиб кетган бўлади. Флектив тиллардан фарқли ўларок, агглютинатив тилда леммалаштириш жараёни анча осон. Сўзшаклнинг грамматик шаклсиз қисми (ўзак ёки негиз) леммага тенг.

МОРФОЛОГИК ТЕГ – морфологик разметка учун изоҳлар мажмуи.

МОШКОВА КУТУБХОНАСИ КОРПУСИ¹¹ – 680 миллион сўздан иборат рус матнлари корпуси. Корпус Мошкова кутубхонаси матнлари асосида А.Сокирко томонидан тузилган. Лексик бирликларни DDC корпус менежеридан фойдаланиб сўз туркуми, морфологик характеристикаси асосида қидириш мумкин. Шунингдек, М1 мезони асосида ҳисобланган

¹¹ <http://aot.ru/search1.html>

биграммалар (54 млн.) кидируви сервиси ҳам мавжуд. Корпус 40дан ортик турли тил корпусларидан тузилган Инглиз лингвистик хизмати (Lexical Computing Ltd.) томонидан тижорат операциялари асосида фойдаланиш ҳукукини беради. Улар орасида бир қатор рус корпуслари, Интернет матнларидан Wacky технологияси асосида тайёрланган 20 миллиард ҳажмли корпус аҳамаиятга молик. Инглиз тадқиқотчилари ва Масарика университети ходимлари ҳамкорлигида Sketch Engine корпус менежери¹² ишлаб чиқилган, мазкур корпус ноёб имкониятларга эга. Стандарт кидирув тизими, конкорданс билан бирга синтактик қолипларни ҳам аниқлаб беради, частотали луғатни шакллантиради, лексик бирликларнинг ўзаро бирикиш имкониятларини кўрсатиб, лексик-семантик майдонда гуруҳлайди (*Қаранг:* Захаров В.П. Корпусная лингвистика в России. CrossLingua'2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.)

МУАЛЛИФЛИК КОРПУСИ – Корпус материали алоҳида муаллиф асарларидан иборат корпус. Ундан корпус, конкорданс ва электрон кутубхона сифатида фойдаланиш мумкин. Ҳозирги ўзбек тили бошқа тиллар каби ўз тарихий тараққиёт йўлида бир қанча ўзгаришларга учраган. Тараққиёт давридаги ҳар бир босқичнинг ўзига хос хусусиятини тадқиқ этиш мақсадида тилнинг тараққиёт босқичлари ажратилган, ўрганилган; бу тадқиқотларда давр тилининг ўзига хос хусусияти маълум бир адиб ижоди мисолида ёритилган. Ҳар бир давр ижодкорлари ижодий меросини тўлик қамраб олувчи бир қанча лингвистик тадқиқот мавжуд бўлса-да, улар махсус тизимлаштирилган база шаклини олмаган. Ахборот технологиялари асида бундай яхлит тизимнинг ўзбек (туркий) ёзма манбалари учун мавжуд эмаслиги ачинарли. Тилшунослигимиз тарихида алоҳида муаллиф асари тил хусусиятини ўрганишга бағишланган махсус тадқиқотлар талайгина. Руний, уйғур ёзуви асосида яратилган ёдгорликларнинг тил хусусиятига бағишланган бир қанча асарлар майдонга келди. Бу тадқиқот ишларида қадимги туркий тил хусусиятларининг айрим морфологик, синтактик,

¹² <http://sketchengine.co.uk/>

фонетик хусусиятлари таҳлил қилинди, муаллиф тили лексикаси атрофлича тавсифланди. Бундай тадқиқотлар тилнинг ривожланиш жараёни, ўзгаришларни (маъно торайиши, кенгайиши, историзм, архаизм, неологизмлар ҳаракатини) кузатиш учун қулай. Мана шундай хусусият ҳар бир муаллиф, давр тили учун тадқиқ этилган бўлса-да, улар алоҳида-алоҳида монографияларда, тадқиқот ишларида мавжуд. Агар бу тадқиқотлар Миллий корпусда яхлит бир тизим асосида ўз аксини топса, исталган лексеманинг тарихий тараққиёт жараёнини турли муаллиф тили мисолида ёрқин тасаввур қилиш имконияти пайдо бўлади. Тилшунослигимиз олдида турган ғалдаги вазифа шу тадқиқотларни бирлаштириш асосида корпус яратишдан иборат. Бу вазифани амалга ошириш икки босқични ўз ичига олади: тилимиз тарихий тараққиёти даврида тил хусусиятлари тадқиқ қилинган адиблар ижодий мероси ва тадқиқот натижасини корпус сифатида шакллантириш (1); шу пайтгача тил хусусияти ўрганилмаган ёзувчи, шоирларнинг асарлари корпусини тузиш (2). Биринчи вазифа ўрганилган миллий-маданий меросимизнинг электрон шаклда яхлит система(корпус шакли)да сақланиши, ундан кейинчалик турли таълимий, тадқиқий мақсадда фойдаланиш имконини беради. Иккинчи вазифани амалга ошириш тил хусусияти ўрганилмаган ёзувчи, шоирлар асарига бағишланган турли филологик тадқиқотлар олиб бориш учун электрон база ҳамда миллий-маданий мерос, тилни электрон шакл(корпус формати)да сақлаш, фойдаланиш, келгуси авлодга тўлиқ етказишни таъминлайди. Адабий тил ривожига салмоқли ҳисса қўшган буюк ёзувчилар асарларининг лексикасини асраб қолиш, нафақат асраш, балки бутун қўлами билан тавсифлаш, тил эгалари учун намуна вазифасини ўтайдиган модел шаклига келтириш адабий тилни асраш, ривожлантириш омили бўлиб хизмат қилади. Бу борада ўз миллати миллий, адабий тили асосчиси ҳисобланган бутун дунёга машҳур адиблар Шекспир, Данте, Гёте, Пушкинларнинг адабий мероси асосида қилинган ишлар (тузилган корпуслар) намуна сифатида хизмат қилади (Қаранг: Потемкин С.Б. Авторский корпус и словарь языка Антона Чехова. (Электрон ресурс: // <https://istina.msu.ru>)). Бу

сирага адабий тилнинг энг оптимал ташувчиси сифатида тан олинган ёзувчилар Шандор Петефи, Генрик Ибсен, Адам Мицкевич, Карел Чапекларнинг ишларини киритиш ўринли. Жаҳон тилшу-нослигида корпус лингвистикаси анча ютуқларга эришган бўлса-да, муаллифлик корпуси тузиш тажрибаси охириги 5 йил ичидагина тўлароқ шаклланди, бу борада анча натижаларга эриша олди. Бундай корпус тузиш маълум давр адабий тили, хусусан, лексик таркиби, услубий хусусиятини тадқиқ этиш учун материал вазифасини ўташдек қимматли вазифани адо этади.

МУАЛЛИФЛИК ЛЕКСИКОГРАФИЯСИ – жаҳон лексикографиясининг алоҳида бир йўналиши. У муаллиф тили луғатини тузишнинг назарий асослари, муаллифлик идиостилини ўрганиш тажрибаси, муаллифлик электрон луғати тузишнинг ўзига хос жиҳати, миллий корпусдан ва адабий тил корпусоти корпусидан муаллифлик лексико-графиясида фойдаланиш тажрибаси, янги турдаги луғатлар тузиш ғоялари, муаллифлик лексикографияси асосида замонавий адабий тилдаги ўзгаришларни қиёсий-таҳлилий тадқиқ этиш каби муаммолар билан шуғулланади¹³. Муаллифлик лексикографияси – жаҳон, шунингдек, рус лексикографиясида шиддат билан ривожланиб бораётган соҳа. Унинг тараққиёт натижаси ўлароқ кўплаб муаллифлик электрон луғат ва корпуслар яратилди. М.В.Ломоносов номидаги Москва Давлат университетида “Крылов масаллари тили луғати”, “Трибодов комедиялари тили луғати”, “Гоголнинг “Ревизор” комедияси тили луғати”, “Фонвизин комедиялари тили” каби луғатлар яратилиб, бундай ишларга намуна сифатида “А.С.Пушкин тили луғати”¹⁴ хизмат килди. Бу тажриба рус лексикографияси соҳасида оммалашиб, бошқа олий таълим муассасаларида ҳам шундай тадқиқотлар бошланди. “А.Н.Островский асарлари тили луғати”¹⁵, “М.В.Ломоносов асарлари тили луғати”, “М.Кузмин

¹³ Қаранг: www.ruslang.ru/seminar_aut_lexocorg020413 сайтидаги “Теория и практика авторской лексикографии” семинар материаллари

¹⁴ “Словарь языка Пушкина” в четырех томах, создан в Институте русского языка АН СССР под ред. акад. В.В.Виноградова (издан в 1956–1961 гг.

¹⁵ Хуснутдинов А.А.. Словарь языка А.Н.Островского” как один из вариантов авторского словаря” доклад (Ивановский государственный университет). Электронный ресурс: www.ruslang.ru/images/seminar_aut_lexocorg/husnutdinov_doklad.pdf

шеърлари конкорданси”, “А.Мицкевичнинг “Пан Тадеуш” поэмаси тили идеографик луғати” каби ишлар шулар жумласидан. Бу соҳадаги тадқиқотлар ичидан О.В.Кукушкина, А.А.Поликарпов, Е.В. Суровцевалар томонидан тузилган “А.П.Чехов бадий асарлари тили луғати”¹⁶ диққатга сазовор.

МУАЛЛИФЛИК ЛЕКСИКОГРАФИЯСИ АСОС ВА ҲОСИЛАСИ - муаллифлик лексикографияси маҳсули, одатда, 2 қисм: асос ҳамда маҳсулотдан иборат бўлади. Асос – корпус ёки конкорданс, маҳсулот – луғат. Кейинги пайтда лексикогра-фияда (фақат корпус лингвистикасида эмас) конкорданс тушунчаси кенг оммалашди. *Конкорданс* у ёки бу тил ифода воситасининг контекстда барча қўлланиш ҳолати манбасига ҳавола қилувчи рўйхат маъносида корпус лингвистикасида қўлланилади. Шунингдек, конкорданс атамаси бирор китоб ёки асар калит сўзларининг контекстга ҳавола қилувчи алифбо тартибдаги рўйхати маъносини ҳам ифодалайди (Қаранг: Бобунова М.А. Конкорданс XXI в.: новая старая форма // Вопр. лексикографии. –2016. №2 (10) С.41-54.). Бундай конкорданснинг илк намунаси XIII асрда 500 монах иштирокида тузилган. Унда иш ҳажми улканлиги учун (жуда муҳим конкорданслар: Веда, Инжил, Куръон ёки Уильям Шекспир асарлари) катта жамоалар биллашиб ишлашган. Бундан ташқари, конкорданс амалий тилшуносликда таржима жараёнида, лексикографияда, тилни ўқитиш, ўрганиш ҳамда матн таҳлилида ҳам қўлланилади (Қаранг: Якубовский К.И., Якубовская К.А. Обзор современных лингвистических технологий и систем // Вестник МГУП. 2015.– №2 С.315-319.). Корпусдан ташқари конкорданс сўзлар рўйхатини тузиш, бир сўзнинг турлича ишлатилишини қиёслаш, ибора ҳамда идиомаларни қидириш, тадқиқ этиш, калит сўзларни таҳлил қилиш, турли сўз ва бирикмалар таржимасини излаш ёки қўлланиш частотасини аниқлаш каби лингвистик амалларни бажаришда қўлланилади (Столяров А.И. Словарь-конкорданс и его применение в рамках корпусной лингвистики // Гуманитарные научные исследования. – 2017. № 2 [Электронный ресурс]. URL: <http://human.snauka.ru/2017/02/21074> (дата обращения: 29.09.2017). Конкорданс-луғат – ҳар бир сўзга қисқача контекст (семантик изоҳ эмас!)

