

**FARG‘ONA DAVLAT UNIVERSITETI HUZURIDAGI
ILMIY DARAJALAR BERUVCHI
DSc.03/30.12.2019.Fil.05.02 RAQAMLI ILMIY KENGASH**

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT O‘ZBEK TILI VA
ADABIYOTI UNIVERSITETI**

ABJALOVA MANZURA ABDURASHETOVNA

O‘ZBEK TILI ONTOLOGIYASINI YARATISH TAMOYILLARI

10.00.11 – Til nazariyasi. Amaliy va kompyuter lingvistikasi

**FILOLOGIYA FANLARI DOKTORI (DSc)
DISSERTATSIYASI AVTOREFERATI**

Toshkent – 2022

Doktorlik (DSc) dissertatsiyasi avtoreferati mundarijasi

Оглавление автореферата докторской (DSc) диссертации

Content of the abstract of doctoral (DSc) dissertation

Abjalova Manzura Abdurashetovna О‘zbek tili ontologiyasini yaratish tamoyillari	5
Abjalova Manzura Abdurashetovna Principles of creating an ontology of the Uzbek language	38
Abjalova Manzura Abdurashetovna Принципы создания онтологии узбекского языка	72
E‘lon qilingan ishlar ro‘uxati Список опубликованных работ List of published works.....	78

**FARG‘ONA DAVLAT UNIVERSITETI HUZURIDAGI
ILMIY DARAJALAR BERUVCHI
DSc.03/30.12.2019.Fil.05.02 RAQAMLI ILMIY KENGASH**

**ALISHER NAVOIY NOMIDAGI TOSHKENT DAVLAT O‘ZBEK TILI VA
ADABIYOTI UNIVERSITETI**

ABJALOVA MANZURA ABDURASHETOVNA

O‘ZBEK TILI ONTOLOGIYASINI YARATISH TAMOYILLARI

10.00.11 – Til nazariyasi. Amaliy va kompyuter lingvistikasi

**FILOLOGIYA FANLARI DOKTORI (DSc)
DISSERTATSIYASI AVTOREFERATI**

Toshkent – 2022

Fan doktori (DSc) dissertatsiyasi mavzusi O'zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi Oliy attestatsiya komissiyasida B2021.4DSc/Fil364 raqam bilan ro'yxatga olingan.

Dissertatsiya Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universitetida bajarilgan.


Dissertatsiya avtoreferati uch tilda (o'zbek, ingliz, rus (rezyume)) Farg'ona davlat universitetining veb-sayti (www.fdu.uz) va "ZiyoNet" Axborot-ta'lim portali (www.ziynet.uz)da joylashtirilgan.

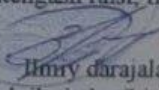
Ilmiy maslahatchi:	Dadaboyev Hamidulla Aripovich filologiya fanlari doktori, professor
Rasmiy opponentlar:	Hoshimov G'anijon Mirzaaxmetovich filologiya fanlari doktori, professor Urinbayeva Dilbar Bazarovna filologiya fanlari doktori Eshref Adali texnika fanlari doktori, professor
Yetakchi tashkilot:	O'zbek tili, adabiyoti va folklori instituti

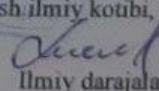
Dissertatsiya himoyasi Farg'ona davlat universiteti huzuridagi filologiya fanlari bo'yicha ilmiy darajalar beruvchi DSc.03/30.12.2019.Fil.05.02 raqamli Ilmiy kengashning 2022-yil "5" aprel soat 11 dagi majlisida bo'lib o'tadi (Manzil: 100151, Farg'ona shahri, Murabbiylar ko'chasi, 19-uy. Tel: (99873) 244-66-02; faks: (99873) 244-44-01; e-mail: info@fdu.uz).

Dissertatsiya bilan Farg'ona davlat universitetining Axborot-resurs markazida tanishish mumkin (159 raqam bilan ro'yxatga olingan) (Manzil: 100100, 100151, Farg'ona shahri, Murabbiylar ko'chasi, 19-uy. Tel: (99873) 244-71-28.

Dissertatsiya avtoreferati 2022-yil 24 mart kuni tarqatildi.
(2022-yil 24 mart dagi raqamli reyestr bayonnomasi).


Sh. Iskandarova
Ilmiy darajalar beruvchi Ilmiy
kengash raisi, filol. f.d., professor


M. Zokirov
Ilmiy darajalar beruvchi Ilmiy
kengash ilmiy kotibi, filol. f.n., dotsent


A. Mamajonov
Ilmiy darajalar beruvchi Ilmiy
kengash qoshidagi ilmiy
seminar raisi, filol. f.d., professor

KIRISH (doktorlik (DSc) dissertatsiyasi annotatsiyasi)

Dissertatsiya mavzusining dolzarbligi va zarurati. Jahonda inson faoliyatining bilim talab qiladigan zamonaviy sohalari rivojlanishi kompyuter texnologiyalari rolining oshishi bilan belgilanadi. Bugungi kunda axborot oqimi sezilarli darajada ko‘paymoqda, endi ularni saqlash, taqdim etish, formallashtirish va tartibga solish, shuningdek, avtomatik qayta ishlashning yangi usullarini izlash zarurati yuzaga kelmoqda. Shu bois turli amaliy maqsadlarda qo‘llanilishi mumkin bo‘lgan keng qamrovli bilim bazalariga qiziqish ortib bormoqda. Ayniqsa, inson omilisiz matndan har qanday ma‘lumotni chiqarib oluvchi neyro-to‘rlarga asoslangan tizimlarga ehtiyoj katta. XX asr yarimida butunjahon tarmog‘i bilan bir qatorda Semantik vebalar paydo bo‘ldi, unda gipermatnli sahifalardagi elementlar semantikasi haqida axborot tashuvchi qo‘shimcha teglar bilan ta‘minlandi. Semantik vebning ajralmas qismi ontologiya tushunchasi bo‘lib, u so‘zlar tarmog‘idan iborat leksik ma‘lumotlar bazasi hisoblanadi.

Dunyo amaliy tilshunosligi va kompyuter lingvistikasida axborot qidiruv imkoniyatini oshirish, avtomatik tarjima tizimlarini takomillashtirish, matnlarni komponent tahlil qilishga erishish, tilning lug‘at boyligini aks ettirish maqsadida yaratilgan lingvistik ontologiyalar tezauruslardan tildagi semantik munosabatlarning o‘zaro tarmoqlanishi bilan ajralib turdi. Natijada WordNet ingliz tili leksik ma‘lumotlar bazasi asosida boshqa ko‘plab tillar resurslarini yaratish keng tus oldi. Shu sababli tabiiy tilni qayta ishlash (NLP), tilni lingvistik modellashtirish, so‘z turkumlarini teglash, tilda semantik munosabatlar to‘plami (sinonimiya, meronimiya, giperonimiya, antonimiya)ni shakllantirishga alohida e‘tibor qaratildi.

Hozirda o‘zbek tilshunosligi va kompyuter lingvistikasida sun‘iy intellekt uchun tilni formallashtirish, lingvistik modellarni yaratish, kiberleksikografiyada amaliy natijalarga erishish, ayniqsa “...davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini ta‘minlash”¹, o‘zbek tilini Internet jahon axborot tarmog‘ida ommalashtirish, unda munosib o‘rin egallashini ta‘minlash², grammatik va semantik tahlil tizimlarini yaratish, buning uchun o‘zbek tilida tabiiy tilni qayta ishlash borasida tadqiqotlar amalga oshirilmoqda, amaliy loyihalar bajarilmoqda. Natijada amaliy tilshunoslik va kompyuter lingvistikasining matnlarni qayta ishlash, nutq sintezatorini yaratish, nutqni tanish, mashinali ta‘lim, tabiiy tilni qayta ishlash (NLP), kompyuter tarjimasini, korpusshunoslik sohasi, kompyuter leksikografiyasi va lingvodidaktikasi yo‘nalishlarining rivojlanishiga zamin yaratilmoqda. Endilikda o‘zbek tilining xalqaro maqomini yuksaltirish, milliy tilimizning imkoniyatlarini kengaytirish, tarjimon dastur va tizimlarini yaratish, global tarmoqda o‘zbek tilida axborot qidiruvi imkoniyatlarini oshirish maqsadida jahon tajribasiga tayangan holda semantik munosabatlarga asoslangan, so‘zlar tarmog‘iga ega, o‘zida o‘zbek tilining butun leksik boyligini jamlagan UzNet ontologik tizimini yaratish dolzarb

¹ Ўзбекистон Республикаси Президентининг 2020 йил 20 октябрдаги “Мамлакатимизда ўзбек тилини янада ривожлантириш ва тил сиёсатини такомиллаштириш чора-тадбирлари тўғрисида”ги ПФ-6048-сон Фармони.
<https://lex.uz/ru/docs/5058351>

² Mazkur Farmon.

hisoblanib, bu borada ilmiy tadqiqotlarni chuqurlashtirish zarurati o'z yechimini kutmoqda.

O'zbekiston Respublikasi Prezidentining 2016-yil 13-maydagi PF-4997-son "Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universitetini tashkil etish to'g'risida", 2017-yil 7-fevraldagi PF-4947-son "O'zbekiston Respublikasini yanada rivojlantirish bo'yicha Harakatlar strategiyasi to'g'risida», 2019-yil 21-oktabrdagi PF-5850-son "O'zbek tilining davlat tili sifatidagi nufuzi va mavqei tubdan oshirish chora-tadbirlari to'g'risida"gi farmonlari, 2020-yil 20-oktabrdagi PF-6084-son "Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida"gi Farmoni bilan tasdiqlangan "2020-2030-yillarda o'zbek tilini rivojlantirish va til siyosatini takomillashtirish Konsepsiyasi"da belgilangan davlat tilining zamonaviy axborot texnologiyalari va kommunikatsiyalariga faol integratsiyalashuvini ta'minlash ustuvor yo'nalishida belgilangan vazifalar, 2021-yil 17-fevraldagi "Sun'iy intellekt texnologiyalarini jadal joriy etish uchun shart-sharoitlar yaratish chora-tadbirlari to'g'risi"dagi PQ-4996-son Qarori hamda O'zbekiston Respublikasi Vazirlar Mahkamasining 2019-yil 12-dekabrda 984-son «Davlat tilini rivojlantirish departamenti to'g'risidagi Nizomni tasdiqlash haqida», 2020-yil 29-yanvardagi «O'zbekiston Respublikasi Vazirlar Mahkamasi huzuridagi Terminlar komissiyasining faoliyatini tashkil qilish chora-tadbirlari to'g'risida»gi 40-son qarorlari va sohaga oid boshqa me'yoriy-huquqiy hujjatlarda belgilangan vazifalarni amalga oshirishda ushbu dissertatsiya muayyan darajada xizmat qiladi.

Tadqiqotning respublika fan va texnologiyalar taraqqiyotining ustuvor yo'nalishlariga mosligi. Mazkur tadqiqot respublika fan va texnologiyalari rivojlanishining I. "Axborotlashgan jamiyat va demokratik davlatni ijtimoiy, huquqiy, iqtisodiy, madaniy, ma'naviy-ma'rifiy rivojlantirish, innovatsion iqtisodiyotni rivojlantirish" ustuvor yo'nalishiga muvofiq bajarilgan.

Dissertatsiya mavzusi bo'yicha xorijiy ilmiy-tadqiqotlar sharhi³

Jahon amaliy tilshunosligi va kompyuter lingvistikasi va axborot texnologiyalari sohalarida WordNet asosidagi lingvistik resurslar ta'minotini o'rganish va yaratishga yo'naltirilgan ilmiy izlanishlar yetakchi ilmiy markazlar va oliy ta'lim muassasalarida, jumladan, Princeton universiteti, Sent-Luis universiteti (AQSh), Janubiy Afrika universiteti, Janubiy Afrika raqamli til resurslari markazi (Pretoriya, Janubiy Afrika), Vlore universiteti (Albaniya), Milliy elektronika va kompyuter texnologiyalari markazi (Yaponiya), Tailand hisoblash lingvistikasi laboratoriyasi (Tailand), Basklar mamlakati universiteti, Kataloniya Texnika universiteti (Kataloniya), Hindiston statistika instituti, Hindiston Texnologiya Instituti (Hindiston), Bolgariya tili instituti, Bolgariya Fanlar akademiyasi (Bolgariya), Patras universiteti (Gretsiya), Ontologiya, tillarni qayta ishlash va

³ Dissertatsiya mavzusi bo'yicha xorijiy ilmiy-tadqiqotlar sharhi www.princeton.edu, <https://aclanthology.org/W98-0709.pdf>, <https://pythonprogramming.net/wordnet-nltk-tutorial/>, <http://wordnet.ru/>, <http://www.globalwordnet.org>, <https://scholar.google.com>, <https://www.researchgate.net>, <https://www.aclweb.org/anthology/L16-1207.pdf>, <https://acikerisim.isikun.edu.tr/>, www.navoiy-uni.uz, <http://uzschoolcorpara.uz/> va boshqa manbalar asosida amalga oshirildi.

elektron gumanitar fanlar laboratoriyasi, Tayvan milliy universiteti, Xitoy-Taypey Respublikasi (Tayvan), Zagreb universiteti (Xorvatiya), Charlz universiteti, Rasmiy va amaliy tilshunoslik instituti (Chexiya), Sprogteknologi markazi, Kobenhavns universiteti, Daniya tili va adabiyoti jamiyati (Daniya), Vrije Universiteti (Niderlandiya), Amsterdam universiteti (Niderlandiya), Sheffild universiteti (Angliya), Tartu universiteti (Estoniya), Helsinki universiteti (Finlyandiya), Parij diderot universiteti, Avignon universiteti, Memodata kompyuter laboratoriyasi (Fransiya), Tübingen Universiteti (Germaniya), Hayfa universiteti (Isroil), (Mumbay, Hindiston), Seged universiteti (Vengriya), Kompyuter tilshunosligi instituti, Bruno Kessler jamg'armasi, Aloqa va axborot texnologiyalari markazi, Inson tili texnologiyalari guruhi, Verona universiteti (Italiya), Yaponiya Milliy axborot-kommunikatsiya texnologiyalari instituti (Yaponiya), Pusan milliy universiteti (Janubiy Koreya), Kurdiston universiteti (Kurdiston), Latviya universitetining matematika va informatika instituti (Latviya), Kiril va Mefodiya universiteti, Staffordshire universiteti (Angliya), Malta universiteti (Malta), Moldova Fanlar akademiyasi Matematika instituti va Katmandu universiteti (Moldova), Bergen universiteti (Norvegiya), Shahid Beheshti universiteti, Eron telekommunikatsiya tadqiqotlari markazi (Tehron, Eron), Tehron universiteti, NLP laboratoriyasi (Tehron, Eron), Vroslav fan va texnologiya universiteti (Polsha), Adam Miskevich universiteti (Poznan, Polsha), Coimbra universiteti (Portugaliya), Getulio Vargas universiteti, Amaliy matematika maktabi (Rio-de-Janeyro, Braziliya), Lissabon universitetining tilshunoslik markazi (Portugaliya), Aleksandru Ioan Kuza universiteti, Ruminiya akademiyasi, Buxarest sun'iy intellekt instituti, Sun'iy intellekt instituti, Ruminiya akademiyasi (Ruminiya), Sankt-Peterburg universiteti, Moskva davlat universiteti (Rossiya), Belgrad universitetining Matematika fakulteti (Serbiya), Moratuva universiteti (Moratuva, Shri-Lanka), Lyublyana universiteti, Jozef Stefan instituti (Sloveniya), Gothenburg shved universiteti (Shvetsiya), Turk tili va nutqni qayta ishlash markazi (Turkiya), INHA universiteti, Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti, O'zbekiston Milliy universiteti (O'zbekiston)da olib borilmoqda.

Dunyo tilshunosligi va axborot texnologiyalari sohalarida til hamda dunyo bilimlarini aks ettiruvchi ontologiyalarni yaratish mezonlari, axborot-qidiruv tizimi samaradorligini oshiruvchi tezauruslarni yaratish tamoyillari, ontologiya va tezauruslarni yaratish tadqiqiga oid olib borilgan izlanishlardan quyidagi ilmiy natijalar olingan: ontologiyani yaratishning nazariy aspektlari (Princeton universiteti, AQSh, Moskva davlat universiteti, Rossiya, Ontologiya, tillarni qayta ishlash va elektron gumanitar fanlar laboratoriyasi, Tayvan, Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universiteti, O'zbekiston), lingvistik ontologiya va tezauruslar tushunchalari tadqiq etilgan ularning mushtarak hamda farqli jihatlari yoritilgan (Sent-Luis universiteti, AQSh, Moskva davlat universiteti, Rossiya, Turk tili va nutqni qayta ishlash markazi, Turkiya), turlari tavsiflangan (Moskva davlat universiteti, Buxarest sun'iy intellekt instituti, Sun'iy intellekt instituti, Ruminiya akademiyasi), matnning avtomatik tahlil va

tabiiy tilni qayta ishlashdagi ahamiyati yoritilgan (Yaponiya Milliy axborot-kommunikatsiya texnologiyalari instituti, Yaponiya, Tübingen Universiteti, Germaniya, Kompyuter tilshunosligi instituti, Bruno Kessler jamg'armasi, Aloqa va axborot texnologiyalari markazi, Inson tili texnologiyalari guruhi, Verona universiteti, Italiya), axborot-qidiruv tizimidagi o'rni, ontologiyalarning til va dunyo bilimlarini aks ettirishi tahlil qilingan (Tehron universiteti, NLP laboratoriyasi, Eron, Hindiston statistika instituti, Hindiston Texnologiya Instituti, Hindiston), Princeton WordNet bazasi zamirida til ontologiyasini yaratish imkoniyatlari va nazariy metodologiyasi (sharhning 1-xatboshida keltirilgan barcha muassasalarda) lingvistik ontologiya asosida tabiiy tilni qayta ishlash texnologiyasi (Vroslav fan va texnologiya universiteti, Polsha, Sheffild universiteti, Angliya), ko'p tilli bilimlarni izlash va olish tizimlarini yaratish uchun turkiy tillarning elektron tezaurusi ishlab chiqish ustida ishlar bajarilmoqda (Qozoq universiteti, Almata).

Jahon tilshunosligida lingvistik ontologiyani yaratish va WordNet zamiridagi lingvistik resurlarni takomillashtirish, axborot-qidiruv tizimi imkoniyatini oshirish, milliy ontologiyalar va ko'p tilli ontologiyalarni yaratish, zamonaviy axborot texnologiyalari vositalarida tabiiy tilni qayta ishlash, til korpuslari asosida leksik ma'lumotlar bazalarini yaratish va takomillashtirish, mashina tarjimai uchun multitilli ontologiyani yaratish kabi yo'nalishlarda tadqiqotlar olib borilmoqda.

Muammoning o'rganilganlik darajasi. Xorij amaliy tilshunosligi va kompyuter lingvistikasi sohalarida lingvistik (semantik) ontologiyalarni yaratish masalasi ko'plab olimlar tomonidan o'rganilgan.

Tilshunoslik ontologiyasi mazkur sohada yangi tushuncha emas, ammo zamonaviy talqinda ular XX asrning oxiridan boshlab qo'llanila boshlandi.

Lingvistik ontologiyani yaratish, iyerarxiyal bazani shakllantirish, sohalar bo'yicha terminologik bazalarni yaratish, semantik munosabatlarni tadqiq etish, tezaurus lug'atlar asosida til ontologiyasini yuzaga keltirish borasida jahon kompyuter lingvistikasi va axborot texnologiyalari sohalarida J.Miller⁴, K.Fellbaum⁵ (Princeton WordNet mualliflari), A.Burgen, O.Bodenreider; K.Kunze, A.Vagner⁶ (nemis tili ontologistlari), L.Bentivogli, E.Pianta (hind tili ontologistlari), M.Buenaga Rodriguez, J.Gomez-Hidalgo⁷ (bolgar tili

⁴ Miller G. Nouns in WordNet. In: Fellbaum, C (ed) WordNet – An Electronic Lexical Database. – The MIT Press. 1998. – pp.23-47.; Miller G., Fellbaum C. Morphosemantic links in WordNet. – Traitement automatique de langue, 44.2. 2003. – pp. 69-80.; Miller G., Hristea F. WordNet Nouns: Classes and Instances. – Computational linguistics, Volume 32, Number 1. 2006. – pp.1-3.; Miller K. Modifiers in WordNet. In: Fellbaum, C (ed) WordNet – An Electronic Lexical Database. – The MIT Press. 1998. – pp .47-68.

⁵ Fellbaum Ch. A Semantic Network of English Verbs. – In: Fellbaum, C (ed) WordNet – An Electronic Lexical Database. – The MIT Press. 1998. – pp. 69-104.

⁶ Kunze C., Wagner A. Integrating GermaNet into EuroWordNet, a multilingual lexically semantic database. In: Sprache und Datenverarbeitung – International Journal for Language Data Processing. Bonn. 1999.

⁷ Burgun A., Bodenreider O., Aubry M., Mosser J. Dependence relations in Gene Ontology: A preliminary study. Workshop on The Formal Architecture of the Gene Ontology. – Leipzig, Germany, May 28-29. 2004.; Buenaga Rodriguez M., Gomez-Hidalgo J., Diaz-Agudo B. 1997 Using WordNet to complement training information in text categorization // In Proceedings of the 2nd International Conference on Recent Advances in Natural Language Processing (RANLP 1997), Bulgaria. 1997. – pp. 150-157.

ontologislari), N.Guarino⁸, P.Guaretta (formal ontologiya va axborot tizimlari bo'yicha mutaxassis) samarali faoliyat olib borishgan.

Rus tilshunoslaridan N.V.Lukashevich (RuThez – rus tili ontologiyasini yaratish bo'yicha loyiha rahbari, ontologist)⁹, V.B.Dobrov (fan sohasi matnlari uchun terminologik iboralar bazasini shakllantirish)¹⁰, I.B.Azarova (RussNet lingvistik resursini asosiy yaratuvchisi va axborot-qidiruv tizimlari ontologiyasi bo'yicha tadqiqot ishlarini amalga oshirgan)¹¹, O.A.Nevzorova (yangi fan sohalari ontologiyalarini ishlab chiqish texnologiyasi)¹², A.S.Narinyan (TEON: Tezaurus+Ontologiya loyihasi asosida ontologik tadqiqotlarni amalga oshirgan)¹³, B.B.Morkovkin (ideografik lug'at tuzgan), shuningdek, Markaziy Osiyodan A.Sharipbayev¹⁴ (qozoq tilini qayta ishlash va turkiy tillar ontologiyasini yaratish bo'yicha izlanish olib bormoqda) faoliyat olib borishgan.

