ЖИЗЗАХ ДАВЛАТ ПЕДАГОГИКА ИНСТИТУТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
PhD.03/04.06.2020.Fil.113.02 РАҚАМЛИ ИЛМИЙ КЕНГАШ
ЖИЗЗАХ ДАВЛАТ ПЕДАГОГИКА ИНСТИТУТИ

ХИДИРОВ ОТАБЕК ЖЎРАБОЕВИЧ

МИЛЛИЙ КОРПУС УЧУН ПАРСИНГ ДАСТУРИ ЯРАТИШНИНГ ЛИНГВИСТИК АСОСЛАРИ

10.00.01 – Ўзбек тили

ФИЛОЛОГИЯ ФАНЛАРИ БЎЙИЧА ФАЛСАФА ДОКТОРИ (PhD)
ДИССЕРТАЦИЯСИ АВТОРЕФЕРАТИ

Жиззах – 2021

УДК: 811.512.133.93

Филология фанлари бўйича фалсафа доктори (PhD) диссертация автореферати мундарижаси

Оглавление автореферата диссертации доктора философии (PhD) по филологическим наукам

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

ЖИЗЗАХ ДАВЛАТ ПЕДАГОГИКА ИНСТИТУТИ
ҲУЗУРИДАГИ ИЛМИЙ ДАРАЖАЛАР БЕРУВЧИ
PhD.03/04.06.2020.Fil.113.02 РАҚАМЛИ ИЛМИЙ КЕНГАШ

ЖИЗЗАХ ДАВЛАТ ПЕДАГОГИКА ИНСТИТУТИ

ХИДИРОВ ОТАБЕК ЖЎРАБОЕВИЧ

МИЛЛИЙ КОРПУС УЧУН ПАРСИНГ ДАСТУРИ ЯРАТИШНИНГ ЛИНГВИСТИК АСОСЛАРИ

10.00.01 – Ўзбек тили

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

Жиззах – 2021

Фалсафа доктори (PhD) диссертацияси мавзуси Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясида №B2020.4.PhD/Fil24 рақам билан рўйхатга олинган.

Диссертация Жиззах давлат педагогика институтида бажарилган.

Диссертация автореферати уч тилда (ўзбек, рус, инглиз (резюме)) Илмий кенгашнинг веб-саҳифасида (www.jspi.uz) ва "Ziyonet" Ахборот таълим портали (www. ziyonet.uz)да жойлаштирилган.

Илмий раҳбар:	**Менглиев Бахтиёр Ражабович** филология фанлари доктори, профессор
Расмий оппонентлар:	**Ўринбоева Дилбар Бозоровна** филология фанлари доктори, доцент
	Абжалова Манзура Абдурашетовна филология фанлари бўйича фалсафа доктори (PhD), доцент
Етакчи ташкилот:	**Қарши давлат университети**

Диссертация ҳимояси Жиззах давлат педагогика институти ҳузуридаги илмий даражалар берувчи PhD.03/04.06.2020.Fil.113.02 рақамли Илмий кенгашнинг 2021 йил "11" ноябр соат 10 даги мажлисида бўлиб ўтади. (Манзил: 130100, Жиззах шаҳри, Шароф Рашидов шоҳ кўчаси, 4. Тел: (+99872) 226-13-57, 226-21-73, факс: (+99872) 226-46-56; e-mail: jspi_info@umail.uz). Жиззах давлат педагогика институти. Бош ўқув бино, 2-қават, маърузалар зали)

Диссертация билан Жиззах давлат педагогика институтининг Ахборот-ресурс марказида танишиш мумкин (28 рақами билан рўйхатга олинган). (Манзил: 130100, Жиззах шаҳри, Шароф Рашидов шоҳ кўчаси, 4. Тел: (+99872) 226-13-57, 226-21-73, факс: (+99872) 226-46-56.

Диссертация автореферати 2021 йил "29" 10 куни тарқатилди.
(2021 йил "29" 10 даги 9 рақамли реестр баённомаси).

А.Э.Маматов
Илмий даражалар берувчи илмий кенгаш раиси, филол.ф.д., профессор

Ф.Э.Ибрагимова
Илмий даражалар берувчи илмий кенгаш илмий котиби, филол.ф.н., доцент

У.Қосимов
Илмий даражалар берувчи илмий кенгаш қошидаги Илмий семинар раиси, филол.ф.д., доцент

КИРИШ (фалсафа доктори (PhD) диссертацияси аннотацияси)

Диссертация мавзусининг долзарблиги ва зарурати. Жаҳон тилшунослигида компьютер ва корпус лингвистикаси муаммоларига эътибор XX асрнинг иккинчи ярмидан бошлаб жадаллашди, XXI аср бошларида катта ҳажмли тил корпуслари пайдо бўлмоқда. Автоматик таржима, электрон луғат, лингвистик корпуслардан фойдаланиш имконияти янада кенгаймоқда. Мазкур янгиланишлар ахборот технологияларини тилшуносликка татбиқ этиш билан боғлиқ истиқболли илмий йўналишлар пайдо бўлишига йўл очди. Бу эса тил бирликларини корпус материали сифатида теглаш тамойилларига бўлган эҳтиёжни кун тартибига қўймоқда.

Дунё тилшунослигида XXI асрга келиб, корпус лингвистикасини ўрганиш ҳаракати жадал тус олмоқда. Компьютер лингвистикаси йўналишида автоматик таржима сифатини яхшилаш, тил бирликларини теглаш назарияси, алгоритми ҳамда лингвистик таъминотини яратиш, матн таҳлил қилиш (таггинг, парсинг, спелккер) дастурлари тузиш жаҳон компьютер лингвистикасида долзарб масалага айланиб бормоқда. Тилшуносликда, хусусан, компьютер лингвистикаси соҳасида корпус бирликларини синтактик теглаш, тег моделлари, теглаш алгоритмини ишлаб чиқиш, шу асосда автоматик синтактик теглаш (парсинг) дастурини ишлаб чиқишга зарурат сезилмоқда.

Истиқлол йилларида компьютер лингвистикасида автоматик таржима, сунъий интеллектнинг ўзбек тилини тушуниш ва қайта ишлашига эришиш борасида қатор ишлар амалга оширилмоқда. Бинобарин, "...давлат тилининг софлигини сақлаш, уни бойитиб бориш ва аҳолининг нутқ маданиятини ошириш; давлат тилининг замонавий ахборот технологиялари ва коммуникацияларига фаол интеграциялашувини таъминлаш"[1] бугунги кунда ўзбек тилшунослиги олдида турган долзарб вазифадир. Мамлакатимизда давлат тилига эътибор давлат сиёсатининг устувор йўналишларидан бири даражасига кўтарилди. Корпус лингвистикаси истиқболли илмий йўналиш эканлигини инобатга олган ҳолда ўзбек тили миллий корпусини яратиш, лингвистик моделларни тузиш сингари масалаларни замонавий илмий тамойиллар асосида тадқиқ этиш фанимиз олдида турган долзарб вазифалардан биридир.

Ўзбекистон Республикаси Президентининг 2016 йил 13 майдаги ПФ-4997-сон "Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетини ташкил этиш тўғрисида", 2017 йил 7 февралдаги ПФ-4947-сон "Ўзбекистон Республикасини янада ривожлантириш бўйича Ҳаракатлар стратегияси тўғрисида", 2019 йил 21 октябрдаги ПФ-5850-сон "Ўзбек тилининг давлат тили сифатидаги нуфузи ва мавқеини тубдан

[1]Ўзбекистон Республикаси Президенти Шавкат Мирзиёевнинг 2020 йил 20 октябрдаги "Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида"ги ПФ-6084-сон фармони // https://lex.uz/docs/5058351

ошириш чора-тадбирлари тўғрисида", 2020 йил 20 октябрдаги ПФ-6084-сон "Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида"ги фармонлари, 2019 йил 4 октябрдаги ПҚ-4479-сон "Ўзбекистон Республикасининг "Давлат тили ҳақида"ги Қонуни қабул қилинганлигининг ўттиз йиллигини кенг нишонлаш тўғрисида" қарорлари ҳамда бошқа меъёрий-ҳуқуқий ҳужжатларда белгиланган вазифаларни амалга оширишда ушбу тадқиқот муайян даражада хизмат қилади.

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

Муаммонинг ўрганилганлик даражаси. Жаҳон тилшунослигида корпус соҳасидаги мақсадли тадқиқотлар XX асрнинг 40-йилларида Блумфильд, Фрайс ва Бонджерслар томонидан бошланган[2]; Н.Френсис ва Г.Кучера илк марта корпус тузиш принципларини ишлаб чиққан[3]. Рус тилшунослигида В.П.Захаров, А.Б.Кутузов, Е.В.Недошивина, В.В.Риков, В.А.Плунгянлар корпус, унинг турлари, корпус тузиш ва теглаш тамойил–лари борасида тадқиқот олиб боришган[4]. Корпус бирликларини синтактик теглаш муаммолари, парсинг дастурлари яратиш Д.Бибер, С.Конрад, Р.Реппен[5], Э.Финеган[6], М.В.Копотев, Г.Б.Гурин[7], И.М.Ножов[8], Ю.Д.Апресян, И.М.Богуславский, Л.Л.Иомдин[9]тадқиқотлари предмети бўлган.

---

[2]Блумфильд Л. Язык. – Москва: Прогресс, 1968. – 608 с.; Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – New York:Xarcourd,1952. -304 p.; Bongers H. The history and principles of Vocabulary control. – Woerden: WOCOPI, 1947. - 442 p.

[3]Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – Москва, 1967.-150 с.; Синклер Д. Предисловие к книге "Как использовать корпуса в преподавании иностранного языка"/ Д. Синклер [Электронный ресурс]. – Режим доступа: http://www/ruscorpora.ru/corpora-infro.html, свободный.

[4] Захаров В.П. Корпусная лингвистика: Учебно-методическое пособие. – Санкт-Петербург, 2005. – 69 с. – С. 27.; Кутузов А.Б. Корпусная лингвистика. – (Электрон ресурс). Лицензия Creative commons Attribution Share-Alike 3.0 Unported (Электрон ресурс) - //lab314.brsu.by/kmp-lite/kmp-video/CL/CorporeLingva.pdf; Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – Санкт-Петербург, 2006.-126 с.; Рыков В.В. Курс лекций по корпусной лингвистике. URL: http://rykov-cl.narod.ru/c.html; Плунгян В. Зачем мы делаем Национальный корпус русского языка? "Отечественные записки" 2005, №2. http://magazines.russ.ru/oz/2005/2/2005_2_20-pr.html

[5]Biber D., Conrad S., Reppen R. Corpus linguistics: Investigating language structure and use. – Cambridje University Press, 1998.-189 p.

[6]Finegan E. Language: its structure and use. – N.Y.: Harcourt Brace College Publishers, 2004.-152 p.

[7]Копотев М.В., Гурин Г.Б. Принцы синтаксической разметки хельсинского аннотированного корпуса русских текстов ХАНКО (электрон ресурс): www.dialog-21.ru.

[8]Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы) сегментации русского предложения. – М.: АКД, 2003.-152 с.

[9]Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистическое обеспечение системы ЭТАП-2. – М.: Наука, 1989.-190 с.

Туркий тилларнинг автоматик синтактик разметкаси масаласи Т.Френсис, Ж.Вашингтон, Ч.Чўлтекин, А.Макажанов[10], В.П.Желтов, П.В.Желтов[11]ларнинг тадқиқотида кун тартибига қўйилган.

Ўзбек тилшунослигида компьютер лингвистикаси соҳасида бир қанча тадқиқотлар амалга оширилган. Жумладан, А.Қ.Пўлатов, С.М.Муҳаммедов, С.Муҳаммедова, Д.Б.Ўринбоева, Н.З.Абдурахмонова, А.М.Норов ва бошқаларнинг изланишлари лингвостатистик, лексикографик муаммоларни компьютер ёрдамида ечиш масалаларига бағишланган[12].

Ўзбек тили матнларини синтактик теглаш муаммолари махсус тадқиқ предмети бўлмаган, аммо айрим изланишларда масаланинг муайян қирралари ҳақида фикр билдирилган. Чунончи, ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш[13], ўзбек-инглиз тили машина таржимасининг лингвистик таъминоти[14], ўзбек тили муаллифлик корпусини тузиш тамойиллари[15], ўзбек тили бирликларини графематик таҳлил қилиш муаммолари[16], тил корпуси лингвистик базасини тузиш тамойиллари[17], ўзбек тили атов бирликларини семантик теглашнинг

---

[10]Francis M. Tyers, Jonathan Washington, Çağrı Çöltekin, Aibek Makazhanov. Оценка критериев морфо-синтаксической разметки для тюркских языков в проекте «UNIVERSAL DEPENDENCIES» // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 1. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 380 с. – Б. 356-362.

[11]Желтов В.П., Желтов П.В. Синтаксический анализатор национального корпуса чувашского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 1. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 380 с. – Б. 304-315.

[12]Муҳаммедов С.М. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дисс... канд. фил. наук.–Ташкент, 1980.-180 с.; Бабанаров А. Разработка принципов построения словарного обеспечения турецко-русского машинного перевода: Автореф. дисс... канд. фил. наук. – Ленинград, 1981.-221 с.; Айымбетов Н.К. Опыт лингвостатистического анализа лексики и морфологии каракалпакского публицистического текста: Автореф. дисс... канд. фил. наук.– Ташкент, 1987.-180 с.; Муҳаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма.–Тошкент, 2006.–45 б.; Ўринбоева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010. – 121 б.; Жуманазарова Г.У. Фозил Йўлдош ўғли достонлари тилининг лингвопоэтикаси: Фил. фан. док. дис... автореф.–Тошкент, 2017.-52 б; Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (содда гаплар мисолида): Фил.фан.бўйича фалсафа доктори (PhD)...дис. автореф. – Тошкент, 2018.-52 б; Пўлатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.-175 б.; Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. - 136 б.

[13]Муҳаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. –Тошкент, 2006.-98 б.;

[14]Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида). Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, 2018. – 52 б.

[15]Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари. Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – 250 б.

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

[17]Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси. Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2019. – 140 б.;

лингвистик асослари[18], ўзбек тили морфологик анализаторининг лингвистик таъминоти муаммолари[19]монографик планда ўрганилган. Шунингдек, ўзбек-инглиз параллел корпуси тузиш муаммолари[20], ўзбек тилидаги отларни автоматик таҳлил қилиш[21], сўз ясалишининг формал моделлари[22] ҳақида қатор мақолалар нашр этилган. Бу тадқиқотлар компьютер лингвистикаси муаммоларини тадқиқ этганлиги билан долзарблик касб этган, аммо ўзбек тили парсинг дастури – матнни автоматик синтактик таҳлил қилиш муаммоси кун тартибига махсус қўйилмаган. Диссертацияни тайёрлаш жараёнида юқорида санаб ўтилган илмий изланишлар атрофлича ўрганилди, зарур ўринларда уларга муносабат билдирилди ва улардан тадқиқотда фойдаланилди.

Тадқиқотнинг диссертация бажарилган олий таълим муассасасининг илмий-тадқиқот ишлари режалари билан боғлиқлиги. Диссертация Жиззах давлат педагогика институти илмий-тадқиқот ишлари режасининг 1-сон «Филология фанларининг долзарб муаммолари ва уларни амалиётга жорий этишнинг янги педагогик технологиялари» (2018-2020 йй) мавзуси доирасида бажарилган.

Тадқиқотнинг мақсади ўзбек тилида синтактик бирликларни теглаш дастурини яратишнинг лингвистик асослари бўйича тавсиялар ишлаб чиқиш ҳамда ўзбек тилида сўз бирикмаси ва гапларни автоматик аннотациялашнинг лингвистик таъминотини яратишдан иборат.

Тадқиқотнинг вазифалари.

парсинг дастурларининг ўхшаш/фарқли томонлари ва имкониятларини тадқиқ этиш;

синтактик тег категориялари, тег моделларини ишлаб чиқиш, дастурлаш учун тавсиялар тайёрлаш;

ўзбек тилида синтактик муносабат ва синтактик бирлик ҳақидаги тадқиқотларнинг синтактик теглашдаги аҳамиятини ўрганиш;

содда ва қўшма гап синтаксиси борасидаги тадқиқотларни умумлаштириш ҳамда уларни парсинг дастури лингвистик базаси сифатида яратиш.

---

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

[19]Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Филология фанлари доктори диссертацияси автореферати. – Фарғона, 2021. – 78 б.

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

[21]Орхун М. Computational analysis of uzbek nouns / Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018» (Труды конференции) – Ташкент, 2018 – 320 с

[22]Турсунов А.Вопросы словообразования в формальных моделях тюркских языков (на примере узбекского языка) // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018» Труды конференции) – Ташкент, 2018 – 320 с.

Тадқиқотнинг объекти сифатида ўзбек тили синтактик бирликлари танланган.

Тадқиқотнинг предметини синтактик теглашнинг лингвистик тег ва моделлари ташкил этади.

Тадқиқотнинг усуллари. Тадқиқот мавзусини ёритишда таснифлаш, тавсифлаш, статистик ҳамда компонент таҳлил усулларидан фойдаланилди.

Тадқиқотнинг илмий янгилиги қуйидагилардан иборат:

жаҳон корпус лингвистикасидаги Penn Treebank, SynTagger, Link Grammar Parser, ХАНКО каби синтактик таҳлил дастурларининг тобе-ҳокимлик муносабати, тизимли схема грамматикаси, анъанавий синтактик таълимотлардан фойдаланишдаги ўхшаш, шунингдек, лингвистик жиҳатдан изоҳлаш ёки изоҳлай олмаслиги каби фарқли жиҳатлари аниқланган;

тилшунослар А.Ғуломов, М.Асқарова, Ғ.Абдурахмоновларнинг сўз бирикмаси бўйича, Н.Маҳмудов, А.Нурмонов, А.Бердиалиевларнинг семантик синтаксис ва валентлик ҳақидаги, Ҳ.Неъматов, М.Қодировларнинг йиғиқ гап ва ихчам гап тўғрисидаги қарашлари ўзбек тили корпуслари учун парсер дастурлари ишлаб чиқишда назарий асос вазифасини ўтаганлиги далилланган;

ўзбек тилида сўзнинг мослашув, битишув, бошқарув каби бирикиш усуллари, гапнинг мазмунига, тузилишига кўра турлари бўйича синтактик тег категориялари, сўз бирикмаси, содда ва қўшма гапларнинг якка ва жуфт теглари, ўзбек тили синтактик бирликларини анализ ва синтез қилувчи лингвистик моделлар ишлаб чиқилган;

ўзбек тилида боғланган, боғловчисиз, эргашган, мураккаб қўшма гаплар ҳамда кўчирма гапли қўшма гап конструкциясининг лингвистик синтактик теглар тизими ва сўзнинг бирикиш усули, содда, қўшма, мураккаб гап турларининг даража-уядошлик ва тур-жинс асосидаги кидирув параметрлари ишлаб чиқилган.