¹⁶ Кукушкина О.В., Поликарпов А.А., Суровцева Е.В. “Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова (с электронным приложением) как особый лексикографический продукт”. (Электрон ресурс: www.ruslang.ru/seminar_aut_lexicorg)

келтирилдиган луғатнинг алоҳида бир типи. Матнни ўрганишда конкорданснинг қидирув, эвристик, аналитик функцияларидан фойдаланиш, индекс-сация, қиёслаш каби имкониятини ишга солиши мумкин. Қидирув функцияси берилган сўз ёки бирикмадан фойдаланган ҳолда матндаги керакли лавҳани зудлик билан топишга ёрдам беради (Қаранг: Труфанов М.В. Применение цифрового словаря с корпусом для работы над лингвистическими проектами по немецкому языку // Социально-экономические явления и процессы. 2011.– №8 С.229-232.). Конкорданснинг ушбу функциясидан цитата танлаб олиш ёки уларни текширишда, матннинг оригиналлигини ўрганишда ҳам фойдаланиш мумкин. Одатда, конкорданс хрестоматия тузувчилари томонидан кўпроқ ишлатилади: у матнда ибора, идиомаларни қидириш, атамаларни таржима қилишда муҳим. Конкорданснинг эвристик функцияси тилшунос ёки оддий китобхонга матннинг янги талқини, шарҳини кўрсатишда, аналитик функцияси эса тилнинг турли бирликлари: лексема, калит сўз, уларнинг матнда қўлланиш частотаси устида таҳлил олиб бориш, индексация функцияси матнни нашрга тайёрлашда сўзнинг рўйхати, индексини тузиш, қиёслаш функцияси матн ичидаги барча коннотациялар, сўз қўллаш ҳолатини таққослаш имконини беради. Луғат мақоласи оддий луғатдан фарқли равишда конкорданс сўзга мисол келтиради, яъни қўлланиш контекстини беради. Бу билан конкорданс турли дифференциал, шунингдек, изоҳли луғатлар яратишда асос вазифасини ўтайди. Ф.М.Достоевский, Ф.И.Тютчев, А.П.Чехов, А.С.Грибоедов, С.А.Есенин, С.Т.Аксаков, М.В.Ломоносов асарлари конкорданси (Қаранг: “Частотный словарь-конкорданс публицистики Ф. М. Достоевского”. (Данный словарь создан на основе полного собрания сочинений писателя в 30-ти томах, включая Дневник писателя.); “Частотный словарь-конкорданс словоформ языка Ф. И. Тютчева”. (Создан по изданию: Тютчев Ф.И. Полное собрание стихотворений / Сост., подгот. текста и примеч. А.А. Николаева. – Л.: Сов. писатель, 1987. – 448 с.); “Метапоэтический словарь-конкорданс драматического текста А.П.Чехова”. (В словаре представлено исследование А.П. Чеховым собственного драматического творчества.); “Алфавитно-частотный конкорданс текстов Грибоедова”. (Словарь содержит все слова, встретившиеся во всех опубликованных текстах Грибоедова, взятых из наиболее авторитетных источников.); “Конкорданс к стихам А.С.Пушкина”. (Расширенный словарь языка писателя, в котором на каждое слово и словоформу Пушкина приводится полный текст всех стихотворных произведений.); “Частотно-семантический конкорданс служебных слов в поэзии Сергея Есенина”. (Словарь представляет собой первое полное лексикографическое описание служебных слов в лирике великого русского поэта С.А. Есенина.); “Конкорданс произведений

С.Т.Аксакова”. (Может быть применим при реконструкции художественного мира С.Т.Аксакова, тематического своеобразия, мотивно-образной системы прозаических произведений писателя.); “Конкорданс к текстам Ломоносова”. (Является частью проекта по созданию электронного научного издания “Ломоносов”.) шундай ишлар жумласидан. Конкорданс анъанавий лексикографияга нисбатан катта ҳажмли маълумотлар базасини ўзида сақлаш имкониятига эга. Конкорданс бирлиги матндаги парадигматик муносабатни кузатиш, ижодкор меросининг турли парчаларида сўз қўлланиш ҳолатини миқдорий таҳлил қилиш, унга хос синтактик қурилмалар ҳақида хулосалар чиқариш каби вазифаларни бажара олади. Луғат-конкорданс – лингвистик, адабиётшуносликка оид тадқиқотлар бажаришда замонавий, самарали восита.

МУЛЬТИМОДАЛ КОРПУС – мулоқот иштирокчиси видеоёзувини ҳам ўз ичига олган, мимика, қўл, кўз, қош ҳаракати ва ҳ. белгилар асосида разметкаланган корпус (Қаранг: Захаров В.П. Корпусная лингвистика в России. CrossLingua'2014 < 27 – 30. 05,14 < Crimea, Simferopol – Alushta.). Ўзаро эмоционал стратегия, ихтилоф, мулоқот одоби, нутқий тўхтам, бошқа ҳодисаларни ўрганиш воситаси бўлиши мумкин. Мижоз билан ишлайдиган ходим учун ўрганиш материали, мультипликатор ҳамда режиссёр учун эмоционал реакция маълумотлар базаси сифатида хизмат қилади. Рус тили эмоционал корпуси (РЭС)¹⁷ – шундай корпус намунаси. 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. Корпус ELAN дастури асосида разметка-ланган, параллел корпус шаклини олган, диалоглар маъновий гуруҳларга бўлинган ҳолда жойлаштирилган. Масалан, рус тили эгаси нутқи билан рус тилини ўрганаётган хитой тили вакили нутқи айни бир мавзудаги диалог (“Танишув”, “Аэропортдаги регистрацияда”) асосида қиёсланади. Ушбу корпус биринчи навбатда лингводидактик мақсадни, хусусан, типик хатоларни топиш, машғулотлар давомида шу хатоларни

¹⁷ <http://www.harpia.ru/rec>

нутқдан четлаштириш йўлини излаш ҳамда талабанинг мустақил ишига кўмаклашишни кўзлайди. Юқоридаги корпус намунаси мультимедиали корпус асосида имо-ишора, кўшимча эмоционал ҳолатларнинг (оғзаки) нутққа таъсирини ўрганиш воситаси эканлигини кўрсатади.

ОВОЗЛИ КОРПУС намунаси сифатида Рус тили миллий корпуси таркибидаги оғзаки нутқ корпуси¹⁸ни санаб ўтиш жоиз (Қаранг: Захаров В.П. Корпусная лингвистика в России. CrossLingua'2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.). Бу корпус мультимедиали, диалектал, бадий, акцентологик корпуслар каби миллий корпус таркибидан жой олган. Санкт-Петербург давлат университети филология факультети филологик тадқиқотлар институти томонидан ишлаб чиқилган оғзаки корпус ҳам шундай корпус намунасидир. “Бир кунлик нутқ” деб номланувчи бу корпус – предмети кундалик сўзлашув нутқи, у тил эгасининг кундалик мулоқотдаги жонли тилини ўрганади. Бошқа овозли корпуслардан асосий фарқи жонли нутқдан ёзиб олинганлиги, у лаборатор ёки бошқа махсус шароитда ёзиб олинган нутқдан фарқ қилади. Санкт-Петербургда яшовчи турли ёш, ижтимоий қатламга мансуб аҳолидан (30 киши) иборат информантлар гуруҳи танлаб олинган, улар бир кун давомида диктофон билан юришга розилик берган, шундай табиий шароитда уларнинг кундалик мулоқот нутқи ёзиб олинган. Информант овоз ёзиш жараёни борасида йўриқнома билан таништирилган, психологик тест, ижтимоий сўровномадан ўтказилган. Информантлар билан бирга кун давомида мулоқотда бўлган шахс (қариндош, дўст, ҳамкасб, таниш, нотаниш) нутқи ҳам киритилган, натижада, корпус ҳар хил ёшдаги шахс, турли соҳа вакиллари нутқини қамраб олган. Овозли ёзув 500 соатни ташкил этади (2013 йил ҳолатида). Бундай овозли корпус ёзиб олинган материалнинг лингвистик динамикасини ўрганиш имконини беради. Жумладан, стандарт статистик метод асосида микдорий ўзгаришларни ўрганиш (1); лексик, грамматик қатор частотаси, синтактик структура, у ёки бу акустик ходиса, просодик контурнинг семантика, суҳбат мавзусини

¹⁸ <http://ruscorpора.ru/sbornik2008/07.pdf>

сутканинг маълум қисми, мулоқот шароити билан боғлиқ ҳолатни таҳлил қилиш (2); шахс касбининг маиший ҳаётига таъсирини таҳлил қилиш (3); тил ташувчисининг табиий нутқ жараёнидаги ўртача артикуляцион темпи ҳақида маълумот олиш (4) каби ҳодисалар тадқиқи учун характерли.

ОҒЗАКИ АМЕРИКАНЧА ИНГЛИЗ ТИЛИ КОРПУСИ (Corpus of Spoken American English (1991)) – 1 млн. сўз қўллаш ҳолати, 80 соат овозли ёзувдан иборат корпус; Map Task Corpus (1991, Глазго университети, Шотландия) 147000 сўз қўллаш ҳолати, 16 соат овозли ёзувдан иборат. Оғзаки корпуслар ёзма корпусга нисбатан кам миқдордаги сўзшаклни камраб олади. Бу фақат маълумотни йиғишнинг қийинлиги учун эмас, балки просодик тадқиқот учун кам миқдор ҳам етарли эканлиги билан боғлиқ. Масалан, интонацияни ўрганиш учун 100.000 сўздан иборат корпус етарли. Бундай корпуслар монологик, диалогик нутқни ўз ичига олади. Материал йиғиш учун радио, телевидение ёзувлари, социологик, социолингвистик сўров натижалари-дан фойдаланилади. Бугунги кунда (XX асрнинг 70-йиллардан фарқли ўлароқ) овозни яширин усулда ёзиб олиш этик меъёрга тўғри келмайди. Одатда, респондент ҳақида атрофлича тўлиқ маълумот (ёзув ўрни, респондентнинг нима билан шуғулланаётганлиги, вақти, сана, иштирокчилар сони, жинси, ёши, суҳбатнинг фавқуллода(табиий)лик даражаси, мавзу, иштирокчининг этник мансублиги, асосий (она) тили, касби, ижтимоий мавқеи ва ҳ.) йиғилади. Энг машаққатли иш транскрипциялаш жараёни, бир соатлик овозли ёзувнинг орфографик транскрипцияси, минимал интонацион разметкасини тайёрлаш учун 10 соат вақт керак бўлади. Агар матн ТЕИ (Text Encoding Initiative) асосида тўлиқ разметкаланса, бу ишга 25 соат вақт сарфланиши мумкин. Разметкасиз оғзаки корпус ҳеч қандай аҳамиятга эга эмас, чунки, ҳеч бўлмаганда, бир вақтдаги нутқнинг тўхтаб давомийлиги, урғу ва интонацияси кўрсатилиши лозим. Баъзан матнга изоҳлар ҳам қўшилади. Лондон-Лунд корпуси тўлиқ разметкаси оғзаки корпус разметкалари учун стандарт вазифасини ўтапти. Корпус материали неча тилда берилишига кўра бир ва кўп тилли деб таснифланади

(Қарағ: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf). ОҒЗАКИ корпусга Санкт-Петербургда тузилган “Одинь речевой день” корпуси мисол бўлади.

ОҒЗАКИ КОРПУС – фақат оғзаки матндан иборат корпус.

ОЛМОШ-ЛЕММА – олмош-лексема.

ОТ-ЛЕММА – от-лексема.

ПАРАЛЛЕЛ КОРПУС – (1) бир мавзуга оид икки тилдаги матнли корпус. Биринчи типдаги корпус “параллел корпус” (parallel corpora) деб номланиб, маълум бир таржиманинг турли аспектини ўрганиш учун қўлланилади. Бундай корпуслар сирасига Канада парламенти йиғини (инглиз/француз) матнлари корпусини киритиш мумкин. Параллел корпус ўз навбатида яна 2 турга бўлинади. Бу мосланган (aligned), мосланмаган (not aligned) корпусдир. “Мосланган” атамаси корпусда таржима бирликлари орасида бир-бирини такозо этувчи аниқ алоқа мавжудлигини билдиради. Бундай корпусдан у ёки бу гапнинг қандай таржима қилинганини дарров топамиз. Бу турдаги корпус таржимон учун аҳамиятли, чунки унда ноёб ресурс – “таржима хотира”си (translation memory) мавжуд; (2) бир-бирининг таржимаси бўлган матнли корпус.