⁸ Guarino N. Formal Ontology and Information Systems. In N. Guarino, editor, Proceedings of the 1st International Conference on Formal Ontologies in Information Systems, FOIS'98, Trento, Italy, IOS Press. 1998. – pp. 3-15.; Guarino N., Giaretta P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. Mars (ed.) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995. IOS Press, Amsterdam. 1995. – pp. 25-32.

⁹ Лукашевич Н.В., Автоматизированное формирование информационно-поискового тезауруса по общественно-политической жизни России // НТИ. Сер.2. – 1995. – N 3. – С.21-24.; Лукашевич Н.В., Добров Б.В. Тезаурус русского языка для автоматической обработки больших текстовых коллекций // Компьютерная лингвистика и интеллектуальные технологии. Труды Международного семинара Диалог'2002 / Под ред. А.С.Нариньяни. – Т.2. – М.: Наука – 2002.– С.338-346.; Лукашевич Н.В., Добров Б.В. Отношения в онтологиях для решения задач информационного поиска в больших международных текстовых коллекциях. Девятая национальная конференция по искусственному интеллекту с международным участием КИИ-2004 (28 сентября-2 октября 2004 г., Тверь). Труды коференции. В 3-х т. – Т2. – М.: Физматлит, 2004. – С.544-551.; Лукашевич Н.В. Моделирование отношения ЧАСТЬ-ЦЕЛОЕ в лингвистических и онтологических ресурсах. Информационные технологии. – 2007. – N 12.; Лукашевич Н.В. Проблемы установления родовидовых отношений в лингвистических онтологиях. – Материалы Всероссийской конференции «Знания – Онтологии – решения» (ЗОНТ-07). – С.211-220.;

¹⁰ Добров Б.В., Лукашевич Н.В., Сыромятников С.В. Формирование базы терминологических словосочетаний по текстам предметной области. Труды пятой всероссийской научной конференции “Электронные библиотеки: Перспективные методы и технологии, электронные коллекции. – 2003. – С. 201-210.; Добров Б.В., Лукашевич Н.В. Онтологии для автоматической обработки текстов: описания понятий и лексических значений. Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции “Диалог'2005 / Под ред. И.М. Кобозевой, А.С. Нариньяни, В.П. Селегея. – М.: Наука, 2005. – С.138-142.; Добров Б.В., Лукашевич Н.В. Вторичное использование лингвистических онтологий: изменение в структуре концептуализации. Восьмая Всероссийская научная конференция «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Владимир-Суздаль, 16-18 октября 2006 г.). 2006.; Добров Б.В., Лукашевич Н.В. Транзитивные нетаксономические отношения в онтологическом моделировании. Труды симпозиума Онтологическое моделирование. Институт проблем информатики РАН, 2008. – С.229-259.

¹¹ Азарова И.В., Митрофанова О.А., Синопальникова А.А. Компьютерный тезаурус русского языка типа WordNet // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог'2003. М., 2003. – С . 43-50.; Азарова И.В., Синопальникова А.А., Яворская М.В. Принципы построения wordnet-тезауруса RussNet. Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог'2004. М., 2004. – С. 542-547.; Азарова И.В., Синопальникова А.А., Смирж П. Представление устойчивых лексических сочетаний в компьютерном тезаурусе RussNet. Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог'2005. М., 2004. – С. 11-16.

¹² Добров Б.В., Лукашевич Н.В., Невзорова О.А. Технология разработки онтологий новых предметных областей. Труды Казанской школы по компьютерной лингвистике TEL-2002. Выпуск 7. / Под ред. В.Г.Бухараева, В.Д.Соловьева, Д.Ш.Сулейманова – Казань: Отечество, 2002. – С.90-106.

¹³ Нариньяни А.С. Кентавр по имени ТЕОН: Тезаурус+Онтология. Труды Международной конференции ДИАЛОГ-2001. – Т.1. – М., 2001.– С.184-188.

¹⁴ Вычислительная обработка казахского языка. Сборник научных трудов / Под редакцией Рахимовой Д.Р. – Алматы: Қазақ университеті, 2020. – 147 с.

Lingvistik ontologiyani yaratish va qo‘llash bo‘yicha ko‘plab ishlar Rossiyada va xorijda amaliy tilshunoslik sohasida olib borilgan va hozirda jadal ravishda davom etmoqda. Ushbu tadqiqotda J.Miller va K.Fellbaum P. Butelaar, F.Chimiano, P.Haaza, B.V.Dobrov, N.V.Lukashevich, O.Narinyan, I.Azarovanning eng muhim tadqiqotlarini ko‘rib chiqamiz. Shuningdek, eng yirik lingvistik ontologiyalar – SUMO, OMEGA, DOLCE, Princeton WordNet, KeNet, RuThez va RussNet kabi ontologiya hamda tezauruslar o‘rganildi, tahlil qilindi, zarur o‘rinlarda munosabat bildirildi. Natijada o‘zbek tili ontologiyasini yaratish bo‘yicha metodologik bilimlar o‘zlashtirildi, muayyan sohaga xos bo‘lgan ma’lumotlarni qayta ishlashga imkon beruvchi tajribalar orttirildi. Buning natijasida esa “O‘zbek tili ontologiyasini yaratish konsepsiyasi” ishlab chiqildi.

O‘zbek tili lingvistik ontologiyasini yaratish bo‘yicha izlanishlar o‘zbek kompyuter lingvistikasi (O‘KL) sohasida amalga oshirilmoqda. Asosan, WordNet tizimi asosida tezaurus lug‘atlar va ontologik resurslarni yaratish borasida bir qancha maqolalar e‘lon qilingan. O‘KL o‘z taraqqiyoti davomida ko‘plab nazariy va amaliy tadqiqot ishlari bilan boyitib borilmoqda. XX asr oxirlarida shakllangan O‘KL ikki bosqichli taraqqiyot davriga ega deyish mumkin. Birinchi bosqich davriy nashrlar va bir qancha badiiy adabiyotlarda eng ko‘p foydalanilgan so‘zlar chastotasini aniqlab beradigan dastur yordamida ilmiy tadqiqot ishlari va bir qancha chastotali lug‘atlarning yaratilishi, o‘quv adabiyotlarining yaratilishi bilan belgilanadi. Ushbu davrni, o‘z o‘rnida, ikki bosqichga ajratish mumkin: XX asrning oxiridan XXI asrgacha bo‘lgan davr. Uning namoyondalari S.Muhamedov, T.Sodiqov, H.Arziqulov, M.Ayimbetov, S.Rizayev,¹⁵ hisoblanadi. Ikkinchi bosqich esa 2001-yildan Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universitetida “Kompyuter lingvistikasi” laboratoriyasi ochilishi va fanning ta’lim tizimida joriy etilishi bilan belgilanadi hamda XX asrning 20-yillarigacha bo‘lgan davrni o‘z ichiga oladi. Mazkur natijalar O‘KL rivojlanishiga salmoqli hissa qo‘shgan A.Po‘latov sa’y-harakatlari bilan amalga oshirilgan. Bu davrda aynan kompyuter lingvistikasi o‘quv adabiyotlarining yaratilishi¹⁶ va tadqiqotlarning rivojlanishiga erishilgan¹⁷ va A.Po‘latov, A.Rahimov, S.Muhamedova, N.Jo‘rayeva,

¹⁵ Мухамедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дисс. ...канд.филол.наук. – Тошкент, 1980. – 25 с.; Мухамедов С.А., Пиотровский Р.Г. Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов. – Т.: Фан, 1986.; Садыков Т. Проблемы моделирования тюркской морфологии. – Фрунзе, 1987.; Арзикулов Х.А., Пиотровская К.Р. Информатика и переработка текста средствами вычислительной техники (учебное пособие). – Самарканд, 1986.; Айымбетов М.К. Проблемы и методы квантитативно-типологического измерения близости тюркских языков (на материалах каракалпакского, казахского и узбекского языков): Автореф.дисс. ...д-ра филол.наук. – Т., 1997. – 47 с.; Айымбетов М.К. Проблемы и методы квантитативно-типологического измерения близости тюркских языков (на материалах каракалпакского, казахского и узбекского языков): Автореф.дисс. ...д-ра филол.наук. – Т., 1997. – 47 с.; Ризаев С. Ўзбек тилишунослигида лингвостатистика муаммолари (монография). – Тошкент: Фан. 2005. – 295 б.; Shu muallif. Ўзбек тилининг лингвостатистик тадқиқи: Филол. фан. док. ...дисс. автореф. – Тошкент, 2008. – 50 б.; Айымбетов М.К. Квантитативная типология тюркского текста (сборник избранных статей). – Нукус.: Илим. 2012.

¹⁶ Muhamedova S. Kompyuter lingvistikasi (metodik qo‘llanma). – Toshkent, 2007.; Po‘latov A., Muhamedova S. Kompyuter lingvistikasi (o‘quv qo‘llanma). – Toshkent, 2008 – 98 b.; Pўлатов А. Компьютер лингвистикаси. – Тошкент: Академнашр, 2011. – 520 b.; Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent: Akademnashr. 2011. – 160 b.

¹⁷ Пўлатов А.К., Алиходжаев Б., Джураева Н. Разработка программы компьютерного анализа и синтеза глаголов узбекского языка // O‘zMU xabarlari. – Toshkent, 2002. №2. – С. 17-19.; Po‘latov A.Q., Mo‘minova T.,

U.Dusimova, N.Valiyeva, M.Abjalova, N.Abdurahmonova tomonidan ba'zi bir nazariy tadqiqot ishlari amalga oshirildi.

2018-yillarga kelib O'KLda jadal ravishda nomzodlik va doktorlik ishlarining amalga oshirilishi o'z davrining ikkinchi bosqichiga o'tishida zamin bo'ldi. Bu davrda o'zbek tili mualliflik korpusini tuzish tamoyillari¹⁸, o'zbek-ingliz tili mashina tarjimasining lingvistik ta'minoti¹⁹, janriy-lisoniy va lingvostatistik tadqiqot masalalari²⁰, o'zbek tilidagi matnlarni avtomatik tahrir va tahlil qilish dasturining lingvistik modullari²¹, til korpusi lingvistik bazasini tuzish tamoyillari²², o'zbek tili atov birliklarini semantik teglashning lingvistik asoslari²³, o'zbek tili morfologik analizatorining lingvistik ta'minoti muammolari²⁴ monografik planda o'rganilgan. O'KLda kompyuter lingvistikasining bir qancha yo'nalishlarida e'tiborga molik ishlar qilingan bo'lsa-da, o'zbek tili ontologiyasini tuzish masalasi bo'yicha maxsus tadqiqot ishi bajarilmagan.

XXI asrning 20-yillariga kelib, amaliy natijalarga ega tadqiqotlarning amalga oshirilishi, turli lingvistik dasturlar (avtomatik tahrir va tahlil (Ubuntu Linux tizimi uchun matnlarni avtomatik tahlil qilish dasturi²⁵), transliteratsiya, mobil ilovalari)ning yaratilishi (1) va Respublikamizning oliy ta'lim muassasalari (ToshDO'TAU, O'zMU, SamDCHTI, UrDU)da Kompyuter lingvistikasi mutaxassisligi bo'yicha magistratura bosqichining ochilishi (2) bilan KL o'z taraqqiyotining yangi bosqichiga o'tdi. Alisher Navoiy nomidagi Toshkent davlat o'zbek tili va adabiyoti universitetida bir guruh mutaxassis olimlar sa'y-harakati natijasida O'zbek tilining ta'limiy korpusi²⁶ yaratilishi bilan O'zbek tili Milliy korpusiga tamal toshi qo'yildi.

“O'zbek tili ontologiyasini yaratish tamoyillari” mavzusidagi mazkur tadqiqot UzNet leksik ma'lumotlar bazasini yaratishning samarali yondashuvlarini tanlashga va “O'zbek tili lingvistik ontologiyasini yaratish konsepsiyasi”ni shakllantirishga xizmat qiladi.

Tadqiqotning dissertatsiya bajarilgan oliy ta'lim yoki ilmiy-tadqiqot muassasasining ilmiy-tadqiqot ishlari rejalari bilan bog'liqligi. Tadqiqot

Po'latova I.O. Dunyoviy o'zbek tili (O'zbek tilida fe'lning shakllari va ularning rus, ingliz tillarida berilishi). – Toshkent: Universitet, 2003. – 404 b.; Muhamedova S. O'zbek tilidagi harakat fe'llari asosida kompyuter dasturlari uchun lingvistik ta'min yaratish. – Toshkent, 2006. – 80 b.

¹⁸Xamroyeva Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Filol.fan.bo'yicha falsafa dokt. (PhD)...diss. – Qarshi, 2018. – 250 b.

¹⁹Abduraxmonova N.Z. Inglizcha matnlarni o'zbek tiliga tarjima qilish dasturining lingvistik ta'minoti (Sodda gaplar misolida): Filol.fan.bo'yicha falsafa dokt. (PhD)...diss. aytoref. – Toshkent, 2018. – 52 b.

²⁰O'rinboyeva D. Xalq og'zaki ijodi: janriy-lisoniy va lingvostatistik tadqiqot muammolari: Filol. fan. dokt. (DSc) ...diss. avtoref. – Samarqand, 2019. – 74 b.

²¹Abjalova M.A. O'zbek tilidagi matnlarni tahrir va tahlil qiluvchi dasturning lingvistik modullari (Rasmiy va ilmiy uslubdagi matnlar tahriri dasturi uchun): Filol.fan.bo'yicha falsafa doktori (PhD)...diss. – Farg'ona, 2019. – 164 b.

²²Eshmo'minov A.A. O'zbek tili milliy korpusining sinonim so'zlar bazasi: Filol.fan.bo'yicha falsafa dokt. (PhD)...diss. – Qarshi, 2019. – 140 b.

²³Ahmedova D. Atov birliklarini o'zbek tili korpuslari uchun leksik-semantik teglashning lingvistik asos va modellari: Filol.fan.bo'yicha falsafa dokt. (PhD)...diss. – Buxoro, 2020. – 156 b.

²⁴Xamroyeva Sh. O'zbek tili morfologik analizatorining lingvistik ta'minoti. Filol.f.d. diss. avtoref. – Farg'ona, 2021. – 78 b.

²⁵Abjalova M.A. Tahrir va tahlil dasturlarining lingvistik modullari. Monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – 176 b.

²⁶AM-FZ-201908172 raqamli “O'zbek tili ta'limiy korpusini yaratish” mavzusidagi amaliy loyiha doirasida bajarilgan.

Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyot universitetida bajarilayotgan “O‘zbek tilining ta’limiy korpusini yaratish” mavzusidagi AM-FZ-201908172 raqamli amaliy loyihasi doirasida hamda “O‘zbek tili ontologiyasini yaratish konsepsiyasi” asosida bajarilgan.

Tadqiqotning maqsadi raqamli texnologiya va sun’iy intellekt uchun lingvistik ontologiyalarni yaratish tamoyillari va leksik ma’lumotlar bazalari tuzilishini tadqiq etish hamda UzNet – o‘zbek tili ontologiyasini yaratish konsepsiyasini shakllantirishdan iborat.

Tadqiqot vazifalari:

ma’lumot qidirishda birlamchi manbalardan tezaurus va ontologiya konseptlarini tavsiflash, ahamiyatini yoritish va lingvistik ontologiya ta’rifini muayyanlashtirish;

ingliz tilining yirik leksik ma’lumotlari bazasi – WordNet lingvistik ontologiyasining tuzilishini o‘rganish, muammolarini tahlil qilish, foydalanish imkoniyatlari va boshqa tillar uchun tayanch tizim vazifasini bajarishi omillarini yoritib berish;

WordNet tizimidagi turk tili (KeNet) va rus tili (PyTez, RussNet) leksik resurslarini yaratish jarayonidagi tajribalarni o‘rganish hamda EuroWordNet multilingval formal ontologiya loyihasining ustuvor yo‘nalishlarini tadqiq etish natijasida o‘zbek tili ontologiyasini tuzishning umumiy tamoyillarini shakllantirish;

Lingvistik ontologiyalar klassifikatsiyasini tahlil qilish;

Leksik ma’lumotlar bazalari uchun semantik munosabatlar, jumladan, giperonimiya, xolo-meronimiya, sinonimiya, antonimiya hodisalari va ularning tasniflarini tahlil qilish va UzNet ontologik tizimini yaratishda o‘zbek tiliga tatbiq etish;

tadqiq etilgan ma’lumotlar asosida “O‘zbek tili ontologiyasi konsepsiyasi”ni loyihalash;

UzNet o‘zbek tili ontologiyasining lingvistik resurslari bazasini shakllantirish.

Tadqiqotning obyekti sifatida “O‘zbek tilining izohli lug‘ati”, o‘zbek tilidagi sinonimlar va o‘zbek tili ta’limiy korpusi bazasidagi lisoniy ma’lumotlar tanlandi.

Tadqiqotning predmetini WordNet leksik ma’lumotlar bazasi, lingvistik ontologiyalarni yaratish texnologiyasi, WordNet, KeNet, RuTez leksik ma’lumotlar bazalaridagi semantik munosabatlar, UzNet lingvistik ontologiyasini yaratish konsepsiyasi tashkil qiladi.

Tadqiqot usullari. Tadqiqot mavzusini yoritishda tavsiflash, chog‘ishtirish, komponent tahlil, statistik, modellashtirish, substansial tahlil kabi metodlardan foydalanildi.

Tadqiqotning ilmiy yangiligi quyidagilardan iborat:

lingvistik ontologiya va tezaurus tushunchalarining semantik munosabatlarni qamrab olishi bo‘yicha integral hamda leksik birliklarning sohalararo tarmoqlanishi bo‘yicha differensial jihatlari falsafiy va lisoniy tamoyillar asosida muvaffaqiyatli aniqlanishi dalillangan;

lingvistik ontologiyada keng ko‘lamli sohalararo terminologik bazaning semantik tarmoqlanishiga asoslanishi aks etgan matematik modellar va leksik ma’lumotlar bazasi sifatida ontologiyaga qo‘yilgan konseptuallashtirish talablari aniqlangan;

ingliz tili Princeton WordNet ontologik tizimi asosida sinset (sinonimik to‘plam)lardagi semantik munosabatlar asnosidagi o‘zbek tili ontologiyasi (UzNet)ni yaratishning tizimli texnologiyasi, unda lingvistik va ekstralingvistik ma’lumotlarni taqdim etishdagi tamoyillar asoslangan;

lingvistik ontologiyalardagi birlamchi elementlar – sinonim so‘zlar to‘plami (sinsetlar)ni yaratishning formal prinsiplari, tur va jins (giponim-giperonim), butun-bo‘lak (xolonim-meronim) munosabatlaridagi assimetriya, refleksivlik, tranzitivlik va merosiylik xususiyatlari aniqlangan, butun-bo‘lak munosabatlarini yuzaga keltiruvchi funksional, gomeo‘lchovli va alohidalangan qismlar asosida turlanishi dalillangan;

o‘zbek tili ontologiyasini yaratish konsepsiyasining milliy lingvistik xususiyatlari KeNet – turk tili maxsus taksonomik va semantik veb tizimining struktural texnologiyasi bilan qiyoslangan holda asoslangan;

UzNetda so‘z ma’nolari izohini berishning pog‘onali tartibi, so‘z turkumlarini teglashning qoidalarga asoslangan va stoxastik (ehtimollik nazariyasi va statistik) usullari, o‘zbek tili ontologiyasida kvazisinonimlar to‘plamini yaratish va lingvistik sinkretizmni aniqlashning aktual jihatlari hamda shakliiy munosabatlarga doir ma’lumotlar bazasi tarkibini tuzish prinsiplari dalillangan.

Tadqiqotning amaliy natijalari quyidagilardan iborat:

UzNet lingvistik ontologiyasi sinset (ma’nodosh so‘zlar) to‘plamlari uchun “O‘zbek tilidagi sinonim so‘zlar bazasi” yaratilib, unga mualliflik guvohnomasi olingan²⁷;

UzNet o‘zbek leksik ma’lumotlar bazasi uchun omonim so‘zlar²⁸, antonim so‘zlar²⁹ va paronim so‘zlar³⁰ bazalari yaratilib, ularga mualliflik guvohnomalari olingan;

o‘zlashma so‘zlar turkumini teglash va o‘zbek tilidagi 11 000 o‘zlashma so‘zning izohini berish uchun o‘zlashma so‘zlarning urg‘uli bazasi yaratilgan va unga mualliflik guvohnomasi olingan³¹;

o‘zbek tili ontologiyasining lemmatizatsiya jarayoni uchun o‘zbek tilining morfologik ma’lumotlar bazasi hamda orfografik qoidalar tizimi yaratilgan;

yaratilgan formal orfografik qoidalar, so‘z yasalishi lug‘atining ma’lumotlari bazasi, paronim so‘zlar bazasidan AM-FZ-201908172 raqamli “O‘zbek tilining ta’limiy korpusini yaratish” mavzusidagi amaliy loyihada, o‘zbek tilining universal grammatikasi uchun ishlab chiqilgan “Fonetika” bo‘limi ma’lumotlari va nutqiy kompetensiyani shakllantirishga yo‘naltirilgan topshiriqlar to‘plami PZ-

²⁷ O‘zbek tilidagi sinonim so‘zlarning ma’lumotlar bazasi. Guvohnoma № BGU 00380. – Toshkent, 2019.

²⁸ O‘zbek tilidagi omonim so‘zlarning ma’lumotlar bazasi. Guvohnoma № BGU 00381. – Toshkent, 2019.

²⁹ O‘zbek tilidagi antonim so‘zlarning ma’lumotlar bazasi. Guvohnoma № BGU 00390. – Toshkent, 2020.

³⁰ O‘zbek tilidagi paronim so‘zlarning ma’lumotlar bazasi. Guvohnoma № BGU 00469. – Toshkent, 2021.

³¹ O‘zbek tilidagi o‘zlashma so‘zlarning urg‘uli bazasi. Guvohnoma № BGU 00404. – Toshkent, 2020.

2020042022 raqamli “Turkiy tillarning lingvodidaktik elektron platformasini yaratish” olima ayollar grantida foydalanilgan.

Tadqiqot natijalarning ishonchliligi o‘zbek tili ontologiyasini yaratish muammosining aniq qo‘yilganligi, ishning o‘rganilish doirasi belgilab olinganligi, flektiv va agglyutinativ tillar doirasida leksik ma’lumotlar tizimlarini yaratish tajribalarining o‘rganilganligi, qolaversa, korpus va kompyuter lingvistikasida metodologik jihatdan asoslangan nazariy qarashlar, umummetodologik asoslarga ega materiallarga tayanib, ilmiy xulosalarga kelingani, tadqiqotda qo‘yilgan vazifalarni hal qilish borasidagi nazariy fikr va xulosalarning amaliyotga joriy etilganligi, olingan natijalarning vakolatli tashkilotlar tomonidan tasdiqlanganligi bilan izohlanadi.