Тадқиқотнинг амалий натижалари қуйидагилардан иборат:

жаҳон тилшунослигида синтактик теглаш назарияси ва амалиёти, синтактик разметка, унинг имкониятлари, синтактик разметкаланган корпуслар ва уларнинг дастурий таъминоти асослаб берилган;

ўзбек тили корпусида сўз бирикмасини синтактик теглаш тамойиллари ишлаб чиқилган;

ўзбек тилида содда йиғиқ гапларни синтактик теглаш ва лингвистик модел ишлаб чиқишда лисоний-синтактик қолипларнинг ўрни очиб берилган ҳамда ЛСҚлар асосида лингвистик моделлар тузилган;

ўзбек тилида содда ёйиқ гапларни синтактик теглашда асос бўлувчи назарий маълумотларга биноан содда ёйиқ гапнинг тег ва моделлари ишлаб чиқилган;

ўзбек тилида боғланган, боғловчисиз ва эргашган қўшма гапларнинг синтактик моделлари ишлаб чиқилган;

ўзбек тилида мураккаб қўшма гапларни синтактик теглашнинг назарий асослари белгиланган.

Тадқиқот натижаларининг ишончлилиги ўрганилган материалларнинг ўзбек тили табиатидан келиб чиққан ҳолда хулосалар қилишга ёрдам берганлиги, уларнинг асосли эканлиги, методологик мукаммаллиги, ўзбек тили корпуси парсинг (синтактик таҳлил) дастурини ишлаб чиқиш тамойилларини белгилашда амалда исботланган манбаларга таянилганлиги билан изоҳланади.

Тадқиқот натижаларининг илмий ва амалий аҳамияти. Тадқиқот натижаларининг илмий аҳамияти ўзбек тили корпуслари учун парсинг дастурлари яратишнинг назарий асосларини ишлаб чиқишда, компьютер ва корпус лингвистикаси йўналишида тадқиқотлар яратишда, илмий-назарий манба сифатида хизмат қилиши билан белгиланади.

Тадқиқот натижаларининг амалий аҳамияти амалий филология бўлимлари: компьютер лингвистикаси, корпус лингвистикаси, компьютер лексикографияси, табиий тилни қайта ишлаш (NLP) каби фанларни ўқитишда дастур, режалар тузиш, мавзуларни баён этишда манба вазифасини ўташи, ўзбек тили парсинг – синтактик таҳлил дастури учун лингвистик таъминот бўла олиши билан изоҳланади.

Тадқиқот натижаларининг жорий қилиниши. Миллий корпус учун парсинг дастури яратишнинг лингвистик асослари тадқиқи бўйича олинган илмий натижалар асосида:

жаҳон корпус лингвистикасидаги Penn Treebank, SynTagger, Link Grammar Parser, ХАНКО каби синтактик таҳлил дастурларининг тобе-ҳокимлик муносабати, тизимли схема грамматикаси, анъанавий синтактик таълимотлардан фойдаланишдаги ўхшаш, шунингдек, лингвистик жиҳатдан изоҳлаш ёки изоҳлай олмаслиги каби аниқланган фарқли жиҳатлари, шунингдек, тилшунослар А.Ғуломов, М.Асқарова, Ғ.Абдурахмоновларнинг сўз бирикмаси бўйича, Н.Маҳмудов, А.Нурмонов, А.Бердиалиевларнинг семантик синтаксис ва валентлик ҳақидаги, Ҳ.Неъматов, М.Қодировларнинг йиғиқ гап ва ихчам гап тўғрисидаги қарашлари ўзбек тили корпуслари учун парсер дастурлари ишлаб чиқишда назарий асос вазифасини ўташи ҳақидаги хулосаларидан PZ-20170927147 рақамли "Қадимий туркий ёзувлар ва XIII асргача бўлган фолклор тадқиқи" мавзусидаги фундаментал тадқиқот лойиҳасида фойдаланилган (Алишер Навоий номидаги Тошкент давлат ўзбек тили ва адабиёти университетининг 2021 йил 08 апрелдаги 04/1-1238-сон маълумотномаси). Натижада фундаментал тадқиқот лойиҳасининг қадимий туркий ёзувларни автоматик қайта ишлаш усулларига бағишланган бобининг бойитилишига хизмат қилган;

ўзбек тилида сўзнинг мослашув, битишув, бошқарув каби бирикиш усуллари, гапнинг мазмунига, тузилишига кўра турлари бўйича синтактик тег категориялари, сўз бирикмаси, содда ва қўшма гапларнинг якка ва жуфт теглари, ўзбек тили синтактик бирликларини анализ ва синтез қилувчи

лингвистик моделлар ҳақидаги назарий фикрлардан 63-11/41 рақамли "Фориш қадриятлари, анъаналари, урф-одатлари, удумларини сақлаш ва улар билан кенг жамоатчиликни таништириш" мавзусидаги амалий тадқиқот лойиҳасида фойдаланилган (Ўзбекистон Республикаси Тожик миллий-маданий маркази Жиззах вилояти бўлимининг 2021 йил 05 майдаги 39-сон маълумотномаси). Натижада амалий тадқиқот лойиҳасининг Фориш қадриятлари, анъаналари, урф-одатлари, удумларини автоматик қайта ишлаш ва теглаш усулларига бағишланган бобининг бойитилишига хизмат қилган;

ўзбек тилида боғланган, боғловчисиз, эргашган, мураккаб қўшма гаплар ҳамда кўчирма гапли қўшма гап конструкциясининг лингвистик синтактик теглар тизими ва сўзнинг бирикиш усули, содда, қўшма, мураккаб гап турларининг даража-уядошлик ва тур-жинс асосидаги кидирув параметрларига доир хулосаларидан ФА-А1-Г007 "Қорақалпоқ нақл-мақоллари лингвистик тадқиқот объекти сифатида" мавзусидаги амалий тадқиқот лойиҳасида фойдаланилган (ЎзРФА Қорақалпоғистон бўлимининг 2021 йил 17 январдаги 17.01/112-сон маълумотномаси). Натижада тадқиқотнинг содда ва қўшма гапларда сўз ясалиш типи бўйича фарқланувчи бирликларнинг қўлланилиши тадқиқи бўлимининг бойитилишига имкон берган.

Тадқиқот натижаларининг апробацияси. Мазкур тадқиқот натижа–лари 5 та халқаро, 5 та республика анжуманида муҳокамадан ўтказилган.

Тадқиқот натижаларининг эълон қилинганлиги. Диссертация мавзуси бўйича 14 та илмий иш нашр эттирилган, жумладан, Ўзбекистон Республикаси Вазирлар Маҳкамаси ҳузуридаги Олий аттестация комиссиясининг докторлик диссертациялари асосий илмий натижаларини чоп этиш учун тавсия этилган илмий нашрларда 4 та илмий мақола, улардан 1 таси хорижий журналларда чоп қилинган.

Диссертациянинг тузилиши ва ҳажми. Диссертация кириш, уч боб, хулоса, фойдаланилган адабиётлар рўйхати ва иловадан иборат. Умумий ҳажми 128 саҳифани ташкил этади.

ДИССЕРТАЦИЯНИНГ АСОСИЙ МАЗМУНИ

Кириш қисмида мавзунинг долзарблиги ва зарурати асосланган, тадқиқотнинг республика фан ва технологиялари ривожланишининг устувор йўналишларига боғлиқлиги кўрсатилган, мақсад ва вазифалари берилган, объекти ҳамда предмети тавсифланган, илмий янгилиги ва амалий натижалари баён қилинган, натижаларнинг илмий ҳамда амалий аҳамияти очиб берилган, жорийланиши, апробацияси, нашр этилган ишлар ва диссертация тузилиши бўйича маълумотлар келтирилган.

Диссертациянинг биринчи боби **"Синтактик бирликларни теглашнинг назарий асослари"** деб аталган. Ушбу бобда синтактик теглаш назарияси, амалиёти, синтактик разметка ва унинг корпусни теглашдаги

имкониятлари, синтактик разметкаланган корпусларнинг дастурий таъминотлари хусусида фикр юритилади. Бобнинг *"Жаҳон тилшунослигида синтактик теглаш назарияси ва амалиёти хусусида"* деб аталувчи биринчи фаслида синтактик теглаш борасида жаҳон компьютер лингвистикасида қилинган ишларнинг умумий тавсифи келтирилади. Синтактик теглаш – матннинг синтактик таҳлилига тегишли теглар мажмуи, морфологик таҳлил асосига қуриладиган парсинг натижаси. Разметканинг бу кўриниши лексик ва бошқа синтактик қурилмалар (содда гап, қўшма гап, кўчирма гап) орасидаги синтактик алоқани кўрсатади[23]. Илк корпусларни яратишда синтактик разметкалаш ноавтоматик бўлса, кейинги авлод корпусларининг синтактик разметкаси парсинг дастури асосида, автоматик/ярим автоматик тарзда амалга оширилган. Синтактик разметканинг турли усуллари мавжуд: бири гапда сўз богланишининг шажара усули бўлса, иккинчиси матн бирликларига синтактик тег бириктириш орқали амалга оширилади. 1993 йилда Ланкастер-Осло/Берген (ЛОБ) ва Британия миллий корпуси (BNC) муаллифи Ж.Леич томонидан 1993 йилда тузилган аннотациялаш постулатлардан бири –тил белгиларини аниқ, тушунарли тавсифлаш принципи эътиборга молик. Унинг фикрига кўра, умумфойдаланишга мўлжалланган корпуснинг разметкаси уч принципга мувофиқ келиши керак.

1. Разметка (корпус аннотацияси) фойдаланувчи учун қўлланма ёки кўрсатма шаклида мавжуд бўлган таҳлил схемасига асосланган бўлиши ҳамда ҳар бир параметр ундан жой олиши керак.

2. Фойдаланувчи учун очиқ корпус разметкаси "назарий жиҳатдан нейтрал" бўлиши лозим: разметка параметрлари барча учун тушунарли бўлган тушунчалар тизимидан иборат бўлиши талаб этилади. Агар корпус аниқ бир лойиҳа учун мўлжалланган бўлса, уни разметкалашда махсус, айнан муаллифга хос ҳамда умумқабул қилинган таснифдан фойдаланиш лозим: бунда ҳам тузувчидан у ёки бу тил назариясига таяниш талаб қилинади.

3. Корпус аннотацияси схемаси ким томонидан, қайси аудиторияга мўлжалланганлиги аниқ, равшан кўрсатилиши лозим, чунки корпусдан фойдаланишда турли юридик ва техник чегаралар мавжуд[24].

Демак, синтактик теглар тизимини ишлаб чиқиш учун компьютер технологиялари ютуқлари билан бирга ўзбек тилшунослигида синтаксис бўйича яратилган назариялар асосида корпуснинг парсер дастурини ишлаб чиқиш мумкин.

Бобнинг *"Синтактик разметка ва унинг турли корпуслардаги имкониятлари"* номли иккинчи фаслида синтактик аннотация (разметка) турлари тадқиқ этилади. Ж.Лич томонидан яратилган синтактик аннотациялаш принципи ҳақидаги назарий материалларни кузатишимиз шуни кўрсатдики, синтактик аннотацияси мавжуд корпуснинг аудиторияси

---

[23] Қаранг: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011.-154 с.

[24] Leech, G. Corpus annotation schemes / G. Leech Literary and Linguistic Computing, 1993. – 8/4. – P. 275-281.

кенг бўлади; бундай корпуснинг ахборот тизимлари билан алоқа ўрнатиш имконияти кенгроқ бўлади. Шу билан бирга, маркировканинг изчиллиги учун барча жавобгарликни муаллиф зиммасига юкламайдиган, мавжуд таснифларга асосланиб, корпус тузишга ёндашув тил тавсифларидаги бўшлиқни аниқлашга, тилга бўлган ёндашувлардаги нуқсон, қарама-қаршиликларни аниқлашга имкон беради.

И.М.Богуславский матн разметкаси махсус тег – маркер билан амалга оширилишини таъкидлайди ҳамда тегларни якка (1), контейнер тег (2)га ажратади. Якка тег матн бирлиги (сўз) ҳақида ахборот беради, контейнер тег эса разметка тизимида сақланадиган матн структураси тўғрисидаги ахборотни ташийди.

1. Матнни гапга ажратиш жуфт контейнер теглар воситасида амалга оширилади: <С> : </С>. Очилувчи тег яна бир параметрга эга бўлиши мумкин, бу гап идентификатори <С ИД=идентификатор>. Ушбу тег –матн таркибидаги гаплар орасидаги муносабатни ифодаловчи изоҳ.

2. Матнни лексик элементларга ажратиш жуфт контейнер теглар билан амалга оширилади: <W> : </W>. Сўз ҳам ўз идентификаторига эга бўлиши мумкин <W ИД=идентификатор>.

3. Сўзнинг морфологик характеристикаси якка тег билан ёзилади: <НОМ>; улар контейнер теглар ичида жойлашади. <НОМ> тегининг 4 та майдони мавжуд: ИД – идентификатор, ЛЕММА – сўзнинг луғатдаги шакли (лексема), POS – сўз туркуми, FEAT – морфологик характеристикалар.

4. Гапнинг синтактик структураси тўғрисидаги ахборот <НОМ> теги ичида жойлашувчи алоҳида белги – DOM билан ифодаланади: <НОМ ДОМ= идентификатор / алоқа типи>. Идентификатор синтактик тобе сўзга ишора қилса, алоқа типи ҳоким ва тобе сўз ўртасидаги синтактик муносабат типини акс эттиради.

Формализм етарли мослашувчанликка эга: у нафақат тайёр тузилмани, балки матннинг оралиқ ҳолатини қайд этиш имконини ҳам беради. Хусусан, битта контейнер <W> : </W> теги орасига бир неча <НОМ> тегларини киритиш орқали сўзшакл морфологик анализининг бир неча варианти ҳақидаги ахборотни битта разметка таркибида сақлашга эришиш мумкин. <НОМ> теги таркибида бир қанча DOM тегларини киритиш билан шажара тузилишини сақлаш мумкин.

Бобнинг учинчи фасли *"Синтактик разметкаланган корпуслар ва уларнинг дастурий таъминоти борасида айрим мулоҳазалар"* деб аталган. Ушбу бўлимда разметкаланган матнда лингвистик ахборот типи: морфологик, синтактик ахборот, унинг синтактик разметкадаги аҳамияти, *SynTagger синтактик таҳлил дастури* ўрганилади. Синтактик таҳлил алгоритми ишлаб чиқилганда қўшимча фильтр яратиш ҳам талаб этилган: 2-4 аъзодан ташкил топган ушбу восита таҳлил қилинаётган гапни потенциал тармоқлар воситасида таҳлилдан ўтказади. Бундай тажриба натижасини корпуснинг кейинги қисмини куришда ҳам қўллаш мумкин, чунки янги,

автоматик равишда қурилган гапларни таҳлил қилиш янада осонлашади. О.И.Бабина, Н.Ю.Дюминлар томонидан таклиф этилган *автоматик синтактик разметка модули* (SynTagger) матннинг синтактик жиҳатдан бир-бирига бўйсунувчи, тобе-ҳоким бўлак бўлиб келган лексик бирликни қавслар билан бириктирилган қўшилма сифатида ўз ичига олади (Қаранг: 3-расм).

Фойдалунувчи синтактик блокнинг боши ва охирини белгилаши, унинг типи (отли бирикма, феълли бирикма, сонни ифодаловчи бирикма)ни аниқлаши тавсия этилади. SynTagger модули морфологик разметка мавжуд бўлган тақдирда автоматик равишда турли хилдаги синтактик структура гуруҳларини ажратишга имкон беради. Ундан турли функционал услуб ёки лаҳжанинг ўзига хослигини кўрсатувчи синтактик тадқиқотларда фойдаланиш мумкин.

Жаҳон корпуслари, хусусан, инглиз тили корпуслари орасида ҳам синтактик разметкаланган корпуслар мавжуд бўлиб, улар ҳам ўзига хос парсинг дастурларига эга. Улар орасида Penn Treebank[25] таркибидаги воситалари бошқа парсерлар учун намуна бўла олади, у синтактик таҳлил натижалари аниқ чиқадиган энг мукаммал парсер. Инглиз тилининг синтактик аннотацияланган тарихий корпуслари ҳам мавжуд:Penn Parsed Corpus of Middle English (PPCME), Penn Chinese Treebank, Penn Korean Treebank, Prague Dependency Treebank, Arabic Syntactic/Predicate-Argument annotation.

Кузатишларимиз шуни кўрсатдики, ушбу синтактик таҳлил дастурлари – парсерлар турли лойиҳалар учун "олтин стандарт" намунаси сифатида хизмат қила олади, чунки уларда синтактик таҳлил методларига тўғри ёндашилган. Бу синтактик таҳлил тизимлари ўзбек тили синтактик таҳлил дастурини яратиш учун зарурий тажриба майдони бўлиб хизмат қилади. Юқорида санаб ўтилган парсер (синтактик таҳлил тизим)ларни ўрганар эканмиз, синтактик таҳлил тизими қандай таркибий қисмлардан ташкил топиши, синтактик таҳлил тегларини ишлаб чиқиш учун қандай лингвистик билимлар керак бўлишини кузатдик. Демак, ҳар бир тилдаги синтактик разметка тизимини ишлаб чиқиш учун ўша тилнинг синтактик қурилишини моделлаштириш талаб этилади. Моделлаштиришдан кейинги босқич синтактик теглар тизимини тузиш, сўнгги қадам эса матн тил бирликларига синтактик тегларни бириктиришдир.

Тадқиқотнинг иккинчи боби **"Сўз бирикмаси ва содда гапларни синтактик теглашнинг назарий асослари"** деб номланган. Унда ўзбек тили корпусида сўз бирикмаларни синтактик теглаш, ўзбек тилида содда йиғиқ гапларни синтактик теглашда лисоний-синтактик қолип ва лингвистик моделнинг ўрни, содда ёйиқ гапларни синтактик теглаш масалалари таҳлилга тортилган. *"Ўзбек тили корпусида сўз бирикмаларни синтактик теглаш хусусида"* деб номланган биринчи фаслида синтактик теглаш тамойиллари,

[25] https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

ўзбек тилида сўз бирикмаси ва унинг тадқиқи, синтактик теглашда синтактик алоқа, синтактик муносабат ва синтактик валентлик тушунчаси хусусида фикр билдирилган.