ПАРСИНГ – автоматик синтактик разметкаш дастури. Рус ва инглиз тилларининг, умуман, корпус лингвистикаси ривожланган тилларнинг парсинг дастурлари яратилган.

ПРОСОДИК РАЗМЕТКА – оҳанг белгисини билдирувчи разметка. Овоз транскрипция қилинган корпусда урғу, оҳангни ифодаловчи изоҳ мавжуд бўлади. Разметканинг дискурс деб ҳам аталувчи ушбу тури шарҳ, изоҳ, эслатма, такрорлардаги тўхтамларни билдириш учун ишлатилади.

ПУБЛИЦИСТИК КОРПУС – публицистик матнлар йиғиндиси. XX аср охири рус тили газета матнлари корпуси¹⁹, сиёсий метафоралар корпусларини санаш мумкин.

¹⁹ <http://www.philol.msu.ru/~lex/corpus>

РАВИШ-ЛЕММА – равиш-лексема.

РАЗМЕТКА – изоҳ, тег. Корпусни оддий электрон тўплам ёки виртуал кутубхонадан фарқловчи, матн устида турли лингвистик амалларни бажаришга имкон берувчи энг асосий омил – унинг разметкаси. Аввалги бобда корпус ёрдамида бажарилиши мумкин бўлган ишлар ҳақида айтиб ўтилди. Бу вазифаларни бажариш учун корпуснинг ўзи мавжуд бўлиши етарли бўлмай, корпус материалига турли қўшимча лингвистик изоҳ ҳам бўлиши керак. В.П.Захаров, Б.Кутузов, С.А.Шаров ишларида разметка ва унинг корпусдаги аҳамиятини атрофлича ёритиб берилган (Қаранг: Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf). З.П.Захаров раҳбарлиги остида Л.В.Северюхина ишларида (Қаранг: Л.В. Северюхина. “Моделирование логико-понятийной области корпусной лингвистики”. (Дип.раб; рук. доц. В.П. Захаров), В.В.Рыковнинг шу мавзудаги маърузаларида ҳам разметка масаласи атрофлича ёритилган. Ишларнинг ҳар бирида разметканинг алоҳида жиҳати таърифланган (Қаранг: В.В.Рыков маърузалари билан <http://rykov-cl.narod.ru/c.html> сайтида танишиш мумкин.). Шуларга таяниб, “Лингвистик аннотация ёки корпус разметкаси (инг. linguistic markup) – корпусдаги матннинг қисми бўлмаган, лекин шу матнга бириктирилган қўшимча маълумот (метаахборот)” деб хулоса қилиш ўринли. Бу қўшимча маълумотнинг энг оддий кўриниши сўз туркуми ҳақидаги ахборот. У қуйидаги шаклда берилиши мумкин: Осмон (*om*) остидаги (*om*) ҳамма (*олмош*) нарса (*om*) омонатдир (*сифат*). (Лао Цзи). Разметка корпусни автоматик таҳлил қилишда қўл келади. Матндаги сўзларни бир марта туркумга ажратсак, исталган тадқиқотни бажаришда (масалан, корпусдаги барча сифатларни ажратишда) фойдаланиш учун қўл келади. Разметка масаласига тўхталишдан олдин разметка тизимлари тарихига назар ташлаш фойдадан холи бўлмайди. Ўтган асрнинг 80-йилларида SGML (Standard Generalized Markup Language) номи остидаги электрон матнлар разметкаси стандарти қабул қилинган эди. Бу стандарт дастлаб типография саноати учун ишлаб чиқилган бўлса-да, тез орада бошқа соҳаларда ҳам қўлланила бошланди. SGMLнинг моҳияти турли матн

мухаррирларида терилган хужжатни таҳрирлаш, таҳлил қилиш, ўзгартиришдан иборат.

РАЗМЕТКАЛАНГАН КОРПУС (индексланган, аннотацияланган; русча таггированные) – қўшимча изоҳнинг мавжудлигига кўра корпуснинг бир тури. Инглизча tag (изоҳ) сўзидан олинган. Бундай корпусда разметка характеридан келиб чиққан ҳолда сўз ҳамда гапга морфологик, синтактик, семантик, просодик ва ҳ. изоҳ (корпус тили билан *тег*) қўшилган бўлади.

РАЗМЕТКАЛАНМАГАН КОРПУС (индексланмаган, аннотацияланмаган; русча нетаггированные.) – қўшимча изоҳнинг мавжудлиги-га кўра корпуснинг бир тури. Корпус матнларига ҳеч қандай изоҳ (морфологик, синтактик, семантик, просодик) берилмаган корпус; электрон матнлар базаси.

РЕПРЕЗЕНТАТИВЛИК – зарурий ҳажм, миқдор, жанр, услуб, муаллифга тегишли матнлар пропорциясининг тўғри танланиши. Репрезентативлик хусусиятини корпус тузиш жараёнида ҳам, ундан фойдаланиш жараёнида ҳам сақлаб туриш корпуснинг аҳамиятини таъминловчи асосий омил ҳисобланади. Корпус маълум тилнинг белгиланган даврда мавжуд бўлган барча турдаги матн намуналарини ўзида жамлаган бўлиши, ҳажм жиҳатдан тўғри пропорцияни ташкил этиши лозим. Корпус тилнинг хусусиятларини ўрганишда қанчалик аҳамиятли бўлмасин, унинг фақат бадий асарлар билан чегараланиб қолиши тўғри эмас. У турли мавзудаги (ижтимоий-сиёсий мавзудан токи спорт мавзусигача) газета, журнал материали, фаннинг барча соҳасига тегишли илмий, илмий-оммабоп асарлар, реклама, шахсий ёзишмалар матнidan таркиб топган бўлиши лозим. Тилда қандай жараён кечаётганлигини аниқ тасаввур қилиш учун корпус қамровини янада кенгайтириш, ёзма нутқ билан бирга оғзаки нутқ материалидан фойдаланиш мақсадга мувофиқ. Бундай корпус ёрдамида тил тараққиёти натижасида содир бўлган, қутилаётган ўзгариш ҳақида аниқ хулоса чиқариш мумкин (Қаранг: www.ruscorpora.ru/ Плунгян В. Зачем мы делаем корпус?). Корпуснинг мана шу хусусияти унинг репрезентативлигидир.

РУДАКИЙ КОРПУСИ²⁰ – разметкаланмаган субкорпус; Тоҷик тили миллий корпуси таркибида жойлашган. Рудакий асарларининг махсус қидирув тизимига бўйсундирилган электрон йиғиндиси. Форс-тоҷик мумтоз адабиёти асосчиси Абу Абдулло Рудакий меросини яхлит қўринишда тўплаш, рудакийшуносликни ривожлантириш ҳамда янги тадқиқотлар олиб бориш учун қулайлик яратиш мақсадида шоир таваллудининг 1150 йиллигига тайёргарлик жараёнида Рудакий корпусига асос солинган. Корпус асосини шоирнинг асарлари матни ташкил этган. Замонавий ахборот технологиялари ёрдамида Рудакий ғазаллари матни бўйича бир нечта луғат тузилган. Частотали луғат, камайиш, ортиб бориш алифбо частотали луғати, Рудакий ижоди матнлари терс луғатлари шулар жумласидан. Корпус таркибидан Рудакий ижодий мероси-нинг барча намуналари, библиографияси, Рудакий ҳаёти, ижодига бағишланган асарлар, шоир ижоди турли аспектлари тадқиқи натижалари ўрин олган. Рудакий шеърлари, улар асосидаги қўшиқларнинг аудио, видео ёзувлари тўпланган; корпус доим тўлдириб борилади. Бундай самарали меҳнат натижасида Рудакий шеърляти мухлислари унинг нодир ижодий мероси билан тўлик, эркин танишиш имконига эга бўлишди. Ҳатто корпусга рудакийшунослик видео дарслари серияси ҳам киритилган. Корпусни ташкил этишда Ахборот ва коммуникацион технологиялари мактаби 12-18 ёшли аъзоларидан иборат ёшлар қатлами кизгин фаоллик кўрсатишган. Ушбу корпус электрон кутубхона шаклида фаолият кўрсатади, разметкаланмаган. Шундай бўлса-да, айна бир ижодкор мероси, шу соҳадаги илмий тадқиқотларни ўрганиш мақсадида фойдаланиш учун жуда қулай ахборот манбаи.

РУС ТИЛИ ГАЗЕТАЛАР КОМПЬЮТЕР КОРПУСИ²¹ (Компьютерный корпус текстов русских газет конца хх в.) – 2000-2002 йилларда Москва Давлат университети филология факультетида Умумий ва компьютер лексикологияси ҳамда лексикографияси лабораториясида

²⁰ <http://termcom.tj>.

²¹ www.philol.msu.ru/~lex/corpus

тузилган²² корпус. Корпус материали сифатида 1994-1997 йилларда Россиянинг 13 номдаги газеталари тўлиқ сонлари асос қилиб олинган, матнлардаги сўз қўллаш ҳолати 1 миллионни ташкил этади. Унга кундалик, ноқундалик (“МН”, “Новая газета”), “чап” (“Завтра”, “Правда”, “Правда-5”) ва “Ўнг” марказий ҳамда ҳудудий, умумий ҳамда касбий (масалан, “Литературная газета”) газеталар кирган. Бу тамойил газета материали билан боғлиқ турли хилдаги матн (масалан, турли тур, жанр) бирликлари, улар ўртасидаги муносабатни объектив ҳамда ишонарли картина асосида кузатиш имконини беради. Корпус Москва Давлат университети филология факультети Умумий ва компьютер лексикологияси ҳамда лексикографияси лабораториясининг Диктум-1 системаси асосида тузилган. Бу тизим ёрдамида корпус бирлик ва матнларига турли хилдаги маркерлар билан автоматик, ярим автоматик белги қўйилган: матн – газета манбаси, матн ҳажми, жанри, нашр санаси ва х.; сўз қўлланиш ҳолати – грамматик, лексик, морфем ва бошқа категория бўйича белгиланган. Корпуснинг Интернетга мўлжалланган кўргазмалари вариантлари тайёрлаш учун корпуснинг автоматик лемматизация, морфологик квалификация, шунингдек, лексема, сўзшаклнинг морфем сегментацияланиши имконига эга бўлган 200 000га сўз қўлланиш ҳолати ҳажмини қамраб олган лавҳа (фрагмент) ажратиб олинган. Корпуснинг тўлиқ варианты устида иш олиб бориляпти.

РУС ТИЛИ МИЛЛИЙ КОРПУСИ²³ (Национальный корпус русского языка) – оммабоп рус матнлари корпуси; 2004 йил 29 апрелда очилган. Корпус тузиш 2001 йилда Москва, Петербург, Воронеж каби шаҳарлардаги бир гуруҳ тилшунослар ҳамкорлигида бошланган, асосий тузувчилар – РФА Рус тили институти, РФА Тилшунослик институти ҳамда “Яндекс” компанияси. 2007 йилда 140 млн. сўз қўллаш ҳолатини қамраб олган ёзма, оғзаки, бадий, диалектал матнлар йиғиндисини ташкил этган (Қаранг: *Қаранг:* Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution

²² <http://www.philol.msu.ru/~lex/corpus>

²³ <http://ruscorpora.ru>.

2013 йил март ойи ҳолатида ҳажми 500 миллион сўз бўлган. Бу корпус доимий тарзда янгиланиб, тўлдирилиб борилади. Унинг яна бир хусусияти – матннинг жанрий хилма-хиллиги; рус тилида қўлланидиган барча соҳа услуб ва жанрларини (илмий, расмий-идоровий, публицистик, черков, бадий, кундалик сўзлашув, оғзаки, электрон коммуникация) қамраб олган (Қаранг: Қаранг: Захаров В.П. Корпусная лингвистика в России. CrossLingua'2014 < 27 — 30. 05.14 < Crimea, Simferopol — Alushta.). Корпус морфологик, семантик разметкага эга; Web sayt орқали эркин фойдаланиш мумкин. Бу корпуснинг машҳур ва оммабоплиги, бошқа корпуслардан фарқ қилувчи асосий белгиси *мультимедиа*ли корпус²⁴ ҳамда *семантик разметка*га²⁵ эга эканлигида.