Tadqiqot natijalarining ilmiy va amaliy ahamiyati. Tadqiqot natijalarining ilmiy ahamiyati o‘zbek tili lingvistik ontologiyasining nazariy asoslarini ishlab chiqishda, lingvistik ontologiya uchun semantik munosabatlarni o‘rnatishning metodologik asosini belgilashda, UzNet ontologiyasini takomillashtirish borasidagi tadqiqotlarni olib borishda ilmiy-nazariy manba sifatida xizmat qiladi.

Tadqiqot natijalarining amaliy ahamiyati axborot qidiruvi tizimlari va mashina tarjimai uchun muhim manba bo‘lishi, tadqiqot materiallaridan “Kompyuter lingvistikasi” ilmiy-tadqiqot markazi faoliyatini olib borishda, oliy o‘quv yurtlarida “Kompyuter lingvistikasi”, “Ontologiyalar va semantik tizimlar”, “Korpus lingvistikasi”, “Mashina tarjimai”, “Parallel korpuslar” kabi yondosh fanlardan ma’ruzalar o‘qish, darslik va qo‘llanmalar yaratish, elektron lug‘atlar (tarjima, tezaurus) tuzish jarayonida foydalanish mumkinligi bilan izohlanadi.

Tadqiqot natijalarining joriy qilinishi.

O‘zbek tili ontologiyasi konsepsiyasi bo‘yicha olingan ilmiy va amaliy natijalar asosida:

lingvistik ontologiya va tezaurus tushunchalarining semantik munosabatlarni qamrab olishi bo‘yicha integral hamda leksik birliklarning sohalararo tarmoqlanishi bo‘yicha differensial jihatlari falsafiy va lisoniy tamoyillar asosida muvaffaqiyatli aniqlanishiga oid natijalaridan Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetida bajarilgan I-OT-2019-42 raqamli “O‘zbek va ingliz tillarining elektron (inson qiyofasi, fe‘l-atvori, tabiat va milliy timsollar tasviri) poetik lug‘atini yaratish” mavzusidagi fundamental ilmiy loyihada foydalanilgan (Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetining 2021-yil 26-noyabrdagi 04/1-2339-son ma’lumotnomasi). Natijada inson qiyofasi, fe‘l-atvori, tabiat va milliy timsollar tasvirida xolonim (butun), meronim (qism), giperonim (jins), giponim (tur) so‘zlar o‘rni va ahamiyatini yoritishga; o‘zbek va ingliz tillaridagi ko‘p ma’noli so‘zlarning ma’noviy munosabatlarini ochib berishga erishilgan;

lingvistik ontologiyalardagi birlamchi elementlar – sinonim so‘zlar to‘plami (sinsetlar)ni yaratishning formal prinsiplari, tur va jins (giponim-giperonim), butun-bo‘lak (xolonim-meronim) munosabatlaridagi assimetriya, refleksivlik, tranzitivlik va merosiylik xususiyatlari, butun-bo‘lak munosabatlarini yuzaga keltiruvchi funksional, gomeo‘lchovli va alohidalangan qismlar asosida

turlanishiga doir tadqiqot materiallari va natijalaridan “O‘zbek tilini o‘qitish metodikasi” o‘quv qo‘llanmasining “O‘quvchilarda kompetensiyalarni shakllantirish tamoyillari”, “Leksik-semantik munosabatlar” va “So‘z turkumlari va ularni tasniflash tamoyillari” nomli paragraflarida foydalanilgan (Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetining 2021-yil 26-noyabrdagi 04/1-2341-son ma‘lumotnomasi). Natijada nutqiy kompetensiya va lingvistik kompetensiyalarni shakllantirish omillarini yoritishga, sinonimiya va giponimiya munosabatlarini misollar bilan ochib berishga, leksik-semantik munosabatlarning lisonni boyitishdagi ahamiyatini namoyon etishga, so‘z turkumlarini tasniflash zaruratini yoritib berishga erishilgan;

UzNetda so‘z ma‘nolari izohini berishning pog‘onali tartibi, so‘z turkumlarini teglashning qoidalarga asoslangan va stoxastik (ehtimollik nazariyasi va statistik) usullari, o‘zbek tili ontologiyasida kvazisinonimlar to‘plamini yaratish va lingvistik sinkretizmni aniqlashning aktual jihatlari hamda shakliy munosabatlarga doir ma‘lumotlar bazasi tarkibini tuzishga doir tadqiqot natijalari va materiallaridan “Hozirgi o‘zbek tili” darsligida foydalanilgan (Oliy va o‘rta maxsus ta‘lim vazirligining 2020-yil 4-maydagi 285-son buyrug‘iga asosan 285-078 raqamli ruxsatnoma). Natijada darslikda sinonimiya va giponimiya hodisalarini misollar bilan yoritishga, turlarini ochib berishga, so‘zlarning semantik tasnifi va leksikolizatsiya hodisasini yoritishga erishilgan;

lingvistik ontologiyada keng ko‘lamli sohalararo terminologik bazaning semantik tarmoqlanishiga asoslanishi aks etgan matematik modellar va leksik ma‘lumotlar bazasi sifatida ontologiyaga qo‘yilgan konseptuallashtirish talablaridan AM-FZ-201908172 raqamli “O‘zbek tili ta‘limiy korpusini yaratish” mavzusidagi amaliy loyihada foydalanilgan (Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili va adabiyoti universitetining 2021-yil 26-noyabrdagi 04/1-2340-son ma‘lumotnomasi). Natijada o‘zbek tilining ta‘limiy korpusida sinonim, antonim so‘zlari bazalari qatorida omonim va paronim so‘zlar bazasi asosidagi tizimli qidiruv natijasining chiqishiga erishilgan;

UzNet o‘zbek tili ontologiyasida semantik munosabatlarni shakllantirishda paronim so‘zlarning leksikografik bazasini yaratish maqsadida “O‘zbek tilidagi paronimlar so‘zlar lug‘ati” nashr qilingan (ISBN 978-9943-7870-6-3). Natijada o‘zbek tilidagi paronimlar bazasi yaratilib, o‘zbek tili ontologiyasi uchun lingvistik ta‘minot shakllantirilgan;

o‘zbek tilidagi o‘zlashma so‘zlar, ularning turkumi, izoh(lar)i, asliy til haqidagi ma‘lumotlar, o‘zlashmaning sinonim(lar)i, zid ma‘nolari haqidagi ma‘lumotlar, nutqni tanish va nutqiy kompetensiyani rivojlantirishga, nutqiy g‘alizliklarni bartaraf etishga xizmat qilishi maqsadida belgilab qo‘yilgan o‘zlashmaning urg‘uli bazasi “O‘zbek tili o‘zlashma so‘zlarining urg‘uli lug‘ati”ni nashr etishda foydalanilgan (O‘zbek tilidagi o‘zlashma so‘zlarning urg‘uli lug‘ati [Matn]: o‘quv-uslubiy lug‘at. – Toshkent: Nodirabegim, 2021. – 988 b. ISBN 978-9943-6940-9-5). Natijada o‘zlashma so‘zlarning urg‘uli bazasi shakllantirilib, o‘zbek nutq sintezatorini yaratishda lingvistik ta‘minot vazifasini o‘tagan.

Tadqiqot natijalarining aprobasiyasi. Mazkur tadqiqot natijalari 14 ta xalqaro, 22 ta respublika ilmiy-amaliy anjumanda e’lon qilingan.

Muallifning quyidagi

<https://scholar.google.com/citations?user=ZSEZYo8AAAAJ&hl=ru>,

<https://www.researchgate.net/profile/Manzura-Abjalova>,

<https://www.linkedin.com/in/manzura-abjalova-0125b21ba/>,

<https://orcid.org/0000-0002-1927-2669> ilmiy platforma profillarida ilmiy ishlari muhokama qilingan.

Tadqiqot natijalarning e’lon qilinganligi. Dissertatsiya mavzusi bo’yicha 48 ta ilmiy ish, jumladan, O’zbekiston Respublikasi Oliy attestatsiyasi komissiyasining doktorlik dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan ilmiy nashrlarda 17 ta maqola (shundan 6 tasi xorijiy jurnalda), 5 ta mualliflik guvohnomasi, 1 ta Skopus bazasiga indekslangan nufuzli xalqaro konferensiyada hamda Respublika va xorijiy konferensiyalarda 20 ta ilmiy maqola va tezis e’lon qilingan. Natijalar 2 ta lug’at va 2 ta monografiyadan o’rin olgan.

Dissertatsiyaning tuzilishi va hajmi. Dissertatsiya kirish, to’rt asosiy bob, xulosa, glossariy va foydalanilgan adabiyotlar ro’yxatidan iborat bo’lib, hajmi 266 sahifani tashkil qiladi. Dissertatsiyaga 33 sahifali ilova biriktiriladi.

DISSERTATSIYANING ASOSIY MAZMUNI

Kirish qismida mavzuning dolzarbligi va zarurati asoslangan, tadqiqotning respublika fan va texnologiyalari rivojlanishining ustuvor yo’nalishlariga bog’liqligi ko’rsatilgan, dissertatsiya mavzusi bo’yicha xorijiy ilmiy-tadqiqotlar sharhi, muammoning o’rganilganlik darajasi tahlili, maqsad va vazifalari berilgan, obyekt va predmeti tavsiflangan, ilmiy yangiligi va amaliy natijalari bayon qilingan, natijalarning ilmiy va amaliy ahamiyati ochib berilgan, joriylanishi, aprobatsiyasi, nashr etilgan ishlar va dissertatsiya tuzilishi bo’yicha ma’lumotlar keltirilgan.

Dissertatsiyaning birinchi bobi “**Lingvistik ontologiya – leksik ma’lumotlar bazasi**” deb nomlangan bo’lib, uch bo’limdan iborat. “*Lingvistik ontologiya va tezauruslar tahlili*” nomli birinchi bo’limda tadqiqot jarayonida ko’ndalang bo’lgan dastlabki masala – ontologiya va tezaurus konseptlari tahlil qilingan, WordNet – ingliz tili leksik ma’lumotlari tizimi va rus tilining ontologik tizimlari tahlil qilingan, ularning tuzilishi, asosiy elementlari va imkoniyatlari yoritilgan.

So’nggi vaqtlarda tibbiy, ilmiy, bank-moliya, siyosiy qidiruv kabi axborot qidirishning ixtisoslashgan turlari tobora muhim ahamiyat kasb etmoqda va bunday axborot tizimlarining sifatini ta’minlashda fan sohalaridagi bilimlarning o’rni muhim. Umuman, matnga avtomatik ishlov berishning zamonaviy usullari yordamida dasturiy tizimlarga til va dunyo haqidagi bilimlarni kiritish qiyin vazifadir. Buning yechimi esa til va dunyo to’g’risidagi bilimlarning maxsus yaratilgan manbalar (tezaurus, ontologiyalar)da aks etishi bilan bog’liq, bunday manbalarda o’n minglab so’zlar va iboralarning tavsifi, boshqa so’z va birliklar bilan semantik munosabatga kirishish va mantiqiy xulosa chiqarish imkoniyatlari

bo‘ladi. Ulardan foydalanilganda, odatda, so‘zlarning ko‘pma’nolilik, omonimlik va polifunksionallik xususiyatlari avtomatik tarzda hal qilinadi.

Ontologiya (yun. *ontos* [ὄντος] – borliq va *logiya* [λόγος] – ta’limot) aslida falsafa bo‘limi, borliq haqidagi ta’limot. Borliqning umumiy asoslari, prinsiplari, uning shakllari va qonuniyatlarini tekshiradi. Ontologiya terminini nemis faylasufi R.Goklenius³² fanga 1613-yil kiritgan, so‘ngra X.Volf (1679 – 1754) o‘z darsligida qo‘llagan (1730-yil) bo‘lsa-da, dastlab yunon faylasuflari uning turli talqinlarini bayon etganlar³³ va “ontologiya” ma’nodoshi sifatida “metafizika” terminini ham qo‘llashgan. Markaziy Osiyoning Kindiy, Zakariyo, Roziy, Forobiy, Ibn Sino singari mutafakkirlari yunon faylasuflaridan farqli ravishda ontologik ta’limotni butunlay yangi bosqichga ko‘tarishdi. Masalan, Forobiy ontologiyaga yagona borliqning mohiyatini ochib beruvchi ta’limot sifatida yondashgan³⁴.

Ontologiyada aloqalar va munosabatlar birlamchi hisoblanadi. Mazkur terminning mana shu xususiyati uning boshqa sohalarida ham keng qo‘llanilishiga sabab bo‘ldi. Demak, “ontologiya” atamasi ko‘plab sohalarida qo‘llaniladi va ikki xil ma’noga ega: 1) “borliq” va “mohiyat”ni o‘zida namoyon etuvchi falsafiy tushuncha; 2) elementlarning mazmunini tavsiflaydigan, ular o‘rtasida tarmoqli munosabat o‘rnatilgan tizim. Ontologiyaga semantik tarmoq sifatidagi qarashlar XX asrning 90-yillar oxirlarida boshlangan. Lingvistik ontologiya (LO) til borlig‘i va mohiyati haqidagi fan sifatida kamdan-kam tilga olinadi. LO lisoniy borliqni tahlil qilish orqali tilning mohiyatini ochib berishga xizmat qiladi. Borliqning asosiy sohalariga tabiat, jamiyat va ong kiradi³⁵. Lingvistik ontologiyalarda ham tabiiy til boyligi, undan foydalanish imkoniyati va lison qamrab olinadi.

Falsafada ontologiyaning *predmeti* mohiyat, kompyuter lingvistikasida esa bilim sohaları hisoblanadi. Falsafada ontologiyaning *obyekti* inson, kompyuter lingvistikasida esa uning obyekti sinsetlar, ya’ni so‘zlarning ma’nodosh qatorlari sanaladi.

Tezaurus (yun. “xazina”) muayyan so‘zning leksik-semantik, kontekstual ma’nolarini qamragan lug‘at hisoblanadi, umuman olganda, maxsus terminologiyadir³⁶. Tezauruslar – muayyan fan sohalarini tavsiflashning eng samarali vositalaridan biridir³⁷. Ayrim manbalarda tezaurus ideografik (semantik) lug‘atga tenglashtiriladi. Ideografik lug‘atda lug‘at maqolalari glossema (bosh so‘z)ning odatdagidek alfavit tartibida emas, balki uning ma’nolari bo‘yicha (bosh so‘z yoki iboraning leksik ma’nosi) shakllantiriladi. Alfavit tartibidagi lug‘atlar muayyan bir so‘z haqida biror narsani bilib olishga xizmat qilsa, ideografik lug‘at ma’lum bir tushunchaga asoslangan ma’lumotlarni o‘zida mujassamlashtiradi,

³² Goklenius R. *Lexicon philosophicum*. Francofurti, 1613.

³³ <https://uz.wikipedia.org/wiki/Ontologiya>

³⁴ To‘rayev B.O. Borliq: mohiyati, shakllari, xususiyati: monografiya/ B.O.To‘rayev; maxs. muharrir M.N.Abdullayeva, O‘zR FA I.Mo‘minov nomidagi Falsafa va huquq instituti. – Toshkent: Falsafa va huquq instituti nashriyoti (FHIN), 2011. – 128 b.

³⁵ To‘rayev B.O. Borliq: mohiyati, shakllari, xususiyati: monografiya/ B.O.To‘rayev; maxs. muharrir M.N.Abdullayeva, O‘zR FA I.Mo‘minov nomidagi Falsafa va huquq instituti. – Toshkent: Falsafa va huquq instituti nashriyoti (FHIN), 2011. – B. 5.

³⁶ [Тезаурус — Википедия \(wikipedia.org\)](https://ru.wikipedia.org/wiki/Тезаурус)

³⁷ [Тезаурус – Википедия \(wikipedia.org\)](https://ru.wikipedia.org/wiki/Тезаурус)

ya'ni muayyan tushunchani qanday so'zlar yordamida ifodalanishi ko'rsatiladi. Ideografik lug'atda so'zdan tushuncha (so'z → tushuncha)ga o'tilmaydi, balki tushunchadan so'zlar (tushuncha → so'zlar)ga tomon fikr harakati yo'naltiriladi. Masalan, *oila* konsepti *ota, ona, farzand, o'g'il, qiz, aka, uka, opa, singil* so'zlarini qamrab oladi. Mazkur lug'atdan foydalanish natijasida o'quvchida muayyan tushunchani turli so'zlar yordamida ifodalash, mantiqiy fikr yuritish, mental idrok etish kompetensiyalarini shakllantiradi yoxud rivojlanishiga turtki beradi.

LO til imkoniyatini borligicha, hamma soha bo'yicha qamrab olsa, tezauruslar muayyan to'plam yoxud yo'nalish, sohaga oid tushunchalar munosabati bilan cheklanadi. LOlar asosida so'zlar va so'z birikmalari bo'lganligi bois ular leksik ma'lumotlar bazasi deyiladi. Bunda lingvistik ontologiyaga faqat leksemalardan iborat tizim emas, balki tabiiy til leksemalari qamroviga ega tizim sifatida qaraladi. Shu bois ham **LOga nisbatan leksik ma'lumotlar bazasi** parafrazasi qo'llaniladi.

Birinchi bobning "*WordNet – lingvistik ontologiyalar uchun tayanch baza*" nomli bo'limida ingliz tili imkoniyatlarini o'zida namoyon etish maqsadida yaratilgan WordNet leksik ma'lumotlar bazasi (ayrim manbalarda *ochiq elektron ontologiya*, ayrim manbalarda *tezaurus*³⁸ deyiladi) tuzilishi, tarkibi va imkoniyatlari, uning asosida yaratilgan ko'p tilli EuroWordNet formal ontologiyasi va GlobalWordNet tizimi tarkibi ochib berilgan.

Princeton WordNet (PWN) leksik ma'lumotlar bazasi (LMB)ni yaratish ishlari 1984-yilda J.Miller va K.Fillbaumlardan tomonidan boshlangan bo'lib, 1995-yildagina WordNet'dan Internet tarmog'ida erkin foydalanish imkoniyati paydo bo'ldi va u matnlarni avtomatik qayta ishlashga mo'ljallangan dasturiy ta'minotlar bo'yicha tadqiqotlarning jadallashishiga turtki berdi. Aslida, WordNet inson xotirasi modeli sifatida psixolingvist J.Miller tomonidan yaratilgan. Shu o'rinda savol tug'iladi: nega aynan tilshunos emas, psixolingvistlar WordNet uchun tamal toshini qo'ydi? Sababi so'z tavsiflarini taqdim etish yuzasidan chiqarilgan ko'plab xulosalar psixolingvistik eksperimentlar bilan bog'liq bo'lgani bois, inson xotirasi va miya neyronlari tarmoqlari imitatsiyasi sifatida ingliz tili uchun WordNet tarmoqli leksik ma'lumotlar bazasi ishlab chiqiladi. Ammo WordNet psixolingvistlardan ko'ra kompyuter lingvistlari qiziqishlarini uyg'otdi.

J.Miller WordNet rivojlanishi omillarini quyidagi 3 farazda mujassamlashtirdi³⁹ (uch gipoteza): 1) **ajraluvchanlik gipotezasi**: tabiiy tilning leksik tarkibiy qismi tavsifini ajratish va alohida o'rganish mumkin. Muayyan yo'nalish, masalan, mashina tarjimai lingvistik bazasi uchun so'z turkumlarini teglash maqsadida barcha turkumlar bazasi alohidalanilishida mana shunday yirik leksik ma'lumotlar bazasi qo'l keladi; 2) **"namuna" gipotezasi (patterning hypothesis)**: tilda o'z formal izohiga ega shunday so'zlar borki, bunday izohlarni tildagi aksariyat so'zlarga qo'llash imkoniyati mavjud. Bunday tavsiflar, asosan, ma'nodosh so'zlarga muvofiq keladi. Shu bois ham WordNet asosini sinonim so'zlar tashkil etadi; 3) **qamrab olish gipotezasi (comprehensiveness)**

³⁸ http://db4.sbras.ru/elbib/data/show_page.phtml?20+1531 _ Тезаурус WordNet; Лукашевич Н. В. Тезаурусы в задачах информационного поиска – М., 2010. – 396 с.

³⁹ http://db4.sbras.ru/elbib/data/show_page.phtml?20+1531 _ Тезаурус WordNet

hypothesis): lug‘aviy birliklar qamrovi keng elektron lug‘at. Matnlarni avtomatik qayta ishlash dasturlarida kompyuter lug‘atlaridan samarali foydalanish uchun lug‘at juda katta hajm va qiymatga ega bo‘lishi zarur hisoblanadi.

WordNet bazasi ot turkumiga oid so‘zlar, fe’llar, sifatlar va ravishlar “sinset” deb nomlangan kognitiv sinonimlar to‘plamida guruhlashtirilgan bo‘lib, har birining semantik munosabatlari quyidagicha: *otda* sinonim, giperonim-giponim, *fe’lda* sinonim, giperonim-giponim, *sifatda* antonim, *ravishda* uning semantik guruhlari va valentliklari aks etgan. Sinsetlarda har bir turkumdagi birlik alohida tushunchani ifodalaydi. Masalan “book” so‘zining *rule book (qoida kitobi) / record book (yozuv kitobi) / volume (jild, tom) / Book (Kitob (atoqli ot)) / record (yozuvlar) / reserve (zaxira)* tushunchalari mavjud. Ular, o‘z navbatida, tarkibiy ma’nodoshlariga ega. Jumladan: *accumulation (to‘plash), aggregation (jamlash), assemblage (yig‘ish), collection (to‘plash, kolleksiya)* – ot turkumi; *section (bo‘lim), subdivision (bo‘linma)* – ot turkumi; *product (mahsulot), production (ishlab chiqarish)* – ot turkumi; *schedule (dastur, jadval)* – ot turkumi; *reserve (zaxira), hold (ushlab turish)* – fe’l turkumi; *put down (qo‘ymoq), enter (kiritmoq)* – fe’l turkumi; *record (yozmoq) - put down (yozib qo‘shmoq), enter (kiritish)* – fe’l turkumi; *record book (yozuvlar kitobi) – record (yozuvlar)* – ot turkumi; *account book (kirim-chiqim kitobi), book of account (hisob kitobi), ledger (ro‘yxatga olish kitobi), leger* – ot turkumi; *playscript (ssenariy), script-publication (qo‘l yozma)* – ot turkumi. Mazkur tarmoqda ayon bo‘lyaptiki, “record” so‘zi ham ot turkumi, ham fe’l turkumiga mansub bo‘lib, ot va fe’l sinonimik qatorlarini birlashtiruvchi tugun hisoblanadi.