Синтактик теглаш – тил корпусида синтактик бирлик қидирувини юзага чиқарувчи омил. Тил корпусида қидирувнинг содда кўриниши –лемма, сўзшакл, ибора ва коллокация қидируви. Ушбу белгилар қарийб барча аннотацияланган тил корпусларида мавжуд. Синтактик ва семантик белги асосидаги қидирувни шакллантириш эса анча мураккаб жараён, шу сабабли ҳамма корпусларда ҳам қидирувнинг бу кўриниши бўлмайди. Ҳар қандай аннотация турида бўлгани каби синтактик теглар тизимини ишлаб чиқиш учун ҳам компьютер технологиялари ютуқлари билан бирга ўзбек тилшунослигида синтаксис бўйича яратилган назария (қараш)лар асосида корпуснинг парсер дастурини ишлаб чиқиш мумкин. Синтаксис бўйича яратилган барча назарий материалларни ўрганиш, умумлаштириш, қиёслаш ҳамда синтактик теглар яратишда ёндашувни тўғри танлаш муҳим. Ҳар бир тадқиқотчи тилни моделлаштиришда объектив/субъектив сабабларга кўра маълум нуқтаи назарни қўллаб-қувватлаши мумкин: тил структурасидан келиб чиқиб, тилни формаллаштириш назарияси маълум бир тилга қўлланса, бошқа тилга тўғри келмаслиги мумкин[26]. Олдинги айтилганидек, минималлаштириш концепциясига асосланадиган бўлсак, теглар тизимига фақат зарурий ахборотни киритиш мақсадга мувофиқ бўлади.

Ўзбек формал грамматик талқинлари асослари А.Ғуломов, М.Асқарова, Ғ.Абдураҳмоновларнинг ишлари[27] орқали кенг оммалашди. Ўтган асрнинг 70-йилларига келиб, формал таҳлил усулига таянган ўзбек синтактик назарияси ва ўзбек тили синтактик қурилишининг талқини тугал шаклланди, 1976-йилда нашр этилган илмий грамматика[28]да умумлаштирилди. Формал синтактик тавсиф натижаларига таяниб, ўзбек тили синтаксиси бўйича амалга оширилган ҳамда катта илмий самара берган йўналишлар сифатида систем-структур таҳлил усулларига таяниб олиб борилган семантик синтаксис ва валентлик бўйича тадқиқотларни алоҳида санаб ўтиш лозим. Ушбу тадқиқотлардаги назарияларни умумлаштирган ҳолда синтактик бирликларни фарқлашнинг назарий асослари сифатида қуйидагиларни ажратиш мумкин:

1) корпусда қидирувнинг содда кўриниши сифатида лемма, сўзшакл, ибора ва коллокация фарқланса-да, аннотацияли корпусларда морфологик белги асосидаги қидирув етакчилик қилади. Синтактик ва семантик белгилар

---

[26]Бабина О.И., Дюмин Н.Ю. Автоматизация лингвистической разметки корпуса текстов // http://helling100.narod.ru/pubs/Automation Babina Dyumin.pdf

[27] Ғуломов А. Синтаксис ва пунктуация бўйича машқлар тўплами. – Т.: Ўздавнашр, 1938; 1939, 1947. - 257 б.; Ўзбек тили грамматикаси. 2-қисм. Синтаксис. –Т.: Ўздавнашр, 1940. - 245 б.; Ғуломов А. Ўзбек тилида аниқловчилар. –Т.: Ўздавнашр, 1941. - 238. б.; Ўзбек тили грамматикаси. 2-қисм. Синтаксис. –Т.: Ўздавнашр, 1944-1960. - 265 б; Ғуломов А. Содда гап. Ҳозирги замон ўзбек тили курси бўйича материаллар. –Т.: ЎзФАнашр, 1955. - 235 б.

[28]Ўзбек тили грамматикаси. 2-том. Синтаксис. – Т.: Фан, 1976.-261 б.

асосидаги қидирув нисбатан мураккаб жараён бўлганлиги сабабли ҳамма корпусларни ҳам синтактик/семантик теглаш имкони, воситаси мавжуд эмас;

2) синтактик теглар тизимини ишлаб чиқишда замонавий автоматик таҳлил воситалари/дастурлар билан бирга ўзбек тилшунослигида синтаксис бўйича мавжуд қарашлар асосида ўзбек тили корпуслари учун парсер дастурини ишлаб чиқиш мумкин. Бунда синтактик бирликлар ҳақида яратилган барча назарий материалларни ўрганиш, умумлаштириш ҳамда синтактик теглар яратишда ёндашувни тўғри танлаш муҳим;

3) синтактик бирликларга турлича ёндашув мавжуд, шу сабабли корпус бирлигини теглашда ўзбек тили синтактик бирликлар тизимини аниқлаш ҳам асосий вазифа саналади;

4) корпусда гап ёки бирикма эканлиги ташқи белгилари асосида фарқланмайдиган бирикувларни аниқлашнинг махсус алгоритми, фильтри ишлаб чиқилиши лозим. Бунда гап/сўз бирикмасини аниқлашнинг бир компонентли ёки икки компонентли эканлигига асосланиш натижа беради.

Шунингдек, бу бўлимда синтактик алоқа, синтактик муносабат ва синтактик валентлик тушунчаларининг фарқи ёритилади; синтактик теглашдаги ўрни тавсифланади. Тадқиқот давомида ўзбек тилида сўз бирикманинг лисоний-синтактик қолипларига асосланган синтактик теглаш тизими ишлаб чиқилди.Бу борада С.Назарова томонидан ишлаб чиқилган ЛСҚ[29]ларга алоҳида эътибор қаратиш муҳим. С.Назарова ЛСҚларнинг [$W^{морфологик\ восита}$ _ $W^{морфологик\ восита}$] кўринишдаги инвариант, [$W^{қаратқич\ келишиги}$ _ $W^{эгалик\ қўшимчаси}$], [$Исм^{қаратқич\ келишиги}$ _ $Исм^{эгалик\ қўшимчаси}$], [$От^{қаратқич\ келишиги}$ _ $От^{эгалик\ қўшимчаси}$], [$От_{атоқли}^{қаратқич\ келишиги}$ _$От_{турдош}^{эгалик\ қўшимчаси}$] каби вариантини ажратади. Албатта, тил корпусида сўз бирикмасини теглаш муаммоси умумий (инвариант) ЛСҚлар билан ўз ечимини топмайди, балки сўз бирикмаларни аниқлашда нисбатан аниқроқ қолиплар талаб этилади. Ушбу қолиплар С.Назарова томонидан тадқиқ этилган, юзлаб нутқий ҳосилаларда синаб кўрилган, умумлаштирилган. Шунинг учун биз исм+исм қолипли сўз бирикмаларнинг моделини шу қолиплар асосида тузишимиз мумкин. Бунинг учун, аввало, қолип таркибидаги исмларнинг турини англатувчи қисмларни маълум белги билан, тобеланишни кўрсатиб турувчи морфологик воситаларнинг махсус белгиларини танлаб оламиз. Бунда от = N, сифат = Adj, сон = Num, турдош от = N^{sub}, олмош = Pr, равиш = Prv, ҳаракат номи = Ger теглари билан; қаратқич келишиги = Case (ёки Cs), эгалик қўшимчаси = Possessive (ёки Pos) теглари билан белгиланади. Шундан келиб чиқиб, тил корпуси учун сўз бирикмаларни синтактик теглашнинг исм+исм қолипи учун қуйидаги моделларни таклиф қилиш мумкин:

1) [N^{Cs}→ N^{Pos}]: *китобнинг варағи;*
2) [N^{Cs}→ Adj^{Pos}]: *дарахтнинг мўрти;*

---

[29]Назарова С. Бирикмаларда сўзларнинг эркин боғланиш омиллари: филол.фан. номз.... дисс. автореф. – Тошкент: 1997. – Б 26.

3) [Adj ^{Cs}→ N ^{Pos}]: *гулнинг/қизилининг ҳиди;*
4) [Adj ^{Cs}→ Adj ^{Pos}]: *олманинг/каттасининг чучуги;*
5) [N турдош ^{Cs}→ Adj ^{Pos}]: *гулнинг биттаси;*
6) [Num ^{Cs}→ Num ^{Pos}=]: *ўннинг ярми;*
7) [N ^{Cs}→ Ger ^{Pos}=]: *Отабекнинг қайтиши;*
8) [Ger ^{Cs}→ N ^{Pos}=]: *уялишининг ўрни;*
9) [Ger ^{Cs}→ Ger ^{Pos}=]: *олмоқнинг бермоғи;*
10) [N ^{Cs}→ Adjдош ^{Pos}=]: *юракнинг тўхтагани;*
11) [Adjдош ^{Cs}→N ^{Pos}=]: қўрққаннинг кўзи;
12) [N^{Cs}→ Prv^{Pos}=]: *меҳнатнинг кеча-кундузи;*
13) [Pr ^{Cs}→ N ^{Pos}=]: *менинг ватаним;*
14) [Prv ^{Cs}→ N ^{Pos}=]: *ҳозирнинг ҳузури.*

*"Ўзбек тилида содда йиғиқ гапларни синтактик теглаш: лисоний-синтактик қолип ва лингвистик модел"*деган фаслида содда гапни теглаш моделлари тавсифланган. Кузатишларимиз асосида гап бўлаклари ҳамда гапнинг маъно муносабатига кўра турларини теглаш борасида қуйидаги хулосаларга келдик:

1) тиниш белгилари орқали фарқлаш имкони мавжуд бўлган гапга шундай ахборотдан иборат тег бириктирмаслик ҳам мумкин;

2) икки хил гап бўлаги бўлиши мумкин бўлган бирликларга иккилик тег бириктириш талаб этилади;

3) синтактик теглар тизимини белгилашда интерфейснинг қулайлигини инобатга олиш зарур;

4) синтактик теглар тизимини тузишда матнга автоматик ишлов беришга эришиш, қўлда ишлов беришни имкон қадар камайтириш принципига амал қилиш лозим.

"Ўзбек тилидаги содда ёйиқ гапларни синтактик теглаш масалалари" деган учинчи фаслида содда ёйиқ гапларни теглашга доир модел ва қолиплар ишлаб чиқиш усуллари таҳлилга тортилган. Тадқиқотчи Ш.Хамроева ўзбек тили муаллифлик корпусини тузишнинг лингвистик асосларини тадқиқ этар экан, корпус материалларини синтактик теглаш масаласига йўл-йўлакай муносабат билдиради ҳамда содда гапларни синтактик теглаш бўйича айрим тавсияларни беради[30]. Унинг фикрича, матн синтактик разметкасининг энг катта ахборот базасини ташкил этувчи изоҳлар гап қурилишига оид маълумотлар йиғиндисидир. Гап синтаксиси гапни қайси жиҳатдан ўрганса, теглаш жараёнида шу белгиларнинг барчасини қамраб олиш ўринли. Чунки тузилган корпус кейинчалик синтаксис, бошқа бўлимлар билан боғлиқ турли тадқиқотлар олиб боришда маълумотлар базаси сифатида хизмат қилиши керак. Ш.Хамроева томонидан ишлаб чиқилган теглар тизимини мукаммаллаштириш бўйича қуйидаги таклифларни берамиз:

[30] Хамроева Ш.М. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD) дис. – Бухоро, 2018. – 250 б.

1) Грамматик марказлар сонига кўра гапнинг турларини аниқлаш учун <СГ>, </СГ> ҳамда <ҚГ>, </ҚГ> дан бири олинади. Содда гаплар синтактик теглар тизимини шакллантираётганимиз боис фақат <СГ>, </СГ> теги бизга етарли бўлади.

2) Гапнинг ифода мақсадига кўра турини фарқлаш учун "дарак гап" = <дг>, "сўроқ гап"= <сг>, "буйруқ гап"= <бг> тегларини киритамиз.

3) Эганинг иштирокига кўра турини аниқлаш учун "эгали гап" = <Е+>, "эгасиз гап" = <Е->; эгасиз гапларнинг "шахси номаълум гап" - <ш.н.г>, "атов гап" = <а.г>, "семантик-функционал шаклланган гап" = <с.фш.г> каби белгиларни қўшиш мумкин.

4. Бош, иккинчи даражали бўлакнинг иштирокига кўра "йиғиқ гап" = <йг>, "ёйиқ гап" = <ёг> каби изоҳ ҳам тег сирасидан ўрин олади.

5. Гап билан грамматик алоқага киришмайдиган бўлакларнинг мавжудлигига кўра "Ундалма" = <у>, </у>, "киритма" = <к>, </к>теглари киритилиши мумкин.

Диссертациянинг учинчи боби **"Ўзбек тилида қўшма гапларни синтактик теглашнинг назарий асослари"** деб номланган.

Бобнинг *"Ўзбек тилида боғланган ва боғловчисиз қўшма гапларни синтактик теглаш ва моделлаштириш"* номли биринчи фасли боғланган ва боғловчисиз қўшма гапларни синтактик теглаш модел ва теглар тизими тадқиқига бағишланган. Ҳар бир тилда синтактик бирликлар ўзига хос характерга эга. Шу сабабли ўзбек тилидаги қўшма гаплар учун ҳам синтактик теглар тизимини ишлаб чиқиш лозим. Жаҳон компьютер лингвистикаси тажрибасида синтактик теглаш муаммолари ёритилган махсус тадқиқотлар мавжуд бўлмаса-да, айрим масалалар таҳлилига бағишланган мақолалар мавжуд[31]. Аммо катта ҳажмли корпуслар таркибида ҳам қўшма гапларнинг тегланмагани – қўшма гап турларига оид сўров/қидирувни амалга ошириш имконини йўқлиги ҳали бу борада қилиниши керак бўлган, ечимини кутаётган масалаларнинг кўплигини кўрсатади. Ш.Хамроева қўшма гапларнинг умумий теглари ҳақида қуйидагиларни ёзади[32]: "Қўшма гапнинг боғловчи воситасига кўра "боғланган ҚГ" = <БҚГ>, </БҚГ>, "эргашган ҚГ" = <ЭҚГ >, </ЭҚГ>, "боғловчисиз ҚГ" = <Б-сиз ҚГ>, </Б-сиз ҚГ> турларини изоҳлаш бундай бирликларни ажратиш ҳамда изоҳлаш учун алоҳида белги вазифасини ўтайди.

Қўшма гапларнинг мукаммал теглар тизимини ишлаб чиқиш учун ўзбек тилшунослигида ҳозирга қадар тўпланган назарий материал ва таснифларга таяниш мақсадга мувофиқ.

---

[31]Апресян Ю. Д., Богуславский И. М., Иомдин Б. Л. и др. Синтаксически и семантически аннотированный корпус русского языка: современное состояние и перспективы // Национальный корпус русского языка: 2003-2005. М.:Индрик, 2005, 193-214 ; Сичинава Д. В. Обработка текстов с грамматической разметкой: инструкция разметчика // Национальный корпус русского языка: 2003-2005. Результатыиперспективы. – М., 2005, 136-154

[32]Хамроева Ш.М. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. буйича фалсафа доктори (PhD) дис. – Бухоро, 2018. – 250 б. – Б. 155.

Қисмлари тенг боғловчилар, *бўлса, эса* сўзлари, *-у (-ю), -да* юкламалари ёрдамида боғланган гаплар боғланган қўшма гаплар ҳисобланади. Таркибидаги содда гапнинг ўзаро мазмун муносабатига кўра боғланган қўшма гап 1) бириктирув боғловчиси ёрдамида БҚГ; 2) зидлов боғловчилари ёрдамида БҚГ; 3) айирув боғловчиси ёрдамида БҚГ; 4) бўлса, эса сўзлари ёрдамида БҚГ; 5) инкор муносабатли боғланган БҚГ каби турларга бўлинади[33]. Демак, ўзбек тилида боғланган қўшма гапларнинг синтактик теглар тизимига дастлаб <b.BQG>, <z.BQG>, <a.BQG>, <yo.s.BQG>, <i.BQG> каби белгилар киритилади.

Зидлов ва бириктирув муносабатли боғловчилар боглаган қўшма гапларни боглашда *-у (-ю)* юкламалари бирдек қўлланади. Бундай пайтда боғловчи восита орқали қўшма гап турини аниқлаш моделлари грамматик омонимияга дуч келади. Шу сабабли бундай ҳолларда қўшма гапларни аниқлашнинг моделини қуйидагича шакллантириш лозим.

1. Агар [Wpm] + ва/ ҳам/ -у/ -ю/ -да+[Wpm] бўлса = <a.BQG>. Бунда [Wpm] = содда гап.

2. Агар [Wpm] + аммо/ лекин/ бирок/ -у/ -ю/ -да + [Wpm] бўлса = <z.BQG>. Бунда [Wpm] = содда гап.

Бу икки моделда ҳам омонимлик юзага келади, натижада автоматик синтактик таҳлил *-у, -ю, -да* юкламалари билан богланган қўшма гапларни фарқлай олмайди. Бунинг учун фойдаланувчи автоматик таҳлил натижасини бирма-бир таҳлил қилиши керак. Келажакда ўзбек тилининг семантик таҳлил моделлари яратилса, шаклий ноаниқлик мазмуний моделлар асосида аниқланишига эришилади.

Айирув боғловчиси такрор келганда қўшма гап таркибидаги содда гаплар тиниш белги билан ажратилади. Модел ва тегларини қуйидагича белгилаш мумкин бўлади:

1. [Гоҳ...Wpm] + [гоҳ...Wpm] бўлса = <a.BQG>.
2. [Ё/ёки...Wpm] + [ёки...Wpm] бўлса = <a.BQG>.

Аммо айирув боғловчиси якка қўлланган ҳолда тиниш белги тушиб қолганда модел бошқача тус олади.

Бўлса, эса сўзлари боғланган қўшма гап қисмларини боғлаш билан бирга, улар ўртасида қиёслаш, зидлаш муносабати мавжудлигини ифодалайди. Бундай гаплардаги боғловчи воситалар [...са + ...бўлса/эса] шаклида жойлашади. Синтактик модели ва тегни [Wpm...са] + [бўлса/эса... Wpm] = <yo.s.BQG> шаклида белгилаш мумкин.

Инкор муносабатли богланган қўшма гап қисми ўзаро такрор қўлланган *на* инкор юкламаси орқали боғланади; гап орасига вергул қўйилади. Инкор

[33]Замонавий ўзбек тили. Синтаксис / Муаллифлар жамоаси. Масъул муҳаррирлар Ҳ.Ғ.Неъматов, Р.Сайфуллаева. ЎзР Олий ва ўрта махсус таълим вазирлиги, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети. – Тошкент: Мумтоз сўз, 2011. – 312 б. – Б. 226-235.

боғловчиси ҳам, одатда, содда гапнинг бошида келади[34]. Модел ва теги қуйидагича бўлади: [На...Wpm]+ [на ... Wpm] = <i.BQG>.

Боғланган қўшма гап қисмларининг ўзаро алоқаси орқали қиёслаш, бириктирув, айирув, сабаб ва натижа, изоҳлаш мазмун-муносабати ифодаланади[35]. Афсуски, корпуснинг синтактик белги асосидаги қидируви ёки синтактик теглар тизимига бундай муносабатни кўрсатувчи параметрни кирита олмаймиз.