РУС ТИЛИНИНГ МАШИНА ФОНДИ – 1985 йилда СССРда академик А.П.Ершов ташаббуси билан тузиш ишлари бошланган. Жуда кенг қўламли лойиҳа, уни тузишда 40дан ортик ташкилот ҳамкорлик қилган. Рус тили институти, Москва, Санкт-Петербург, Харьков, Саратов университетлари шулар жумласидан. Фонднинг вазифасига рус тилини ўрганиш, амалий ишланмаларни бажариш учун керакли матний, лексикографик, грамматик маълумотлар базасини яратиш, уларни электрон шаклда сақлаш, ташиш кирган. Айни вақтда, лингвистик тадқиқотлар ўтказиш учун бир неча дастур ҳам ишлаб чиқилган. 1985-1992 йилларда рус тилининг машина фонди архитектураси ва концепцияси ишлаб чиқилган; XIX–XX рус адабиётининг барча манбалари, рус тилининг асосий луғатлари, қисқа академик грамматикалар электрон шаклга келтирилган; шеърият, бадий адабиёт, ижтимоий-сиёсий, техникага оид матнлар корпуси яратилган. 1991 йилга келиб машина фондини ташкил этиш ишларининг иқтисодий таъминоти сусайиб кетган, кейинроқ иш батамом тўхтаб қолган.

РУС ТИЛИНИНГ МУЛЬТИМЕДИАЛИ КОРПУСИ²⁶
(Мультимедийный русский корпус (МУРКО)) – 1930-2000 йиллардаги

²⁴ (<http://ruscorpora.ru/search-murco.html>)

²⁵ (<http://ruscorpora.ru/corpora-sem.html>)

²⁶ <http://ruscorpora.ru/search-murco.html>

кинофильм лавҳаларидан ташкил топган корпус. Видеоряд, аудиоряд, овознинг матнли расшифровкаси, кадрда учрайдиган имо-ишоралар каби параллел параметрларни қамраб олган. Корпус ичидаги қидирув нафақат талаффуз қилинаётган матн, балки имо-ишора (бошни қимирлатиш, елка қоқиш ва х.) ҳамда нутқий ҳаракат тури (розилик, пичинг, кесатик, киноя ва х.) орқали ҳам амалга оширилади. Қидирув натижаси видеолавҳани томоша қилиш ҳамда эшитиш имконини беради.

РУС ТИЛИНИНГ УПСАЛЛА КОРПУСИ (Упсальский корпус русского языка²⁷) – рус тилидаги илк корпус; 1980 йилларда Швециянинг Упсалла университетида асос солинган. 1 миллион сўз қўлланиш ҳолатида, турли жанрдаги 600 матндан иборат. Корпус замонавий (ҳозирги) рус тилининг ҳолатини ақс эттириб туриши лозим. Ушбу корпусни тузишнинг асосий мақсади, аввало, адабий тилни кўрсатиш сифатида белгиланган, шунинг учун корпусга оғзаки нутқ намуналари киритилмаган. Корпусга 1985-1989 йилги махсус матнлар (матндан парча эмас, балки тўлиқ ҳолда), 1960-1988 йилги бадиий адабиёт намуналари киритилган. Махсус матнларни танлашда матн мавзун, бадиий адабиёт намунасини танлашда муаллифнинг машҳурлигига алоҳида эътибор қаратилган. Матнлар лотин алифбосида киритилган; бугунги кунда ушбу корпус Тюбинген университетида 1990-2000 йилларда махсус SFB 441 илмий-тадқиқот сектори доирасида тузилган “Рус матн-ларининг Тюбинген корпуси” (“Тюбингенские корпусы русских текстов²⁸”) таркибида жойлашган. Корпус морфологик аннотацияланган, қидирув тизими сўз шакли ҳамда морфологик белгилар асосида амалга оширилиши, қидирув натижаси эса KOI8 ёки Windows-1251 коди асосида кириллча ҳамда лотинча ҳарфлар билан унинг транслитерацияси асосида бўлиши мумкин. Қидирув катта корпусларни бошқаришга мўлжалланган Штутгарт университети Тилни автоматик қайта ишлаш институти ишлаб чиққан SQR дастури билан амалга оширилади (Қаранг: Захаров В.П. Корпусная

²⁷ www.slaviska.uu.se/korpus.htm

²⁸ www.sfb441.uni-tuebingen.de/b1/rus/korpora.html

лингвистика в России. CrossLingua'2014 < 27 — 30. 05.14 < Crimea, Simferopol — Alushta.). Мазкур корпус Интернетдаги очик, фойдаланиш эркин бўлган, илк аннотацияланган корпус; якунланган(Қаранг: История корпусной лингвистики. (Электрон ресурс) - <https://www.myfilology.ru/177/istoriya-korpusnoj-lingvistiki/>).

СЕМАНТИК РАЗМЕТКА - ягона стандарт шакли бўлмаса ҳам, харф, рақам ёки фақат рақамдан иборат кодлардан иборат разметка. Биринчи харф ёки рақам умумий семантик маънони, кейинги белги эса сўз маъносини янада махсуслаштирувчи кичик семантик гуруҳни ифодалайди. Семантик разметка нафақат сўз, балки кўплаб бирикмаларни ҳам семантик гуруҳларга бирлаштиради, бундай пайтда турли бирикувдаги бир маънони билдирувчи бирикмалар битта белги билан кодланади. Идиоматик бирлик (ибора) таркибидаги сўзлар миқдорини билдирувчи ахборот ҳам разметкадан жой олади. Семантик разметка корпусдаги сўз маъносининг ихтисослашуви, омонимлик, синонимлик, маъновий гуруҳга ажратиш каби муаммоларни ҳал қилади (Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011.). Ру тили миллий корпусида ҳар бир сўзга уч хил – сўз разряди, лексик-семантик тавсиф, деривацион изох берилган.

СЕМАНТИК ТЕГ – семантик разметка учун изоҳлар мажмуи.

СИНТАКТИК РАЗМЕТКА – синтактик таҳлил, парсинг (ингл. parsing) натижаси, у компонентларининг грамматик структурасига асосланади. Гапдаги бўлақлар орасидаги синтагмалар график ва шажара тарзида, матнда эса улар отли, феълли ҳамда мураккаб бирикмаларни, содда, қўшма гапни кўрсатувчи очилувчи, ёпилувчи кавслар ёрдамида кўрсатилади. Синтактик разметкага эга корпуслар treebanks номи билан оммалашган. Морфологик разметка каби таҳлилни тезлаштириш мақсадида синтактик разметка ҳам соддалаш-тирилди, натижада бу усул skeleton parsing номини олди.

СИНТАКТИК ТЕГ – синтактик разметка учун изоҳлар мажмуи.

СИФАТ-ЛЕММА – сифат-лексема.

СОН-ЛЕММА – сон-лексема.

СУБКОРПУС – маълум бир корпуснинг автоном қисми.

СЎЗ БИРИКМА (COLLOCATIONS) – сўровни қаноатлантирувчи бирикма; (конкрет сўзшакл) конкордансда алоҳида рангда белгиланади.

СЎЗШАКЛ – сўзнинг матнда қўлланган шакли ёхуд танланган матндаги морфологик бирлик.

ТАГГИНГ – автоматик морфологик разметкалаш дастури.

ТАДҚИҚИЙ КОРПУС – мақсадига кўра тил ҳаракати (мавжудлиги)нинг турли аспектини ўрганувчи корпус. Корпуснинг бу типи кенг миқёсдаги лингвистик топшириқларни бажаришга мўлжалланган.

ТАДҚИҚОТНИНГ КОРПУС МЕТОДИ – тадқиқот натижаларини корпус материаллари асосида текшириш, шу асосда хулоса чиқариш; тадқиқот учун корпусдан материал танлаш усули. Корпусдан тадқиқот учун эркин фойдаланиш анъанасини Браун корпуси бошлаб берган.

ТАҚЛИД-ЛЕММА – тақлид сўз туркумига мансуб лексема.

ТАСДИҚЛОВЧИ ФИЛЬТР (P-filter) – сўров билан мос келувчи сатрнинг конкорданс рўйхатида берилиши.

ТАЪЛИМИЙ КОРПУС – хорижий ва она тили таълими воситаси сифатида қараладиган, тил ўрганиш ва ўргатиш усул, имкониятларига эга корпус.

ТЕГ – лингвистик восита. SGML тег концепциясини олиб кирди. “Тег” (ингл. tags) – матндаги ишчи изоҳ, у матн ҳақидаги маълумотни камраб олади. Корпус ёрдамида статистик ҳисоб жараёнида тилимизда мавжуд сўзларнинг частота-сини аниқлаш, турли маълумотларни олиш мумкин. Масалан, ҳар бир сўз билан билан ёнма-ён унинг туркуми белгиланган бўлса, тилда турли нуткий вазиятда сўз туркумларининг қўлланилиш даражасини аниқлаш мумкин. Лингвистик разметка ҳар бир сўзнинг маълум кодга эга бўлиши билан характерланади. Ушбу код *тег*, сўзни кодлаш эса *таггинг* (ингл. tagging)²⁹ дейилади. Айни пайтда матнга лингвистик ва бошқа маълумотларни қўшишнинг умумэтироф қилинган стандарти мавжуд эмас,

²⁹ <http://rykov-cl.narod.ru/c.html>.

лекин Text Encoding Initiative (TEI) махсус халқаро лойиҳаси разметканинг стандарт воситасини ишлаб чиқишга мўлжалланган. Бунинг учун ҳужжат разметкасининг бутун халқаро қабул қилинган тили – SGML ва XML мавжуд. XML кенгайтмаси³⁰ XML (инг. Extensible Markup Language) базаси асосида қурилган ҳамда луғат тег ва атрибути, коидалар базасини қамраб олган аниқ грамматика тили. SGML³¹ эса ҳужжат учун разметка тилини аниқлайдиган метатил (метаязык). Анъанавий теглар жуфт ҳолда қўлланувчи (очилувчи, ёпилувчи) учбурчак қавсдан иборат бўлади. Масалан, <f> очилувчи, </f> ёпилувчи тег, очилувчи тег бераётган маълумотнинг ёпилганлигини билдиради. <ds> Бу, биринчи навбатда сизга боғлиқ </ds>, деди у.

Бу, биринчи навбатда сизга боғлиқ, -деди у матни “кўчирма гап” (инг., direct speech) маъносини берувчи <ds>, унинг ичида келган биринчи навбатда киритмаси ажратувчи белгиси (тег) орқали ажратилган. Теглар якка ҳолда (жуфтлашмай) ишлатилиши ҳам мумкин. Масалан, оғзаки корпусда маълум ўринда тўхтамни билдириш учун <pause> теги ишлатилади. Бу якка тег ҳисобланади. Тегнинг ўзи фойдаланувчига кўринмайди, разметқаланган матнни кўрсатаётган дастур тегни қабул қилинган, айни шу матн учун қўллашга келишилган шаклда кўрсатиши мумкин.

ТЕГНИНГ ЛИНГВИСТИК МОДЕЛИ – тег қисқартмаси. Корпус таркибидаги сўзга бириктириладиган изоҳ шартли қисқартма шаклида бўлади, корпус тузувчи бу шартли қисқартмани ўзи танлайди. Масалан: аслий сифат = [ас. с.], нисбий сифат = [нис. с.]. Разметка кўриниши: *Разметка кўриниши: йирок <йирок> [сиф.], [ас. с.], [од.д.], [с.с.], [т.с.], [хус. ЛМГ].*

ТЕРМИНОЛОГИК КОРПУС – бевосита жонли тил асосида терминлар луғати яратиш мақсадида тузилган корпус.

³⁰ <https://ru.m.wikipedia.org/wiki/XML>.

³¹ <https://ru.m.wikipedia.org/wiki/SGML>.

ТИЖОРАТ КОРПУСИ (пуллик корпус) – фойдаланиш on-line режимда ишлаш рухсатномасини талаб қиладиган ёки фойдаланувчи корпуснинг диск вариантини сотиб олиш орқали ишлатиши мумкин бўлган корпус. Корпусни олишдан олдин корпус аннотацияси билан танишиш, ҳатто синон вариантда корпуснинг маълум қисми билан ишлаб кўриш мумкин.