1996-yil martidan 1999-yil sentyabrigacha Yevropa Komissiyasi tomonidan moliyalashtirilishi natijasida WordNet’ning ko‘p tilli versiyasi – EuroWordNet yaratildi⁴⁰. Mazkur tizim formal ontologiya hisoblanadi va unda ingliz, daniya, ispan, italyan, nemis, fransuz, chex va eston tillarining WordNet lug‘atlari, Princeton WordNet 1.5 versiyasi birlashtirilgan. EuroWordNet lug‘atlari tijoriy mahsulotlar hisoblanadi.

Birinchi bobning “RuTez lug‘aviy ma’lumotlar bazasining dasturiy ta’minoti” nomli ikkinchi bo‘limida RuTez va RussNet lingvisti resurlari tahlil qilingan. RuThes – rus tilidagi tezaurus (ayrim manbalarda ontologiya⁴¹). Axborot tadqiqotlari markazi tomonidan avtomatik indeksatsiya qilish vositasi sifatida 1994-yildan yaratila boshlangan va hozirgi kungacha tarkibi ishlab chiqishda davom etmoqda. RuTez tezaurusining rivojlanishi ijtimoiy-siyosiy tezaurusning rivojlanishi bilan boshlangan⁴². 45 ming tushunchalar, 107 ming so‘z va iboralar, 177 ming sinonimik munosabatlarni o‘z ichiga oladi.

RuTez quyidagi to‘rtta tamoyilga asoslanadi va shunga binoan to‘rtta XML faylidan tashkil topgan⁴³: 1) tushuncha – `concepts.xml`; 2) tushunchalar o‘rtasidagi

⁴⁰ P.Vossen. Building a multilingual database with wordnets for several European languages. <http://www.illc.uva.nl/EuroWordNet/>

⁴¹ <https://nlpub.ru/PyTez>; <https://new.labinform.ru/pub/ruthes/index.htm>

⁴² Лукашевич Н.В., Салий А.Д., Тезаурус для автоматического рубрицирования и индексирования: разработка, структура, ведение // НТИ. Сер.2. - 1996. - N 1. – С.1-6.

⁴³ <https://nlpub.ru/PyTez>

munosabat – relations.xml ; 3) matn kiritish elementi – text_entry.xml; 4) tushunchalar va matnli kiritmalar o‘rtasidagi munosabat – synonyms.xml.

Dissertatsiyaning “**Lingvistik ontologiyalarni yaratish texnologiyasi va mezonlari**” deb nomlangan ikkinchi bobida keng ko‘lamli bilim sohalarda matnlarni qayta ishlash uchun foydalanishga mo‘ljallangan lingvistik ontologiya modeli, formal sxemalari va shu model asosida ishlab chiqilgan aniq resurslar tavsiflanadi. Model tarzida keng ko‘lamli predmet sohalarida bilimni tavsiflashga qaratilgan uch paradigma: axborot-qidiruv tezauruslari, WordNet turiga mansub tezauruslar, ontologiyalar inobatga olingan. Tushunchalar orasidagi munosabatlar tizimiga alohida e‘tibor qaratilgan. Shuningdek, turk tilining WordNet tizimidagi KeNet resursi tuzilishi, tarkibi, yuzaga kelgan muammolarini yoritish asnosida o‘zbek tili ontologiyasini yaratish texnologiyasi va mezonlari aniqlashtirib berilgan.

Mazkur bobning birinchi – “*KeNet – turk tili ontologiyasining yaratilish tamoyillari*” deb nomlangan bo‘limida turk tili uchun keng qamrovli WordNet, ya‘ni KeNet⁴⁴ va uning yaratilishi atroflicha yoritib berilgan va tahlil qilingan. KeNet ham ichki semantik munosabatlarga ega bo‘lib, 76 757 ta sinsetni o‘z ichiga oladi. PWN (Princeton WordNet) bilan tillararo aloqa orqali bog‘langan. Ushbu bo‘limda Sinset yaratishda sinset tarkibiga kiruvchilarning semantik aloqasida yuzaga kelgan asosiy muammo sababli sinsetlardagi sinonimik munosabatlarni boshqarish uchun ikkita jarayon – *birlashish jarayoni* va *ajratish jarayoni* bajarilganligi tushuntiriladi. KeNet`da birlashish jarayonida birlashtirilishi kerak bo‘lgan turli xil sinset to‘plamlari aniqlangan, identifikatsiya qilingan va sinsetlar bir to‘plam sifatida guruhlangan. Sinsetlar to‘plamini birlashtirishda uchta narsa juda muhim sanaladi⁴⁵: 1) har bir ma‘lumotlar to‘plami yagona va o‘ziga xos izoh/tavsifga ega bo‘lishi; 2) har bir ma‘lumotlar to‘plamida sinonim a‘zosi sifatida haqiqiy sinonimlarning mavjud bo‘lishi; 3) har bir ma‘lumotlar to‘plamida birinchi tarkibiy qismi reprezentativlangan (o‘zida lingvistik ma‘lumotni ifodalovchi, ko‘rsatuvchi, grammatik xarakterli) bo‘lishi kerak. Noto‘g‘ri birlashtirishning uchta asosiy omili bor: 1) ma‘no, izohlarning turli xilligi; 2) POS, ya‘ni so‘z turkumlarini teglash bilan bog‘liq muammo; 3) morfologiyaga bog‘liq muammolar⁴⁶. Dissertatsiyada mazkur omillar izohlangan. Ma‘noga bog‘liq bo‘lgan muammo shundaki, semantik jihatdan bir-biriga yaqin bo‘lgan, ammo aynan sinonim bo‘lolmaydigan (aynan emas yoki bir-biri o‘rnida qo‘llanilmaydigan) so‘zlar sinsetlarning tarkibiy qismi sifatida berilgan. Masalan, *dere (irmoq)* va *irmak nehir (daryo)* kabi o‘xshash ma‘nolarga ega bo‘lgan otlar ideografik jihatdan o‘rinli joylashgan bo‘lsa-da, bir-biriga semantik muvofiqligi jihatdan noto‘g‘ri birlashtirilgan. POS muammosi ma‘no jihatdan bir-biriga muvofiq, ammo turli xil so‘z turkumlariga mansub bo‘lgan sinset tarkibiy qismlari

⁴⁴ KeNet qisqartma nomidagi Ke qismi turkcha “kelime” (kalom, so‘z) so‘zining birinchi bo‘g‘ini hisoblanadi.

⁴⁵ Özge Bakay and others. TurkishWordNet KeNet. Global Wordnet Virtual Conference. 2021. January. – P. 166. https://www.researchgate.net/publication/348264475_Turkish_WordNet_KeNet

⁴⁶ Bakay O., Ergelen O., and Yildiz O.T. 2019. Problems caused by semantic drift in wordnet synset construction. In UBMK.

uchraganda yuzaga keladi. Masalan, *güç* (kuch, quvvat) va *güç* (qarshilik) soʻzlari ot turkumiga oid, *güç* (qiyin) sifat turkumiga oid soʻz bir oʻzakka birlashtirilib, unga izohlar umumiy holda berilgan. Bunday holat natijasi sinset aʼzolarining notoʻgʻri birlashtirilishiga sabab boʻladi. Morfologiya bilan bogʻliq muammo esa soʻzning turli morfologik shakllari bitta sinset tarkibiy qismlari sifatida notoʻgʻri birlashtirilganda yuzaga kelgan. Masalan, bir xil asosga ega *sopalamak* (savalamoq) – sof feʼl, aniq nisbat, feʼlning noaniq shaklida va *sopalanmak* (savalanmoq, kaltaklanmoq) – majhul nisbatdagi feʼl turli xil bajaruvchilariga egaligini bildirishi bilan maʼno farqlariga ega, morfologik koʻrsatkichlari ham farqli boʻlsa-da bir xil guruhda birlashtirilgan. KeNet`da bunday shakllar ham ajratilib, ular uchun alohida maʼlumotlar toʻplami yaratilgan.

Ikkinchi bobning “*Leksik maʼlumotlar bazalari klassifikatsiyasi (tasnifi)*” deb nomlangan boʻlimida manbalarda lingvistik ontologiyaga berilgan 11 yondashuvli taʼriflar tahlil qilinadi va natijada biz tomonidan maqbul taʼrif muayyanlashtiriladi: “*Ontologiya – til va dunyo bilimlariga asoslangan, sohalararo terminlarni qamrab olgan va ular oʻrtasidagi munosabatlar asosida shakllantirilgan tarmoqli leksik maʼlumotlar bazasi*”. Mazkur boʻlimda, shuningdek, ontologiyalar tasnifi tahlil qilinadi va ikki yondashuv asosida yuzaga kelgan uning quyidagi turlari yoritib beriladi: *metaontologiyalar* – predmet sohalariga bogʻliq boʻlmagan eng umumiy tushunchalarni tavsiflaydi; *soha ontologiyasi* – sohaning formal tavsifi; u odatda meta-ontologiyada (agar foydalanilsa) aniqlangan tushunchalarni aniqlashtirish va/yoki predmet sohasining umumiy terminologik bazasini aniqlash uchun ishlatiladi; *aniq vazifalar ontologiyasi* – vazifa yoki muammo bilan bogʻliq umumiy terminologik bazani belgilaydigan ontologiya; *tarmoq ontologiyalari*, koʻpincha, soha yoki vazifadagi obyektlar tomonidan bajarilgan harakatlarning yakuniy natijalarini tavsiflash uchun qoʻllaniladi.

Ikkinchi bobning “*Ontologik tarmoq – leksik maʼlumotlar tizimi modellari*” nomli boʻlimida hozirgi vaqtda axborot qidiruv va axborot-tahlil tizimlarida foydalaniladigan dunyo va til haqidagi bilimlarni oʻz ichiga olgan resurslarning uchta asosiy paradigmasi muhokama qilingan. Bular: tezauruslar, WordNet (tezaurusi) va formal ontologiyalar⁴⁷.

Formal ontologiyalar – axborot qidirish ilovalari uchun kompyuter resurslarining zamonaviy paradigmasi⁴⁸. Ushbu tizim keng koʻlamli ontologik resurslarni qurishga asoslangan Semantik veb (Semantic Web)⁴⁹ konsepsiyasini ilgari surish natijasida yuzaga kelgan⁵⁰. Ammo formal ontologiyalar tarafdorlari fikriga qarshi ularoq S.Nerinburg tabiiy tildagi strukturlashtirilmagan matnlarni ulardagi koʻp maʼnolilik, omonim va polifunksionallik hodisalari bilan avtomatik qayta ishlashni aksiomatik nazariyalar yordamida amalga oshirish qiyin, deya

⁴⁷ Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – Москва: МГУ, 2011. – 512 с.

⁴⁸ Qarang: <http://globalwordnet.org/resources/wordnets-in-the-world/>

⁴⁹ Berners-Lee, T. The Semantic Web / T. Berners-Lee, J. Handler, O. Lassila // Scientific American - 2001. – V. 284. – No 5. – P. 28-37.

⁵⁰ Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования, том 5, №1(15)/2015. – С. 48.

munosabat bildiradi⁵¹. Shu bois matnlarni avtomatik qayta ishlash uchun ontologiyaning maxsus turlari (terminologik yoki maxsus ontologiyalar) ishlab chiqilmoqda⁵². Bir jihatdan, formal ontologiyalarda tushunchalar to‘liq o‘z aksini topmaydi. Boshqa tomondan, matnni tahlil qilishda, ko‘pincha, ontologiyaga asoslangan formal mantiqiy xulosa zarur hisoblanadi, chunki izchil matnda ma’lumotlarning hajmi aniq ko‘rsatilmagan bo‘ladi⁵³. Bundan tashqari, matnlarni avtomatik qayta ishlashda ontologiyalardan foydalanilgani sababli til birliklari tushunchalari va bilim sohasi atamalari bilan jiddiy aloqador bo‘lgan *lingvistik ontologiya* deb ataluvchi tushuncha paydo bo‘ldi⁵⁴.

N.V.Lukashevich ontologiyaning rasmiy ta’rifini quyidagi formulada ifodalaydi:

$$O = \langle C, E, At, R, A \rangle,$$

bunda: C – ontologiyadagi tushunchalar (sinflar), E – ontologiyadagi ekzemplarlar (atoqli otlarni turdosh otlar munosabatdoshi sifatida qayd qilinishi), At – ontologiya tushunchalari va ekzemplarlari atributlari, R – tushunchalar o‘rtasidagi munosabatlar, A – ontologiya aksiomalari⁵⁵.

Ontologiyaning mashhur rasmiy modellaridan biri quyidagicha ifodalanadi⁵⁶:

$$O = \langle L, C, F, G, H, R, A \rangle,$$

bu yerda: $L = L_C \cup L_R$ – ontologiya lug‘ati. Unda L_C tushunchasi uchun leksik birliklar (belgilar) to‘plami hamda L_R munosabatlari uchun belgilar jamlanmasi bo‘ladi;

C – ontologiyadagi tushunchalar to‘plami; F va G – L ga tegishli $\{l_j\}$ leksik birliklar to‘plamlarini ushbu ontologiya tushunchalari va munosabatlari jamlanmasi bilan bog‘laydi; H – munosabatlar (aloqalar)ning taksonomik tabiatini belgilaydi, bunday hollarda ontologiya tushunchalari $H \subset C \times C$ ko‘rinishidagi norefleksiv, atsiklik, tranzitiv munosabatlar bilan bog‘liq bo‘ladi; R – ontologiya tushunchalari o‘rtasidagi notaksonomik munosabatlarni anglatadi; A – ontologiya aksiomalari to‘plami. Tabiiy til ontologiyasi, jumladan, o‘zbek tili ontologiyasini yaratishda ushbu modelga asoslanish maqsadga muvofiq sanaladi.

Tadqiqot ishining **“Leksik ma’lumotlar bazalarida semantik munosabatlar”** nomli uchinchi bobida sinonimlar to‘plami – sinsetlarni, ya’ni muayyan tilning WordNet tizimini yaratish jarayonlari tushuntirilgan, xolonimiya

⁵¹ Nirenburg, S. What’s in a symbol: Ontology, representation, and language / S. Nirenburg, Y. Wilks // Journal of Experimental and Theoretical Artificial Intelligence. - 2001. - V. 13(1). - P. 9-23.

⁵² Ландэ, Д.В. Подход к созданию терминологических онтологий / Д.В. Ландэ, А.А. Снарский // Онтология проектирования. 2014. № 2(12). – С. 83-91.; Sowa, J. Building, Sharing and Merging Ontologies. – <http://www.jfsowa.com/ontology/ontosar.htm> .

⁵³ Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования, том 5, №1(15)/2015. – С. 48-49.

⁵⁴ Magnini, B. Merging Global and Specialized Linguistic Ontologies / B. Magnini, M. Speranza // Proceedings of OntoLex. - 2002. - P. 43-48.; Veale, T. A context-sensitive framework for lexical ontologies / T. Veale, Y. Hao // Knowledge Engineering Review. 2007. Vol. 23(1). – P. 101-115.

⁵⁵ Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования, том 5, №1(15)/2015. – С. 49.

⁵⁶ Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования, том 5, №1(15)/2015. – С. 50.

va giperonimiya munosabatlar konsepsiyasi, ularni tarkibiy modellashtirish tamoyillari va ahamiyati, turlari borasida tadqiq natijalari keltirildi.

WordNet.princeton.edu tizimida quyidagi turkumlar semantik tarmoqlangan: *ot* (sinonim, giperonim-giponim), *fe'l* (sinonim, giperonim-giponim), *sifat* (antonimi), *ravish* (semantik guruhi, valentliklari).

Mazkur bobning “*Sinset to‘plamlari – leksik ma’lumotlar tizimining birlamchi elementlari*” nomli birinchi bo‘limida sinset to‘plami va uni yaratish bosqichlari yoritildi. Ma’nodosh birliklar qatori va turkumlari belgilangan so‘zlar lingvistik ontologiyada eng muhim element / atribut (axborot texnologiyalari tili bilan aytganda)lar hisoblanadi. LMBda sinonim so‘zlar qatori to‘plam hisoblanadi va ular tarmoq orqali bir-biri bilan bog‘langani uchun “*sinset*” deyiladi.

Sinsetlarni, jumladan, o‘zbek tili ontologiyasi sinonimlar to‘plamini yaratishda, asosan, quyidagi resurslar zarur hisoblanadi: 1) sinonimlar lug‘ati – ma’nodosh so‘zlar to‘plamini yaratish va /yoki sinset tarkibini to‘ldirish uchun; 2) izohli lug‘at – so‘zlar izohini berish va giperonim munosabatini bergilash uchun. M-n: stul ta’rifi beriladi, uning mebel predmeti ekanligi belgilanadi; 3) Milliy korpus – lug‘atga kirmagan lug‘aviy ma’nodoshlar, kvazisinonimlar, matniy sinonimlarni topish uchun; 4) parallel korpuslar – ishonchli lug‘aviy, matniy tarjima ekvivalentini topish uchun.

E’tiborli jihati, PWN’da dastlab so‘z turkumlariaro munosabat o‘rnatilmagan. Bu dasturiy ilovalarda jiddiy muammolarni keltirib chiqarganligi sababli, EuroWordNet loyihasida so‘z turkumlari o‘rtasida qo‘shimcha munosabatlar kiritilgan: 1) xpos-sinonimiya – so‘z turkumlari sinonimiyasi; 2) xpos-antonimiya – so‘z turkumlari antonimiyasi; 3) xpos-giponimiya – so‘z turkumlari giponimiyasi. Shunday qilib, 2.5.1-bo‘limda *adornment2* (bezatish jarayoni) va *adorn1* (bezatish) sinsetlari o‘rtasidagi munosabatlarni so‘z turkumlari sinonimiyasi munosabati bilan izohlash mumkin⁵⁷. Bunday munosabat sinonimiyaning kvazisinonim deb nomlangan turi hisoblanadi.

III bobning “*Xolonim – meronim munosabati: modellashtirish tamoyilari*” nomli bo‘limida xolónim (qad. yunoncha ὅλος = «butun» + ὄνομα = «nom») muayyan tushuncha(lar)ning butun holati (ko‘rinishi)⁵⁸ – butun va merónim (qad. yunon. μέρος = «qism» va ὄνομα = «nom») boshqa tushunchaning tarkibiy qismi⁵⁹, meronim ayrim adabiyotlarda partonim⁶⁰ (lot. pars, chiq.k. – partis = «qism») – qism munosabati va ularning turlari tahlil qilindi.

M.V.Nikitinning so‘zlariga ko‘ra, meronimik munosabatlar lug‘atni semantik tartiblashda eng muhim omil, shuning uchun u ajralmas iyerarxik tuzilma sifatida

⁵⁷ Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – Москва: МГУ, 2010. – С. 72.

⁵⁸ https://ru.wikipedia.org/wiki/Мероним_и_холоним

⁵⁹ <https://kartaslov.ru/значение-слова/мероним>

⁶⁰ Глобина Л.В. Лексико-семантическое поле партитивной лексики в современном русском языке: дис. ... канд.филол.наук. – Воронеж, 1995. – 205 с.; Коннова М.Н. Введение в когнитивную лингвистику : учебное пособие. Изд. 2-е, перераб. — Калининград : Изд-во БФУ им. И. Канта, 2012. – 313 с.; Материнская О.В. Система меронимов в немецком и английском языках: дис. ... д-ра филол.наук. – Донецк, 2013. – 403 с.; Колодько Д.А. К вопросу о классификации меронимов // Научные записки Национального университета «Острожская академия». Серия «Филология»: сборник научных трудов. – Острог, 2015. Вып.51. – С. 226-228;

namoyon bo‘ladi. M.V.Nikitin meronimiyaga shunday tavsif beradi: “Shubhasiz, butun-bo‘lak munosabatlari butun dunyoni pastdan yuqoriga, mikro dan makrokosmosga, elementar zarralardan galaktikalarga qadar qamrab oladi. Ular har xil daraja murakkabligidagi narsalarni qismlar-butun, elementlar-tizimning turli bosqichli iyerarxiyaga keltirgan holda o‘z ichiga oladi. Bu munosabatlar umumiy va global hisoblanadi”⁶¹.

Shu o‘rinda qayd etib o‘tish joiz: sinekdoxa ham butun-bo‘lak munosabatiga asoslanadi. Xolonimiya bilan farqli jihati shundaki, sinekdoxa ma’no ko‘chishining bir turi hisoblansa, xolo-meronimiyada to‘g‘ri ma’noda funksional va fiziologik yoki jisimiy tuzilishi bo‘yicha qism-bo‘lak munosabatlari mavjud bo‘ladi.

Partonimiya hodisasini o‘rganish lug‘at boyligi tizimining «sir»larini ochish, so‘z ma’nolarini aniq va to‘g‘ri izohlash nuqtayi nazardan ahamiyatlidir⁶². Bu hodisa o‘zbek tilshunosligida H.Ne‘matov, R.Rasulov, B.Qilichev, H.Jamolxonov⁶³ kabi tilshunoslarning ishlarida ma’lum darajada o‘z talqinini topgan. Lingvistik ontologiyani yaratishda butun-bo‘lak munosabatlarini o‘rganishda xorijlik mutaxassislardan Ye.Layzi, D.Kruz, R.Chaffin, Ye.Vinston, D.German, V.Storey; rus olimlaridan M.Nikitin, N.Lukashevich, Ye.Materinskaya, Yu.Rusina, D.Kolodko ishlari fundamental asos bo‘ldi.

Klassik mereologiyada butun-qism (bo‘lak) munosabatida 3 ta aksioma keltiriladi⁶⁴. Formulalarda P – butun, x, y, z – qismlar.

1. Refleksivlik. Hamma narsa o‘zining tarkibiy qismi hisoblanadi. ($P.S.$ yoki oddiygina $P=x,y$)

2. Antisimmetriya: hech narsa o‘zining tarkibiy qismlarining bo‘lagi bo‘la olmaydi. ($P.S. P \neq x \leftarrow p, y \leftarrow p$)

3. Tranzitivlik – o‘tkazuvchanlik: qismlarning qismlari butunning ham qismlari hisoblanadi. ($P.S. x \leftarrow y, y \leftarrow z = P \leftarrow z$)

Butun-qism munosabatlarining mazkur aksiomalar tizimi, odatda, asosiy mereologiya (базовая мереология, ground mereology) deb ataladi⁶⁵.