Боғловчисиз қўшма гапларни (оҳанг) тиниш белги боғлашини инобатга олсак, уларнинг модел ва тегларини қуйидаги тартибда белгилаш мумкин бўлади:

1. [Wpm] <,> [Wpm] бўлса = <Б-сиз ҚГ>.
2. [Wpm] <:> [Wpm] бўлса = <Б-сиз ҚГ>.
3. [Wpm] <–> [Wpm] бўлса = <Б-сиз ҚГ>.
4. [Wpm] <;> [Wpm] бўлса = <Б-сиз ҚГ>.

Ажратилган кесимли содда гаплар борки, улар [Wpm] <–> [Wpm] қолипига тўғри келади: биринчи содда гап <–>кесим. Бундай ажратилган кесимлар дастур томонидан йиғиқ гап сифатида аниқланади, натижада грамматик омонимия келиб чиқади. Масалан: *Каттагина, чирик тўнгак орқасида мукка тушган гавдага кўзи тушди –югурди (Ойбек)....пайқамай қолди: бакларга тегдими, моторигами ёки бомбалар солинган кассеткаларгами –билмади.* Ҳозирча бундай тузилишли ҚГни теглаш орқали автоматик аниқлаш имкони йўқ. Келгусида ўзбек тилида онтологик моделларнинг мукаммаллашиши структур/грамматик кўпмаънолилик ва омонимияни аниқлашга ёрдам бериши мумкин.

Бобнинг *"Ўзбек тилида эргашган қўшма гапларни синтактик теглаш ва моделлаштириш"* деб номланувчи иккинчи фаслида ўзбек тилида эргашган қўшма гапларнинг боғловчи воситаларига кўра турининг лингвистик теглар тизими, ўзбек тилида эргашган қўшма гапларни моделлаштиришда маъно тури асосидаги ёндашув масалалари таҳлил этилган.

ЭҚГнинг тег тизимини ишлаб чиқишда ўзбек тилшунослигида ЭГнинг таснифлари ҳақидаги хулосаларга асосландик. Қисмлари эргаштирувчи боғловчи ёки шундай боғловчи вазифасидаги сўз ёрдамида боғланган ҚГэргашган қўшма гап ҳисобланади. ЭҚГ таркибидаги гапни ифодаловчи тегни қуйидагича белгилаймиз: бош гап=[bG]; эргаш гап =[eG].

ҚГ синтактик моделларини ишлаб чиқишда шаклий восита – қўшма гап таркибидаги содда гапларни боғловчи восита муҳим рол ўйнайди. Шу сабабли ЭГни боғловчи воситалар ҚГ турини автоматик аниқлаш учун асосий восита вазифасини бажаради. Мавжуд қоидалар эргаш гапни

[34]Замонавий ўзбек тили: Синтаксис. / Муаллифлар жамоаси. Масъул муҳаррирлар Ҳ.Ғ.Неъматов, Р.Сайфуллаева. ЎзР Олий ва ўрта махсус таълим вазирлиги, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети. – Тошкент: Мумтоз сўз, 2011. – 312 б. – Б. 256-264.
[35]Кўрсатилган манба. Б. 264.

автоматик теглаш, модел ва тегларини ишлаб чиқишда мустаҳкам назарий асос бўлади. Шу сабабли ЭГни боғловчи воситасига кўра қуйидаги турларини кўриб чиқамиз:

1. Кўмакчили қурилмалар ёрдамида ЭҚГ. Бош гапда ифодаланган мазмуннинг сабабини билдирувчи ЭГ бош гапга *шунинг учун, шу сабабли, шу туфайли сингари* кўмакчили қурилмалар билан боғланади. Улар бош гап таркибида келади. Шу асосда бундай қўшма гапларнинг синтактик моделларини қуйидагича тузиш мумкин. Агар: [eG] + <KQ>[bG] бўлса, [eG] + [bG] = <EQG> / <сабаб-EQG>. Ушбу тегда ЭҚГ ҳамда унинг маъно жиҳатдан сабаб маъносини ифодалайди.

2. *Деб* сўзи ёрдамида ЭҚГ. Бош гапда ифодаланган мазмуннинг мақсади, сабабини билдирувчи ЭГ бош гапга кўпинча **деб** сўзи ёрдамида боғланади, эргаш гапларнинг кесими III шахс буйруқ майли шаклидаги феъллар билан ифодаланади. Бундай гаплар таркибидаги *деб* сўзи *учун* кўмакчиси билан маънодош, шунинг учун бир-бири билан эркин алмаша олади[36].Шу асосда бундай ҚГнинг синтактик модели қуйидагича бўлади. Агар: [bG] + <…sin deb Q>[eG] бўлса, [eG] + [bG] = <EQG> / <мақсад-EQG>. Ушбу тегда ЭҚГнинг мақсад маъносини ифодалаган қўшма гап эканлиги ҳақидаги ахборот мавжуд.

3. Шарт майли воситасида ЭҚГ. Эргаш гапнинг кесими шарт майли шаклидаги феъллар орқали ифодаланганда, шарт майли қўшимчаси эргаш гапни бош гапга боғловчи восита ҳам саналади. Бундай қўшма гапларнинг синтактик моделини қуйидагича тузиш мумкин. Агар: [eG...ca] + [bG] бўлса, [eG] + [bG] = <EQG> / <шарт-EQG>. Ушбу тегда шарт маъносини ифодалаб келган ЭҚГ эканлиги ҳақида ахборот мавжуд.

4. Кўрсатиш олмошли ЭҚГ. Бош гап таркибидаги кўрсатиш олмошининг маъносини изоҳлаш учун қўлланган ЭГ бош гапга *-ки* юкламаси ёрдамида боғланади. Бу юклама бош гап кесими таркибида бўлади ва ЭГ бош гапдан вергул билан ажратилади. Бундай восита ёрдамида боғланишни <olmoshQ> теги билан белгилаймиз. Ушбу тартибдаги боғланиш тури синтактик тегларини қуйидагича тузиш мумкин. Агар: [bG\shu\shuni\shunga\shunday...ki] + [eG] бўлса, [bG] + [eG] = <EQG> / <шарт-EQG>. Ушбу тегда бош гапдаги кўрсатиш олмоши маъносини ифодалаган эргашган қўшма гап эканлиги ҳақидаги ахборот мавжуд.

5. Нисбий сўзли ЭҚГ. ЭГ таркибида қўлланувчи *ким, нима, қанча, қанчалик, қандай, қаер* каби сўроқ олмошлари ва бош гап таркибида унга жавоб бўлиб келувчи *шу, ўша, шунча, шунчалик, шундай* каби олмошлар бир-бирига нисбатан қўлланганлиги, бири иккинчисини тақозо этганлиги учун нисбий сўзлар ҳисобланади. ЭГнинг кесими шарт майли шаклидаги феъллар

---

[36]Замонавий ўзбек тили. Синтаксис. / Муаллифлар жамоаси. Масъул муҳаррирлар Ҳ.Ғ.Неъматов, Р.Сайфуллаева. ЎзР Олий ва ўрта махсус таълим вазирлиги, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети. – Тошкент.: Мумтоз сўз, 2011. – 312 б. – 288-293.

билан ифодаланади[37]. Бундай восита ёрдамида боғланишни <nisbiy so`zQ> теги билан белгилаймиз. Ушбу тартибдаги боғланиш тури синтактик тегларини қуйидагича тузиш мумкин. Агар: [eG\ *ким\ нима\ қанча\ қанчалик\ қандай\ қаер*...sa] + [eG\ *шу\ ўша\ шунча\ шунчалик\ шундай*] бўлса, [eG] + [bG] = <EQG> / <nisbiy so`z-EQG>. Ушбу тегда бош гапдаги нисбий сўз (олмош)ни изоҳлаш маъносини ифодалаган ЭҚГ эканлиги ҳақидаги ахборот мавжуд.

Юқоридаги таҳлилдан келиб чиқиб, ЭҚГнинг қидирув ойнасида қуйидаги параметрни жойлаштириш мумкин:

1. Кўмакчили қурилмалар ёрдамида ЭҚГ.
2. *Деб* сўзи ёрдамида ЭҚГ.
3. Шарт майли воситасида ЭҚГ.
4. Кўрсатиш олмошли ЭҚГ.
5. Нисбий сўзли ЭҚГ.

Шунингдек, ЭҚГнинг қандай боғловчи восита билан шаклланишига кўра автоматик таҳлилни амалга ошириш учун модел ва теглар тизимини қуйидаги шаклда ифодалаймиз:

1) Агар: [eG] + <KQ>[bG] бўлса, [eG] + [bG] = <EQG> / <сабаб-EQG>.

2) Агар: [bG] + <...sin deb Q>[eG] бўлса, [eG] + [bG] = <EQG> / <мақсад-EQG>.

3) Агар: [eG...sa] + [bG] бўлса, [eG] + [bG] = <EQG> / <шарт-EQG>.

4) Агар: [bG\shu\shuni\shunga\shunday...ki] + [eG] бўлса, [bG] + [eG] = <EQG> / <шарт-EQG>.

5) Агар: [eG\ *ким\ нима\ қанча\ қанчалик\ қандай\ қаер*...sa] + [eG\ *шу\ ўша\ шунча\ шунчалик\ шундай*] бўлса, [eG] + [bG] = <EQG> / <nisbiy so`z-EQG>.

Шундай қилиб, ўзбек тилидаги ЭҚГнинг автоматик қидирув тизими учун тег ва моделларни эргаш гапни боғловчи воситалар асосида ишлаб чиқиш мумкин. Боғловчисиз ҳамда эргашган қўшма гапларнинг жами 87 та модел, жумладан, эргаш гапнинг бош гапдаги қайси бўлакни изоҳлашига кўра эргашган қўшма гап турларининг қидирувини йўлга қўйиш учун боғловчи восита турига асосланган 67 та модел ишлаб чиқилди.

Бобнинг *"Ўзбек тилидаги мураккаб қўшма гаплар ва кўчирма гапли қўшма гап конструкциясининг автоматик анализ ва синтези"* номли учинчи фаслида мураккаб қўшма гап ҳамда кўчирма гапли қўшма гап конструкциясини автоматик таҳлил қилиш моделлари ҳақида сўз боради.

Мураккаб қўшма гаплар бир неча эргаш гапли, бир неча бош гапли, аралаш мураккаб қўшма гап ҳамда қисмлари уюшган мураккаб қўшма гап каби тўрт гуруҳга бўлинади[38]. Қидирув тизимида бундай гапни аниқлашда

---

[37]Кўрсатилган манба. – 295-296.

[38]Замонавий ўзбек тили. Синтаксис / Муаллифлар жамоаси. Масъул муҳаррирлар Ҳ.Ғ.Неъматов, Р.Сайфуллаева. ЎзР Олий ва ўрта махсус таълим вазирлиги, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети. – Тошкент: Мумтоз сўз, 2011. – 312 б.– Б. 463.

боғловчисиз ҚГ ёки боғланган ҚГ моделларидан фойдаланиш мумкин: мураккаб гапнинг чегараси битта бутунлик сифатида тиниш белгиси ёрдамида белгиланади. Фақат боғловчисиз қўшма гапда "нуқтадан нуқтагача" принципи иккита содда гапдан иборат қўшма гапни аниқласа, мураккаб қўшма гапда уч ва ундан ортиқ содда гапдан ташкил топган бирликни ажратиб олади. Бундай усул тенг боғловчи билан боғланган мураккаб қўшма гапларга нисбатан ҳам қўлланади: мураккаб ҚГ учун махсус модел тузишга зарурат қолмайди. Тенг боғловчи воситасида боғланган бир неча қўшма гап "нуқтадан нуқтагача" оралиқда аниқланади. Аммо қидирув ойнасида мураккаб қўшма гапнинг турини ажратиш мумкин. Бунда дастур икки босқичли таҳлилни қуйидаги алгоритм асосида амалга оширади:

1) "нуқтадан нуқтагача" оралиқдаги бирликни ажратади;

2) боғловчи восита миқдорига кўра "қўшма гап" (шунингдек, унинг тури) ёки "мураккаб қўшма гап" эканлиги ҳақида хулоса чиқаради.

Юқоридаги тасниф асосида синтактик таҳлил дастури мураккаб қўшма гап қидирув интерфейсига қуйидаги параметрларни киритиш мумкин:

- ☐ Бир неча эргаш гапли қўшма гап
- ✓ тўғридан-тўғри тобеланиш (биргалик эргашиш)
- ✓ кетма-кет эргашиш
- ☐ Бир неча бош гапли мураккаб қўшма гап
- ☐ Аралаш мураккаб қўшма гап
- ☐ Қисмлари уюшган мураккаб қўшма гап.

Булар орасида аралаш мураккаб қўшма гап ҳам тиниш белги, ҳам боғловчи восита асосида таҳлил қилинадиган бирлик саналади. Бундай қўшма гапларнинг автоматик қидируви ҳам ажратишнинг "нуқтадан нуқтагача" принципи, шунингдек, қўшма гапларнинг боғловчи воситасига асосланган моделларига таянади. Тил корпусида, бошқа синтактик бирликда бўлганидек, кўчирма гапли қўшма гапнинг автоматик қидирувини ҳам йўлга қўйиш мумкин. Бунинг автоматик таҳлили моделларини тузишда кўчирма гапли конструкция таркибидаги содда гапнинг кесимлик кўрсаткичи ҳамда тиниш белгилари позициясидан ўринли фойдаланиш мумкин.

Кўчирма гапли қўшма гапнинг ушбу турғун (аниқ) конструкцияси автоматик таҳлил жараёнини бирмунча осонлаштирилади.Кўчирма гапда, кўчирилган гапнинг ўрнига қараб, тиниш белгисининг ишлатилиши турлича: кўчирилган гап дарак гап бўлиб, муаллиф гапидан олдин келса, ундан кейин вергул ва тире қўйилади: *"Юринг, мен ўша томонга бораман", – деди Комила.* Сўроқ ва ундов белгиси қўштирноқ ичида қолади. Ушбу позициянинг модели ва тегини қуйидагича белгилаш мумкин: "WPm", –WPm = <K>, <M>.

а) кўчирилган гап муаллиф гапидан кейин келса, муаллиф гапидан кейин икки нуқта қўйилади: *Маърузачи бундай деди: "Ўсиб бораётган авлодга*

китоб худди мактаб каби керак". Ушбу позициянинг модели ва тегини қуйидагича: WPm: "WPm" = <M>: <K>.

б) муаллиф гапи кўчирилган гап ичида келса, тиниш белгиси қуйидагича қўйилади: "Бизнинг қишлоғимизда, – деди Фазлиддин, – киши зерикмайди". Бунда муаллиф гапи ва кўчирма гаппнинг тугалланмай қолгани автоматик таҳлилда чалкашликни келтириб чиқариши мумкин. Шу сабабли қўштирноқ ичидаги қисм муаллиф гапи, яъни "WPm", –WPm, –"WPm" =<M>,<K>, <M> деб белгиланади. Аммо ушбу тег (эҳтимолий белги)нинг қанчалик тўғри натижа бериши парсинг дастури ишга тушгач аниқланади.

Демак мураккаб қўшма гап ҳамда қўшма гапли конструкция, асосан, тиниш белгилар воситасида таҳлил қилинади: гап чегараси тиниш белгидан тиниш белгигача тамойили билан белгиланади.

ХУЛОСА

1. Синтактик теглаш матннинг автоматик синтактик таҳлилини амалга оширишга ёрдам берувчи асосий жараён бўлиб, синтактик теглар мажмуидан иборат бўлади, синтактик қурилмалар орасидаги синтактик алоқани кўрсатади. Илк корпусларни яратишда синтактик разметкалаш қўлда амалга оширилган бўлса, кейинги авлод корпуслари синтактик разметкаси парсинг дастури асосида, автоматик/ярим автоматик тарзда амалга оширилиши аниқланган.

2. Синтактик разметкада қўлланиладиган асосий алгоритм – синтактик декомпозиция бўлиб, у бўшлиқ (пробел), тиниш белгилари асосида жумлани фарқлайди. Аммо бу усул мукаммал эмас, чунки ажратувчи сифатида ишлатиладиган белги нафақат жумланинг охирида, балки ўртасида ҳам ишлатилиши мумкин. Сарлавҳа, бўлимни ажратиб кўрсатувчи бирлик, расм, жадвал номи, колонтитул каби бирликлар гап бўлмаса-да, гап кўринишида ифодаланиши талаб этилади.

3. Синтактик разметкаланган матнда морфологик разметка билан бирга ҳар бир гапга унинг синтактик структураси тўғрисидаги изоҳ ёзилади. Синтактик разметкада теглар якка ҳамда контейнер тегга ажратилади. Якка тег матн бирлиги (сўз) ҳақида ахборот беради, контейнер тег эса разметка тизимида сақланадиган матн структураси тўғрисидаги ахборотни ташийди. Матнни гапга ажратиш жуфт контейнер теглар воситасида амалга оширилади. Идентификатор синтактик тобе сўзга ишора қилса, алоқа типи ҳоким ва тобе сўз ўртасидаги синтактик муносабат типини акс эттириши аниқланган.

4. Ҳар бир тилдаги синтактик разметка тизимини ишлаб чиқиш учун ўша тилнинг синтактик қурилиши моделлаштирилиши талаб этилади. Моделлаштиришдан кейинги босқич синтактик теглар тизимини тузиш, сўнгги қадам эса матн тил бирликларига синтактик тегларни бириктиришдир. Шунингдек, синтактик теглар тизимини белгилашда интерфейснинг

қулайлигини инобатга олиш зарурати мавжуд. Синтактик теглар тизимини тузишда матнга автоматик ишлов беришга эришиш, қўлда ишлов беришни имкон қадар камайтириш принципига амал қилиш тавсия этилади.

5. Аннотацияланган корпусларда морфологик белги асосидаги қидирув етакчилик қилади. Синтактик ва семантик белгилар асосидаги қидирув нисбатан мураккаб жараён бўлганлиги сабабли ҳамма корпусларни ҳам синтактик/семантик теглаш имкони, воситаси мавжуд эмас.

6. Ўзбек тили корпусларининг парсер дастури, унинг синтактик теглар тизимини ишлаб чиқишда замонавий автоматик таҳлил дастурлари ҳамда ўзбек тилшунослигида синтаксис бўйича мавжуд назарий материалларни ўрганиш, умумлаштириш ҳамда синтактик теглар яратишда ёндашувни тўғри танлаш муҳим. Корпусда гап ёки бирикма эканлиги ташқи белгилари асосида фарқланмайдиган бирикувларни аниқлашнинг махсус алгоритми, фильтри ишлаб чиқилиши таклиф этилади.