ТИЛ КОРПУСИ – маълум тилнинг белгиланган даврдаги, хилма-хил жанр, ранг-баранг услуб, ҳудудий ҳамда ижтимоий вариантдаги матнларининг электрон шаклдаги махсус дастурий таъминот асосидаги йиғиндиси; муайян тилнинг ўзига хос хусусияти, вариантларини акс эттирувчи бир неча белги асосида танлаб олинган электрон шаклдаги матн парчалари, лингвистик тадқиқот учун асос вазифасини ўтайдиган тизим. Инглиз тилида linguistic corpus ёки text corpus каби атамалар билан қўлланилади (Қаранг: Кутузов А.Б. Курс “Корпусная лингвистика”. – Москва, 2008. – 26 с.). Унга “Корпус – маълум мақсадда йиғилган матнлар мажмуини ташкил этувчи тил бирликларининг йиғиндиси”, “корпус – алоҳида тилнинг ёзиб олинган оғзаки ҳамда ёзма матнлари асосидаги лингвистик маълумотлар мажмуи”, “корпус – исталган табиий (реал) тилдаги электрон шаклда сақланадиган ёзма ёки оғзаки, компьютерлаштирилган кидирув тизимига жойлаштирилган матнлар йиғиндиси” каби таърифлар берилган. Жаҳон тилларининг жуда кўпчилиги мукамаллик даражаси, матнни (илмий) қайта ишлаш имконияти билан фарқ қилувчи ўз миллий корпусларига эга. Бугунги кунда лингвистик тадқиқот ва амалий топшириқлар ечими учун тил корпуслари замонавий тилшуносликнинг инкор этиб бўлмас иш қуролига айланди, чунки корпус турли лингвистик топшириқларни ечишга хизмат қилади (Қаранг: Недошвина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург, 2006. –26 с.). Корпус оддий электрон кутубхона (мас., “Библиотека Максима Мошкова”, “Русская виртуальная библиотека”, “Фундаментальная электронная библиотека”, “ZiYONET кутубхонаси”, “Хуршид Даврон кутубхонаси” кабилар)дан фарқ қилади. Бундай электрон кутубхоналар халқнинг ижтимоий-сиёсий,

маънавий, иқтисодий ҳаётини акс эттирувчи бадий ҳамда публицистик асарларни нисбатан тўлароқ қамраб олишни мақсад қилади. Электрон кутубхона матнлари тилшунослик нуқтаи назаридан лексик, морфологик, грамматик, семантик белгилар асосида ишлов берилмаганлиги сабабли лингвистик тадқиқотларда фойдаланиш ноқулай. Электрон кутубхона илмий тадқиқот материали базасини тайёрлаш учун тузилмайди; у маълум бир миллат ёки бутун инсониятга тегишли миллий-маънавий меросни тўплаш, марказлаш-тириштиришни мақсад қилган бўлади. Тил корпуси эса электрон кутубхонадан фарқли равишда муайян тилни тадқиқ қилиш, ўрганиш ва ўргатиш учун зарурий, фойдали ва қизиқарли матнлар тўпланишини назарда тутди. Бундай матнлар эса нафақат бадий адабиёт намунаси, балки телефондаги оддий суҳбат ёзуви ёки ижара шартномаси матни бўлиши ҳам мумкин. Корпусни электрон кутубхона-дан ажратиб турувчи биринчи фарқ ундаги матннинг хусусияти (қўшимча маълумот билан бойитилганлиги) ҳисобланади ҳамда бу белги корпуснинг алоҳида қисми – корпус аннотацияси (рус., разметка)ни ташкил этади.

ТОЖИК ТИЛИ МИЛЛИЙ КОРПУСИ – Тожикистон Республикаси Фанлар академияси Информацион ва коммуникацион технологиялар маркази Ахборот ва коммуникацион технологиялари мактаби ҳамкорлигида яратилган тожик тили миллий корпуси. Корпус таркибида энг машхур форс тожик адабиёти намояндалари Рудакий, Фирдавсий, Умар Хайём, Ҳофиз, Румий ижоди корпуси, замонавий адабиётнинг турли жанрлари, тожик радио ва телевидениеси орқали эфирга узатилган материал матни; турли газета, журнал, информацион бюллетень материаллари, илмий, илмий-оммабоп, ишлаб чиқаришга оид расмий-административ, юридик матнлар, тожик тилининг изоҳли, таржима луғатлари матнлари, ижтимоий аҳамиятга эга матнлар(ёзишма, мулоқот)дан таркиб топган.

ТЎЛИҚ МАТНЛИ КОРПУС – матн ҳажмига кўра корпус тури. Тўлиқ матнли корпусга маълум муаллиф асарлари корпуси, қисқа матнлар

корпусини мисол келтириш мумкин. Газета сарлавҳалари корпуси ҳам шунга мисол бўлади.

ТУРҒУН КОРПУС – корпус таркибининг тўлдириб борилишига кўра корпуснинг бир тури. Дастлабки корпуслар тилнинг турғун ҳолати акси, тилнинг маълум бир даврини қамраб олган. Нисбатан аҳамиятли замонавий корпус саналган Британия миллий корпуси (BNC) ҳам турғун корпусдир. Бундай корпуснинг типик намунаси Муаллифлик (авторлик) корпуси, яъни бир ёзувчи асарлари йиғиндиси саналади. Баъзи лингвистик амалларни бажариш учун (масалан, сўз частотасини аниқлаш) турғун корпус талабга жавоб беролмай қолади. Чунки фақат чегараланган давр қамраб олинган бўлади (Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. –Иркутск: ИГЛУ, 2011.). Шу эҳтиёждан келиб чиққан ҳолда кейинчалик тўлдириб бориладиган динамик (мониторинг) корпус ҳам пайдо бўла бошлади.

УНДОВ-ЛЕММА – ундов сўз туркумига мансуб лексема.

УНИВЕРСАЛ КОРПУС – яратилиш мақсадига кўра корпус кўриниши; турли жанрдаги матнларни қамраб олади; миллий корпуслар универсал корпуслар сирасига қиради.

Ф.М.ДОСТОЕВСКИЙ МАТНЛАРИ КОРПУСИ – Достоевский асарлари тили луғати тузиш манбаи сифатида яратилган, корпус тузиш параметрларига луғат мақолаи тузилиши ва таркиби асос бўлган корпус. Луғат Достоевский лексика-сини тўлиқ қамраб олишни мақсад қилганлиги сабабли корпусга муаллифнинг барча бадиий, публицистик асарлари киритилган. Унда матн сақлавиш бирлиги алоҳида асар (романдан тортиб хатгача) шаклида белгиланган. Бугунги кунда барча электрон ташувчи воситаларда унинг бадиий асарлари, публицистикаси, эпистоляр мероси мавжуд; асар ва саҳифага ҳавола қилувчи минимал разметка билан ASCII форматида сақланади. Корпус асосида бир нечта луғат яратилган, муаллиф идиомалари луғатига асос бўлган идиоматик маълумотлар базаси ҳам тузилган. Достоевский корпуси CD-ром шакли (“Достоевский: Тексты, исследования, материалы”)га ҳам эга.

ФЕЪЛ-ЛЕММА – феъл-лексема.

ХАЛҚАРО ИНГЛИЗ ТИЛИ КОРПУСИ³² (The International Corpus of English (ICE)) – ўнлаб университетларнинг ҳамкорликдаги лойиҳаси, ҳар бири миллионлаб сўз қўллаш ҳолатини қамраб олган, 20та параллел ички корпусдан иборат. Корпусдан инглиз тили давлат тили ёки иккинчи тил ҳисобланган давлатларнинг (Австрия, Канада, Янги Зеландия ва х.) ўзига хос хусусиятини ўрганиш мумкин. Бу корпусни таҳлил қилиш учун махсус мукамал дастурий таъминот ишлаб чиқилган.

ХЕЛЬСИНКИ АННОТАЦИЯЛАНГАН КОРПУСИ (Хельсинкий аннотированный корпус³³ (ХАНКО)) – 2000 йилларда Хельсинки университетида “Рус тилининг функционал синтаксиси” (раҳбар проф. А. Мустайоки) лойиҳасининг бир қисми сифатида иш бошлаган, доимий тўлдириб борилаётган корпус. Ҳажми 100 минг сўзшаклни ташкил этган, унинг асосини 2001 йил “Сарҳисоб”и журнали мақолалари ташкил қилган. Корпус³⁴дан интернетда эркин фойдаланиш мумкин. Морфологик, синтактик разметка ҳамда шу параметрли кидирув мавжуд. Корпуснинг ўзига хослиги маълумотлар лингвистик тавсифининг пухта ишланган ҳамда автоматик разметка натижаси тўлиқ, бирма-бир қўлда текшириш имконияти мавжудлигида. Бу билан грамматик омонимияни тўлиқ кўрсатишга эришилади. Корпуснинг синтактик разметкаси гап бўлаги, грамматик алоқани аниқлаб бериш вазифасини бирга олиб бориши лозим. Бу корпус бошқаларидан фарқли ўларок, шу вазифани жуда яхши бажаради. Корпусда тахминан 2000 бирлик атрофида мураккаб гап бўлақларини ҳосил қилувчи турғун бирикма ҳам аниқланган. Семантик категорияни белгилайдиган терминлар асосида семантик разметкани амалга ошириш кўзда тутилмоқда. Корпус, энг аввало, таълимий мақсадни кўзлайди. 2000 йилларда Лидс университети(Буюк Британия)нинг таржимонлик тадқиқот-лари марказида турли тилларнинг (инглиз, араб, хитой, француз, немис, итальян, япон, испан,

³² <http://www.ucl.ac.uk/english-usage/ice/>

³³ [http:// www.ling.helsinki.fi/projects/hanco/](http://www.ling.helsinki.fi/projects/hanco/).

³⁴ Ўша манба.

поляк ва ҳ.) катта ҳажмли корпуслари тузилган³⁵. Улар орасида рус тили корпуси³⁶ ҳам мавжуд. 116 миллион сўзни қамраб олган. Рус тили миллий корпусининг бу версияси асосида “Рус тилининг частотали луғати” (О.Н.Ляшевская, С.А.Шаров таҳрири остида) фойдаланувчилар қўлига етиб борди. Бу сайтда Рус газеталари корпуси (2001-2004 йиллар, 76 миллион сўз), Рус Интернет матнлари корпуси (160 миллион сўз), бизнес-корпус (12 миллион сўз) ҳамда санаб ўтилганларни жамлаган яхлит корпус ҳам мавжуд.

ЧАСТОТАЛИ РЎЙХАТ – интерфейсидаги корпусдаги сўзларнинг частотасига ишора берувчи ойна.

ШЕЪРИЙ МАТНЛАР КОРПУСИ – шеърий матнлардан таркиб топган корпус. Тоҷик тили миллий корпуси таркибидаги Рудакий, Ҳофиз, Фирдавсий, Саъдий корпуслари шеърий корпус саналади.