Ko‘plab mualliflarning ta’kidlashicha, lingvistik tahlilda butun-bo‘lak munosabatlar transitivligi bilan jiddiy muammolar paydo bo‘ladi. Masalan, *qo‘l – bu dirijor qismi* (a’zosi), *dirijor esa orkestr bo‘lagi*, ammo yuqoridagi tranzitivlik (3) aksiomasiga binoan “*qo‘l – orkestr qismi*” deyish g‘alizlikni yuzaga keltiradi⁶⁶. N.V.Lukashevich tranzitivlik bilan bog‘liq bunday muammolar butun-qism munosabatlarida har xil turdagi obyektlarning aralashishi bilan bog‘liqligini

⁶¹ Никитин М.В. Курс лингвистической семантики: учебное пособие для студентов, аспирантов и преподавателей лингвистических дисциплин в школах, лицеях, колледжах и вузах. СПб.: Научный центр проблем диалога, 1996. – С. 442.

⁶² Jamolxonov H. Hozirgi o‘zbek adabiy tili: Darslik. – Toshkent: Talqin, 2005. – B. 147.

⁶³ Begmatov E., Ne‘matov H., Rasulov R. Leksik makrosistema va uning tadqiq metodikasi (Sistem leksikologiya tezislari) // O‘zbek tili va adabiyoti. 1989, № 6. – B.35-40.; Qilichev B. O‘zbek tilida partonimiya. Filol.fan.nomz...dis. – Toshkent, 1997.; Jamolxonov H. Hozirgi o‘zbek adabiy tili: Darslik. – Toshkent: Talqin, 2005. – 260 b.

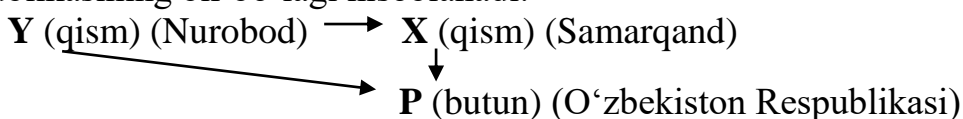
⁶⁴ Simons P. (1987). Parts. A study in Ontology. Oxford University Press. – 390 p.; Varzi A. (2006). A Note on Transitivity of Parthood // Applied Ontology, 1:2, pp. 141-146.

⁶⁵ Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – Москва: МГУ, 2011. – 512 с.

⁶⁶ Лукашевич Н.В. Отношения часть-целое: теория и практика. // «Нейрокомпьютеры: разработка, применение». – Москва: Радиотехника, 2013. – С. 9.

ta'kidlaydi⁶⁷. Vinston va boshq. (Winston et al) ishida tranzitivlik bilan bog'liq muammolar quyidagicha izohlanadi: "bir turdagi munosabat qo'llanilganda, butun qism har doim ham o'tuvchi bo'ladi. Biroq meronimiyaning har xil munosabatlari aralashganda, transitivlikda muammo yuzaga keladi"⁶⁸. Yana bir misol: *barg* – *daraxtning bir qismi*, *daraxt* – *o'rmonning bir bo'lagi*, lekin *barg* *o'rmonning bir qismi* deyish g'alati bo'ladi.

D.Kruz esa o'z ishida yaxshi shakllangan iyerarxiya bir xil turdagi elementlardan iborat bo'lishini ta'kidlaydi⁶⁹. D.Kruz fikrini misol yordamida tushuntiramiz: butun-qism munosabatida agar bir element geografik nom yoki hudud bo'lsa, u holda boshqa elementlar ham shu turga mansub bo'lishi kerak. Masalan, Nurobod tumani Samarqand viloyatining bir qismi, Samarqand O'zbekiston Respublikasining bir bo'lagi, demak, Nurobod O'zbekiston Respublikasining bir bo'lagi hisoblanadi.



Shunday qilib, agar meronimiya elementi jismaniy obyekt bo'lsa, meronimiyaning boshqa barcha elementlari ham jismaniy bir xil bo'lishi kerak. Agar bitta element mavhum ot bo'lsa, unda boshqalari ham shunday turda bo'lishi kerak. Bizningcha, tranzitivlikda butun-qism munosabatini "to'liq tranzitivlik" va "qizman tranzitivlik" tarzida guruhlab olinsa, maqsadga muvofiq bo'ladi. Shunda qisman tranzitivlikda matnda uchragan, ammo qism-butun munosabatida g'aliz hisoblangan birliklar o'z aksini topadi. Umuman olganda, bunday holatning yuzaga kelishi kundalik turmushda "qism" tushunchasining torayishi bilan bog'liq.

Endi yuqorida keltirilgan misolga qaytamiz: *dirijorning qo'li* – *dirijor* – *orkestr*. Bunda qo'lning massasi orkestr massasining bir qismi ekanligi, dirijorning qo'li orkestr egallagan joyning bir qismidaligini ko'rishimiz mumkin. Agar dirijorning qo'li shikastlangan bo'lsa, u orkestrning ishlashi bilan bog'liq muammolarga olib kelishi mumkin (hatto jiddiy bu orkestr uchun fojea ham hisoblanadi). *Barg* – *daraxt* – *o'rmon* munosabatlarini ham xuddi shunday izohlash mumkin. N.V.Lukashevich shu masalaga to'xtalar ekan, "qism" tushunchasini talqin qilish uchun qo'shimcha shartlar qo'yish joizligini ta'kidlaydi, ya'ni bu qism funksional bo'lishi kerakligi haqidagi qo'shimcha talab va hokazolar, albatta, o'tuvchanlikni yuzaga keltirmasligi mumkin⁷⁰, deydi. Bunday holda, Butunning o'z funksiyalari va qismning butun bajaradigan vazifalarni to'ldiruvchi funksialar tranzitivlikni yuzaga keltirmaydigan yoxud g'alizlikka olib keluvchi omillar bo'lishi mumkin. Mana shunday farqlanishga asosan leksik ma'lumotlar bazasida biz yuqorida taklif qilgan ikki toifa guruhlashni amalga oshirish mumkin.

⁶⁷ O'sha manba.

⁶⁸ Winston M., Chaffin R, Herrmann D. 1987. A Taxonomy of Part-Whole Relations // Cognitive Science, 11, – pp. 417-444.

⁶⁹ Cruse D. 1986. Lexical Semantics. Cambridge. University Press. – 310 p.; Winston M., Chaffin R, Herrmann D. 1987. A Taxonomy of Part-Whole Relations // Cognitive Science, 11, – pp. 417-444.

⁷⁰ Лукашевич Н.В. Отношения часть-целое: теория и практика. // «Нейрокомпьютеры: разработка, применение». – Москва: Радиотехника, 2013. – С. 11.

1. *Funksional qismlar* fazoviy holat va vaqt vaziyatida funksiyasi bilan chegaralangan bo'ladi. Misol uchun, *dastali piyola* (chashka) *tutqichi* ushlar vazifasini bajarsa, joyning chegaralangan qismida joylashishi mumkin.

2. *Gomeo'lovli qismlar* butun mansub bo'lgan turni ifodalaydi, ya'ni qismlar butunga aynan o'xshaydi yoki o'zi mansub bo'lgan butunga to'g'ridan to'g'ri tegishli bo'ladi. Masalan, *ushoq – non, bo'lak – tort*. Gomomerik yoxud gomeo'lovli bo'lmagan qismlar esa butundan farq qiladi, masalan, *daraxt – o'rmon, stol – mebel* kabi.

3. *Alohidalgan qismlar* butundan nisbatan ayri holda turadi. Masalan, *tortma – stol* (tortmani stoldan ajratish mumkin), *tutqich – chashka* (bir-biridan ajratib bo'lmaydigan qismlar).

Mazkur 3 belgi kombinatsiyasi asosida qism va butun munosabati yuzaga kelishining oltita turi ajratiladi (dissertatsiyada xolonimiya turlari misollar bilan batafsil yoritildi).

Wordnet ma'lumotlar bazasida muayyan bir tushuncha bilan bog'liq meronim munosabatlarning uch turi aniqlandi⁷¹: 1) qisman meronim: “*g'ildirak*” “*mashina*”ning bir qismidir; 2) ishtirokchining meronimi: “*mashina*” – “*tirbandlik*” ishtirokchisi; 3) modda (moddiy) meronimi: “*g'ildirak*” “*kauchuk*”dan tayyorlangan.

Turli ontologiyalar bazalari va ularga doir tadqiqotlarni o'rganish jarayonida butun-qism munosabatlarini o'rnatish tamoyillari va talqinlari hamma leksik bazalarida turlicha va o'rganilgan manbalarda ham mushtarak me'yor mavjud emasligi kuzatildi. Shunday bo'lsa-da, meronimik (partonimik) munosabatda butun va qismni bog'lovchi belgi ularning bir xil funktsiya bajarishi bilan belgilanishini qayd etib joiz.

Uchinchi bobning “*Ontologik lug'atlarda gipo-giperonim munosabatlari*” nomli bo'limida jins va tur munosabati tadqiq etilgan. Tilning lug'at boyligini tizim (sistema) sifatida o'rganishda leksemalarning gipo-giperonimik munosabatlariga tayanish muhim ahamiyatga ega: u tabiat va jamiyatdagi narsa-predmetlarning, voqea-hodisalarning tildagi nomlari bo'lgan leksemalarning ma'nolarini va shu ma'nolar orqali borliqdagi narsa-hodisalarning o'zlari haqidagi tushuncha-tasavvurlarni umumlashtirish va farqlash imkonini beradi⁷². Shuningdek, leksik turkumlash natijasida yuzaga kelgan so'zlarning giponimik bog'lanishlari lug'atni iyerarxik tashkil etishning eng muhim usuli sifatida qaraladi.⁷³

Giponimiya hodisasi fransuz tilshunosi Dj. Layonz, Sh. Balli, V.G. Gak; rus tadqiqotchilaridan Yu.N. Karaulov, A. Vejvskaya, D.N. Shmelev, A.A. Ufimseva,

⁷¹ <https://www.greelane.com/ru/what-is-a-meronym-1691308/>

⁷² Bu haqda qarang: Новиков Л.А. Семантика русского языка. – М.: Высшая школа, 1982, – С.136-142.; Ne'matov H, Rasulov R. O'zbek tili sistem leksikologiyasi. – T.: O'qituvchi, 1995, 111-123-b.; Qo'chqortoyev I. So'zlarning leksik-semantik to'dalari haqida. // ToshDU ilmiy asarlari, 359-chiqishi. – Toshkent, 1969.; Расулов Р. Лексико-семантические группы глаголов состояния и их валентность. – Тошкент, 1991.; Сафарова Р. Гипонимия в узбекском языке. АКД. – Тошкент, 1990.

⁷³ Котцова Е.Е. Гипонимические связи глаголов и существительных в лексической системе русского языка. // Вестник Нижегородского университета им. Н.И.Лобачевского, 2011. № 6 (2), – С. 324-327.

L.A.Novikov, M.V. Nikitin, Ye.Ye.Kotsov tomonidan keng tadqiq etilgan.⁷⁴ O‘zbek tilshunosligida giponimik munosabatlarning ajratilishi va keng jamoatchilik hukmiga havola etish Rohatoy Safarovaning ilmiy tadqiqotlari bilan bog‘liq. Ushbu ishda jins-tur munosabatlari o‘zbek tilida ilk bor tadqiq etilgan⁷⁵. R.Safarova o‘zbek tilida 1000 ga yaqin hayvonlar nomini o‘nta mazmuniy guruhga ajratib, ular o‘rtasidagi giponimik (jins-tur) munosabatlarni ochib bergan. Shuningdek, D.Ahmedova o‘z tadqiqotida atov birliklarini teglashda leksik-semantik munosabatlarning o‘rni borasida to‘xtalib o‘tgan. Bu borada qilingan tadqiqotlarni tahlilga tortgan.⁷⁶

So‘zlarning giponimik qatori giperonim va giponimdan iborat. Tarkibi kengroq so‘zni giperonim, tor mazmundagi tegishli so‘zni giponim deb atashadi: idishlar (giperonim), tarelka, chashka (giponimlar). Yanayam aniqroq aytganda, giperonim jins, ya‘ni dominant yohud A.Nurmonov aytganlaridek, uya; giponim – ma‘lum jins turlarining nomlarini ifodalovchi hamda o‘zining semantik tarkibida jins ma‘nosini ifodalovchi so‘zni birliktirib kelgan, semantik jihatdan giperonimga nisbatan boy, ammo o‘z o‘rnida qaram bo‘lgan lug‘aviy birlik, uyadosh.

A.Sobirovning fikricha, jins-tur (giper-giponimik) munosabati leksik sathdagi semantik maydonlarning asosini tashkil qiladi. Uzv (a‘zo)larni uyalarga, uyalarni ulardan kattaroq to‘dalarga, to‘dalarni guruhlarga, guruhlarni semantik maydonlarga birlashtirish chog‘ida jins-tur munosabati yuzaga chiqib boraveradi.⁷⁷ Ayon bo‘lganidek, kichikroq narsalar o‘zidan kattaroq narsalarning ichiga darajama-daraja kirib boraveradi. Paradigma ichidagi har bir leksema giponimlik mavqeiga ega bo‘ladi. O‘z navbatida har bir giponim bir qancha leksemalarni bir yerga jamlay oladi va u jamlangan guruh ichida boshqalariga nisbatan giperonim bo‘lib qoladi.

giperonim (jins / uya)	giponimlar (tur / uyadosh)
<i>nasr</i>	<i>qatra, felyeton, ocherk, novella, qissa, povest, roman</i>
<i>nazm</i>	<i>g‘azal, tuyuq, ruboiy, fard, masnaviy, qasida, oq she‘r, poema, doston</i>
<i>drama</i>	<i>drama, komediya, tragediya, tragikomediya</i>
<i>adabiyot</i>	<i>nasr, nazm, drama</i>
<i>san‘at</i>	<i>Adabiyot, haykaltaroshlik, rassomchilik, kino, teatr</i>

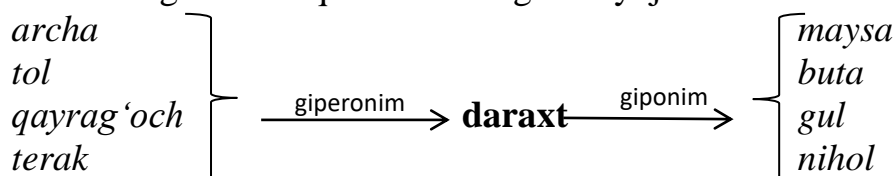
⁷⁴ Уфимцева А.А. Семантика слова // Аспекты семантических исследований. М., 1980. С. 5-80; Степанов Ю.С. Имена, предикаты, предложения (Семиологическая грамматика). М., 1981. – 360 с.; Новиков Л.А. Семантика русского языка. – Москва: Высшая школа. 1982. – С.241-243.; Никитин М.В. Основы лингвистической теории значения. – Москва: Высшая школа, 1988. – С.73-87.; Котцова Е.Е. Гипонимия в лексической системе русского языка (на материале глагола): Автор. дисс...д.ф.н. – Архангельск, 2010. – С. 10.

⁷⁵ Safarova R. O‘zbek tilida giponimiya. // O‘zbek tili va adabiyoti. – T., №1. 1987.; Сафарова Р. Гипонимия в узбекском языке (на материале общепотребительных зоонимов): Автор.дисс.учен.ст.канд.ф.н. – Ташкент, 1990. – 20 с.; Safarova R. Leksik-semantik munosabat turlari. – Toshkent: O‘qituvchi, 1996.

⁷⁶ Ahmedova D. Atov birliklarini o‘zbek tili korpuslari uchun leksik-semantik teglashning lingvistik asos va modellari: Filol.fan.bo‘yicha falsafa doktori (PhD)...diss. – Buxoro, 2020. – 145 b.

⁷⁷ Sobirov A. O‘zbek tilining leksik sathini sistemalar sistemasi tamoyili asosida tadqiq etish. – Toshkent: Ma‘naviyat, 2004. – 104 b.

Bir tomondan, *daraxt* tushunchasi umuman, bir butun, aniq, real tushuncha sifatida *buta* va *maysa* soʻzlari bilan bogʻlanadi. *daraxt* giperonimi jins tushunchasini ifodalovchi soʻz sifatida shu jinsning turlarini ifodalovchi soʻzlar bilan bogʻlanadi. Ikkinchi tomondan, tur tushunchasini ifodalovchi lugʻaviy birlik oʻrnida qoʻllanishi mumkin, bunda *daraxt* soʻzi giponim hisoblanadi. Lekin jins tushunchasini ifodalovchi soʻz giperonim, tur tushunchasining nomini ifodalovchi lugʻaviy birliklar giponimlar oʻrnida qoʻllanganda tur maʼnosini toʻliq va aniq tarzda ifodalamaydi. Shuning uchun ham ongimizda mavjud boʻlgan tur tushunchasining tilda aniq ifodalanishiga ehtiyoj seziladi.



WordNet modelidagi har bir soʻz turkumi oʻz munosabatlar toʻplamiga ega. Eʼtibor qiladigan boʻlsak, turli kompyuter ilovalarida, aksariyat hollarda, sinonimiya, antonimiya, giponimiya (giperonimiya), meronimiya (butun-boʻlak) munosabatlari oʻzaro oʻrnatilgan otlardan foydalaniladi. Ot sinsetlari oʻrtasidagi munosabatlarning asosiysi jins-tur aloqasi⁷⁸ boʻlib, tur sinset giponim, jins sinset esa giperonim deb ataladi. Bular tranzitiv iyerarxik munosabatlar boʻlib, ularga nisbatan *is A* munosabatlar nomi ham qoʻllaniladi. Agar ingliz tilida soʻzlashuvchilar “*An X is a (kind of) Y*” tipidagi jumalarni meʼriy holat deb hisoblashsa, X sinseti Y sinsetining giponimi boʻladi deyiladi⁷⁹. Bunday holda sinsetlar oʻrtasidagi munosabatlar iyerarxik tuzilmani hosil qiladi. Jins-tur munosabatlariga tayangan iyerarxik tizimlarni yaratishda, odatda, yuqori turuvchi tushunchalarning xususiyatlari quyidagilariga meros boʻlib oʻtadi deb hisoblanadi, yaʼni merosiylik xususiyati yuzaga keladi. WordNet modelidagi otlar ana shunday merosiylikka ega iyerarxik tizim shaklida tashkil qilingan: har bir sinset uchun uning oʻziga xos turdosh tushunchasi, uning giperonimini topish maqsadida tizimli bogʻlanishlar amalga oshirilgan.

Uchunchi bobning “*Leksik maʼlumotlar bazasida antonimik munosabat turlari*” deb nomlangan boʻlimida antonimlarning tushunchalarni ifodalashiga koʻra kontrar korrelyatlar, vektorli korrelyatlar, konversivlar, pragmatik antonimlar, enantiosemiya, enantiema; tuzilishiga koʻra turli asosli va bir asosli; til va nutq nuqtayi nazaridan lingvistik (uzual) va matniy (kontekstual, nutqiy, okkazional); harakat nuqtayi nazaridan mutanosib va nomutanosib; soʻzlarning shakl va maʼno munosabatiga koʻra turlari tadqiq etildi.

Dissertatsiyaning toʻrtinchi bobi “**UzNet lingvistik ontologiyasini yaratish tamoyillari**” deb nomlanadi. Unda UzNet lingvistik ontologiyasini yaratish konsepsiyasi loyihalashtirilgan, UzNet bazasida soʻz turkumlarini teglash masalasi

⁷⁸ Miller, G. Nouns in WordNet / G. Miller // WordNet – An Electronic Lexical Database. – The MIT Press, 1998. – P. 23-47.

⁷⁹ Miller, G. 1998.

yoritilgan, UzNet'da sinsetlar bazasi va sinonimayzer imkoniyati haqida batafsil soʻz yuritilgan.

“UzNet lingvistik ontologiyasini yaratishni loyihalash konsepsiyasi” boʻlimida xalqaro WordNet tizimida havolasi berilgan UzWordNet lingvistik resursi tahlil qilingan, uning ontologiya talablariga muvofiq emasligi asoslangan.

Tabiiy til tushunchalari, soʻzlar va ifodalar orasidagi munosabatlarni shakllantirishda tilning ijtimoiy tabiati bilan bogʻliq muammolar yuzaga keladi. Bundan tashqari, ontologiya tushunchalari oʻrtasidagi munosabatlarni aniq va muqim holda oʻrnatish ham tilning pragmatik jihati bilan bogʻliq qiyin masala hisoblanadi, ayniqsa, “leksik birlik – tushuncha (yoxud uning izohi)” aloqalarini tashkil qilishda ham muammolar yuzaga keladi. Bu masalada, tabiiy tildagi har bir leksik birlik (LB)ning toʻliq va mukammal izohlari shakllantirilishi talab qilinadi. Buning uchun, avvalo, (1) butun soʻz boyligi jamlanishi, (2) har bir LBning turkumi aniqlanishi va (3) oʻsha LBning aniq hamda ixcham izoh(lar)i darajalangan (yoxud asosiy tushunchani anglatishi jihatidan tartibli raqamlangan) tarzda shakllantirilishi darkor. Mazkur tamoyil umumiylik kasb etadi va aynan shu tamoyil asosida oʻzbek tili ontologiyasini shakllantirish kutilgan natijalarni beradi.

Tadqiqotimiz jarayonida lingvistik ontologiyani yaratish talablari, axborot qidiruv tezaurusini yaratish tamoyillari oʻrganilib chiqildi. Tezauruslarda bir sohadagi barcha leksik boyluk iyerarxiyasi yaratilsa, lingvistik ontologiyada butun til leksikasi tarmogʻi yuzaga kelishi zarur hisoblanadi. Shu bois tezauruslarni yaratish nisbatan muvaffaqiyatli va ontologiyaga nisbatan qisqa vaqtda amalga oshiriladi. Soʻzlar tarmogʻini noldan yaratish yoki avval mavjud boʻlganini kengaytirish esa bir necha bosqichlarni oʻz ichiga oluvchi va inson mehnatidan hamda avtomatlashtirilgan tizimlardan keng foydalanishni talab qiladigan mashaqqatli jarayondir. Shu bois Oʻzbek tili ontologiyasi – UzNet tizimini yaratish uchun konsepsiya ishlab chiqildi.

“OʻZBEK TILI ONTOLOGIYASINI YARATISH KONSEPSIYASI”

Tarkibi:

- 1. Lingvistik ontologiyani yaratishning konseptual asosi tadqiqi**
- 2. UzNet tizimi uchun leksikografik baza yaratish**
- 3. Soʻz turkumlarini teglash**
- 4. Soʻzlarning shakl va semantik munosabatini belgilash**
- 5. Sinsetlar (lugʻaviy, maʼnoviy va matniy sinonimlar)**

Har bir bosqich bir necha tarkibiy bosqichlarga ega.