7. Синтактик теглашда синтактик алоқанинг тури инобатга олинади ҳамда теглар тизимидан синтактик алоқани ифодалайдиган тег жой олади. Тил корпуси материалини синтактик теглашда синтактик алоқа ва синтактик муносабат фарқланиши лозим. Шаклан фарқлашнинг имкони йўқ: боғловчи воситалар бир хил. Шу сабабли корпус разметкасида синтактик алоқа ва синтактик муносабатни фарқловчи тег киритилиши талаб этилади. Корпус синтактик разметка тизимида коллокатив қидирув жуда муҳим: Коллокатив қидирувни ташкил этиш учун корпус бирликлари сўз бирикмаси ва сўз қўшилмасини ажратувчи тегга эга бўлиши талаб қилинади.

8. Морфологик кўрсаткичсиз тобе-ҳокимлик муносабатини аниқлаш учун лисоний-синтактик қолип ҳамда морфологик теглар тизимидан фойдаланиш ўринли, чунки лисоний-синтактик қолипда қайси сўз туркуми ўзаро бирикиши кўрсатиб берилган. Корпус бирликлари морфологик тегланган бўлса, тег ва лисоний-синтактик қолип асосида морфологик кўрсаткичсиз синтактик муносабатли бирикмаларни ҳам автоматик теглаш имкони пайдо бўлади. Албатта, корпусда сўз бирикмасини теглаш муаммоси умумий (инвариант) ЛСҚлар билан ҳал этилмайди: сўз бирикмаларни аниқлашда аниқроқ қолиплар талаб этилади. Шундан келиб чиқиб, сўз бирикмаларни синтактик теглашнинг исм+исм қолипи учун 14 та модел тавсия этилди.

9. Корпусда билвосита белгилар асосида аниқланиши мумкин бўлган бирликлар тегланмаслиги ҳам мумкин: чунки бундай бирликлар тегишли тиниш белгилари орқали фарқланади. Тиниш белгилари орқали фарқлаш имкони мавжуд бўлган гапга шундай ахборотдан иборат тег бириктирмаслик ҳам мумкин; икки хил гап бўлаги бўлиши мумкин бўлган бирликларга иккилик тег бириктириш таклиф этилади.

10. Қўшма гапни боғловчи воситалар қўшма гап турини автоматик аниқлашнинг бирламчи воситаси бўла олади: синтактик теглаш дастури (парсер) содда гаплар чегарасини улар орасидаги тиниш белги ҳамда

боғловчи восита орқали аниқлайди. Автоматик таҳлил дастури боғловчи воситалари омонимияни келтириб чиқарувчи қўшма гапларни фарқлай олмайди. Бунинг учун фойдаланувчи автоматик таҳлил натижасини бирма-бир таҳлил қилиши керак. Ўзбек тилининг семантик таҳлил моделлари яратилса, шаклий ноаниқлик мазмуний моделлар асосида аниқланади.

НАУЧНЫЙ СОВЕТ № PhD.03/04.06.2020.Fil.113.02 ПО ПРИСУЖДЕНИЮ УЧЁНОЙ СТЕПЕНИ ПРИ ДЖИЗАКСКОМ ГОСУДАРСТВЕННОМ ПЕДАГОГИЧЕСКОМ ИНСТИТУТЕ

ДЖИЗАКСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ ИНСТИТУТ

ХИДИРОВ ОТАБЕК ЖУРАБОЕВИЧ

ЛИНГВИСТИЧЕСКИЙ ОСНОВЫ СОЗДАНИЯ ПРОГРАММЫ ПАРСИНГА ДЛЯ НАЦИОНАЛЬНОГО КОРПУСА

10.00.01 – Узбекский язык

АВТОРЕФЕРАТ ДИССЕРТАЦИИ ДОКТОРА ФИЛОСОФИИ (PhD) ПО ФИЛОСОФСКИМ НАУКАМ

Джизак – 2021

Тема диссертации доктора философии (PhD) зарегистрирована по номером №B2020.4.PhD/Fil24 в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан.

Диссертация выполнена в Джизакском государственном педагогическом институте.

Автореферат диссертации размещен на трёх языках (русском, узбекском, английском (резюме)) на веб-странице samdu.uz Научного совета и на Информационно-образовательном портале ZiyoNET (www.ziyonet.uz).

Научный руководитель:	**Менглиев Бахтиёр Ражабович** доктор филологических наук, профессор
Официальные оппоненты:	**Уринбаева Дилбар Бозоровна** доктор филологических наук, доцент
	Абжалова Манзура Абдурашетовна доктор философии по филологическим наукам (PhD), доцент
Ведущая организация:	**Каршинский государственный университет**

Защита диссертации состоится на заседании научного совета № PhD.03/04.06.2020.Fil.113.02 присуждающий ученую степень доктора наук при Джизакском Государственном педагогическом институте 10⁰⁰ часов "11" 11 2021 года. (Адрес: 130100, г.Джизак, проспект Шарофа Рашидова, дом 4. Тел.: (+99872) 226-13-57, 226-21-73, факс: (+99872) 226-46-56; e-mail: jspi_info@umail.uz). Джизакский госдуарственный педагогический институт, 2-этаж, лекционный зал).

С диссертацией можно ознакомиться в Центре информационных ресурсов Джизакского Государственного педагогического института (зарегистрирована под № 23 Адрес: 130100, г.Джизак, проспект Шарофа Рашидова, дом 4. Тел.: (+99872) 226-13-57, 226-21-73, факс: (+99872) 226-46-56

Автореферат диссертации распространен "29" 10 2021 года.
(протокол реестра под номером 9 от «29» 10 2021 года).

А.Э.Маматов
Председатель научного совета по присуждению научной степени, доктор филологических наук, профессор

Ф.Э.Ибрагимова
Секретарь научного совета по присуждению научной степени, доктор филологических наук, доцент

У.Касымов
Председатель научного семинара при научном совете по присуждению научной степени, доктор филологических наук, доцент

ВВЕДЕНИЕ (аннотация диссертации доктора философии (PhD))

Актуальность и необходимость темы диссертации. В мировой лингвистике внимание к проблемам компьютерной и корпусной лингвистики усилилось со второй половины XX века, в начале XXI века появились крупномасштабные языковые корпусы. Еще более расширилась возможность использования автоматического перевода, электронного словаря, лингвистического корпуса. Вышеуказанные нововведения привели к появлению перспективных научных направлений, связанных с применением информационных технологий в лингвистике. Это поставило на повестку дня потребность разработки принципов пометки языковых единиц как материала корпуса.

К XXI веку в мировой лингвистике усиливается движение изучения корпусной лингвистики. Повышение качества автоматического перевода в области компьютерной лингвистики, разработка теории, алгоритмов и лингвистического программного обеспечения для тегирования языковых единиц, разработка программ анализа текста (tagging, parsing, spelkker) стало актуальной проблемой в мировой компьютерной лингвистике. В лингвистике, в частности в области компьютерной лингвистики, существует потребность в разработке синтаксической разметки единиц корпуса, моделей меток, алгоритмов разметки и на этой основе разработки программы автоматической синтаксической разметки (синтаксического анализа).

За годы независимости в компьютерной лингвистике был проделан ряд работ для достижения автоматического перевода, понимания и обработки узбекского языка с помощью искусственного интеллекта. Следовательно, «... сохранить чистоту государственного языка, обогатить его и повысить речевую культуру населения; Обеспечение активной интеграции государственного языка в современные информационные технологии и коммуникации – актуальная задача, стоящая сегодня перед узбекской лингвистикой"[39]. В нашей стране внимание к государственному языку поднялось до уровня одного из приоритетов государственной политики. Учитывая, что корпусная лингвистика является перспективным научным направлением, одной из актуальных задач, стоящих перед нашей наукой, является изучение таких вопросов, как создание национального корпуса узбекского языка, создание лингвистических моделей на основе современных научных принципов.

Данное исследование в определенной степени способствует реализации задач, поставленных в указах Президента Республики Узбекистан от 13 мая 2016 года №УП-4997 «Об организации деятельности Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои», от 17 февраля 2017 года №УП-4997 «О стратегии действий по дальнейшему развитию Республики Узбекистан», от 21 октября 2019 года

[39] Указ Президента Республики Узбекистан Шавката Мирзиеева от 20 октября 2020 года № ПФ-6084 «О мерах по дальнейшему развитию узбекского языка и совершенствованию языковой политики в нашей стране» // https://lex.uz/docs/5058351

№УП-5850 «О мерах по кардинальному повышению роли и авторитета узбекского языка в качестве государственного языка», указ Президента Республики Узбекистан от 20 октября 2020 года №УП-6084 «О мерах по дальнейшему развитию узбекского языка и совершенствованию языковой политики в стране», в постановлении от 4 октября 2019 года ПП-4479 «О широком праздновании тридцатой годовщины принятия Закона Республики Узбекистан «О государственном языке» и других нормативных актах.

Соответствие исследования приоритетным направлениям развития науки и технологий республики. Исследование выполнено в рамках приоритетного направления развития науки и технологий республики: I. «Формирование системы инновационных идей и способов их реализации в социальном, правовом, экономическом, культурном, духовном и образовательном развитии информированного общества и демократического государства».

Степень изученности проблемы. Целенаправленные исследования корпуса в мировой лингвистике начались в 1940-х годах XX века Блумфильдом, Фрайсом и Бонжерсом.[40]; Н. Фрэнсис и Г. Кучера первыми разработали принципы построения корпуса[41]. В русской лингвистике В.П. Захаров, А.Б. Кутузов, Е.В. Недошивина, В.В. Риков, В.А. Плунгянс провели исследования корпуса, его типов, принципов формирования и тегирования корпуса.

В русской лингвистике В.П. Захаров, А.Б. Кутузов, Е.В. Недошивина, В.В. Риков, В.А. Плунгянс провели исследования корпуса, его типов, принципов формирования и тегирования корпуса. Проблемы синтаксической разметки корпусных единиц, разработка программ парсинга был предметом исследования Д.Бибера, С.Конрада, Р.Реппена[42], Э.Финегана[43], М.В. Копотева, Г.Б. Гурина[44], И.М. Ножова[45], Ю.Д. Апресяна, И.М. Богуславского, Л.Л. Иомдина[46].

Вопрос автоматической синтаксической разметки тюркских языков стоит на повестке дня исследований Т.Фрэнсиса, Дж. Вашингтона, Ч. Чолтекина, А. Макаджанова[47], В. П. Желтова, П. В. Желтова[48].

---

[40] Блумфильд Л. Язык. – Москва: Прогресс, 1968. – 608 с.; Fries Ch.C. The structure of English. An introduction to the construction of English sentences. – New York: Xarcourd,1952. -304 p.; Bongers H. The history and principles of Vocabulary control. – Woerden: WOCOPI, 1947. - 442 p.

[41] Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка. – Москва, 1967.; Синклер Д. Предисловие к книге "Как использовать корпуса в преподавании иностранного языка"/ Д. Синклер [Электронный ресурс]. – Режим доступа: http://www/ruscorpora.ru/corpora-infro.html, свободный.

[42] Biber D., Conrad S., Reppen R. Corpus linguistics. Investigating language structure and use. – Cambridje University Press, 1998.

[43] Finegan E. Language: its structure and use. – N.Y: Harcourt Brace College Publishers, 2004.

[44] Копотев М.В., Гурин Г.Б. Принцы синтаксической разметки хельсинского аннотированного корпуса русских текстов ХАНКО (электрон ресурс): www.dialog-21.ru.

[45] Ножов И.М. Морфологическая и синтаксическая обработка текста (модели и программы) сегментации русского предложения. – М.: АКД, 2003.

[46] Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистическое обеспечение системы ЭТАП-2. – М.: Наука, 1989.

[47] Francis M. Tyers, Jonathan Washington, Çağrı Çöltekin, Aibek Makazhanov. Оценка критериев морфо-синтаксической разметки для тюркских языков в проекте «UNIVERSAL DEPENDENCIES» // Пятая

В узбекском языкознании был проведен ряд исследований в области компьютерной лингвистики. В частности, исследования А.К. Пулатова, С.М. Мухаммедова, С.Мухаммедова, Д.Б. Уринбаева, Н.З. Абдурахманова, А.М. Норова и др. посвящены проблемам решения лингвистических, лексикографических задач с помощью компьютеров[49].

Проблема синтаксической разметки узбекских текстов не была предметом специальных исследований, но в некоторых исследованиях были даны комментарии по отдельным аспектам вопроса. А именно, создание лингвистического программного обеспечения для компьютерных программ на основе глаголов действия[50], лингвистическое сопровождение узбекско-английского машинного перевода[51], принципы узбекского авторского корпуса[52], проблемы графического анализа единиц узбекского языка[53], принципы лингвистического корпуса лингвистического корпуса[54], семантическое тегирование узбекского языка назывных единиц измерения[55], проблемы лингвистического обеспечения морфологического анализатора узбекского языка[56] изучались в монографическом плане. Также, опубликован

Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 1. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 380 с. – Б. 356-362.

[48] Желтов В.П., Желтов П.В. Синтаксический анализатор национального корпуса чувашского языка // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. В 2-х томах. Т 1. – Казань: Издательство Академии наук Республики Татарстан, 2017. – 380 с. – Б. 304-315.

[49] Мухаммедов С.М. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дисс... канд. фил. наук.–Ташкент, 1980.; Бабанаров А. Разработка принципов построения словарного обеспечения турецко-русского машинного перевода: Автореф. дисс... канд. фил. наук. – Ленинград, 1981.; Айымбетов Н.К. Опыт лингвостатистического анализа лексики и морфологии каракалпакского публицистического текста: Автореф. дисс... канд. фил. наук – Ташкент, 1987.; Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма.–Тошкент, 2006. – 45 б.; Ўринбоева Д.Б. Ўзбек фольклори матнларининг лингвостатистик тадқиқи. – Тошкент: Фан, 2010. – 121 б.; Жуманазарова Г.У. Фозил Йўлдош ўғли достонлари тилининг лингвопоэтикаси. Фил. фан. док. дис... автореф.–Тошкент, 2017.; Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (содда гаплар мисолида): Фил.фан.бўйича фалсафа доктори (PhD) дис. автореф. – Тошкент, 2018.; Пулатов А. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011. – 175 б.; Норов А. Компьютер лингвистикаси асослари. – Қарши, 2017. – 136 б.

[50] Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. –Тошкент, 2006.;

[51] Абдурахмонова Н.З. Инглизча матнларни ўзбек тилига таржима қилиш дастурининг лингвистик таъминоти (Содда гаплар мисолида): Филол.фан.бўйича фалсафа доктори (PhD)...дис. афтореф. – Тошкент, 2018. – 52 б.

[52] Хамроева Ш. Ўзбек тили муаллифлик корпусини гузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2018. – 250 б.

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

[54] Эшмўминов А.А. Ўзбек тили миллий корпусининг синоним сўзлар базаси: Филол.фан.бўйича фалсафа доктори (PhD)...диссер. – Қарши, 2019. – 140 б.,

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

[56] Хамроева Ш. Ўзбек тили морфологик анализаторининг лингвистик таъминоти. Филология фанлари доктори диссертацияси автореферати. – Фарғона, 2021. – 78 б.

ряд статей по проблемам составления узбекско-английского параллельного корпуса[57], автоматического анализа имён существительны в узбекском языке[58], формальных моделей словообразования[59]. Эти исследования актуальны тем, что исследуют проблемы компьютерной лингвистики, но программа синтаксического анализа узбекского языка – проблема автоматического синтаксического анализа текста – не стоит в повестке дня. В процессе подготовки диссертации указанные научные исследования были подробно изучены, при необходимости обработаны и использованы в исследованиях.

Связь исследования с планами научно-исследовательских работ высшего учебного заведения где выполнялась диссертация. Исследование входит в состав научно-исследовательского плана №1 «Изучение узбекского языка с точки зрения языка и речи, вопросы языкового образования», который изучается на кафедре узбекского языка и литературы Джизакского государственного педагогического института. (2018-2020г.).

Целью исследования является создание программы для разметки синтаксических единиц на узбекском языке – лингвистической основы развития синтаксического анализа, методов выделения синтаксических единиц на узбекском языке, моделей тегов, лингвистической поддержки автоматической разметки.

Задачи исследования.

изучить сходства / различия и возможности парсинговых программ;

синтаксические категории тегов, разработка моделей тегов, подготовка рекомендаций по программированию;

изучить важность исследования синтаксических отношений и синтаксического единства в узбекском языке при синтаксической маркировке;

обобщить исследования синтаксиса простых и сложных предложений и сформировать их в качестве лингвистической основы программы синтаксического анализа.

В качестве объекта исследования были выбраны синтаксические единицы узбекского языка.

Предмет исследования – лингвистические теги и модели синтаксических тегов.

Методы исследования. Для освещения темы исследования использовались методы классификации, описания, статистического и компонентного анализа.

---

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

[58] Орхуи М. Computational analysis of uzbek nouns / Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018». (Труды конференции) – Ташкент, 2018. – 320 с.

[59] Турсунов А.Вопросы словообразования в формальных моделях тюркских языков (на примере узбекского языка) // Шестая Международная конференция по компьютерной обработке тюркских языков «TurkLang-2018» Труды конференции) – Ташкент, 2018. – 320 с.

Научная новизна исследования состоит из следующих:

программы синтаксического анализа, такие как Penn Treebank, SynTagger, Link Grammar Parser, HANKO в лингвистике мирового корпуса, были определены как подчиненная, систематическая схематическая грамматика, сходство в использовании традиционных синтаксических учений, а также лингвистическая неспособность интерпретировать или интерпретировать;

взгляды лингвистов А.Гуламова, М.Аскаровой, Г.Абдурахманова на фразу, Н.Махмудова, А.Нурмонова, А.Бердалиева на смысловой синтаксис и валентность, Х.Нематова, М.Кадырова на собирательную и лаконичную речь оказались теоретической основой для разработки программ синтаксического анализа для узбекского языкового корпуса;

разработаны методы спряжения слов в узбекском языке, такие как адаптация, спряжение, контроль, синтаксические категории тегов по содержанию, структуре предложений, фраз, единственные и двойные теги простых и сложных предложений, лингвистические модели, анализирующие и синтезирующие узбекские синтаксические единицы;

разработаны параметры поиска как система лингвистических синтаксических тегов и словосочетаний связанных, несвязанных, сопровождаемых, сложных составных предложений узбекского языка и переходных составных предложений, основанная на степени и согласованности простых, составных, сложных предложений, а также типа и пола.