ЭКСТРАЛИНГВИСТИК РАЗМЕТКА – матн ҳақидаги маълумотлардан иборат изоҳ; унинг қуйидаги турлари фарқланади: Матн форматининг ўзига хослигини ақс эттирувчи (боб, хатбоши, қисм ва ҳ.) разметка (1). Матн ва унинг муаллифига тегишли маълумотни ифодаловчи разметка (2). Муаллиф ҳақидаги маълумот ном, ёш, жинс, яшаган даври каби маълумотларни билдириши мумкин. Матн ҳақидаги маълумот эса матн (асар) номи, тили, ёзилган ҳамда нашр этилган йилини қамраб олади. Бундай маълумотларнинг мавжудлиги базада деталлаштирилган қидирувни амалга ошириш имконини беради. Экстралингвистик разметка ёки метааҳборот (русча: метаданные) – ташқи-интеллект-туал маълумотларни қамраб олувчи, библиографик, типологик, тематик, социологик тавсифни; формал-структур разметка(матн, бўлим, боб, қисм, абзац, гап)ни, шунингдек, техник-технологик разметка (кодировка, манбанинг электрон версиясини қайта ишлаш санаси)ни бирлаштирувчи разметка тури. Метааҳборот мажмуи фойдаланувчига унинг имкониятларини аниқлаб беради. Бу маълумотларни танлашда корпус тузувчиси тадқиқот мақсади, тилшунослар талаби ҳамда

³⁵ <http://corpus.leeds.ac.uk/>

³⁶ <http://corpus.leeds.ac.uk/ruscorpora.html>

матнга у ёки бу қўшимча белгини қўшиш имконияти билан таниш бўлмағи лозим. Ташқи интеллектуал разметка, биринчидан, тилнинг ўзаро алоқаси, мавжудлигини аниқлаш; иккинчидан, тилнинг ўзига хос хусусиятини ўрганиш учун керак. Тилга таъсир қиладиган иккита: ташқи, яъни нолисоний ҳамда ички омил мавжуд. Дж. Синклер ташқи омилнинг уч гуруҳини ажратади (Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011.): Муаллиф томонидан матн яратишга алоқадорлик омили (1). Матннинг ташқи белгиларига тегишли омил (2). Матннинг яратилишига сабаб бўлувчи, аудиторияга таъсир қилувчи омил (3). Ички омилнинг икки гуруҳи – матннинг мавзу доираси, услубий хусусияти (услуб ва жанр) ажратилади. Масалан, рус тили миллий корпусида қуйидаги метаахборотлар мажмуи мавжуд: Биринчи блок: матн муаллифи: исми, жинси, туғилган вақти (тахминий ёши) (а); матн номи (б); ёзилиш вақти, жойи (аниқ шаҳар ёки давлат бўлиши мумкин) (в); матн ҳажми (г): бадий асарлар учун меъёр сифатида ҳикоя ҳажми камида 5000 сўз; қисса ҳажми 5000дан 15.000 сўзгача; роман 15.000 сўздан ортиқ бўлиши одат тусига кирган. Иккинчи блок: метаизох корпусдаги матннинг 3 асосий кўриниши – бадий матн, нобадий матн, драматик асарларни фарқлашга мўлжалланган. Рус тили миллий корпусида бадий асар учун қуйидаги маълумотлар кўрсатилади:

- матн жанри: автобиографик проза, детектив, болалар адабиёти, тарихий, криминал адабиёт, саргузашт, фантас-тика, юмор, сатира;
- матн типи: автобиографик проза, латифа, детектив, очерк, адабий мактуб, қисса, масал, пьеса, ҳикоя, роман, эртак, триллер, эпопея, эссе ва х.;
- матн хронотопи: тасвирланаётган воқеанинг тахминий вақти ва жойи.

ЭЛЕКТРОН АРХИВ – электрон ташувчи курилмадаги матнлар жамланмаси; стандартланмаган, бир хир шаклга келтирилмаган.

ЭЛЕКТРОН КУТУБХОНА – бир турдаги матнлар келтирилган; стандарт шаклга келтирилган. Халқнинг ижтимоий-сиёсий, маънавий, иқтисодий ҳаётини акс эттирувчи бадий ҳамда публицистик асарларни нисбатан тўлароқ қамраб олишни мақсад қилади (Қаранг: <http://wikipedia.org/wiki/>

Корпусная лингвистика. Статья в интернет-энциклопедии Википедия.). Электрон кутубхона матнлари тилшунослик нуктаи назаридан лексик, морфологик, грамматик, семантик белгилар асосида ишлов берилмаганлиги сабабли лингвистик тадқиқотларда фойдаланиш учун ноқулай.

ЭРКИН КОРПУС – эркин фойдаланиш мумкин бўлган корпус; on-line режимда хоҳлаган пайтда исталган матннинг тўлиқ шакли билан ишлаш имконини беради. Баъзан эркин корпус ҳам корпус маълумотларининг маълум қисмидангина фойдаланиш билан чегараланган бўлади.

ЮКЛАМА-ЛЕММА – юклама-лексема.

GIGAWORD CORPORA³⁷ – инглиз, араб, хитой ва бошқа тилларнинг мониторингли корпуси. Европа Иттифоқи хомийлигида, уни Linguistic Data Consortium компанияси тузган, корпус устида узлуксиз иш олиб боради. Ҳозирда корпус ҳажми 1 миллиард сўз ҳолатида. Матнлар янгиликлар, публицистик-тикани қамраб олган, энг қиммат корпус. Таркибидаги корпуслар ҳақида тўлиқ маълумотни <http://www.ldc.upenn.edu> сайтидан олиш мумкин.

SGML – матнни турли томондан (alternative views) таҳлил қила оладиган разметка типи. В.П.Заҳаров SGML типи разметкасининг ўзига хос имкониятидан бири матннинг турли кўринишини намойиш эта олиш эканлигини таъкидлайди. Бир марта разметкаланган матн қўйиладиган топшириққа биноан турли шаклда берилиши мумкин. Корпусдан фақат кўчирма гапли қўшма гапнинг муаллиф гапини ажратиб олиш керак бўлса, матнни кўриб чиқаётган дастур <ds> белгиси ичига кирган қисми яширади, натижа қуйидагича кўринади: - *деди у*. Ёки кўринаётган матнда муҳим қисм (тадқиқ манбаси) кўк рангда, муаллиф гапи қора рангда берилиши буюрилади. Натижа қуйидагича кўринади: *Бу, биринчи навбатда сизга боғлиқ*, - *деди у*. Бундан ҳам мураккаброк “alternative views”ни кўришимиз мумкин. Масалан, драматик асарда турли қаҳрамон нутқи алоҳида теглар билан белгиланган, керакли ҳолда қаҳрамон нутқи ёки маълум икки қаҳрамон диалоги руйхати битта буйруқ билан ажратиб олиниши мумкин.

³⁷ <http://www.ldc.upenn.edu>

THE LONGMAN CORPUS NETWORK³⁸ – Логман (Longman) компанияси, Ланкастер университети томонидан тузилган корпус; бир неча тижорат корпуси маълумотлар базасидан иборат, 50-100 миллион сўз қўллаш ҳолатини қамраб олган.

FOR AUTHOR USE ONLY

³⁸ <http://www.pearsonlongman.com/dictionaries/corpus/index.html>

Корпусларнинг қиёсий таҳлили

		Рус тилининг Упсала корпуси ³⁹	Рус тили газеталар компьютер корпуси
	Таркиби	1996-2002 йиллар публицистикаси; XIX-XX аср бадий адабиёти	1997 йил газета матнлари
	Ҳажми	25 миллион сўз атрофида	100 минг сўз атрофида
	Сўз ясалиши	+	-
	синтактик	элемент	элемент
	семантик	-	элемент
Қидирув шакли			
	Сўз шакли орқали	+	+
	Лемма орқали	-	+
	Бир неча сўзни баравар	+	-
	Грамматик белги асосида	+	+
0	Семантик белги асосида	-	+
1	Тиниш белги асосида	+	+

³⁹ Ушбу корпус Интернетда “Рус матнларининг Тюбинген корпуси” деган ном билан аталади.

1.2-жадвал

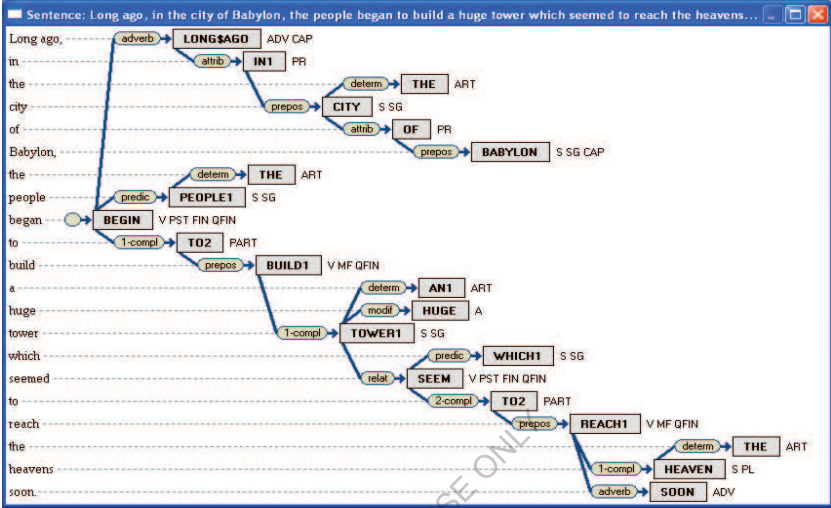
Корпус турлари таснифи

Тасниф белгиси	Корпус турлари
Нутк турига кўра	Оғзаки, ёзма, аралаш
Матн тилига кўра	Русча, инглизча ва ҳ.
Параллеллигига кўра	Бир тилли, икки тилли, кўп тилли
Матннинг ихтисослашувига кўра	Бадиий, диалектал, суҳбат, терминологик, аралаш
Жанрига кўра	Адабий, фольклор, драматик, публицистик
Кириш усулига кўра	Эркин, пулик, ёпик
Мақсадига кўра	Тадқиқий, тавсифий
Изоҳланишига кўра	Кўшимча изоҳли, изоҳсиз
Ўзгарувчанлигига кўра	Ўзгарувчан, турғун
Структурасига кўра	Марказий, архив, маҳаллий
Изоҳ хусусиятига кўра	Морфологик, синтактик, семантик
Матн ҳажмига кўра	Тўлиқ матнли, лавҳа матнли
Хронологик жиҳатига кўра	Синхрон, диахрон
Умумийлигига кўра	Умумий, яқка муаллифли

2.1-жадвал

Муаллиф	Ҳофиз корпуси (<i>Пайкараи Хофиз</i>)	Саъдий Шерозий корпуси (<i>Пайкараи Саъди</i>)	Рудакий корпуси (<i>Пайкараи Рудаки</i>)	Фирдавсий корпуси (<i>"Шоҳнома"и Фирдавси</i>)	Румий корпуси (<i>Пайкараи Мавлоно</i>)
Интерфейс таркиби	1.Биография 2.Ғазаллар 3.Расмлар 4.Луғат	1.Таржимаи ҳол 2.Қасидалар 3.Ғазаллар 4.Рубойлар 5.Китоблар 6.Расмлар 7.Биз билан алоқа	1.Қасида, ғазал, парчалар 2.Рубой 3.Парчалар 4.Рудакий асарлари бўйича On-line тест	1."Шоҳнома"нинг 9 жилди 2.Қидирув 3.Луғат	1.Биография 2.Маснавий 3.Рубой 4.Танланган рубойлар 5.Луғат 6.Суратлар 7. Линклар

(занжирсимон синтактик разметка)



Фойдаланилган адабиётлар рўйхати

I. Норматив-ҳуқуқий ҳужжатлар

1. “Ўзбекистон Республикаси давлат тили ҳақида”ги Қонуни.–Т.: Ўзбекистон, 1989.
2. Ўзбекистон Республикаси Президентининг “Маънавий-маърифий ишлар самарадорлигини ошириш ва соҳани ривожлантиришни янги босқичга кўтариш тўғрисида”ги Қарори // Халқ сўзи, 2017 йил, 29 июль.

II. Ижтимоий-сиёсий адабиётлар

3. Каримов И.А. Юксак маънавият – енгилмас куч. – Тошкент: Маънавият, 2008. – 176 б.
4. Каримов И.А. Она юртимиз бахту иқболи ва буюк келажаги йўлида хизмат қилиш – энг олий саодатдир. – Тошкент: Ўзбекистон, 2015.– 304 б.
5. Мирзиёев Ш.М. Эркин ва фаровон, демократик Ўзбекистон давлатини биргаликда барпо этамиз//Ўзбекистон Республикаси Президенти лавозимидаги киришиш тантанали маросимида бағишланган Олий Мажлис палаталарининг қўшма мажлисидаги нутқ. – Тошкент: Ўзбекистон, 2016. – 56 б.
6. Мирзиёев Ш.М. Буюк келажагимизни мард ва олижаноб халқимиз билан бирга қурамиз. – Тошкент: Ўзбекистон, 2017. – 488 б.
7. Мирзиёев Ш.М. Танқидий таҳлил, қатъий тартиб-интизом ва шахсий жавобгарлик – ҳар бир раҳбар фаолиятининг кундалик қоидаси бўлиши керак.– Тошкент: Ўзбекистон, 2017. – 104 б.
8. Мирзиёев Ш.М. Қонун устуворлиги ва инсон манфаатларини таъминлаш – юрт тараққиёти ва халқ фаровонлигининг гарови. – Тошкент: Ўзбекистон, 2017. – 48 б.