Mazkur tadqiqotda konsepsiya birinchi bosqichining 1.1-, 1.2-, 1.3-qadamlari oʻz aksini topdi. Amaliy jihatdan 2.1-, 2.2-, 2.3-, 2.4-qadamlar leksikografik baza sifatida shakllantirildi. Lingvistik ontologiyani yaratish koʻpbosqichli murakkab va sinchkovlikni talab etadigan jarayon boʻlgani bois yana bir qancha tadqiqot ishlarini talab etadi. Yirik jamoa mehnati katta natijalarni beradi.

Oʻzbek tili ontologiyasi – UzNetning qoʻllanilish sohasi, uning tiplari, tili, ahamiyati, foydalanuvchilari quyidagicha: **I. Tili** – formallashtirilgan oʻzbek tili⁸⁰. **II.**

⁸⁰ Qarang: Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – 176 b.

Tipi: 1) umumlashtirish darajasi bo'yicha: a) umumiy ontologiya; b) soha ontologiyalari; c) ilovalar ontologiyalari; 2) ishlab chiqish chuqurligi bo'yicha: a) glossariy; b) tezaurus; c) taksonomiyalar jamlanmasi; e) mantiqiy cheklovlar to'plami. 3) predmet sohasi bo'yicha: a) rasmiy matnlar ontologiyasi; b) bilim sohalari ontologiyasi; c) sanoat sohasi ontologiyasi; e) moddiy va nomoddiy tushunchalar ontologiyasi. **III. Qo'llanilish holati:** 1) umumiy semantik tizim; 2) bilimlarni boshqarish tizimi; 3) ta'lim texnologiyasi; 4) ekspert tizimlari; 5) tabiiy tilni qayta ishlash vositasi; 6) ilovalar integratsiyasi; 7) modellashtirish va loyihalashtirish asosi. **IV. Ahamiyati:** 1) til o'rganish asosi; 2) sun'iy intellektning birlamchi manbasi; 3) bilimlarni aks ettirish manbasi; 4) turli sohalarni integratsiyasi. **V. Foydalanuvchisi:** 1) shaxs; 2) guruh; 3) tashkilot; 4) davlat.

WordNet tipidagi lingvistik resurlar tilshunoslik an'alariga muvofiq tilning lug'at boyligini tavsiflash uchun yaratilgan. Lekin har qanday axborot tizimi nafaqat umumiy lug'at, balki muayyan fan sohalari va ularning terminologiyalari bilan ham shug'ullanadi. WordNet asosida terminologik resurslarni yaratishga urinishlar tahlil qilinganda xulosa shunday bo'ldiki, WordNet strukturasi terminologiyalarni tavsiflash uchun moslashtirilmagan. So'z turkumlarining alohida tavsifi, bir-biriga bog'liq bo'lmagan ma'nolarning juda katta to'plami, ko'p tarkibli ifodalarni kiritish tamoyillarining yetarli darajada ishlab chiqilmaganligi – bularning barchasi WordNet modeli asosida yaratilgan terminologik resurslarni ishlab chiqish va ulardan foydalanishda muammolarga olib keladi. Shu boisdan har bir til tabiatiga asosan konsepsiya ishlab chiqish madsadga muvofiq sanaladi.

Mazkur bobning “*UzNet bazasida so'z turkumlarini teglash masalasi*” bo'limida sun'iy intellekt tizimini yaratishga bel bog'langan ayni damda tabiiy tilni qayta ishlashda so'zlarning turkumini aniqlash zarurligi asoslangan. Jahon kompyuter lingvistikasidan ushbu lingvo-texnik tahlilning so'zlar turkumini aniqlash – PoS-tagger, ya'ni so'z turkumlarini teglash, shuningdek, uning matnlarni avtomatik qayta ishlash jarayoni bosqichi ekanligi ma'lum. Til korpuslarini yaratish uchun boshlangan dastlabki teglash harakatlari bugungi kunga kelib, matn bilan bog'liq ko'plab dolzarb masalalar yechimini bermoqda. NLP, ya'ni tabiiy tilni qayta ishlash jarayonida ham so'z turkumlarini teglash birlamchi vazifa hisoblanib, buning natijasida omonimlikni, ko'p ma'noli so'zlar semantikasini aniqlash kabi turli lingvistik noaniqlar matn tarkibida tahlil qilinishiga erishiladi.

“O'zbek tilining izohli lug'ati”da *uzoq* so'ziga to'rtta izoh berilgan: 1) oraliq masofasi katta, nisbatan narida joylashgan; olis; 2) vaqt e'tibori bilan davomli katta, ko'p; 3) do'stlik, qarindoshlik, tanishlik va sh.k. jihatdan bog'lanish u qadar yaqin bo'lmagan, yaqin emas; 4) (ko'chma) bevosita aloqasi, munosabati yo'q; aloqasiz⁸¹. Badiiy asarlardagi ifoda bo'yog'i, kundalik turmushda reallashish vaziyatiga muvofiq *uzoq* so'zining qo'shimcha ma'nolari yuzaga keladi: 1) o'tgan davr (tarix); 2) kelajak haqida fikr yuritish, oqibatini o'ylab ish ko'rish (kelajak);

⁸¹ Ўзбек тилининг изохли луғати: 80 000 дан ортиқ сўз ва сўз бирикмаси (А. Мадвалиев таҳрири остида). 4-жилдли. – Тошкент: Ўзбекистон миллий энциклопедияси, 2006. – Б. 268.

3) miqdorni anglatish, ko'p; 4) uzunlik o'lchovi; 5) ishga yuzaki yondashish (yuzakilik); 6) davomli emas, tez fursatda bitadi (qisqa vaqt); 7) uzoqdagi joy.

Gap bo'laklari teglarini identifikatsiyalash bir muncha qiyin jarayon⁸². Sababi o'zbek tilidagi jamiki so'zlarni universal holda 12 turkum doirasida teglash imkoniyati yo'q. So'z uning jumla tarkibida reallashish holati va N-gramma⁸³ so'zlarning semantik valentligiga binoan polifunksional bo'lishi mumkin. Masalan: "Shifoxonaga bemorni keltirishdi" va "Shifoxonaga bemor odamni keltirishdi" jumllarining 1-sida *bemor* so'zi turkumlik belgisi (kim? so'rog'iga javob berayotgan tushum kelishigidagi so'z)ga ko'ra ot turkumi, 2-jumlada esa (qanday? so'rog'iga javob beryapti) sifat turkumi vazifasidagi so'z hisoblanadi. O'zbek tili izohli lug'ati⁸⁴ da mavjud 11 000 o'zlashma so'zlardan 66 ta xuddi shunday polifunksional so'zlar aniqlandi⁸⁵.

So'z turkumlarini teglash quyidagi jarayonlarda zarur sanaladi: 1) korpusda so'zshakllarni grammatik teglashda; 2) lingvistik ontologiyada so'zning izohi va turkumlararo munosabatlarni to'g'ri hamda to'liq shakllantirishda; 3) so'zning ko'p ma'noli va / yoki omonimligini aniqlashda; 4) gapni sintaktik tahlil qilishda ahamiyatli sanaladi. Eng muhimi, ST teglari tabiiy tilni qayta ishlash (Natural Language Processing / NLP) uchun eng birlamchi zaruriy lingvistik element hisoblanadi, shu bois STni teglash NLPda turli xil muammolarni soddalashtirish uchun dastlabki shart sifatida amalga oshiriladi.

O'zbek tili ontologiyasini yaratishda WordNet, EuroWordNet, BabelNet, KeNet, RuTez, RussNet tizimlari imkoniyatlari hamda biznes-jarayoni tadqiq etilib, quyidagi **to'rt tamoyilni** belgilab oldik:

1) keng qamrovlilik: o'zbek tilidagi mavjud leksik birliklar qamrab olinadi. Buning uchun barcha sohalardagi leksikografik manbalarga tayanish maqsadga muvofiq. Shuningdek, qamrov imkoniyatini oshirish uchun o'zbek tili korpuslaridagi matnlarga asoslanilsa, maqsadga muvofiq. Natijada tildagi sinkretik shakllar, kvazisinonimlar, enantiosemlar, polifunksional so'zlar va polisemantik so'zlarni aniqlash imkoni oshadi, pragmatik tahlil natijasida o'zbek tili ontologiyasining lingvomadaniy ma'lumotlarni qamrab olish xususiyati yuzaga keladi.

2) formal til aspekti: tizimga qamrab olingan leksik birliklarning har biri o'z turkumiga ega bo'ladi, so'z turkumlari teglari aniq belgilanadi, lemmatizatsiya jarayoni⁸⁶ amalga oshiriladi, tildagi istisnoli holatlar bartaraf etilib o'zbek tilining formal shakli yaratiladi.

3) muvofiqlik: lingvistik ontologiyaga qo'yilgan talab asosida mutloq sinonimlarda ularning izohlari muvofiqligi belgilanadi. Kvazisinonimlar va matniy

⁸² https://en.wikipedia.org/wiki/Hidden_Markov_model

⁸³ Qarang: Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – B. 73-77.

⁸⁴ O'zbek tilining izohli lug'ati: 80 000 dan ortiq so'z va so'z birikmasi (A. Madvaliyev tahriri ostida). 5 jildli. – Toshkent: O'zbekiston milliy ensiklopediyasi, 2006.

⁸⁵ Qarang: Qurbonova M., Abjalova M. va boshq. O'zbek tili o'zlashma so'zlarining urg'uli lug'ati. [Matn]: o'quv-uslubiy lug'at / M.Qurbonova, M.Abjalova, N.Axmedova, R.To'laboyeva. – Toshkent: Nodirabegim, 2021. – 988 b.

⁸⁶ Qarang: Abjalova M. Tahriri va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A.Abjalova. – Toshkent: Nodirabegim, 2020. – B. 25.

sinonimlar esa muayyan mutloq sinonim sinsetlariga kiritilgansagina shun sinset izohiga aynan keladi.

semantik simmetriya: LO va tezauruslarni boshqa lingvistik tizim va leksikografik manbalardan farqlovchi asosiy xususiyat va ularning qimmatini oshiruvchi omil semantik munosabatlar sanaladi. Shu bois lingvistik ontologiyada har bir sinset tarkibida jins-tur, butun-bo‘lak, antonimiya, omonimiya kabi munosabatlarning o‘rnatilishi leksik birliklar o‘rtasida bog‘lanishni ta‘minlaydi. Natijada tizimda yirik tarmoqlanish vujudga kelib, axborot-qidiruv imkoniyatini oshiradi.

Ma‘lumki, o‘zbek tilida 12 so‘z turkumi (mustaqil so‘z turkumlari: ot, fe‘l, sifat, ravish, son, olmosh; yordamchi so‘z turkumlari: bog‘lovchi, ko‘makchi, yuklama; alohida olingan so‘zlar turkumi: modal, taqlid, undov)ga so‘z yasovchi qo‘shimchalarning qo‘shilishi natijasida 4 so‘z turkumi: ot, fe‘l, sifat, ravish yasaladi. Aniqlangan yasovchi qo‘shimchalar (337 ta: ot yasovchi qo‘shimcha 114 ta, fe‘l yasovchi 58 ta, sifat yasovchi 117 ta, ravish yasovchi qo‘shimcha 48 ta)⁸⁷ sirasida **-an** ravish yasovchi affiks hisoblanadi. Ushbu parametrdan kelib chiqib xulosalash mumkinki, ot turkumiga mansub “*shaxs*” so‘ziga **-an** yasovchi qo‘shimchasi birikishi natijasida yasama ravish hosil qilingan: *shaxs (Ot) ∪ {-an}* => *shaxsan*.

Aksariyat hollarda so‘z turkumlarini teglashda quyidagi usul (metod, algoritmlarga asoslaniladi⁸⁸: 1) qoidalarga asoslangan usul; 2) stoxastik (yoxud statistik) usul.

Qoidalarga asoslangan PoS teglar. Eng azaliy teglash usullaridan biri bu qoidalarga asoslangan POS-teglash sanaladi. Bunda, asosan, Brill usuli qo‘l keladi⁸⁹. Qoidalarga asoslangan teggerlar har bir so‘zni teglash uchun lug‘at yoxud leksikadan foydalanadilar. Agar so‘zda (polifunksional, omonim, ko‘p ma‘noli so‘zlar nazarda tutilmoqda) bir nechta teglar bo‘lsa, unda qoidalarga asoslangan teggerlar gapdagi so‘zning turkumlik tegini to‘g‘ri aniqlash uchun qo‘lda yozilgan qoidalardan foydalanadi. Yanada aniq teglarni berishda so‘zning lingvistik xususiyatlarini undan oldingi va keyingi so‘zlarni tahlil qilish orqali qoidalarga asoslanib belgilash orqali ham amalga oshirilishi mumkin. Masalan, qaratqich kelishigidagi ismga mansub so‘zdan so‘ng kelgan lingvistik birlik egalik qo‘shimchasini olgan ot turkumidagi so‘z hisoblanadi. Masalan, *mening kitobim, akamning uyi, Salimaning ko‘ylagi* kabi. Demak, bunday holda so‘zning ot turkumligi o‘zining oldida kelayotgan qaratqich kelishigidagi ism orqali belgilanadi. Ingliz tilidan misolga e‘tibor qaratamiz: oldingi so‘z artikl bo‘lsa, u holda undan keyin kelayotgan so‘z ot turkumiga oid leksik birlik sanaladi. Masalan, *an egg, a book, the train, the windows* kabi.

PoS teglaridagi bunday holatlar qoidalar shaklida kodlanadi. Ushbu qoidalar quyidagilarni tashkil etishi mumkin: 1. *Lingvistik me‘yorlarga asoslangan qoidalar*. Tilning orfografik qoidalarga asoslangan yuzlab qoidalar umumiy,

⁸⁷ Qarang: Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – B. 122-123.

⁸⁸ <https://www.freecodecamp.org/news/an-introduction-to-part-of-speech-tagging-and-the-hidden-markov-model-953d45338f24/>; <https://coderlessons.com/tutorials/akademicheskii/obrabotka-estestvennogo-iazyka/pometka-chasti-rechi-pos>; <https://habr.com/ru/post/125988/>

⁸⁹ Brill E. 1992. A simple rule-based part of speech tagger //Proceedings of ANLC. – P. 154.

xususi va istisnoli qoidalar bazasi tarzida shakllantiriladi⁹⁰. 2. *Kontekstli shablon qoidalari*, ya'ni gap tarkibidagi ko'chma ma'noga ega so'zning muntazam ravishda konnotativ ma'noda qo'llanishi dastur xotirasida saqlanadi, natijada keyingi jarayonlarda o'sha konnotativ so'z bilan bog'liq noaniqliklar bartaraf etiladi.

Stoxastik teglash usuli chastota yoki ehtimollik (statistika)ka asoslanadi. Shu bois ayrim manbalarda statistik yoxud ehtimolikka asoslangan usul tarzida tushuntiriladi⁹¹. *Chastotali yondashuv*da stoxastik tegetlar so'zning matnda ma'lum bir teg bilan uchrashi ehtimoli asosida grammatik noaniqliklarni bartaraf etadi. Teglarining ketma-ketligi ehtimoli yoxud n-gramma usuli teget berilgan teglar ketma-ketligining qo'llanilish ehtimolini hisoblaydi.

O'zbek tili ontologiyasi leksik birliklari bazasida 84094 til birligi qamrab olingan bo'lib, shundan 47817 ot, 17081 fe'l, 14727 sifat, 2644 ravish, 276 son, 240 olmosh, 65 bog'lovchi, 111 ko'makchi, 18 yuklama, 115 modal, 174 undov va 822 taqlid so'zlar qoidalarga asoslangan usul asosida teglangan.

To'rtinchi bobning "*UzNet'da sinsetlar bazasi va sinonimayzer imkoniyati*" nomli sinonimlar turlari, ularning tadqiqi masalasi, o'zbek tili ta'limiy korpusida sinonimayzer (sinonimzator) dasturining imkoniyatlari ko'rsatilgan va asosiysi, kvazisinonimlar tadqiq etilgan.

Princeton WordNet'ni yaratish tamoyili bo'yicha gapdagi barcha ma'nodoshlar leksik ma'lumotlar bazasida o'z aksini topishi zarur hisoblanadi. Buning uchun, albatta, matnda sinonimlik ehtimoli bo'lgan tushunchalar o'rtasidagi munosabat ham tavsiflanishi talab qilinadi. Ana shunday hodisalardan biri nutq jarayonida voqelanadigan va ayrim hollarda, lug'at tarkibida ham ma'nodoshlar sifatida qayd etish mumkin bo'lgan kvazisinonimlar bugungi kunda o'z tadqiqini kutayotgan dolzarb masalalardan hisoblanadi.

"**kvazi-**" termin elementi lotin tiliga mansub bo'lib, *xayoliy, haqqoniy emas, soxta, xayoliy* degan tushunchalarni anglatadi. Kvázisinónimlar (soxta sinonimlar, qisman ma'nodoshlar⁹²) ma'no jihatidan nisbatan yaqin, ammo hamma vaziyatlarda ham bir-birining o'rnini bosa olmaydigan so'zlar hisoblanadi.

Kvazisinonimlarning quyidagi turlari farqlanadi⁹³: 1) bir-biriga o'xshash ma'noga ega bo'lgan kvazisinonim atamalar, masalan: *uy – bino, iste'dod – daholik*; 2) bir so'zning ma'nosi va ko'lami boshqa sohaga ham tegishli bo'lgan qisman sinonimlar, masalan: *metall – temir*; 3) qarama-qarshi semantikaga ega so'zlarning nutqiy voqelanish jarayonida ma'nodoshlar tarzida qo'llaniladigan sinonim so'zlar (bitta hodisaning miqdoriy xususiyatlari kabi), masalan: *qattiqlik – yumshoqlik, shaffoflik – qorong'ulik*.

⁹⁰ Qarang: Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – 176 b.

⁹¹ <https://www.freecodecamp.org/news/an-introduction-to-part-of-speech-tagging-and-the-hidden-markov-model-953d45338f24/>; <https://habr.com/ru/post/125988/>; https://ru.wikipedia.org/wiki/Частеречная_разметка; https://en.wikipedia.org/wiki/Part-of-speech_tagging#:~:text=In%20corpus%20linguistics%2C%20part%2Dof,its%20definition%20and%20its%20context.

⁹² <https://ru.wiktionary.org/wiki/quasi>

⁹³ [Квazисинонимы – Википедия \(wikipedia.org\)](https://ru.wikipedia.org/wiki/Квazисинонимы)

Ma'lumki lug'aviy sinonimlar, asosan, bir so'z turkumiga mansub yaqin ma'noli leksik birliklar sanaladi. Kvazisinonimlar esa turli so'z turkumlariga oid so'zlarning ma'nodosh bo'lishi bilan ham yuzaga keladi. Masalan: *lemmatizatsiya* (so'zning asos morfemasiga qadar tahlil qilinishi; asosning aniqlanishi) – (*N–ot*) ↔ *lemmatizatsiyalash* – (*V_h – fe'l_{harakat nomi}*); *regeneratsiya* (*N–ot*) ↔ *qayta tiklanish* (*V_h – fe'l_{harakat nomi}*); *dahshat* (*N – ot*) ↔ *zo'r* (*Adj – sifat*) (og'zaki so'zlashuv uslubida). Xulosa qilib aytganda, nutqiy voqelanish jarayonidagi kvazisinonimlar (qisman sinonimlar) ma'no jihatidan yaqin bo'lsa-da, turkumga xoslanishi turlicha bo'ladi. Kvazisinonimlarning semantik munosabatida birikish yoxud kesishish semalari mavjud bo'ladi.

To'rtinchi bobning "*Lingvistik ontologiya uchun lug'atlar ahamiyati va uning leksikografik bazasi*" deb nomlangan bo'limida til ontologiyasi uchun ensiklopedik va filologik lug'atlar ahamiyati, o'zbek tili ontologiyasi uchun zarur leksikografik bazalar va UzNetda qidiruv natijasida foydalanuvchiga taqdim etiladigan ma'lumotlar yoritildi. Mazkur tadqiqot natijasida 11000 o'zlashma so'z, 1100 sinonimik qator, 1836 omonim, 774 antonim juftlik, 561 paronim juftlik bazasi va sinonimayzer dasturi yaratildi.

XULOSA

1. Lingvistik ontologiyaning yaratilishi tabiiy axborotni qayta ishlash sohasidagi zamonaviy tadqiqotlarning istiqbolli yo'nalishi bo'lib, tabiiy til imkoniyatini o'zida aks ettirgan tizim hisoblanadi. Lingvistik ontologiyaning shakllanishi bilan avtomatlashtirilgan tizimlar yordamida bir qator muammolar muvaffaqiyatli ravishda o'z yechimini topa boshladi. Shu bois hozirgi vaqtda ma'lum bir algoritim bo'yicha ishlab chiqilgan ko'plab lingvistik ontologiyalardan muvaffaqiyatli foydalanadigan ilovalar safi va sifati kengayib bormoqda.

2. Ontologiya – til va dunyo bilimlariga asoslangan, sohalararo terminlarni qamrab olgan va ular o'rtasidagi munosabatlar asosida shakllantirilgan tarmoqli leksik ma'lumotlar bazasi. Leksik ma'lumotlar bazasidagi o'zaro munosabatlar tarkibida sinonimik qatorlar birlamchi sanaladi. Qidiruv tizimlari uchun esa giponimiya va xolonimiya ham muhim ahamiyat kasb etadi. Tezauruslar va ontologiyalar tabiiy tilni qayta ishlashda ham zarur manbalardan hisoblanadi.

3. *Lingvistik ontologiya* yoxud *til ontologiyasi* terminlari tilshunoslikka qaraganda, axborot texnologiyalari sohasida ko'proq qo'llaniladi, asosan, matnlarni avtomatik tarzda qayta ishlash uchun ixtisoslashtirilgan axborot qidirish tezaurusi, ya'ni tilning lug'at boyligini o'zida jamlagan, so'zlarning semantik munosabatlari o'rnatilgan (yoxud so'zlar tarmog'iga ega) turli maxsus lingvistik dasturiy ta'minotlarni anglatadi. Lingvistik ontologiyalar til yoki predmet sohasiga oid so'zlarning ko'p qismini qamrab oladi, shu bilan birga tushunchalar o'rtasida munosabat mavjud bo'ladigan ontologik tuzilma hisoblanadi. Shuning uchun lingvistik ontologiyalarni leksik ma'lumotlar bazasining maxsus turi va ontologiyaning alohida tipi sifatida ko'rish mumkin.

4. Formal jihatdan ontologiya – bu tushunchalar to'plami va tushunchalar haqidagi tasdiqlar to'plamidan iborat tizim bo'lib, ular asosida sinflar, obyektlar, munosabatlar, funksiyalar va nazariyalar qurilishi mumkin.