Практические результаты исследования состоят из следующих:

изучаются теоретические основы разметки синтаксических единиц: теория и практика синтаксической разметки в мировой лингвистике; анализируются синтаксическая разметка, ее возможности, синтаксически размеченные трупы и их программное обеспечение;

разработаны принципы синтаксической разметки словосочетаний в корпусе узбекского языка; раскрывается роль лингвистико-синтаксических паттернов в разработке синтаксической разметки и лингвистической модели простых предложений в узбекском языке, а также создаются лингвистические модели на основе лингвистико-синтаксических паттернов;

теги и модели простых предложений в узбекском языке разработаны на основе теоретических данных по синтаксической разметке простых предложений;

разработаны синтаксические модели связанных, несвязанных и последовательных сложных предложений в узбекском языке;

определены теоретические основы синтаксической разметки сложных составных предложений узбекского языка.

Достоверность результатов исследования объясняется тем, что изученные материалы позволили сделать выводы, основанные на характере узбекского языка, их валидности, методическом совершенстве, опоре на проверенные источники при определении принципов разработки синтаксического разбора корпуса узбекского языка. (синтаксический разбор).

Научная и практическая значимость исследования. Исследование служит научно-теоретическим источником в разработке теоретической базы для разработки программ синтаксического разбора корпуса узбекского языка, в создании исследований в области компьютерной и корпусной лингвистики. Практическая значимость исследования заключается в применении программы в преподавании таких дисциплин, как компьютерная лингвистика, корпусная лингвистика, компьютерная лексикография, обработка естественного языка (NLP). Также можно будет разработать поиск по синтаксическим параметрам корпусов, сформированных на узбекском языке, на основе лингвистических моделей, предложенных в данной диссертации.

Внедрение результатов исследований. На основании научных результатов изучения лингвистических основ создания программы синтаксического разбора для Национального корпуса:

выводы о подчиненном подходе программ синтаксического анализа в мировой лингвистике корпуса, таких как Penn Treebank, SynTagger, Link Grammar Parser, XANKO, систематическая схематическая грамматика, сходство в использовании традиционных синтаксических учений, а также лингвистическая интерпретация или неспособность интерпретировать, а также Взгляды А. Гулямова, М. Аскарова, Г. Абдурахманова на лексику, взгляды Н. Махмудова, А. Нурмонова, А. Бердиалиева на семантический синтаксис и валентность, взгляды Г. Нематова, М. Кадырова на программы синтаксического анализа составных и кратких предложений для узбекского языка языковой корпус служит теоретической базой для развития использован в фундаментальном исследовательском проекте №PZ-20170927147 «Изучение древнетюркских надписей и фольклора до XIII века» (справочник Ташкентского государственного университета узбекского языка и литературы при министерстве высшего и среднего специального образования Республики Узбекистан №04/1-1238 от 08.04.2021 г.). В результате это послужило обогащению главы, посвященной методам автоматической обработки древнетюркских надписей;

теоретические мысли о методах спряжения слов в узбекском языке, такие как адаптация, сцепление, контроль, категории синтаксических тегов в соответствии с содержанием и структурой предложения, фразы, единственные и двойные теги простых и сложных предложений, лингвистические модели, анализирующие и синтезирующие узбекские синтаксические единицы использованы в практическом исследовательском проекте №63-11/41 «Сохранение ценностей, традиций, обычаев и традиций Фариша и знакомство с ними» (справочник Джизакского областного филиала Таджикского национально-культурного центра Республики Узбекистан №939 от 5 мая 2021 года.). В результате послужил обогащению главы, посвященной автоматической обработке и маркировке ценностей, традиций, обычаев и ритуалов Фариша;

из выводов о системе лингвистических синтаксических тегов и словосочетаний связанных, несвязанных, сопровождаемых, сложных

составных предложений узбекского языка и конструкция переходных составных предложений, степень и согласованность простых, составных, сложных типов предложений и параметры поиска на основе типа и пола использованы в практическом исследовательском проекте на тему "Каракалпакские предания-пословицы в качестве объекта лингвистического исследования" (справочник Каракалпакского отделения Академии наук Республики Узбекистан № 17.01 / 112 от 17 января 2021 г.). В результате использование в простых и сложных предложениях исследования единиц, различающихся по типу словообразования, дало возможность обогащения исследовательского раздела.

Апробация результатов исследования. Результаты данного исследования обсуждались в 5 международных, 5 республиканских научно-практических конференциях.

Объявление результатов исследования. По теме диссертации опубликовано 14 научных работ, в том числе 4 научных статей в рекомендованных научных публикациях докторских диссертаций Высшей аттестационной комиссии Республики Узбекистан, из них 1 в зарубежных журналах.

Структура и объём диссертации. Диссертация состоит из введения, трёх глав, заключения и списка использованной литературы, общий объем состоит из 128 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ ДИССЕРТАЦИИ

В разделе **"Введение"** обоснованы актуальность и важность темы, показана, что исследование зависит от приоритетов развития науки и технологий в республике, приводятся цели и задачи, описываются объект и предмет, описывается научная новизна и практические результаты, раскрывается научная и практическая значимость результатов, приведены сведения о внедрении, апробации, опубликованных работах и структуре диссертации.

Первая глава диссертации озаглавлена **«Теоретические основы разметки синтаксических единиц»**. В данной главе обсуждаются теория и практика синтаксической разметки, синтаксическая разметка и ее возможности в тэппинге, программном обеспечении для случаев с синтаксической меткой. Первый раздел главы, озаглавленный *«Теория и практика синтаксических тегов в мировой лингвистике»*, дает общее описание работы, проделанной в мировой компьютерной лингвистике по синтаксическим тегам. Синтаксическая разметка - это набор тегов, связанных с синтаксическим анализом текста, результатом синтаксического разбора, основанного на морфологическом анализе. Этот вид макета показывает синтаксическую связь между лексическим и другими синтаксическими устройствами (простое предложение, составное предложение, транслитерация)[60]. В то время как синтаксическая разметка при создании

[60] См: Захаров В.П., Богданова С.Ю. Корпусная лингвистика. –Иркутск: ИГЛУ, 2011.

первых корпусов была неавтоматической, синтаксическая разметка корпусов следующего поколения была сделана автоматически/полуавтоматически на основе программы синтаксического разбора. Существуют разные методы синтаксической маркировки: один - древовидный метод ассоциации слов в предложении, другой - путем присоединения синтаксического тега к текстовым единицам. Один из аннотированных постулатов, разработанных в 1993 г. Дж. Личем, автором книг «Ланкастер-Осло / Берген» (LOB) и Британский национальный корпус (BNC), - это идея четкого и понятного описания языковых знаков. По его мнению, макет корпуса для общепользования должен соответствовать трем принципам.

1. Разметка (аннотация корпуса) должна быть основана на схеме анализа, доступной пользователю в виде руководства или инструкции, и каждый параметр должен включать ее.

2. Открытый макет кейса для пользователя должен быть «в теоретическом отношении нейтральным»: параметры разметки должны состоять из понятной всем системы понятий. Если корпус рассчитан на конкретный проект, необходимо использовать в его обозначении специальную, авторскую и общепринятую классификацию: и в этом случае от разработчика требуется опираться на теорию того или иного языка.

3. Схема аннотации корпуса должна быть четко указана кем, для какой аудитории, потому что существуют разные юридические и технические ограничения на использование корпуса[61].

Таким образом, для разработки системы синтаксических тегов, наряду с достижениями компьютерных технологий, можно разработать программу синтаксического разборатора корпуса, основанную на теориях синтаксиса в узбекской лингвистике.

Во втором разделе главы *«Синтаксическая нотация и ее возможности в различных случаях»* рассматриваются типы синтаксической аннотации. Наше наблюдение за теоретическим материалом по принципу синтаксической аннотации, созданном Дж. Личем, показало, что аудитория корпуса с синтаксической аннотацией будет широкой; возможность такого корпуса для связи с информационными системами будет шире. Вместе с тем подход к построению корпуса, основанный на существующих классификациях, не возлагающий всю ответственность за согласованность выставления оценок на автора, позволяет выявить пробелы в языковых описаниях, недостатки языковых подходов, несоответствия.

И.М.Богуславский подчеркивает, что разметка текста осуществляется специальным тегом - маркером, и делит теги на отдельные (1), контейнерные (2). Один тег предоставляет информацию о текстовой единице (слове), а тег контейнера несет информацию о структуре текста, хранящегося в системе разметки.

1. Разделение текста на предложения выполняется с помощью пары тегов-контейнеров: <C>: </C>. Выпадающий тег может иметь другой

[61] Leech, G. Corpus annotation schemes / G. Leech Literary and Linguistic Computing, 1993. – 8/4. – P. 275-281.

параметр, которым является идентификатор речи <С ИД=идентификатор>. Этот тег представляет собой комментарий, описывающий отношения между предложениями в тексте.

2. Разделение текста на лексические элементы осуществляется парой контейнерных тегов: <W>: </W>. Слово также может иметь собственный идентификатор <WИД=идентификатор>.

3. Морфологические характеристики слова записываются одним тегом: <НОМ>; они помещаются внутри тегов контейнера. Тег <НОМ> имеет 4 поля: ИД – идентификатор, ЛЕММА – лексическая форма слова, POS – группа слов, FEAT – морфологические характеристики.

4. Информация о синтаксической структуре предложения представлена отдельным символом внутри тега <НОМ> - DOM: <НОМ ДОМ = идентификатор / тип связи>. Когда идентификатор относится к синтаксически подчиненному слову, тип связи отражает тип синтаксических отношений между доминирующим и подчиненным словом.

Формализм обладает достаточной гибкостью: он позволяет фиксировать не только готовую структуру, но и промежуточное состояние текста. В частности, вставив несколько тегов <НОМ> между одним тегом контейнера <W>: </W>, можно хранить информацию о нескольких вариантах морфологического анализа слова в едином формате. Можно сохранить древовидную структуру, введя несколько тегов DOM в тег <НОМ>.

Третий раздел главы озаглавлен *«Некоторые комментарии к синтаксически помеченным корпусам и их программному обеспечению»*. В этом разделе рассматриваются типы лингвистической информации в размеченном тексте: морфологическая, синтаксическая информация, ее значение в синтаксическом знаке, программа синтаксического разбора SynTagger. Разработка алгоритма синтаксического разбора также потребовала создания дополнительного фильтра: этот инструмент, состоящий из 2-4 членов, анализирует анализируемое предложение с использованием потенциальных сетей. Результат такого эксперимента также можно применить в построении следующей части корпуса, поскольку легче анализировать новые, автоматически построенные предложения. Модуль автоматической синтаксической разметки (SynTagger), предложенный О.И. Бабиной, Н.Ю. Дюминым, включает в себя лексическую единицу, которая стала синтаксически подчиненной, подчиненной частью текста в виде соединения, заключенного в круглые скобки (см. 3-рисунок).

Рекомендуется, чтобы пользователь определял начало и конец синтаксического блока, определяя его тип (существительное сочетание, глагольное сочетание, сочетание выражающее число). Модуль SynTagger позволяет автоматически различать группы синтаксической структуры при наличии морфологической разметки. Его можно использовать в синтаксических исследованиях, которые показывают специфику различных функциональных стилей или диалектов.

Есть также синтаксически помеченные корпусы среди мировых корпусов, особенно английского корпуса, у которых также есть свои

собственные программы синтаксического разбора. Среди них инструменты Penn Treebank[62] могут служить моделью для других синтаксических анализаторов, наиболее совершенным синтаксическим анализатором, в котором результаты синтаксического разбора проявляются четко. Существуют также синтаксически аннотированные исторические корпуса английского языка: Penn Parsed Corpus of Middle English (PPCME), Penn Chinese Treebank, Penn Korean Treebank, Prague Dependency Treebank, Arabic Syntactic/Predicate-Argument annotation.

Наши наблюдения показали, что эти программы синтаксического разбора – парсеры – могут служить примером «золотого стандарта» для различных проектов, поскольку они правильно подходят к методам синтаксического разбора. Эти системы синтаксического разбора служат необходимой экспериментальной площадкой для создания программы синтаксического разбора узбекского языка. Изучая перечисленные выше синтаксические анализаторы (системы синтаксического разбора), мы наблюдали, из каких компонентов состоит система синтаксического разбора и какие лингвистические знания необходимы для разработки тегов синтаксического разбора. Следовательно, чтобы разработать систему синтаксической разметки на каждом языке, необходимо смоделировать синтаксическую структуру этого языка. Следующим шагом после моделирования является создание системы синтаксических тегов, а последний шаг – присоединение синтаксических тегов к единицам языка текста.

Глава II исследования озаглавлена **«Теоретические основы синтаксической разметки фраз и простых предложений»**. В ней анализируются вопросы синтаксической разметки фраз в узбекском языке, роль лингвосинтаксических моделей и лингвистической модели в синтаксической разметке простых предложений на узбекском языке, синтаксическая разметка простых предложений. В первом разделе, озаглавленном *«О синтаксической маркировке словосочетаний в корпусе узбекского языка»*, обсуждаются принципы синтаксической маркировки, словообразование и ее изучение на узбекском языке, концепция синтаксической коммуникации, синтаксические отношения и синтаксическая валентность в синтаксической маркировке.

Синтаксическое тегирование – это фактор, запускающий поиск синтаксической единицы в теле языка. Самая простая форма поиска в языковом корпусе – это поиск по фразе, слову, фразе и словосочетанию. Эти символы присутствуют почти во всех аннотированных языковых корпусах. Формирование поиска на основе синтаксических и семантических символов – гораздо более сложный процесс, поэтому не все корпусы имеют такой взгляд на поиск. Как и в случае любого типа аннотации, для разработки системы синтаксических тегов, наряду с достижениями компьютерных технологий, можно разработать программу синтаксического разборатора

[62] https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

корпуса, основанную на теориях (взглядах) на синтаксис в узбекской лингвистике. Важно изучить, обобщить, сравнить весь созданный теоретический материал по синтаксису и выбрать правильный подход при создании синтаксических тегов. Каждый исследователь может поддержать определенную точку зрения на языковое моделирование по объективным / субъективным причинам: из-за структуры языка теория языковой формализации может не применяться к другому языку, если она применяется к одному языку[63]. Как упоминалось ранее, исходя из концепции минимизации, имеет смысл включать в систему тегов только необходимую информацию.

Основы узбекской формальной грамматической интерпретации получили широкое распространение благодаря работам А.Гуломова, М.Аскаровой, Г.Абдурахманова[64]. К 70-м годам прошлого века узбекская синтаксическая теория и интерпретация синтаксической структуры узбекского языка, основанная на методе формального анализа, была полностью сформирована, обобщена в научной грамматике[65], опубликованной в 1976 году. По результатам формального синтаксического описания необходимо перечислить исследования семантического синтаксиса и валентности, проведенные на синтаксисе узбекского языка и основанные на методах системно-структурного анализа. Обобщая теории этих исследований, можно выделить следующие теоретические основы для дифференциации синтаксических единиц:

1) хотя лемма, слово, фраза и словосочетание различаются как простые формы поиска в корпусе, поиск на основе морфологических символов ведет в аннотированных корпусах. Поскольку поиск, основанный на синтаксических и семантических символах, является относительно сложным процессом, невозможно синтаксически/семантически пометить все корпуса;

2) при разработке системы синтаксических тегов, наряду с современными инструментами/программами автоматического анализа, можно разработать программу синтаксического разборатора для корпусов узбекского языка на основе существующих взглядов на синтаксис в узбекской лингвистике. Важно изучить и обобщить весь теоретический материал о синтаксических единицах и выбрать правильный подход к созданию синтаксических тегов;

3) существуют разные подходы к синтаксическим единицам, поэтому определение системы синтаксических единиц узбекского языка также является ключевой задачей при выявлении единицы корпуса;

4) должен быть разработан специальный алгоритм, фильтр, для обнаружения соединений, которые невозможно различить по внешним

---

[63] Бабина О.И., Дюмин Н.Ю. Автоматизация лингвистической разметки корпуса текстов // http://helling100.narod.ru/pubs/Automation Babina Dyumin.pdf

[64] Гуломов А. Синтаксис ва пунктуация буйича машқлар тўплами. – Т.: Ўздавнашр, 1938, 1939; 1947); Ўзбек тили грамматикаси. 2-қисм. Синтаксис. –Т.:Ўздавнашр, 1940; Гуломов А. Ўзбек тилида аниқловчилар. –Т.: Ўздавнашр, 1941; Ўзбек тили грамматикаси. 2-қисм. Синтаксис. –Т.: Ўздавнашр, 1944-1960; Гуломов А. Содда гап. Ҳозирги замон ўзбек тили курси буйича материаллар. –Т.: ЎзФАнашр., 1955.

[65] Ўзбек тили грамматикаси. 2-том. Синтаксис. – Т.: Фан, 1976

признакам наличия пробела или соединения в корпусе. Это зависит от того, является ли определение предложения / фразы однокомпонентным или двухкомпонентным..

Также в этом разделе освещаются различия между концепциями синтаксической связи, синтаксической взаимосвязи и синтаксической валентности; описана его роль в синтаксической разметке. В ходе исследования была разработана система синтаксических тегов, основанная на лингвосинтаксических формах слов узбекского языка. В этой связи важно обратить особое внимание на ЛСФ[66], разработанные С. Назаровой. С.Назарова выделяет ЛСФ такой инвариант в виде [W$^{\text{морфологическое средство}}$ _ W$^{\text{морфологическое средство}}$], вариант [W$^{\text{родительный падеж}}$ _W $^{\text{притяжательный аффикс}}$], [Имя$^{\text{родительный падеж}}$ _ Имя $^{\text{притяжательный аффикс}}$], [Существительное $^{\text{родительный падеж}}$ _ Существительное$^{\text{притяжательный аффикс}}$], [Существительное$_{\text{имя собственное}}$ $^{\text{родительный падеж}}$ –Существительное$_{\text{нарицательное}}$ $^{\text{притяжательный аффикс}}$]. Конечно, проблема пометки фразы в языковом корпусе не решается общими (инвариантными) ЛСФ, но для определения фраз требуются относительно более точные шаблоны. Эти закономерности были изучены С. Назаровой, протестированы на сотнях речевых произведений, обобщены. Следовательно, мы можем построить модель образных фраз имя+имя, основанное на этих образцах. Для этого мы сначала выбираем специальные символы морфологических средств, обозначающие подчинение, с определенным характером частей, представляющих тип существительных в форме. В данном случае это обозначается тегами существительное = N, прилагательное = Adj, числительное = Num, нарицательное = N^{sub}, местоиммение = Pr, наречие = Prv, действие = Ger; родительный падеж = Case (или Cs), притяжательный аффикс = Possessive (или Pos). Таким образом, можно предложить следующие модели для модели имя+ имя синтаксической разметки фраз для тела языка:

1) [N ^{Cs}→ N ^{Pos}]: *страница книги;*
2) [N ^{Cs}→ Adj ^{Pos}]: *хрупкое дерево;*
3) [Adj ^{Cs}→ N ^{Pos}]: *аромат цветка/красного;*
4) [Adj ^{Cs}→ Adj ^{Pos}]: *сладость яблоки/большой;*
5) [N нарицательное ^{Cs}→ Adj ^{Pos}]: *один цветок;*
6) [Num ^{Cs}→ Num $^{Pos=}$]: *половина десяти;*
7) [N ^{Cs}→ Ger $^{Pos=}$]: *возвращение Отабека*;
8) [Ger ^{Cs}→ N $^{Pos=}$]: *место стеснения*;
9) [Ger ^{Cs}→ Ger $^{Pos=}$]: *брать и отдавать*;
10) [N ^{Cs}→ Adj $^{Pos=}$]: *остановка сердца*;
11) [Adj ^{Cs}→N $^{Pos=}$]: *испуганный взгляд*;
12) [N^{Cs}→ Prv$^{Pos=}$]: *сутки труда*;
13) [Pr ^{Cs}→ N $^{Pos=}$]: *моя родина*;
14) [Prv ^{Cs}→ N $^{Pos=}$]: *блаженство настоящего.*

---

[66] Назарова С. Бирикмаларда сўзларнинг эркин богланиш омиллари: филол.фан. номз.... дисс. автореф. – Тошкент: 1997. – 26 б.