III. Илмий рисола, монография, дарслик, ўқув ва методик қўлланмалар

9. Абдурахмонов Ғ., Рустамов А. Қадимги туркий тил.–Тошкент: Ўқитувчи, 1982.–168 б.

10. Боровков А.К. Лексика среднеазиатского тефсира XI-XIII вв. ИВЛ.– Москва, 1963.– 268 с.
11. Грудева В.А. Корпусная лингвистика. Учебное пособие. 2-е изд., стереотип.– М.: Флинта, 2012. – 165 с.
12. Добровольский О.Д., Кретов А.А., Шаров С.А. Корпус параллельных текстов: архитектура и возможности использования // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005. – С. 263-296.
13. Замонавий ўзбек тили: Морфология. (Муаллифлар жамоаси. Масъул мухаррирлар: Ҳ.Неъматов, Р.Сайфуллаева.) – Тошкент: Мумтоз сўз, 2008. – 468 б.
14. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с.
15. Захаров В.П., Богданова С.Ю. Корпусная лингвистика.– Иркутск: ИГЛУ, 2011. – 161 с.
16. Камшилова О.Н. Учебный корпус текстов: потенциал, состав, структура. – СПб.: Книжный Дом, 2012.– 58 с.
17. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Прометей, 1990. – 367 с.
18. Casares J. Dissionario ideologico de la lengua Espanola.– Barselona, 1969. – 887 с.
19. Крушевский Н.В. Избранные работы по языкознанию. – М.:Наследие, 1998.–296 с.
20. Курбонова М., Сайфуллаева Р., Бокиева Г., Менглиев Б. Ўзбек тилининг структурал синтаксиси. – Тошкент, 2004. – 143 б.
21. Loriche R. Le mot et l'idee. – Paris, 1965.– 178 p.
22. Малов С.Е. Памятники древнетюркской письменности. – М.-Л. 1951. – 188 стр.
23. Мухин М. Ю. Лексическая статистика и концептуальная система автора: М. Булгаков, В. Набоков, А. Платонов, М. Шолохов. – Екатеринбург, 2010.– 174 с.

24. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб., – 2006. – 26 с.
25. Нейматов Х., Ҳамроева Ш. Лингвистик тадқиқот методикаси, методологияси ва методлари. – Бухоро, 2004. – 89 б.
26. Нейматов Х., Расулов Р. Ўзбек тили систем лексикологияси асослари. – Тошкент: Ўқитувчи, 1995. – 128 б.
27. Поляков А.Е. Словарь языка А.С. Грибоедова.– М.: Языки славянской культуры, 2008.– 408 с.
28. Собиров А. Ўзбек тилининг лексик сатҳини системалар системаси тамойили асосида тадқиқ этиш. – Тошкент: Маънавият, 2004. –140 б.
29. Слово и словарь. Vocabulum et vocabularium:// Сб. науч. трудов по лексикографии. (под ред. В.В.Дубичинского и Т.Ройтера.) — Харьков: Издавництво, 2011 – 230 с.
- 30.Турсунов У., Ўринбоев Б., Алиев А. Ўзбек адабий тили тарихи. – Т.: Ўқитувчи, 1995. – 162 б.
31. Федосюк Ю.А. Что непонятно у классиков или Энциклопедия русского быта XIX века. –М., 1998.– 264 с.
- 32.Чехович О.Д. Бухарские документы XIV в.– Ташкент, 1965.– 366 с.
33. Щербак А.М. Грамматический очерк языка тюркских текстов X-XIII вв. из Восточного Туркестана.– М.–Л., 1961.– 204 с.
34. Щербак А.М. Грамматика староузбекского языка. М.–Л., 1962. – 276 с.
35. Юдахин К.К. Материалы к вопросу о звуковом составе чигатайского языка//Культура и письменность Востока.– книга. 4.– Баку, 1929. – 298 с.

IV.Матбуот материаллари ва илмий мақолалар

36. Боровков А.К. Алишер Навои как основоположник узбекского литературного языка// сб. “Алишер Навои”.– М. – Л., 1946.– 312 с.
37. Бобунова М.А. Конкорданс XXI В.: новая старая форма // Вопр. лексикографии.– 2016.– №2 (10).– С.41-54.

38. Добрушина Н.Р. Информационные технологии в гуманитарном образовании: Национальный корпус русского языка // Вопросы образования. – 2006. – № 4.
39. Добрушина Н.Р., Левинзон А.И. Национальный корпус русского языка в образовании// Cuadernos de Rusística Española. – 2006. – № 2.
40. Кукушкина О.В., Поликарпов А.А., Суровцева Е.В. Электронный корпус текстов художественных произведений А. П. Чехова: принципы организации и возможности лексикографического использования// Слово и словарь. *Vocabulum et vocabularium. Сборник научных трудов по лексикографии. (Под ред. В.В.Дубичинский.) – Вып.:№12.–Харьков-Клагенфурт, 2011.–416 с.
41. Лапонина Л.В., Суровцева Е.В. Тематическая разметка в корпусе художественной прозы А.П.Чехова // Чехов и мировая культура: взгляд из XXI века. Тезисы докладов международной научной конференции (Москва, 29 января–3 февраля, 2010 года) / Сост. и ред. Р.Б.Ахметшин, М.О.Горячева, В.Б.Катаев (отв.ред.).– М.: МГУ, 2010. – 308 с.
42. Махмудов. Н. Абдулла Қаҳҳор ҳикояларининг лингвопоэтикасига доир// Ўзбек тили ва адабиёти, 1987, 4-сон.
43. Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики/ Русский язык в научном освещении. – М., 2008. –№2 (16). – С. 7-20.
44. Рязанова Е.А. Использование языкового лингвистического корпуса в формировании грамматических навыков речи в языковом вузе // Вестник ТГУ.– вып. №10 (114). – 2012.
45. Тенишев Э.Р. Хозайтсвенные записи на древнеуйгурском языке // Сб. “Исследования по грамматике и лексике тюрских языков”.– Ташкент, 1965.– 267 с.
46. Труфанова М.В. Применение цифрового словаря с корпусом для работы над лингвистическими проектами по немецкому языку // Социально-экономические явления и процессы. – Харьков: Видавництво. – 2011. – №8 – с.229-232.

47. Щерба Л.В. Опыт общей теории лексикографии//Известия АН СССР.– ОЛЯ, 1940.– №3.

48. Якубовский К.И, Якубовская К.А. Обзор современных лингвистических технологий и систем // Вестник МГУП. –2015. –№2.– С.315-319.

V.Диссертациялар ва диссертация авторефератлари

49. Бобокалонов Р. Р. Ўзбек тилида семантик-функционал шаклланган сўз-гаплар: Филол. фанлари номзоди... дис. автореф. – Тошкент, 2000. – 19 б.

50. Мухин М.Ю. Лексическая статистика и идиостиль автора: корпусное идеографическое исследование (на материале произведений М. Булгакова, В.Набокова, А. Платонова и М. Шолохова): Автореферат диссертации на соискание ученой степени доктора филологических наук. – Екатеринбург, 2011. – 46 с.

51. Тожиева Г.Н. Ўзбек тили маънавий-маърифий лексикасининг мустақиллик йилларидаги тараққиёти (“Миллий ғоя” атов бирликлари тизими). Фалсафа доктори (PhD) диссертацияси автореферати.– Самарқанд, 2017.– 43 б.

52. Ҳамраева Ё.Н. Ўзбек тилининг идеографик луғатини тузиш тамойиллари (харакат ва ҳолат тушунчасини ифодаловчи сўзлар мисолида). Филол. фан. номз. дисс. –Тошкент, 2009.– 45 б.

53. Рустамов А. Алишер Навоий тилининг фонетик-морфологик хусусиятлари. – Докторлик дис. автореферати.– Тошкент, 1966.– 74 б.

VI.Луғатлар

54. Каримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Алфавитли луғат. – Т.: ЎзМЭ, 2007. – 434 б.

55. Каримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Терс луғат. – Т.: ЎзМЭ, 2007. – 423 б.

56.Каримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Частотали луғат. – Т.: ЎзМЭ, 2007. – 420 б.

57. Идеографический словарь русского языка. (Сост. О.С.Баранов.) – М., 1995. – 4166 статей.
58. Словарь языка Пушкина: в 4 т. / Отв. ред. акад. АН СССР В. В. Виноградов. – 2-е изд., доп. / Российская академия наук. Ин-т рус. яз. им. В. В. Виноградова. – М.: Азбуковник, 2000. – 1296 с.
59. Толковый словарь русского языка с включением сведений о происхождении слов. (Отв. редактор Н.Ю.Шведова.) – М.: Азбуковник, 2011. – 1175 с.
60. Толковый словарь русского языка. В 4 т.; под ред. Д.Н.Ушакова. – М.: Государственный институт СЭ, – Т.1. М., 1935; Т.2. М., 1938; Т.3. М., 1939; Т.4. М., 1940. (Переиздавался в 1947-1948 гг.) - 88 239 статей.
61. Частотный словарь-конкорданс словоформ языка Ф. И. Тютчева”. (Создан по изданию: Тютчев Ф.И. Полное собрание стихотворений / Сост., подгот. текста и примеч. А.А. Николаева.) – Л.: Сов. писатель, 1987. – 448 с.
62. March F.A. March’s Thesaurus Dictionary. – N.Y., 1958.– 1312 p.
63. Roget P.M. Thesaurus of English words and phrases. – Lnd., 1852. – 1258 p.
64. Ахманова О.С. Словарь лингвистических терминов.– М.:СЭ, 1969.– 608 с.
65. Зализняк А.А. Грамматический словарь русского языка. – М.: Русский язык, 1980. – 880 с.
66. Засорина Л.Н. (ред.) Частотный словарь русского языка.–М., 1977.– 936 с.
67. Лингвистический энциклопедический словарь. (Гл. ред. В.Н.Ярцева) – М.: СЭ, 1990.– 686 с.
68. Словарь языка А. С. Грибоедова. Т.1: А – 3 (Сост. А.Е.Поляков.) – М.; Языки славянской культуры, 2008. – 432 с.
69. Частотный словарь рассказов А.П.Чехова. (авт.-сост. А.О.Гребенников; Под ред. Г.Я.Мартыненко.) – СПб., 1999.– 172 с.
70. Частотный грамматико-семантический словарь языка художественных произведений А. П. Чехова. М.:МАКС ПРЕСС, 2012. – 571 с.
71. Шайкевич А.Я., Андриющенко В.М., Ребецкая Н.А. Статистический словарь языка Достоевского. – М., 2003.– 832 с.

VII. Интернет материаллари, электрон ресурслар

77. www.ruslang.ru/seminar_aut_lexocorg250214) Волков С.С., Матвеев Е.М. “О проекте словаря “Риторика М.В.Ломоносова”. Доклад (ИЛИ РАН).
78. www.ruslang.ru/seminar_aut_lexocorg Гик А.В. “О работе “Конкордансом к стихотворениям М.Кузмина”: том четвертый.
79. <http://www.gramota.net/materials/1/2008/8-2/50.html>. Добровольский Д.О., Кретов А.А., Шаров С.А. Корпус параллельных текстов. Архитектура и возможности использования.
80. www.dialog-21.ru/media/2138/zakharov.pdf Захаров В.П. Корпуса русского языка.
81. docplayer.ru/64134811-Lekciya-3-korpusnye-issledovaniya.html Захаров В.П. Корпусные исследования. Лекция 3.
82. <https://www.myfilology.ru/177/istoriya-korpusnoj-lingvistiki>. История корпусной лингвистики.
83. <https://cyberleninka.ru/article/v/lc-tehnologii-v-issledovanii-osvoeniya-yazyka>. Камшилова О.Н. LC-технологии в исследовании освоения языка.
84. http://wikipedia.org/wiki/Корпусная_лингвистика. Статья в интернет-энциклопедии Википедия.
85. www.ruslang.ru/seminar_aut_lexocorg. Кукушкина О.В., Поликарпов А.А., Суровцева Е.В. “Частотный грамматико-семантический словарь языка художественных произведений А.П.Чехова (с электронным приложением) как особый лексикографический продукт”.
86. lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf. Курс “Корпусная лингвистика” (А.Б. Кутузов) Лицензия Creative commons Attribution Share-Alike 3.0 Unported.
87. <http://company.yandex.ru/class/courses/zakharov.xml>. Курс по корпусной лингвистике
88. http://kpfu.ru/staff_files/F1318514280/Nacioanlnyj.korpus.v.praktike.lingivistic_heskih.issledovaniy.i.prepodavaniy.yazyka.Mardanshina.RM.pdf.