5. Tezauruslar matnlar to‘plamidagi yoxud til korpuslaridagi tushunchalarni, maxsus bilim sohasi yoki faoliyat sohasining tushunchalari, ta’riflari va terminlarini qamrab oluvchi leksikografik manbalar hisoblanadi. Lingvistik ontologiya butun til imkoniyatini qamrab olsa, tezauruslar muayyan to‘plam yoxud yo‘nalish, sohaga oid tushunchalar munosabati bilan cheklanadi.

6. Lingvistik ontologiyaning rivojlanishi ulardan yirik dasturiy ta’minot tizimlari uchun bilimlar bazasi komponentlarining qurilish bloklari sifatida foydalanish, shuningdek, obyektga yo‘naltirilgan tizimlardagi obyekt diagrammalari, ma’lumotlar bazasining konseptual sxemalarini qurish bloklari sifatida, matnlarni avtomatik-semantik va pragmatik tahlil qilish, tabiiy tilni qayta ishlash, sun’iy intellekt tizimini yanada rivojlantirish, robototexnika sohasida imkoniyatlarni oshirishda foydalanish mumkin bo‘ladi. Ontologiyalar yordamida turli intellektual, xususan, ekspert tizimlaridagi muammolarni yechish uchun bilim bazalarini shakllantirish mumkin.

7. Tilning lug‘at boyligini tizim sifatida o‘rganishda leksemalarning gipogiperonimik munosabatlariga tayanish muhim ahamiyatga ega: u tabiat va jamiyatdagi narsa-predmetlar, voqea-hodisalarning tildagi nomlari bo‘lgan leksemalarning ma’nolarini va shu ma’nolar orqali borliqdagi narsa-hodisalarning o‘zlari haqidagi tushuncha-tasavvurlarni umumlashtirish va farqlash imkonini beradi. Giponimiya lug‘at boyligidagi leksemalarning pog‘onali (iyerarxik) aloqasidan kelib chiqadigan ma’no munosabatlaridir. Bunday munosabatlarning mohiyati shundaki, torroq tushuncha yoki ma’no ifodalaydigan leksemalar kengroq tushuncha yoki ma’no ifodalaydigan leksemalar bilan tur (giponim) va jins (giperonim) aloqasida bo‘ladi, bunday aloqa birlashtiruvchi (integral) semalar orqali amalga oshiriladi.

8. Informatsion-qidiruv va ontologik tizimlarda aniq ma’lumot olish uchun qat’iy iyerarxik munosabatning o‘rnatilgani maqbul. Biz faqat matniy butun-qism munosabatini “qisman butun-bo‘lak munosabati”ga kirita olamiz.

9. Ontologiya konseptini aniqlashda taklif qilingan yondashuv (ta’rif)lar ontologiyalar va ularning xususiyatlarini muhokama qilish uchun qulay bo‘lgan terminlarning yagona tizimini yaratish va ontologiyaning vazifasi, tuzilishi, munosabatlari va ahamiyatini aks ettiruvchi yagona ta’rifni berishga yordam beradi. Demak, Ontologiya – til va dunyo bilimlariga asoslangan, sohalararo terminlarni qamrab olgan va ular o‘rtasidagi munosabatlar asosida shakllantirilgan tarmoqli leksik ma’lumotlar bazasi.

10. Tezaurusning tuzilishi, uning birliklari va munosabatlarini tavsiflash tamoyillari axborot izlash muammolarini hal qilish sifatini oshirishda muhim ahamiyatga ega. RuTez tezaurusi (tarkiban ontologiya) WordNet tipidagi ontologiya (ayrim manbalarda tezaurus)dan ham, an’anaviy ma’lumot qidirish tezaurusidan ham bir qator muhim rivojlanish tamoyillari bilan farq qiladi. RuTez`da bilimlar va modellarning sifatli muvofiqlashuvi mazkur ontologiya sifatini 10-15 %ga yaxshilaydi, RuTez`da to‘plangan bilimlar bir qator muammolar yechimini topish statistik va mashinada o‘qitish usullarini qo‘llashdan ko‘ra muammoni tezroq va yaxshiroq hal qilish imkonini beradi.

11. Iyerarxiya umumiy ko‘rinishida butun tuzilma bo‘lib, unda elementlar eng yuqoridan pastgacha, umumiydan xususiya, asosiydan ikkinchi darajaga, jinsdan turga tegishlilik munosabatlari bilan o‘zaro bog‘lanadi. Iyerarxiyani qurishda, asosan, bosqichli tamoyilga asoslaniladi, elementlarning qiymatiga muvofiq parallel darajalarda taqsimlanishi nazarda tutiladi. Lingvistik ontologiya sinsetlari aynan shunday iyerarxiyaga asoslanadi va o‘z navbatida, tushunchalar tarmoqlanib bir-biriga bog‘lanadi.

12. Sinsetlar – lingvistik ontologiyaning birlamchi elementlari, sinonimlar to‘plami hisoblanadi. Sinsetlardagi semantik munosabatlar tushunchalarning tarmoq hosil qilishiga olib keladi, so‘z turkumlariaro sinonimiyaning mavjud bo‘lishi kvazisinonimlar deb nomlangan turni hosil qilgan. Bunday tarmoqlanish antiplagiat dasturlari, ekspert tizimlari, nutqni tahlil qilish dasturlari, semantik tahlil dasturlari, axborot-qidiruv tizimlari uchun muhim ahamiyat kasb etadi.

13. Tezauruslarda bir sohadagi barcha leksik boylik iyerarxiyasi yaratilsa, lingvistik ontologiyada butun til leksikasi tarmog‘i yuzaga kelishi zarur hisoblanadi. Shu bois tezauruslarni yaratish nisbatan muvaffaqiyatli va ontologiyaga nisbatan qisqa vaqtda amalga oshiriladi. So‘zlar tarmog‘ini noldan yaratish yoki avval mavjud bo‘lganini kengaytirish esa bir necha bosqichlarni o‘z ichiga oluvchi va inson mehnatidan hamda avtomatlashtirilgan tizimlardan keng foydalanishni talab qiladigan mashaqqatli jarayondir.

14. O‘zbek tili ontologiyasi asosida Internetda samarali qidiruvni tashkil etish, qidirilayotgan lingvistik obyekt bilan bog‘liq uning qism, butun, tur, juns yohud ma’nodoshlari, zid ma’nolari orqali zarur axborotlarni taqdim etish kompyuter lingvistikasining ustuvor vazifalaridan hisoblanadi.

15. Lingvistik ontologiyalar, tezauruslar, axborot-qidiruv tezauruslari, WordNet tipidagi lingvistik resurlarni yaratish texnologiyalarini tadqiq etish jarayonida, ayniqsa, til ontologiyasini yaratish o‘ta mashaqqatli va og‘ir ish jarayoni ekanligi ma’lum bo‘ldi. Shu bois “O‘zbek tili ontologiyasi – UzNet tizimini yaratish konsepsiyasi” asosida tizimli ravishda ish olib borish natijaning muvaffaqiyatli bo‘lishini ta’minlaydi. Mazkur tadqiqotda konsepsiya birinchi bosqichining 1.1-, 1.2-, 1.3-qadamlari o‘z aksini topdi. Amaliy jihatdan 2.1-, 2.2-, 2.3-, 2.4-qadamlar leksikografik baza sifatida shakllantirildi. Lingvistik ontologiyani yaratish ko‘p bosqichli murakkab va sinchkovlikni talab etadigan jarayon bo‘lgani bois yana bir qancha tadqiqot ishlarini talab etadi. Yirik jamoa mehnati katta natijalarni beradi.

16. Axborot-qidiruv, tabiiy tilni qayta ishlash, mashina tarjimasini, sun‘iy intellekt uchun o‘zbek tilini formallashtirish, o‘zbek tilini Internet tiliga aylantirish maqsadida o‘zbek tili ontologiyasini yaratish muhim sanaladi. Bunda WordNet tayanch baza bo‘la olmaydi. Sababi WordNet resurslarini yaratishda ko‘p bosqichlarning avtomatik bajarilishi katta muammolarni yuzaga keltiradi. Chunki bir tabiiy til (ingliz tili) xususiyatlari boshqa bir tabiiy til (jumladan, o‘zbek tili) xususiyatlari bilan mos emas. Shuning uchun UzNetni yaratishda dastlabki bosqichlar qo‘lda bajarilishiga to‘g‘ri keladi. Bu jarayonda qoidalarga asoslangan

metodga tayaniladi, aksariyat ma'lumotlar qayta ishlangandan so'ng stoxastik metod asosida baza boyitiladi.

17. Lingvistik ontologiya uchun so'z turkumlarining teglanishi muhim ahamiyat kasb etadi. Aynan uning ish jarayoni ham grafematik tahlildan so'ng, morfo-tahlilni amalga oshirish bilan davom etadi. Ushbu jarayonda so'zlarning bir ma'noli yoki ko'p ma'noliligi, omonimligi yoxud polifunksionalliligi aniqlanadi. So'zning turkumi va namoyon bo'lish holati aniqlangandan so'ng uning semantik munosabatlari o'rnatiladi. Aynan ushbu tamoyillar aniqligiga erishish uchun ham har bir so'zning turkumi to'g'ri teglanishi zarur sanaladi.

18. Lingvistik ontologiyalar mashina tarjimasi, savol-javob tizimlari, ma'lumot qidirish, bilimlarni olish tizimlari, kompyuter va shaxs o'rtasidagi muloqotni o'tkazish tizimlari, tilni tushunish tizimlari, shuningdek, bilimlarni namoyish qilish, sun'iy intellekt va kompyuter ma'lumotlarini qayta ishlash bilan bog'liq ko'plab muammolarni hal qilishda muhim ahamiyatga ega. Xususan, tilshunoslik sohasida ontologiyalardan matn korpusining semantik annotatsiyasi, mashina tarjimasi, ko'p ma'nolilikni avtomatik hal qilish va kontekstga asoslangan omonimiyani aniqlash, quyi darajadagi ontologik turdagi resurs, lug'at va tezauruslarni yaratishda foydalaniladi. Bundan tashqari, ko'p tilli ontologiyalardan tarjimonlar tomonidan bilimlar va tegishli lug'atni o'z ichiga olgan ma'lumot manbalari sifatida foydalaniladi.

**SCIENTIFIC COUNCIL NO. DSc.03/30.12.2019.Fil.05.02
GRANTING ACADEMIC DEGREES
AT THE FERGANA STATE UNIVERSITY**

**TASHKENT STATE UNIVERSITY OF THE UZBEK LANGUAGE AND
LITERATURE NAMED AFTER ALISHER NAVOI**

ABJALOVA MANZURA ABDURASHETOVNA

PRINCIPLES OF CREATION OF THE UZBEK LANGUAGE ONTOLOGY

10.00.11 - Language Theory. Applied and Computational Linguistics

**DOCTOR OF PHILOLOGICAL SCIENCES (DSc)
DISSERTATION ABSTRACT**

Tashkent - 2022

The topic of the Doctor of Science (DSc) dissertation is registered in the Higher Attestation Commission under the Cabinet of Ministers of the Republic of Uzbekistan under number B2021.4DSc/Fil364.

The dissertation was completed at the Tashkent State University of the Uzbek Language and Literature named after Alisher Navoi.

The abstract of the dissertation is available in three languages (Uzbek, English, Russian (resume)) on the website of Fergana State University (www.fdu.uz) and on the Information and Education Portal "ZiyoNet" (www.ziynet.uz).

Scientific adviser:	Dadaboyev Hamidulla Aripovich doctor of philological sciences, professor
Official opponents:	Khoshimov Ganijon Mirzaaxmetovich doctor of philological sciences, professor Urinbayeva Dilbar Bazarovna doctor of philological sciences, professor Eshref Adali doctor of technical sciences, professor
Leading organization:	Institute of Uzbek Language, Literature and Folklore

The defense of the dissertation will be held at the meeting of the Academic Council of DSc.03 / 30.12.2019. on april "5", 2022 (Address: 100100, 100151, Fergana, Murabbiylar Street, 19. Tel. (99873)244-66-02; fax: (99873) 244-44-01; e-mail: info@fdu.uz).

The dissertation is available at the Information Resource Center of Fergana State University (registered under number 179) (Address: 100100, 100151, Fergana, Murabbiylar Street, 19. Tel: (99873) 244-71-28).

The dissertation abstract was distributed on mart "19", 2022.

(Digital Registration Protocol of 2022 "19" mart)



Sh. Iskandarova
Chairman of the Academic Council
awarding academic degrees,
Doc. Ph. Sc., Professor

M. Zokirov
Academic Secretary of the Academic
Council awarding academic degrees,
Can. Ph. Sc., Associate Professor

A. Mamajonov
Chairman of the Scientific seminar
under the Academic Council
awarding academic degrees,
Doc. Ph. Sc., professor

Introduction (Annotations of doctoral dissertations (DSc))

Relevance of dissertation topic and necessity of it. It should be pointed out that in order to improve modern branches of human activity which requires knowledge it is important to increase of the role of computer technologies worldwide. At present, because the information flow is rising considerably, now there is necessity of protection, officialization and arrangement of information in a sequence and seeking new ways of processing it automatically as well. Due to this, there is an increasing interest in a wide range of information bases which can be used in practical terms. Especially, the demand is really huge for systems based on neural-nets that takes out any kind of information from the text without human interaction. The half of 20th century saw the emergence of the semantic webs together with global branch and it has been provided with extra tags carrying information about element semantics of hypertext pages. Inseparable part of semantic web is the concept of ontology which is lexical database consisting of a branch of words.

Linguistic ontologies which have been created to improve information seeking opportunities, to develop automatic translation systems, to obtain component analysis of texts and to represent language vocabulary in world practical linguistics and computer linguistics is different from thesaurus from the point of mutual branching of semantic relations in terms of the language. As a result, the creation of many other language resources on the basis of WorldNet English language database has spread out. That is why, a special attention has been paid in reorganizing natural language (NLP), modeling the language linguistically, tagging parts of speech and forming the collection of semantic relations in the language (synonymy, meronymy, hypernymy, antonymy).

Nowadays not only a great deal of research is being carried out, practical projects are also being done to formalize the language for artificial intelligence, to develop linguistic models in Uzbek linguistics and computer linguistics, to achieve practical results in cyberlexicography, in particular, “to provide active integration of the state language to information technologies and communications, to be more accurate”⁹⁴, to popularize Uzbek language in cyber global information network, to enable to have its special role in it⁹⁵, to develop grammar and semantic analysis systems and to do so it is required to process natural language in Uzbek language.

As a result, it is clear that a number of directions are being improved, such as, process of texts in practical linguistics and computer linguistics, development of speech synthesizer, speech acknowledgement, machine learning, natural language processing (NLP), computer translation, field of corpus linguistics, computer lexicography and linguo-didactics.

Using world experience, now it is vital to develop UzNet ontological system that not only bases on semantic relations and includes branches of words, but also accumulates all Uzbek language vocabulary in itself. This increases international

⁹⁴Decree of the President of the Republic of Uzbekistan No. PF-6048 of October 20, 2020 “On measures to further develop the Uzbek language and improve language policy in our country”.

⁹⁵This Decree.

status of Uzbek language and expands our national language opportunities. It is also important to develop translator programs and systems and to increase information seeking opportunities of the Uzbek language in global network. Therefore, the matter of broadening scientific research is waiting to be tackled.

This dissertation, to some extent, assists in implementing decrees like decree of the President of the Republic of Uzbekistan of May 13, 2016 No. PF-4997 “On measures to establish the Tashkent State University of Uzbek Language and Literature named after Alisher Navoi (TSUULL)”, of February 07, 2017 No. PF-4947 “About the Strategy of Actions for further development of the Republic of Uzbekistan”, of October 21, 2019 No. PF-5850 “About measures to completely increase prestige and position of the Uzbek language as a state language”, tasks indicated in important direction of providing active integration of the state language to modern information technologies and communications noted in “Conception of further improvement of the Uzbek Language and enhancement of language policy in 2020-2030” which has been confirmed with the decree of October 20, 2020 No. PF-6084 “On measures for further improvement of the Uzbek language in our country and perfecting language policy”, a resolution of February 17, 2021 No. PQ-4996 “On measures to create conditions for a rapid application of artificial intelligence technologies” and the decree of the Cabinet of Ministers of the Republic of Uzbekistan of December 12, 2019 No. 984 “About improvement of the Regulation on the Department for the Development of the State Language”, a resolution of January 21, 2020 No. 40 “On measures to establish activities of the Terminology Commission under the Cabinet of Ministers of the Republic of Uzbekistan” and other responsibilities set out in relevant judicial documents.

The relevance of the research to the priorities of the development of science and technologies in the Republic. This research was carried out in accordance with the first and foremost direction of the Development of Science and Technologies of the Republic, namely “Developing an information society and a democratic state from a social, legal, economic, cultural, spiritual and educational point, developing an innovative economy”.

Review of foreign scientific researches on the dissertation topic⁹⁶

Scientific examinations are being carried out on the study and creation of the supply of WordNet-based linguistic resources in the fields of global applied linguistics and computer linguistics and information technologies in leading research centers and higher education institutions, including Princeton University (PU), St. Louis University (SLU, USA), University of South Africa (USA), South African Center for Digital Language Resources (SACDLR, Pretoria, South Africa), University of Vlora (UV, Albania), National Center for Electronics and Computer Technologies (NCECT, Japan), Laboratory of Computational Linguistics of

⁹⁶Review of foreign scientific research on the topic of the dissertation has been performed on the basis of www.princeton.edu, <https://aclanthology.org/W98-0709.pdf>, <https://pythonprogramming.net/wordnet-nltk-tutorial/>, <http://wordnet.ru/>, <http://www.globalwordnet.org>, <https://scholar.google.com>, <https://www.researchgate.net>, <https://www.aclweb.org/anthology/L16-1207.pdf>, <https://acikerisim.isikun.edu.tr/>, <http://uzschoolcorpara.uz>, www.navoiy-uni.uz, <http://uzschoolcorpara.uz> and other sources.

Thailand (LCLT, Thailand), University of the Basque Country (UBC), Technical University of Catalonia (TUC, Catalonia), Indian Institute of Statistics (IIS), Indian Institute of Technology (IIT, India), Bulgarian Language Institute (BLI), Bulgarian Academy of Sciences (BAS, Bulgaria), Patras University (PU, Greece), Laboratory of Ontology (LO), Language Processing and Electronic Humanitarian Sciences (LPEHS), National University of Taiwan (NUT), Republic of China-Taipei (RChT, Taiwan), University of Zagreb (UZ, Croatia), Charles University (ChU), Institute of Official and Applied Linguistics (IOAL, Czech Republic), Center for Sprog technology (CS), University of Copenhagen (UC), Danish Language and Literature Society (DLS, Denmark), University of Vrije (UV, Netherlands), University of Amsterdam (UA, Netherlands), University of Sheffield (USh, England), University of Tartu (UT, Estonia), University of Helsinki (UH, Finland), University of Paris Diderot (UPD), University of Avignon (UA), Memo Data Computer Laboratory (MDCL, France), University of Tübingen (UT, Germany), University of Haifa (UH, Israel), (Mumbai, India), University of Szeged (US, Hungary), Institute of Computer Linguistics (ICL), Bruno Kessler Foundation (BKF), Center for Communication & Information Technologies (CCIT), Human Language Technology Group (HLTG), University of Verona (UV, Italy), Japan National Institute of Information and Communication Technologies (JNIICT, Japan), Busan National University (BNU) (South Korea), Kurdistan University (KU, Kurdistan), Institute of Mathematics and Informatics of Latvian University (IMILU, Latvia), Cyrillic and Methodius University (CMU), Staffordshire University (SU, England), University of Malta (UM, Malta), Institute of Mathematics of the Moldovan Academy of Sciences and Kathmandu University (IMMASKU, Moldova), Bergen University (BU, Norway), Shahid Beheshti University (ShBU), Iran Telecommunication Research Center (ITRC, Tehran, Iran), University of Tehran (UT), NLP Laboratory (NLPL, Tehran, Iran), Wroclaw University of Science and Technology (WUST, Poland), Adam Mickiewicz University (AMU, Poznan, Poland), Coimbra University (CU, Portugal), Getulio Vargas University (GVU), School of Applied Mathematics (SAM, Rio de Janeiro, Brazil), University of Lisbon Linguistics Center (ULLC, Portugal), Alexandru Ioan Cuza University (AICU), Romanian Academy (RA), Bucharest Institute of Artificial Intelligence (BIAI), Artificial Intelligence Institute (AII), Romanian Academy (RA, Romania), St. Petersburg University (SPU), Moscow State University (MSU, Russia), Faculty of Mathematics (FM), University of Belgrade (UB, Serbia), University of Moratuwa (UM, Moratuwa, Sri Lanka), University of Ljubljana (UL), Josef Stefan Institute (JSI, Slovenia), Gothenburg Sweden University (GSU, Sweden), Center for Turkish Language and Speech Processing (CTLSP, Turkey), INHA University, TSUULL, National University of Uzbekistan (NUUz, Uzbekistan).

Criteria for the creation of ontologies reflecting language and world knowledge in the fields of world linguistics and information technologies, the principles of creating thesauruses that increase the efficiency of the information retrieval system, the following scientific outcomes were obtained from research on

the creation of ontology and thesauruses: Theoretical aspects of the creation of ontology (PU USA, MSU Russia, LO, LPEHS Taiwan, TSUULL Uzbekistan), research of linguistic ontology and thesauruses and their common and distinct sides are highlighted (SLU USA, MSU Russia, TLSPC Turkey), types are described (MSU, BIAI, IAI, RA), the importance of an automatic analysis of the text and natural language processing are stated (JNIICT Japan, UT Germany, ICL, BKF, CCIT, Human Language Technologies Group, VU Italy), its role in the information retrieval system, the reflection of ontologies on language and world science are analyzed (UT Tehran, NLPL Iran, IIS, IIT India), opportunities and theoretical methodology for creating language ontology on the basis of Princeton WordNet (in all institutions listed in the 1st paragraph of the review), natural language processing technology based on linguistic ontology (WUST Poland, USh UK), the tasks are being carried out on the development of Electronic thesaurus of Turkish languages for developing systems of searching and learning multilingual languages (Kazakh University (KU), Almaty).

Research is being carried out in the directions, such as creation of linguistic ontology in world linguistics and improvement of linguistic resources on the basis of WordNet, increase of the opportunities of information retrieval system, creation of national ontologies and multilingual ontologies, processing of natural language by the means of modern information technologies, creation and improvement of lexical databases based on language corpuses, and creation of multilingual ontology for machine translation.