В разделе *«Синтаксическая разметка простых предложений в узбекском языке: лингвистико-синтаксическая модель и лингвистическая модель»* описаны модели разметки простых предложений. На основании наших наблюдений мы пришли к следующим выводам о частях речи и типах предложений в соответствии с их смысловой связью:

1) также можно не прикреплять тег, содержащий такую информацию, к предложению, которое можно отличить по пунктуации;

2) требуется прикрепить двоичный тег к единицам, которые могут быть двумя разными частями речи;

3) при определении системы синтаксических тегов следует учитывать простоту интерфейса;

4) При создании системы синтаксических тегов необходимо добиться принципа автоматической обработки текста, соблюдать принцип минимизации ручной обработки.

В третьем разделе *«Проблемы синтаксической разметки простых предложений в узбекском языке»* анализируются методы разработки моделей и паттернов для разметки простых предложений. Исследователь Ш.Хамроева, изучающая лингвистические основы авторского корпуса узбекского языка, дает пошаговый подход к вопросу синтаксической разметки материалов корпуса и дает некоторые рекомендации по синтаксической разметке простых предложений[67]. По её мнению, комментарии, составляющие самую большую базу данных по синтаксису текста, представляют собой сбор информации о построении предложений. Независимо от того, каким образом синтаксис предложения изучает предложение, целесообразно охватить все эти символы в процессе тегирования. Потому что структурированный корпус должен затем служить базой данных для проведения различных исследований, связанных с синтаксисом, других разделов. Вносим следующие предложения по совершенствованию системы тегов, разработанной Ш.Хамроевой:

4) Для определения типов предложений по количеству грамматических центров один из <ПП>, </ПП> также <СП>, </СП>. Поскольку простые предложения образуют систему синтаксических тегов, нам подойдет только тег <ПП>, </ПП>

5) Чтобы различать тип предложения в соответствии с целью выражения, мы вводим теги «повествовательное предложение» = <ПП>, «вопросительное предложение» = <ВП>, «командное предложение» = <КП>.

6) Чтобы определить тип по наличию подлежащего, можно добавить такие знаки как «притяжательный» = <Е +>, «ownerless» = <Е->; Можно добавить такие символы, как безличные предложения «личное неизвестное предложение» - <ЛНП>, «назывное предложение» = <НП>, «семантически-функционально сформированное предложение» = <СФСП>.

---

[67] Хамроева Ш.М. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол.фан.бўйича фалсафа доктори (PhD) ... дис – Бухоро, 2018. – 250 б.

4) По участию основной, второстепенной части в список тегов также будут включены такие комментарии, как «нераспространенное предложение» = <НП> и «распространенное предложение» = <РП>.

5) В зависимости от наличия частей, не имеющих грамматической связи с предложением, могут быть введены теги «обращение» = <о>, </о>, «ввод» <в>, </в>.

Глава III диссертации озаглавлена **«Теоретические основы синтаксической разметки сложных предложений в узбекском языке».**

Первый раздел главы *«Синтаксическая разметка и моделирование союзных и бессоюзных сложных предложений в узбекском языке»* посвящен исследованию моделей и систем синтаксической маркировки союзных и бессоюзных сложных предложений. У каждого языка есть своеобразные синтаксические единицы. Поэтому необходимо разработать систему синтаксических тегов для сложных предложений на узбекском языке. Хотя в мировой компьютерной лингвистике нет конкретных исследований, посвященных проблемам синтаксических тегов, есть статьи, посвященные анализу некоторых вопросов[68]. Однако и в составе большеобъемных корпусов нетегированность сложных предложений – отсутствие возможности реализации вопроса/поиска принадлежащих к типам сложных предложений показывает, что есть еще много вопросов, ожидающих решения, которые необходимо решить. Ш.Хамроева пишет об общих тегах сложных предложений[69]: "Согласно соединительным средствам сложного предложения, «союзное СП» = <ССП>, </ССП>, «придаточное СП» = <СПП>, </СПП>, «бессоюзное СП» = <Бес-е СП>, </ Бес-е СП> интерпретация видов служит отдельным характером для выделения и интерпретации таких единиц.

Для разработки совершенной системы тегов сложных предложений целесообразно опираться на теоретический материал и классификации, собранные к настоящему времени в узбекской лингвистике.

Слова *бўлса, эса* союзы равных частей предложения объединенные при помощи *-у (-ю), -да* являются сложными предложениями. Сложно предложение связано по отношению взаимосодержанию простого предложения в составе делятся на такие виды как: 1) ССП при помощи соединительного союза; 2) ССП при помощи противительных союзов; 3) ССП при помощи разделительных союзов; 4) ССП при помощи слов *бўлса, эса* сўзлари ёрдамида; 5) ССП связанных отрицательных[70].

---

[68] Апресян Ю. Д., Богуславский И. М., Иомдин Б. Л. и др. Синтаксически и семантически аннотированный корпус русского языка: современное состояние и перспективы // Национальный корпус русского языка: 2003-2005. М.: Индрик, 2005, 193-214.; Сичинава Д. В. Обработка текстов с грамматической разметкой. инструкция разметчика // Национальный корпус русского языка: 2003-2005. Результатыиперспективы. – М., 2005, 136-154.

[69] Хамроева Ш.М. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Филол. фан. бўйича фалсафа доктори (PhD) дис. – Бухоро, 2018. – 250 б. – Б. 155.

[70] Замонавий ўзбек тили: Синтаксис. / Муаллифлар жамоаси. Масъул муҳаррирлар Ҳ.Ғ.Неъматов, Р.Сайфуллаева. ЎзР Олий ва ўрта махсус таълим вазирлиги, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети. – Тошкент: Мумтоз сўз, 2011. – 312 б. – Б. 226-235.

Значит, вначале в систему синтаксических тегов сложных предложения связанных в узбекском языке вводятся такие знаки как <b.BQG>, <z.BQG>, <a.BQG>, <yo.s.BQG>, <i.BQG>.

При связывании союзов, которые связаны противоположными и присоединенными союзами частицы *-у (-ю)* используются одновременно. В такой момент через союзные средства модели определения виды сложного предложения сталкиваются с грамматическими омонимиями. Поэтому в таких случаях модель определения сложных предложений необходимо сформировать следующим образом.

3. Если [Wpm] + ва/ ҳам/ -у/ -ю/ -да+[Wpm] бўлса = <a.BQG>. при этом [Wpm] = прсотое предложение.

4. Если [Wpm] + аммо/ лекин/ бирок/ -у/ -ю/ -да + [Wpm] бўлса = <z.BQG>. при этом [Wpm] = простое предложение.

Даже в обеих этих моделях возникает омонимия, что приводит к автоматическому синтаксическому анализу, который не может различать союзы, связанные с частицами *-у, -ю, -да*. Для этого пользователь должен анализировать результат автоматического анализа один за другим. В будущем, если будут созданы модели семантического анализа узбекского языка, формальная неопределенность будет определяться на основе семантических моделей.

При повторении разделительного союза простые предложения в составе сложного предложения разделяются знаком препинания. Модель и теги можно отметить следующим образом:

3. [Гоҳ...Wpm] + [гоҳ...Wpm] бўлса = <a.BQG>.

4. [Ё/ёки...Wpm] + [ёки...Wpm] бўлса = <a.BQG>.

Однако при одиночном использовании разделительного союза, когда опущен знак препинания модель приобретает другой оттенок.

Слова *бўлса, эса* вместе со связыванием частей союзного сложного предложения, выражают наличие сравнительных, противоречивых отношений между ними. Связующие средства в таких предложениях имеют форму [...са + ...бўлса/эса]. Синтаксическая модель и тег могут быть отметить в форме [Wpm...са] + [бўлса/эса... Wpm] = <yo.s.BQG>.

Часть сложного предложения связанная отрицательного отношения связана ответной или отрицательной нагрузкой; между предложениями ставится запятая. И союз отрицания, обычно, стоит в начале простого предложения[71]. Модель и тег будут следующими: [На...Wpm]+ [на ... Wpm] = <i.BQG>.

Через взаимодействие связанных частей сложного предложения выражаются сравнение, привязанность, раделение, причина и следствие, толкование, отношение содержания[72]. К сожалению, мы не можем ввести

---

[71] Замонавий ўзбек тили. Синтаксис. / Муаллифлар жамоаси. Масъул муҳаррирлар Ҳ.Ғ.Неъматов, Р.Сайфуллаева. ЎзР Олий ва ўрта махсус таълим вазирлиги, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети. – Тошкент: Мумтоз сўз, 2011. – 312 б. – Б. 256-264.

[72] Кўрсатилган манба. Б. 264.

параметр, который указывает на такую связь с поисковой системой на основе синтаксических символов или системой синтаксических тегов.

Если учитывать, что союзы (тона) без союзов перемежаются, можно определить их модель и теги в следующем порядке:

5. [Wpm] <,> [Wpm] бўлса = <Б-сиз ҚГ>.
6. [Wpm] <:> [Wpm] бўлса = <Б-сиз ҚГ>.
7. [Wpm] <–> [Wpm] бўлса = <Б-сиз ҚГ>.
8. [Wpm] <;> [Wpm] бўлса = <Б-сиз ҚГ>.

Есть простые предложения с отделенным разрезом, они подходят к форме [Wpm] <–> [Wpm]: первое простое предложение <–> сказуемое. Такие отдельные разделы определяются программой как составное предложение, что приводит к грамматической омонимии. Например: *Каттагина, чирик тўнгак орқасида мукка тушган гавдага кўзи тушди – югурди (Ойбек)....пайқамай қолди: бакларга тегдими, моторигами ёки бомбалар солинган кассеткаларгами –билмади.* Пока что невозможно автоматически обнаружить СП с такой структурой, нажав. Дальнейшее уточнение онтологических моделей в узбекском языке может помочь выявить структурную / грамматическую двусмысленность и омонимию.

Во втором разделе главы, озаглавленном *«Синтаксическая разметка и моделирование сложных предложений узбекского языка»*, анализируется система лингвистических меток сложных предложений узбекского языка в соответствии с их связующими средствами, семантический типовой подход в моделировании сложных предложений узбекского языка.

При разработке системы тегов СПП мы опирались на выводы о классификации ПП (придаточного предложения) в узбекской лингвистике. Сложное предложение, части которого связаны при помощи придаточного связующего слова или при функции такого связующего является придаточным сложным предложением. В составе придаточного сложного предложения тег выражающий предложение определяем следующим образом: бош гап=[bG]; эргаш гап =[eG].

При разработке синтаксических моделей формирующее средство – играет важную роль связующее средство простых предложений в составе сложных предложений. Поэтому связывающие средства придаточного предложения выполняет функцию основного средства для астоматичекого определения вида сложного предложения. При разработке автоматического тегирования, модели и тегов придаточного предложения существующие правила служат крепкой теоретической основой. Поэтому рассмотрим следующие типы по способам соединения придаточного предложения:

1. Сложноподчинённое предложение при помощи вспомогательных средств. Придаточное предложение означеющее причину содержания выраженного в главном предложении соединяется таких вспомогательных средств как *шунинг учун, шу сабабли, шу туфайли сингари*. Они приходят в составе главного предложения. На этой основе можно составить синтаксические модели таких сложных предложений. Если: [eG] + <KQ>[bG]

бўлса, [eG] + [bG] = <EQG> / <сабаб-EQG>. В данном теге выражает СПП и его причинное значение с точки зрения смысла.

2. Сложноподчинённое предложение при помощи слова *деб*. Придаточное предложение означающее цель, причину смысла выраженного в главном предложении часто связывается при помощи слова **деб**, сказуемое придаточных предложений выражаются глаголами повелительного наклонения III лица. Слова *деб* в составе таких предложений синонимичен со вспомогательным словом *учун*, поэтому могут свободно заменять друг друга[73]. На этой основе синтакчисекая модель такого сложного предложения выглядит следующим образом. Если: [bG] + <…sin deb Q>[eG] бўлса, [eG] + [bG] = <EQG> / <цель-EQG>. В данном теге существует информация о том, что Сложноподчинённое предложение является сложным предложением, которое выражает смысл цели.

3. Сложноподчинённое предложение посредством условного наклонения. При выражении сказуемого сложного предложения через глаголы в форме условного наклонения, условный суффикс также является средством соединения придаточного предложения с главным предложением. Синтаксическую модель таких сложных предложений можно составить следующим образом. Если: [eG...ca] + [bG] бўлса, [eG] + [bG] = <EQG> / <шарт-EQG>. В данном теге существует информация о том, что это сложноподчинённое предложение, которое использовалось для обозначения условия.

4. Сложноподчинённое предложение с указательным местоимением. Придаточное предложение применяемое для разъяснения смысла указательного местоимения в составе главного предложения соединяется с главным предложением при помощи частицы *-ки*. Эта частица содержится в сказуемом главного предложения и придаточное предложение отделяется от главного предложения запятой. Связь при помощи такого средства мы отмечаем тегом <olmoshQ>. Синтаксические теги вида связи такого порядка можно составить следующим образом. Если: [bG\shu\shuni\shunga\shunday...ki] + [eG] бўлса, [bG] + [eG] = <EQG> / <шарт-EQG>. В данном теге существует информация о том, что это сложноподчинённое предложение, которое выражало смысл указательного местоимения в главном предложении.

5. СПП с относительным словом. Такие вопросительные местоимения как *ким, нима, қанча, қанчалик, қандай, қаер* применяемые в составе придаточного предложения и применяемость относительно друг друга такие местоимения как *шу, ўша, шунча, шунчалик, шундай* идущие ему ответом в составе главного предложения, из-за того, что один предусматривает другого, являются относительными словами. Сказуемое придаточного предложения выражается глаголами формы условного наклонения[74]. При

---

[73] Замонавий ўзбек тили: Синтаксис / Муаллифлар жамоаси. Масъул муҳаррирлар ҲҒ.Неъматов, Р.Сайфуллаева. ЎзР Олий ва ўрта махсус таълим вазирлиги, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети. – Тошкент.: Мумтоз сўз, 2011. – 312 б. – 288-293.

[74] Кўрсатилган манба. – 295-296.

помощи такого средства связь мы отмечаем тегом <nisbiy so`zQ>. Вид связи данного порядка можно составить следующим образом. Если: [eG\ *ким*\ *нима*\ *қанча*\ *қанчалик*\ *қандай*\ *қаер*...sa] + [eG\ *шу*\ *ўша*\ *шунча*\ *шунчалик*\ *шундай*] бўлса, [eG] + [bG] = <EQG> / <nisbiy so`z-EQG>. В данном теге существует информация о том, что СПП выразившее значение разъяснения относительного слова (местоимения) в главном предложении.

Исходя из приведенного выше анализа в поле поиска СПП можно поместить следующий параметр:

6. СПП при помощи вспомогательных устройств
7. СПП при помощи слова *деб*
8. СПП посредвтом условного наклонения
9. СПП с указательным местоимением
10. СПП с относительным словом

Также, для реализации автоматического анализа по формированию СПП посредством какого связующего систему модели и тегов выразим в следующей форме:

6) Если: [eG] + <KQ>[bG] бўлса, [eG] + [bG] = <EQG> / <сабаб-EQG>.

7) Если: [bG] + <...sin deb Q>[eG] бўлса, [eG] + [bG] = <EQG> / <мақсад-EQG>.

8) Еслир: [eG...sa] + [bG] бўлса, [eG] + [bG] = <EQG> / <шарт-EQG>.

9) Если: [bG\shu\shuni\shunga\shunday...ki] + [eG] бўлса, [bG] + [eG] = <EQG> / <шарт-EQG>.

10) Если: [eG\ *ким*\ *нима*\ *қанча*\ *қанчалик*\ *қандай*\ *қаер*...sa] + [eG\ *шу*\ *ўша*\ *шунча*\ *шунчалик*\ *шундай*] бўлса, [eG] + [bG] = <EQG> / <nisbiy so`z-EQG>.

Таким образом, теги и модели для системы автоматического поиска СПП на узбекском языке могут быть разработаны на основе средств соединения следующего предложения. Всего было разработано 87 моделей бессоюзных и сложноподчиненных предложений, в том числе, 67 моделей основанных на типе связующего средства для настройки поиска видов сложноподчиненных предложений по разъяснению какой части в главном предложении придаточного предложения.

В третьем разделе главы озаглавленном *«Автоматический анализ и синтез сложносочиненных предложений и построение прямых предложений в узбекском языке»* речь идет о моделях автоматического анализа сложносочиненных предложений и конструкции сложного предложения с прямым предложением.

Сложнсочиненые предложения делятся на четыре группы, такие как несколько придаточных предложений, несколькими главными предложениями, смешанные сложносочиненные предложения и сложносочиненные предложения с объединеными частями[75]. В системе поиска при определении такого предложения можно использовать модели

[75]Замонавий ўзбек тили: Синтаксис / Муаллифлар жамоаси. Масъул муҳаррирлар Ҳ.Ғ.Неъматов, Р.Сайфуллаева. ЎзР Олий ва ўрта махсус таълим вазирлиги, Мирзо Улуғбек номидаги Ўзбекистон Миллий университети. – Тошкент: Мумтоз сўз, 2011. – 312 б. – Б. 463.

бессоюзное СП или союзное СП: граница сложного предложения отмечается при помощи знака препинания в качестве единой целостности. Только в бессоюзном сложном предложении принцип «от точки до точки» если определит сложное предложение, состоящее из двух простых предложений, тогда как в сложносочиненном предложении он различает единицу, состоящую из трех или более простых предложений. Такой метод также применяется к сложносочиненным предложениям, связанным равным союзом: не остается необходимость создания специальной модели ССП. Несколько сложных предложений связанных посредством равного союза определяется на промежутке "от точки до точки". Но в поле поиска можно выделить тип сложносочиненного предложения. При этом программа выполняет двухэтапный анализ на основе следующего алгоритма:

3) Разделяет единицу на промежутке "от точки до точки";

4) Соглано количества связующего средства делается вывод о "сложном предложении" (также, его вид) или "сложносочиненном предложении".