Марданшина Р.М. Национальный корпус в практике лингвистических исследований и преподавании языка.

89. http://www.helsinki.fi/~kopotev/national_corpus.pdf. Национальный корпус русского языка // Вопросы языкознания. – Москва: Наука, 2006. – С. 149-155.

90. www.ruscorpora.ru Плунгян В. Зачем мы делаем корпуса?

91. <https://istina.msu.ru>. Потемкин С.Б. Авторский корпус и словарь языка Антона Чехова.

92. <http://www.ruscorpora.ru/corpora-infro.html>. Синклер Д. Предисловие к книге “Как использовать корпуса в преподавании иностранного языка”.

93. <https://istina.msu.ru>. Сичинава Д.В. К задаче создания корпусов русского языка.

94. www.ruslang.ru/seminar_aut_lexicorg. Слободян Е.А. “Идеографический словарь языка поэмы А.Мицкевича “Пан Тадеуш” в авторской и электронной лексикографии”.

95. <http://ling.ulstu.ru>. Соснина Е.П. Корпусная лингвистика и корпусный подход в обучении иностранному языку.

96. <http://human.snauka.ru/2017/02/21074>. Столяров А.И. Словарь-конкорданс и его применение в рамках корпусной лингвистики // Гуманитарные научные исследования. –2017. – № 2.

97. <http://human.snauka.ru/2017/02/21074>. Столяров А.И. Словарь-конкорданс и его при-менение в рамках корпусной лингвистики // Гуманитарные научные исследования. –2017. – № 2.

98. human.snauka.ru/2017/02/21074. Столяров А.И. Словарь-конкорданс и его применение в рамках корпусной лингвистики.

99. www.actalinguistica.com/arhiv/index.php/als/article/download/297/429. Суровцева Е.В. Словарь тем писателя как разновидность идеографических словарей. на материале художественной прозы и драматургии А.П.Чехова (к постановке проблемы)

100. www.ruslang.ru/images/seminar_aut_lexicorg/husnutdinov_doklad.pdf. Хуснутдинов А.А. “Словарь языка А.Н.Островского” как один из вариантов авторского словаря”.
101. <http://www.ruscopora.ru/corpora-infro.html>. Что такое корпус? Национальный корпус русского языка
102. <https://lamb.viniti.ru>. Шаров С.А. Представительный корпус русского языка в контексте мирового опыта.
103. <https://korpus.sk/attachments>. Шимкова М. Репрезентативность корпуса как лингвистическая проблема.
104. dostoevskij.karelia.ru. Словарь-конкорданс публицистики Ф.М.Достоевского.
105. iie.tpu.ru/smotr/exhibition/cat/1/.../Ковальчук%20А.Н.ТГУ,6курс_к_г_01.04.doc. Ковальчук А.Н. Актуальность использования специализированных лингвистических корпусов для решения практических вопросов юрлингвистики.
106. www.dialog-21.ru/media/2777/dialogue-2015_vol1.pdf. Компьютерная лингвистика и интеллектуальные технологии // Материалы ежегодной Международной конференции “Диалог”, 2014.
107. www.ruslang.ru/seminar_aut_lexicogr020413. Методическое руководство по работе с системой “ИСТОК” и корпусом текстов А.П.Чехова. (Авторы: Кукушкина О.В, Суровцева Е.В, Поликарпов А.А., Варламов А.А., Федотов В.В.).
108. ruscopora.ru/sbornik2008/00.pdf. Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009.
109. www.ict.edu.ru/vconf/files/10376.pdf. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. работа с системой DDC // Языковая инженерия: в поиске смыслов.
110. www.philol.msu.ru/~lex/pdfs/slovar_yazyka_proizvedenij_chehova2012.pdf. Программное средство для работы с корпусами текстов “Компьютерная информационно-исследовательская система для работы с электронными

корпусами текстов “ИСТОК”. Разработчики: Федотов В.В., Кукушкина О.В., Поликарпов А.А.

111. www.ruscorpora.ru/corpora-biblio.html. Резникова Т.И., Копотев М.В. Лингвистически аннотированные корпуса русского языка (обзор общедоступных ресурсов) // Национальный корпус русского языка: 2003– 2005. М.: Индрик, 2005. – С. 31– 61.

112. www.ruscorpora.ru/corpora-biblio.html. Труды международной конференции “Корпусная лингвистика”. (2004, 2006, 2008, 2011, 2013). Санкт-Петербург. – СПб.: СПбГУ.

113. <http://wave9.ru/books/humanitarian/83855-Pushkin-A.-S.-Polnoe-entsiklopedicheskoe-sobranie-.html>. “Александр Сергеевич Пушкин. Полное энциклопедическое собрание сочинений”, изд. ИДДК.

114. enc.biblioclub.ru/Encyclopedia/212471_Slovar_yazyka_A_S_Griboedova/ Алфавитно-частотный конкорданс текстов Грибоедова.

115. library.sakharov-center.ru/vse...bl.../konkordans-k-stikham-a-s-pushkina.html? “Конкорданс к стихам А.С.Пушкина”.

116. www.dialog-21.ru/digests/dialog2009/materials/html/61.htm/ “Конкорданс к текстам Ломоносова”.

117. elar.urfu.ru/bitstream/10995/40478/1/urgu1591_d.pdf “Конкорданс произведений С.Т.Аксакова”.

118. <https://search.rsl.ru/ru/record/01004098867/> “Метапоэтический словарь-конкорданс драматического текста А.П.Чехова”.

119. www.philol.msu.ru/~rly/science/pushkin/ “Электронная энциклопедия языка А.С.Пушкина”

120. dostoevskij.karelia.ru/ “Частотный словарь-конкорданс публицистики Ф. М. Достоевского”.

121. <https://rucont.ru/efd/400419>. “Частотно-семантический конкорданс служебных слов в поэзии Сергея Есенина”.

122. www.philol.msu.ru/~lex/pdfs/slovar_yazyka_proizvedenij_chehova2012.pdf/ Частотный грамматико-семантический словарь языка художественных

произведений А.П.Чехова. (Авторы: Кукушкина О.В., Суровцева Е.В., Лапоница Л.В., Рюдигер Д.Ю)

123. www.lib.cap.ru/shum-bibl/Multimedia.html/ Мультимедиа-энциклопедия “А.С.Пушкин. В зеркале двух столетий”, изд. 1С.

124. www.philol.msu.ru/~lex/kiisa.html КИИСа. Корпусная информационно-исследовательская система. Вып.1. Поэзия и драматургия А.С.Пушкина и Путеводитель по Пушкину. Под. ред. А.А.Полокарпов. - М., 2006.

125. www.crosslingua.cfuv.ru/publications/2014_2_zakharov.pdf. Захаров В.П. Корпусная лингвистика в России. CrossLingua'2014 < 27 – 30. 05.14 < Crimea, Simferopol – Alushta.

VIII. Интернет манбалари

126. <http://www.ruscorpora.ru> . НКРЯ

127. www.ruscorpora.ru

128. <http://khnt.hit.uib.no/icame/manuals/lobman/>

129. <http://www.ucl.ac.uk/english-usage>

130. <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>

131. <http://www.pearsonlongman.com/dictionaries/corpus/index.html>

132. <http://www.natcorp.ox.ac.uk/>

133. <http://www.ucl.ac.uk/english-usage/ice/>

134. <http://www.americannationalcorpus.org/>

135. <http://www ldc.upenn.edu>

136. <http://corpora.phil.spbu.ru>

137. www.slaviska.uu.se/korpus.htm

138. www.sfb441.uni-tuebingen.de/b1/rus/korpora.html

139. <http://www.ruscorpora.ru> . НКРЯ

140. www.philol.msu.ru/~lex/corpus

141. <http://www.philol.msu.ru/~lex/corpus>

142. <http://ruscorpora.ru>.

143. <http://ruscorpora.ru/search-murco.html>

144. [http:// www.ling.helsinki.fi/projects/hanco/](http://www.ling.helsinki.fi/projects/hanco/).

145. [http:// corpus.leeds.ac.uk/](http://corpus.leeds.ac.uk/)
146. <http://corpus.leeds.ac.uk/ruscorpora.html>
147. <http://aot.ru/search1.html>
148. <http://sketchengine.co.uk/>
149. <http://www.philol.msu.ru/~lex/corpus>
150. [http://ruscorpora.ru/sbornik2008/07. pdf](http://ruscorpora.ru/sbornik2008/07.pdf)
151. <http://project.phil.pu.ru/skat>.
152. <http://termcom.tj>.
153. <http://www.uni-koeln.de/phil-fak/englisch/bald/corpora>.
154. http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/history.html.
155. <http://rykov-cl.narod.ru/c.html>.
156. [https://ru.m.wikipedia.org>wiki>XML](https://ru.m.wikipedia.org/wiki/XML).
157. [https://ru.m.wikipedia.org>wiki>SGML](https://ru.m.wikipedia.org/wiki/SGML).
158. <http://rykov-cl.narod.ru/c.html>.
159. <http://www.grsampson.net>.
160. www.sit.tj.
161. www.termcom.tj.
162. <https://ru.m.wikipedia.org>.
163. <http://www.philol.msu.ru/~lex/>
164. lex@philol.msu.ru.
165. [http:// www.ling.helsinki.fi/projects/hanco/](http://www.ling.helsinki.fi/projects/hanco/)



Хамроева Шахло Мирджонова – Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёт университети докторанти. 2018 йилда “Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари” мавзусида филология фанлари бўйича фалсафа доктори (PhD) диссертациясини ёқлаган.

“Лингвистик тадқиқот методикаси, методологияси ва методлари” “Морфологиядан машқлар тўплами”, “Ҳозирги ўзбек адабий тили”дан асосий атамалар луғати”, “Ҳозирги ўзбек адабий тили фанидан таълим технологияси. Академик лицей ва коллеж ўқитувчилари учун методик қўлланма”, “Она тили таълими мақсади ва ўзбек тили доимий анжумани (Ўрта умумтаълим ва ўрта махсус таълим мисолида) ўқув қўлланма”, “Она тили таълимида замонавий педагогик технологиялар (Ўрта умумтаълим ҳамда ўрта махсус таълим мисолида)”, “Корпус лингвистикаси атамаларининг қисқача изоҳли луғати” каби луғат, ўқув ва услубий қўлланмалар муаллифи.

Корпус лингвистикаси бўйича амалга оширган лойиҳалари учун “Абдулла Қаҳҳорнинг муаллифлик корпуси” (Гувоҳнома № 000895), “Абдулла Қаҳҳорнинг “Бемор” ҳикояси корпуси” (Гувоҳнома № 000896), “Ўзбек тилининг миллий ва муаллифлик корпуси учун ЭҲМ дастури” (№ DGU 05735) каби муаллифлик гувоҳномаларини олган.

МУНДАРИЖА

ТИЛ КОРПУСИ ВА УНИНГ АМАЛИЙ АҲАМИЯТИ.....4
Фойдаланилган адабиётлар рўйхати69

FOR AUTHOR USE ONLY

FOR AUTHOR USE ONLY

FOR AUTHOR USE ONLY

**More
Books!**



yes
I want morebooks!

Buy your books fast and straightforward online - at one of world's fastest growing online book stores! Environmentally sound due to Print-on-Demand technologies.

Buy your books online at
www.morebooks.shop

Kaufen Sie Ihre Bücher schnell und unkompliziert online – auf einer der am schnellsten wachsenden Buchhandelsplattformen weltweit! Dank Print-On-Demand umwelt- und ressourcenschonend produziert.

Bücher schneller online kaufen
www.morebooks.shop

KS OmniScriptum Publishing
Brivibas gatve 197
LV-1039 Riga, Latvia
Telefax: +371 686 204 55

info@omniscryptum.com
www.omniscryptum.com

OMNIScriptum



FOR AUTHOR USE ONLY

FOR AUTHOR USE ONLY

FOR AUTHOR USE ONLY