На основе приведенной выше классификации программа синтаксического разбора может включить следующие параметры в интерфейс поиска сложносочиненных предложений:

- ☐ Сложное предложение с несколькими придаточными предложениями
- ✓ прямое подчинение (соподчинение)
- ✓ последовательной подчинение
- ☐ сложносочиненное предложение с несколькими главными предложениями
- ☐ смешанное сложносочиненное предложение
- ☐ Сложное предложение с объединенными частями

Среди них смешанное сложносочиненное предложение считаестя единицей, которая анализируется на основе как знаков препинания, так и связующего средства. И автоматический поиск таких сложных предложений принцип разделения "от точки до точки", а также, опирается на модели основанные на связующем средстве сложных предложений. В языковом корпусе, как и в любой другой синтаксической единице, также можно настроить автоматический поиск составного предложения. При построении моделей автоматического анализа этого можно использовать положение показателя сечения простого предложения и положение знаков препинания в структуре отрывка.

Данная устойчивая (точная) конструкция сложного предложения с прямой речью несколько упрощает процесс автоматического анализа. В прямой речи, использование знака препинания, в зависимости от места прямого предложения, различна: если прямая речь являясь повествовательным предложением, идет перед словами автора, после него ставиттся запятая и тире: *"Юринг, мен ўша томонга бораман", – деди Комила*. Знак вопроса и восклицательный остаются внутри кавычек. Модель

и тег данной позиции можно отметить следующим образом: “WPm”, –WPm = <K>, <M>.

а) если прямая речь идет после слов автора, после слов автора ставится двоеточие: *Маърузачи бундай деди: “Ўсиб бораётган авлодга китоб худди мактаб каби керак”.* Модель и тег данной позиции следующее: WPm: “WPm” = <M>: <K>.

б) если слова авора идут внутри прямой речи, знак препинания ставится следующим образом: “Бизнинг қишлоғимизда, – деди Фазлиддин, – киши зерикмайди”. При этом неоконченность слов автора и прямой речи может привести к неразберихе в автоматическом анализе. Поэтому часть внутри кавычек слова автора, то есть отмечается так “WPm”, –WPm, –“WPm” =<M>,<K>, <M>. Однако насколько точен этот тег (вероятный символ), будет определено после запуска программы синтаксического разбора.

Следовательно, сложносочиненные предложения и конструкция сложных предложений анализируются в основном с помощью знаков препинания: граница предложения определяется по принципу от пунктуации к пунктуации.

ЗАКЛЮЧЕНИЕ

1. Синтаксическое тегирование является основным процессом помогающим в реализации автоматического синтаксического разбора текста, состоит из комплекса синтаксических тегов, показывает синтаксическую связь между синтаксическими устройствами. Если при создании первых корпусов выполнялось вручную, выясненно, что синтаксическая разметка корпусов следующего поколения реализовывалось автоматически/полуавтоматически на основе программы синтаксического разбора.

2. Основной алгоритм применяемый в синтаксической разметке – это синтаксическая декомпозиция, которая различает предложение на основе пустот (пробелов), знаков препинания. Однако этот метод не идеален, потому что символ, используемый в качестве разделителя, может использоваться не только в конце предложения, но и в середине. Такие единицы, как заголовок, единица, разделяющая раздел, изображение, имя таблицы, нижний колонтитул, должны быть выражены в форме предложения, хотя это не так.

3. В синтаксически размеченном тексте, вместе с морфологическим знаком, к каждому предложению пишется комментарий о его синтаксической структуре. В синтаксической записи теги делятся на индивидуальные и контейнерные. Один тег предоставляет информацию о текстовой единице (слове), а тег контейнера несет информацию о структуре текста, хранящегося в системе разметки. Разделение текста на предложения осуществляется с помощью пары тегов-контейнеров. Было обнаружено, что когда идентификатор относится к синтаксически подчиненному слову, тип связи отражает тип синтаксических отношений между доминантным и подчиненным словом.

4. Для разработки системы синтаксической разметки на каждом языке, требуется смоделирование синтаксической структуры этого языка. Следующим этапом после моделирования является создание системы синтаксических тегов, а последний шаг – это присоединение синтаксических тегов к единицам языка текста. Также существует необходимость учитывания удобности интерфейса при определении системы синтаксических тегов. При создании системы синтаксических тегов рекомендуется добиться автоматической обработки текста, следуя принципу минимизации по мере возможности ручной обработки.

5. В аннотированных корпусах лидирует поиск на основе морфологического знака. Поскольку поиск, основанный на синтаксических и семантических знаках, является относительно сложным процессом, возможности, средства синтаксического/семантического тегирования всех корпусов не существует.

6. При разработке программы синтаксического анализатора корпуса узбекского языка, ее системы синтаксических тегов, современных программ автоматизированного анализа, а также изучения и обобщения существующих теоретических материалов по синтаксису в узбекском языкознании важно выбрать правильный подход при создании синтаксических тегов. Предлагается разработать специальный алгоритм, фильтр для обнаружения соединений, не различающихся по внешним признакам наличия разрыва или соединения в организме.

7. При синтаксическом тегировании учитывает тип синтаксического соединения и включает тег, представляющий синтаксическое соединение из системы тегов. Синтаксическая связь и синтаксическая взаимосвязь следует различать в синтаксическом тегировании материала языкового корпуса. Нет возможности форменного различия: связующие средства одинаковы. По этой причине в регистр требуется введения тега, который различает синтаксическую связь и синтаксическое отношение. В синтаксической системе нотации корпуса очень важен разговорный поиск: для организации разговорного поиска требуется, чтобы единицы корпуса имели тег, разделяющий фразу и фразу. В синтаксической системе нотации корпуса очень важен разговорный поиск: для организации разговорного поиска требуется, чтобы единицы корпуса имели тег, разделяющий словосочетание и словосоединение.

8. Уместно использовать лингвистико-синтаксический шаблон и систему морфологических тегов для определения подчиненно-доминантного отношения без морфологического индекса, потому что лингвистико-синтаксический шаблон указывает, какая группа слов связана между собой. Если единицы корпуса морфологически тегированы, можно будет автоматически помечать синтаксически родственные соединения без морфологического показателя на основе тега и лингосинтаксического шаблона. Конечно, проблема пометки фразы в теле не решается общими (инвариантными) лингвистико-синтаксическими шаблонами: требуются

более точные шаблоны в определении фраз. На основании этого было предложено 14 моделей шаблон имя+имя синтаксической разметки фраз.

9. Единицы, которые могут быть идентифицированы на основе косвенных символов в регистре, также нельзя прослушивать: потому что такие единицы выделяются соответствующей пунктуацией. Также можно не прикреплять тег, содержащий такую информацию, к предложению, которое можно отличить по пунктуации; предлагается прикрепить двоичный тег к единицам, которые могут быть двумя разными частями речи.

10. Средства связывающие сложное предложение могут быть основным первичным средством автоматического определения типа сложного предложения: программа синтаксического тегирования (парсер) определяет границы простых предложений между ними с помощью знака препинания и конъюнктива. Программное обеспечение для автоматического анализа не может различить союзы, вызывающие одноименные способы связывания. Для этого пользователь должен анализировать результат автоматического анализа один за другим. При создании моделей семантического анализа узбекского языка формальная неопределенность определяется на основе семантических моделей.

SCIENTIFIC COUNCIL AWARDING SCIENTIFIC
DEGREES PhD.03/04.06.2020.Fil.113.02 AT
JIZZAKH STATE PEDAGOGICAL INSTITUTE

JIZZAKH STATE PEDAGOGICAL INSTITUTE

KHIDIROV OTABEK JURABOEVICH

LINGUISTIC BASIS OF CREATING A PARSING PROGRAM FOR THE NATIONAL CORPUS

10.00.01 – Uzbek language

**DISSERTATION ABSTRACT FOR DOCTOR OF PHILOSOPHY (PhD)
ON PHILOLOGICAL SCIENCES**

Jizzakh – 2021

The theme of the Dissertation of Doctor of Philosophy (PhD) was registered at the Supreme Attestation Commission of Ministers of the Republic of Uzbekistan under number B2020.4.PhD/Fil 24.

The Dissertation has been prepared at Jizzakh State Pedagogical Institute

The Abstract of the (PhD) dissertation is posted in three (Uzbek, Russian, English (resume)) languages on the website of the Scientific Council (jspi.uz) and «ZiyoNet» information and educational portal (www.ziyonet.uz).

Scientific supervisor:	**Mengliev Bakhtiyor Rajabovich** Doctor of Philological sciences, Professor
Official opponents:	**Urinboeva Dilbar Bozorovna** Doctor of Philological sciences, Associate Professor
	Abjalova Manzura Abdurashetovna Doctor of Philological sciences (PhD), Associate Professor
Leading organization:	**Karshi State University**

The defense of the Dissertation will take place on «11» «11» 2021, at «10°°» at the meeting of Scientific Council PhD.03/04.06.2020.Fil.113.02 awarding scientific degrees at Jizzakh State Pedagogical Institute (Address: 130100, Jizzakh, str. Sh.Rashidov, 4. Tel: (872) 226-13-57; fax: (872) 226-46-56; e-mail: jspi info@umail.uz, The main building of Jizzakh State Pedagogical Institute, 2nd Floor, Meeting Hall).

The Dissertation can be reviewed at the Information Resource Centre of Jizzakh State Pedagogical Institute (Address 130100, Jizzakh, str. Sh.Rashidov, 4. Tel: (872) 226-13-57; fax: (872) 226-46-56).

The Abstract of the Dissertation was distributed on «29» 10 2021.
(Mailing report № 9 on «29» 10 2021).

A.E.Mamatov
Chairman of the Scientific Council awarding Scientific degrees, Doctor of Philological sciences, Professor

F.E.Ibragimova
Scientific Secretary of the Scientific Council awarding Scientific degrees, Candidate of Philological sciences, Associate Professor

U.Kosimov
Chairman of the Scientific Seminar at the Scientific Council awarding Scientific degrees, Doctor of Philological sciences, Associate Professor

INTRODUCTION (Abstract of the PhD Dissertation)

The aim of the Research work is to develop recommendations on the linguistic basis of creating a program for tagging syntactic units in Uzbek, as well as to create a linguistic support for the automatic annotation of phrases and sentences in Uzbek.

The object of the Research work. The syntactic units of the Uzbek language were selected as the object of research.

The scientific novelty of the Research work:

different aspects of syntactic analysis programs such as Penn Treebank, SynTagger, Link Grammar Parser, HANKO in world corpus linguistics have been identified, such as subordinate approach, systematic scheme grammar, similarities in use of traditional syntactic teachings, as well as linguistic interpretation or inability to interpret;

Linguists A.Gulamov, M.Askarova, G.Abdurahmanov on the wording, N.Mahmudov, A.Nurmonov, A.Berdialiev's views on semantic syntax and valence, H.Nematov, M.Kadirov's views on compound and concise sentences Uzbek language corpus proved to have served as a theoretical basis in the development of parser programs for;

Methods of conjugation of words in Uzbek, such as adaptation, conjugation, control, syntactic tag categories by type, content, structure, single and double tags of simple and compound sentences, linguistic models analyzing and synthesizing Uzbek syntactic units;

The system of linguistic syntactic tags and word combinations of Uzbek connected, unconnected, followed, complex compound sentences and exemplary compound sentence constructions, degree-coherence and gender-based search parameters of simple, compound, complex sentence types have been developed.

The Implementation of the Research work results. Based on the scientific results of the study of the linguistic basis of the creation of a parsing program for the National Corps:

the subordinate approach of syntactic analysis programs in world corpus linguistics such as Penn Treebank, SynTagger, Link Grammar Parser, XANKO, systematic schematic grammar, similarities in the use of traditional syntactic teachings, as well as linguistic interpretation or inability to interpret, as well as A. Gulyamov, M.Askarova, G.Abdurahmanov's views on vocabulary, N.Mahmudov, A.Nurmonov, A.Berdialiev's views on semantic syntax and valence, H.Nematov, M.Kadirov's views on compound and concise sentences parser programs for Uzbek language corps PZ-20170927147 was used in the fundamental research project "Study of ancient Turkic writings and folklore up to the XIII century" (Tashkent State University of Uzbek Language and Literature named after Alisher Navoi, April 4, 2021, No. 04 / 1-1238). reference). The result served to enrich the chapter of the fundamental research project devoted to the methods of automatic processing of ancient Turkic inscriptions;

Theoretical ideas about the methods of conjugation of words in the Uzbek language, such as adaptation, adhesion, control, syntactic tag categories by type,

sentence, singular and double tags of simple and compound sentences, linguistic models that analyze and synthesize Uzbek syntactic units. Used in the practical research project No. 11/41 "Preservation of Forish values, traditions, customs and traditions and their introduction to the general public" (Reference No. 39 of May 5, 2021 of the Jizzakh regional branch of the Tajik National Cultural Center of the Republic of Uzbekistan). As a result, it served to enrich the chapter of the applied research project devoted to the methods of automatic processing and tagging of Forish values, traditions, customs, and traditions;

FA-A1-G007 on the system of linguistic syntactic tags and method of word combination of connected, unconnected, followed, complex compound sentences and transliterated compound sentences in Uzbek language Used in the practical research project "Karakalpak proverbs as an object of linguistic research" (Handbook of the Karakalpak branch of the Russian Academy of Sciences dated January 17, 2021, No. 17.01 / 112). As a result, the use of units that differ in the type of word formation in simple and compound sentences of the research allowed to enrich the research section.

Publication of Research results. 14 scientific works on the topic of the dissertation were published, including 4 scientific articles in the scientific publications recommended for publication of the main scientific results of doctoral dissertations of the Higher Attestation Commission under the Cabinet of Ministers of the Republic of Uzbekistan, 1 of them in foreign journals.

The outline of the Research work. The Dissertation consists of Introduction, 3 Chapters, General Conclusions, a List of References and Appendices, a total size of 128 pages.

ЭЪЛОН ҚИЛИНГАН ИШЛАР РЎЙХАТИ
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS

I бўлим (I часть; I part)

1. Khidirov O. Problems OF Syntactic Tagging OF Simple Common Sentences In Uzbek Language // The American Journal of Social Science and Education Innovations (ISSN-2689-100x) Published: December 31, 2020/Pages:340-344 (Impact Factor 2020:5.525)

2. Хидиров О. Синтактик разметкаланган корпуслар ва уларнинг дастурий таъминоти борасида айрим мулоҳазалар//Илм сарчашмалари. – Урганч, 2020. – № 9. – Б. 32-35. (10.00.01; №03).

3. Хидиров О. Ўзбек тили корпусида сўз бирикмаларини синтактик теглашнинг назарий асослари // Тафаккур зиёси журнали.– Жиззах, 2020. – № 4. – Б. 184-186. (10.00.01; №29).

4. Хидиров О. Корпус разметкасида ифодаланган лингвистик ахборотлардан фойдаланиш // ЎзМУ хабарлари журнали. – Тошкент, 2021. – № 1. – Б. 282-285. (10.00.01; №15).

5. Хидиров О. Ўзбек тилшунослигида йиғиқ гап ва ихчам гап ҳақидаги назарий умумлашмалар хусусида // MODERN SCIENTIFIC CHALLENGES AND TRENDS. COLLECTION OF SCIENTIFIC WORKS OF THE INTERNATIONAL SCIENTIFIC CONFERENCE.WARSAW, POLAND WYDAWNICTWO NAUKOWE "ISCIENCE" 30 December 2020. – Б. 99-102.

6. Хидиров О. Ўзбек тилида эргашган қўшма гапларни синтактик теглаш ва моделлаштириш // Ўзбек миллий ва таълимий корпусларини яратишнинг назарий ва амалий масалалари : Халқаро илмий-амалий конференция материаллари. – Тошкент, 2021. – Б. 310-313.

7. Хидиров О. Тил корпусида синтактик аннотация (разметка) турлари / Филологиянинг умумназарий масалалари: Республика илмий-амалий конференцияси.– Тошкент, 2020.– Б. 182-185.

8. Хидиров О. Жаҳон тилшунослигида синтактик теглашнинг назарий асослари / Хорижий филология, адабиётшунослик ва таржимашунослик масалалари: Республика илмий-амалий конференцияси.– Жиззах, 2021.– Б. 139-141.

9. Хидиров О. Ўзбек тили корпусида сўз бирикмаларини лисоний-синтактик қолиплар асосида синтактик теглаш / Компьютер лингвистикаси: муаммолар, ечим, истиқболлар: Республика илмий-техникавий конференция.– Тошкент, 2021.– Б. 40-44.

II бўлим (II часть; II part)

10. Хидиров О. Синтактик разметка ва унинг турли корпуслардаги имкониятлари // Тилни ўқитиш ва ўрганишда XXI аср кўникмалари: Халқаро илмий-амалий конференция. – Жиззах, 2020. – Б. 138-140.

11. Хидиров О. Жаҳон тилшунослигида корпус лингвистикаси // Тилни ўқитиш ва ўрганишда XXI аср кўникмалари: Халқаро илмий-амалий конференция. – Жиззах, 2021. – Б. 177-180.

12. Хидиров О. Ўзбек тилида богланган қўшма гапларни синтактик теглашнинг назарий асослари // Ўзбек мутафаккирларининг тил назариясига оид қарашлари: Халқаро илмий-назарий анжуман материаллари. – Тошкент, 2021. – Б. 269-272.

13. Хидиров О. Корпус лингвистикаси ҳақида айрим мулоҳазалар / Ўзбек филологиясининг долзарб муаммолари: таҳлил ва талқинлар. Республика илмий-амалий конференцияси.– Тошкент,2020.– Б. 217-219.

14. Хидиров О. Синтактик теглашда (изоҳлашда) синтактик алоқа, синтактик муносабат тушунчаси / Ўзбек тили ва адабиёти тарихи, тадрижий тараккиётининг ўрганилиши: натижалар ва муаммолар: Республика илмий-амалий анжуман материаллари.–Тошкент, 2021.– Б. 119-120.

Автореферат Самарқанд давлат университетининг
"СамДУ Илмий ахборотномаси" журнали таҳририятида
таҳрирдан ўтказилди (26.10.2021 йил).

2021 йил 27 октябрда босишга рухсат этилди.
Офсет босма қоғози. Қоғоз бичими 60×84 1/16.
"Times" гарнитураси. Офсет босма усули.
Ҳисоб-нашриёт т.: 3,5. Шартли б.т. 3,3.
Адади 100 нусха. Буюртма №28/10.

СамДЧТИ нашр-матбаа марказида чоп этилди.
Манзил: Самарқанд ш., Бўстонсарой кўчаси, 93.