Extent of the study of the problem. The issue of developing linguistic (semantic) ontologies in the fields of foreign applied linguistics and computer linguistics has been studied by many scholars.

The concept of linguistic ontology is not new in this sphere, however, as a modern interpretation it has been used since the late twentieth century.

George A. Miller⁹⁷, Christiane Fellbaum⁹⁸ (authors of Princeton WordNet), A. Burgen, O. Bodenreider; K. Kunze, A. Vagner⁹⁹ (the German language ontologists), L. Bentivogli, E. Pianta (the Italian language ontologists), M. Buenaga Rodriguez, J. Gomez-Hidalgo¹⁰⁰ (the Bulgarian language ontologists), N. Guarino¹⁰¹, P. Guaretta (specialist in formal ontology and information systems)

⁹⁷Miller G. Nouns in WordNet. In: Fellbaum, C (ed) WordNet – An Electronic Lexical Database. – The MIT Press. 1998. – pp.23-47.; Miller G., Fellbaum C. Morphosemantic links in WordNet. – *Traitement automatique de langue*, 44.2. 2003. – pp. 69-80.; Miller G., Hristea F. WordNet Nouns: Classes and Instances. – *Computational linguistics*, Volume 32, Number 1. 2006. – pp.1-3.; Miller K. Modifiers in WordNet. In: Fellbaum, C (ed) WordNet – An Electronic Lexical Database. – The MIT Press. 1998. – pp .47-68.

⁹⁸Fellbaum Ch. A Semantic Network of English Verbs. – In: Fellbaum, C (ed) WordNet – An Electronic Lexical Database. – The MIT Press. 1998. – pp. 69-104.

⁹⁹Kunze C., Wagner A. Integrating GermaNet into EuroWordNet, a multilingual lexical semantic database. In: *Sprache und Datenverarbeitung – International Journal for Language Data Processing*. Bonn. 1999.

¹⁰⁰Burgun A., Bodenreider O., Aubry M., Mosser J. Dependence relations in Gene Ontology: A preliminary study. Workshop on The Formal Architecture of the Gene Ontology. – Leipzig, Germany, May 28-29. 2004.; Buenaga Rodriguez M., Gomez-Hidalgo J., Diaz-Agudo B. 1997 Using WordNet to complement training information in text categorization // In *Proceedings of the 2nd International Conference on Recent Advances in Natural Language Processing (RANLP 1997)*, Bulgaria. 1997. – pp. 150-157.

¹⁰¹Guarino N. Formal Ontology and Information Systems. In N. Guarino, editor, *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems, FOIS'98*, Trento, Italy, IOS Press. 1998. – pp. 3-

effectively worked on the creation of linguistic ontology, the formation of the hierarchical base, the development of terminological bases on the spheres, the study of semantic relations, and in the spheres of world computer linguistics and information technologies for the formation of language ontology based on thesaurus dictionaries.

Among Russian linguists Lukashevich N.V. (RuThes - project manager for the creation of Russian language ontology, ontologist)¹⁰², V.B.Dobrov (formation of the base of terminological expressions for science field texts)¹⁰³, I.B.Azarova (main creator of the linguistic resource RussNet, who carried out researches on the ontology of information retrieval systems)¹⁰⁴, O.A. Nevzorova (technology for developing ontologies of new science fields)¹⁰⁵, A.S.Narinian (conducted ontological researches on the basis of TEON: Thesaurus + Ontology project)¹⁰⁶, B.B.Morkovkin (composed an ideographic dictionary), as well as A.

15.; Guarino N., Giaretta P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In N. Mars (ed.) Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing 1995. IOS Press, Amsterdam. 1995. – pp. 25-32.

¹⁰² Лукашевич Н.В., Автоматизированное формирование информационно-поискового тезауруса по общественно-политической жизни России // НТИ. Сер.2. – 1995. – N 3. – С.21-24.; Лукашевич Н.В., Добров Б.В. Тезаурус русского языка для автоматической обработки больших текстовых коллекций // Компьютерная лингвистика и интеллектуальные технологии. Труды Международного семинара Диалог'2002 / Под ред. А.С.Нариньяни. – Т.2. – М.: Наука – 2002. – С.338-346.; Лукашевич Н.В., Добров Б.В. Отношения в онтологиях для решения задач информационного поиска в больших разнородных текстовых коллекциях. Девятая национальная конференция по искусственному интеллекту с международным участием КИИ-2004 (28 сентября-2 октября 2004 г., Тверь). Труды конференции. В 3-х т. – Т2. – М.: Физматлит, 2004. – С.544-551.; Лукашевич Н.В. Моделирование отношения ЧАСТЬ-ЦЕЛОЕ в лингвистических и онтологических ресурсах. Информационные технологии. – 2007. – N 12.; Лукашевич Н.В. Проблемы установления родовидовых отношений в лингвистических онтологиях. – Материалы Всероссийской конференции «Знания – Онтологии – решения» (ЗОНТ-07). – С.211-220.;

¹⁰³ Добров Б.В., Лукашевич Н.В., Сыромятников С.В. Формирование базы терминологических словосочетаний по текстам предметной области. Труды пятой всероссийской научной конференции «Электронные библиотеки: Перспективные методы и технологии, электронные коллекции. – 2003. – С. 201-210.; Добров Б.В., Лукашевич Н.В. Онтологии для автоматической обработки текстов: описания понятий и лексических значений. Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог'2005 / Под ред. И.М. Кобозевой, А.С. Нариньяни, В.П. Селегея. – М.: Наука, 2005. – С.138-142.; Добров Б.В., Лукашевич Н.В. Вторичное использование лингвистических онтологий: изменение в структуре концептуализации. Восьмая Всероссийская научная конференция «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Владимир-Суздаль, 16-18 октября 2006г.). 2006.; Добров Б.В., Лукашевич Н.В. Транзитивные не таксономические отношения в онтологическом моделировании. Труды симпозиума Онтологическое моделирование. Институт проблем информатики РАН, 2008. – С.229-259.

¹⁰⁴ Азарова И.В., Митрофанова О.А., Синопальникова А.А. Компьютерный тезаурус русского языка типа WordNet // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог'2003. М., 2003. – С. 43-50.; Азарова И.В., Синопальникова А.А., Яворская М.В. Принципы построения wordnet – тезауруса RussNet. Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог'2004. М., 2004. – С. 542-547.; Азарова И.В., Синопальникова А.А., Смрж П. Представление устойчивых лексических сочетаний в компьютерном тезаурусе RussNet. Компьютерная лингвистика интеллектуальные технологии. Труды Международной конференции Диалог'2005. М., 2004. – С. 11-16.

¹⁰⁵ Добров Б.В., Лукашевич Н.В., Невзорова О.А. Технология разработки онтологийных предметных областей. Труды Казанской школы по компьютерной лингвистике TEL-2002. Выпуск 7. / Под ред. В.Г. Бухараева, В.Д. Соловьева, Д.Ш. Сулейманова – Казань: Отечество, 2002. – С.90-106.

¹⁰⁶ Нариньяни А.С. Кентавр по имени ТЕОН: Тезаурус+Онтология. Труды Международной конференции ДИАЛОГ-2001. – Т.1. – М., 2001. – С.184-188.

Sharipbay¹⁰⁷ from Central Asia (who is conducting research on Kazakh language processing and developing the ontology of the Turkic languages) carried out investigations.

A great deal of work on the development and application of linguistic ontology has been carried out in Russia and abroad in the sphere of applied linguistics and now it is going on in a rapid way. In this study we will look through the most important researches of J. Miller and K. Fellbaum, P. Butelaar, F. Chimiano, P. Haaza, B. V. Dobrov, N. V. Lukashevich, O. Narinyan, I. Azarova. Moreover, the largest linguistic ontologies including SUMO, OMEGA, DOLCE, Princeton WordNet, KeNet, RuThes and RussNet ontologies and thesauruses have been studied, analyzed and responded where necessary. As a result, methodological knowledge on the development of the ontology of the Uzbek language has been mastered, and experience that enables to process data relating to a certain field has been gained. Therefore, the "Concept of creating an ontology of the Uzbek language" has been developed.

Investigations on the development of a linguistic ontology of the Uzbek language is being conducted in the sphere of Uzbek computer linguistics (UCL). In particular, a number of articles have been published on the creation of thesaurus dictionaries and ontological resources pertaining to the WordNet system. UCL is being enriched with many theoretical and practical studies during its development. Constituted in the late twentieth century, it can be said that UCL has a two-stage development period. The first stage is defined not only by scientific investigations and generating a number of frequency dictionaries, but by establishing educational literatures and using a program that defines the frequency of the most frequently used words in periodicals and several fictions as well. This period, in turn, might be divided into two stages: the period from the end of the twentieth century to the twenty-first century. Its representatives are S.Mukhamedov, T.Sodikov, H.Arzikulov, M.Ayimbetov, S.Rizayev¹⁰⁸. The second stage is indicated by the opening of the Laboratory of Computer Linguistics at the National University of Uzbekistan named after Mirza Ulugbek from 2001 and the implementation of science in the education system and includes the period up to the 1920s. These achievements were gained through the efforts of A.Pulatov, who made a significant contribution to the improvement of UCL. It was during this time that the creation

¹⁰⁷ Вычислительная обработка казахского языка. Сборник научных трудов / Под редакцией Рахимовой Д.Р. – Алматы: Қазақ университеті, 2020. – 147 с.

¹⁰⁸ Мухамедов С.А. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дисс. ...канд.филол.наук. – Тошкент, 1980. – 25 с.; Мухамедов С.А., Пиотровский Р.Г. Инженерная лингвистика и опыт системно-статистического исследования узбекских текстов. – Т.: Фан, 1986.; Садыков Т. Проблемы моделирования тюркской морфологии. – Фрунзе, 1987.; Арзикулов Х.А., Пиотровская К.Р. Информатика и переработка текста средствами вычислительной техники (учебное пособие). – Самарканд, 1986.; Айымбетов М.К. Проблемы и методы квантитативно-типологического измерения близости тюркских языков (на материалах каракалпакского, казахского и узбекского языков): Автореф. дисс. ...д-ра филол. наук. – Т., 1997. – 47 с.; Айымбетов М.К. Проблемы и методы квантитативно-типологического измерения близости тюркских языков (на материалах каракалпакского, казахского и узбекского языков): Автореф.дисс. ...д-ра филол.наук. – Т., 1997. – 47 с.; Rizaev S. Problems of linguostatics in Uzbek linguistics (monograph). – Tashkent: Fan. 2005. - 295 p .; This author. Linguostatic study of the Uzbek language: Doc. Ph. Sc. ... diss. abstract – Tashkent, 2008. – 50 p .; Айымбетов М.К. Квантитативная типология тюркского текста (сборник избранных статей). – Нукус.: Илим. 2012.

of educational literatures of computer linguistics¹⁰⁹ and enhancement of the researches were achieved¹¹⁰, and some theoretical examinations were undertaken by A.Pulatov, A.Rakhimov, S.Mukhamedova, N.Juraeva, U.Dusimova, N.Valiyeva, M.Abjalova, N.Abdurakhmanova.

By 2018s, a rapid implementation of candidacy and doctoral dissertations in UCL laid the foundation for the transition to the second phase of its period. During this time, the principles of composing the Uzbek language authorship corpus¹¹¹, linguistic provision of Uzbek-English machine translation¹¹², issues of genre-linguistic and lingo-statistical research¹¹³, linguistic modules of the program of automatic editing and analysis of Uzbek texts¹¹⁴, principles of constructing linguistic base of language corpus¹¹⁵, linguistic bases of semantic tagging of Uzbek language nominative units¹¹⁶ and the problems of linguistic provision of the morphological analyzer of the Uzbek language¹¹⁷ have been studied in the monographic plan. Despite the fact that a considerable amount of work has been done in several areas of computer linguistics in UCL, no specific examination has been fulfilled on the issue of compilation of the Uzbek language ontology.

By the 20s of the XXI century, the conduction of practical effects researches, the development of various linguistic programs (automatic editing and analysis (automatic texts analysis program for Ubuntu Linux system¹¹⁸), transliteration, mobile applications) (1) and in higher education institutions of the Republic (TSUULL, NUUZ, SamSIFL, UrSU) CL has entered a new stage of its development with the opening of a master's degree (2) on the specialty of Computer Linguistics. The foundation stone of National Corpus for the Uzbek

¹⁰⁹Mukhamedova S. Computer linguistics (methodical manual). - Tashkent, 2007; Pulatov A., Mukhamedova S. Computer linguistics (textbook). - Tashkent, 2008. - 98 p. ; Pulatov A. Computer linguistics. - Tashkent: Akademnashr, 2011. - 520 p. ; Rakhimov A. Basics of computer linguistics. - Tashkent: Akademnashr. 2011. - 160 p.

¹¹⁰Пўлатов А.К., Алиходжаев Б., Джураева Н. Разработка программы компьютерного анализа и синтеза глаголов узбекского языка // News of the National University of Uzbekistan. - Tashkent, 2002. №2. - P. 17-19. ; Pulatov A.Q., Muminova T., Pulatova I.O. Secular Uzbek language (forms of verbs in Uzbek and their transfer in Russian and English). - Tashkent: University, 2003. - 404 p. ; Mukhamedova S. Development of linguistic software for computer programs based on Uzbek verbs. - Tashkent, 2006. - 80 p.

¹¹¹Khamroyeva Sh. Linguistic bases of creation of the Uzbek language authorship corpus: Doc. of Philosophy on Ph. Sc. (PhD)... diss. - Karshi, 2018. - 250 p.

¹¹²Abdurakhmonova N.Z. Linguistic support of the program for translating English texts into Uzbek (in the form of simple sentences): Doc. of Philosophy on Ph. Sc. (PhD)... diss. abstract. - Tashkent, 2018. - 52 p.

¹¹³Urinbaeva D. Folklore: problems of genre-linguistic and linguistic-statistical research: Doc. Ph. Sc. (DSc) ... diss. abstract. - Samarkand, 2019. - 74 p.

¹¹⁴Abjalova M.A. Linguistic modules of the program of editing and analyzing texts in the Uzbek language (for the program of editing texts in official and scientific style): Doc. of Philosophy on Ph. Sc. (PhD)... diss. - Fergana, 2019. - 164 p.

¹¹⁵Eshmominov A.A. Synonym database of the Uzbek National Corpus: Doc. of Philosophy on Ph. Sc. (PhD)... diss. - Karshi, 2019. - 140 p.

¹¹⁶Akhmedova D. Linguistic bases and models of lexical-semantic tagging of Nominative units for Uzbek language corpuses: Doc. of Philosophy on Ph. Sc. (PhD)... diss. - Bukhara, 2020. - 156 p.

¹¹⁷Khamroyeva Sh. Linguistic support of Uzbek morphological analyzer. Doc. Ph. Sc. diss. abstract. - Fergana, 2021. - 78p.

¹¹⁸Abjalova M.A. Linguistic modules of editing and analysis programs. Monograph / M.A. Abjalova. - Tashkent: Nodirabegim, 2020. - 176 p.

language was laid at the TSUULL with the creation of the Uzbek language educational corpus¹¹⁹ by the efforts of a group of specialists.

This study, entitled “**Principles of creating an ontology of the Uzbek language**”, helps to choose effective approaches of developing UzNet lexical database and to compose the concept of creating a linguistic ontology of the Uzbek language.

The relevance of the research to the plans of scientific examinations of the higher education or scientific research institution where the dissertation was carried out. The research has been carried out within the framework of the practical project No. AM-FZ-201908172 “Creating the educational corpus of the Uzbek language” being performed at TSUULL and pertaining to the concept of “Creating the Uzbek language ontology”.

The purpose of the research is to learn the principles of developing linguistic ontologies for digital technology and artificial intelligence, to analyze the structure of lexical databases and to develop the concept of creating UzNet - the Uzbek language ontology.

Missions of the research:

– to depict the concepts of thesaurus and ontology from primary sources in the search for information, to highlight their significance and define the definition of linguistic ontology;

– to study the large English lexical database that is the structure of WordNet's linguistic ontology, problem analysis, usability, and factors which other languages largely rely on it.

– to learn the experiences in the process of developing Turkish (KeNet) and Russian (RuThes, RussNet) lexical resources in the WordNet system and to compose general principles of the Uzbek language ontology as a result of examining important directions of the EuroWordNet multilingual formal ontology project;

– to analyze the classification of linguistic ontologies;

– to analyze semantic relations for lexical databases, including hyperonymy, holo-meronymy, synonymy, antonymy phenomena and their classifications, and their implementation to the Uzbek language in the development of the ontological system UzNet;

– to project the "Concept of the Uzbek language ontology" on the basis of the examined data;

– to form a database of linguistic resources of UzNet Uzbek language ontology.

As an object of the study the “Explanatory Dictionary of the Uzbek language”, synonyms in the Uzbek language and linguistic information on the basis of the Uzbek language educational corpus have been chosen.

¹¹⁹Carried out in the framework of the practical project “Creation of the educational corpus of the Uzbek language”. No. AM-FZ-201908172

The subject of the research is to develop WordNet lexical database, technology for developing linguistic ontologies, semantic relationships in lexical databases WordNet, KeNet, RuThes and the conception of creating UzNet linguistic ontology.

Research methods. Methods such as description, comparison, component analysis, statistical, modelling, and substantive analysis have been used to cover the research topic.

The scientific novelty of the research includes the following:

- the integral and distinctive aspects of the concepts of linguistic ontology and thesaurus have been shed light on as a result of the explanation of philosophical and linguistic essence of the concept of ontology and the analysis of the opportunities of the thesaurus.

- mathematical models of linguistic ontology and its requirements as a lexical database have been specified;

- the technology of developing the Uzbek language ontology (UzNet) based on defining Princeton WordNet ontological system in English, in which the principles of data presentation are theoretically proved;

- the principles of creating a collection of synonyms (synonyms), appointed as a primary requirement of its existence in linguistic ontologies, the practices of establishing the relationship of species and genus (hyponym-hypernym), whole-part (holonym-meronym) are identified, the features forming species-genus, whole-part relations are proven, as well as their most optimal description is theoretically identified;

- the concept of creation of an Uzbek ontology has emerged as a result of a comprehensive study of the experience of the KeNet project team carried out for the relative Turkish language;

- word meanings in the order of being given in UzNet, methods of tagging parts of speech, the components and structure of the database concerning semantic and formal relations between words are proved.

The practical outcomes of the research are as follows:

- having been developed the "database of synonymous words in the Uzbek language" for synset (synonyms) collections of UzNet Linguistic ontology, a certificate of authorship has been acquired¹²⁰;

- databases of homonyms¹²¹, antonyms¹²² and paronyms¹²³ were created for UzNet Uzbek lexical database and copyright certificates were received;

- stressed database of indirect speech words have been developed in order to tag the group of indirect speech words and provide commentaries for 11 000 indirect speech words in the Uzbek language and a copyright certificate has been obtained¹²⁴;

¹²⁰Database of synonyms in the Uzbek language. Certificate № BGU 00380. – Tashkent, 2019.

¹²¹Database of homonyms in the Uzbek language. Certificate № BGU 00381. – Tashkent, 2019.

¹²²Database of antonyms in the Uzbek language. Certificate № BGU 00390. – Tashkent, 2020.

¹²³Database of paronyms in the Uzbek language. Certificate № BGU 00469. – Tashkent, 2021.

¹²⁴Explanatory base of indirect speech words in the Uzbek language. Certificate № BGU 00404. – Tashkent, 2020.

– a morphological database of the Uzbek language and a system of spelling rules for the lemmatization process of the Uzbek language ontology has been developed;

– formal spelling rules developed, database of the dictionary of word formation, in the practical project on “the formation of educational corpus of Uzbek language” No. AM-FZ-201908172 from the database of paronyms, data of “Phonetics” department produced for universal grammar of the Uzbek language and a set of tasks aimed at developing oral competence was used in the grant No. PZ-2020042022 “Creation of a linguo-didactical electronic platform of Turkic languages” for women scientists.

Reliability of research results are explained through the matter of developing an ontology of the Uzbek language which has been clearly defined, the scope of the investigation of the work that have been defined, and the experiences of creating lexical information systems in inflected and agglutinative languages that have been studied. Furthermore, it provides information about theoretical views in corpus and computer linguistics which are methodologically based, the scientific conclusions gained as a result of general methodological materials, practical implementation of theoretical ideas and conclusions concerning dealing with the tasks assigned in the research and gained results that have been confirmed by authorized organizations.

Scientific and practical significance of research results. The scientific significance of the research results serves as a scientific and theoretical source in the development of theoretical foundations of linguistic ontology of the Uzbek language, in determining the methodological basis for establishing semantic relations for linguistic ontology, and in conducting examinations to improve UzNet ontology.

The practical significance of the research results is explained through the possibility of its usage as being important source for information retrieval systems and machine translation and in managing the activity of the scientific research center “Computer Linguistics” from research materials, as well as, during delivering lectures in higher education institutions on close sciences, such as “Computer linguistics”, “Ontologies and semantic systems”, “Corus linguistics”, “Machine translation”, “Parallel corpuses”, creating textbooks and manuals, compiling electronic dictionaries (translation, thesaurus).

Introduction of research results.

On the basis of the scientific and practical results stemmed from the concept of ontology of the Uzbek language, the followings have been achieved:

– in the compilation of lexicographic sources, the results of the description of the classifications of holonym-meronym relations, genus and species relations and matters of putting them in linguistic ontology were used in fundamental scientific project No. I-OT-2019-42 called “Developing electronic (the images of human face, its character, nature and national symbols) poetic dictionary of the Uzbek and English languages” conducted at the TSUULL. (Reference No. 04 / 1-2339 of November 26, 2021, TSUULL). As a result, the role and significance of

the words like holonym (whole), meronym (part), hypernym (genus), hyponym (species) in human appearance, its character, nature and national symbols has been successfully explained; semantic relations of polysemous words in Uzbek and English languages have been successfully revealed;

– “Database of homonymous words in the Uzbek language” (№ BGU 00381.2019.) and “Database of paronym words in the Uzbek language” (№ BGU 00469. 2021.) developed during the research were used in the practical project named “Development of the educational corpus of the Uzbek language” No.AM-FZ-201908172 (Reference No. 04 / 1-2340 of November 26, 2021, TSUULL). Consequently, the systematic search results have been obtained on the basis of synonyms, antonyms, homonyms and paronyms in the educational corpus of the Uzbek language;

– materials and practices stemmed from the research on classification of semantic words in the Uzbek language, factors generating quasi-synonyms and their types, features forming the whole-part relations were applied in the paragraph called “Principles of developing competences in students” in the 1st section of the textbook “Methodology of teaching the Uzbek language” (Almaty: Evero, 2021) printed in the authorship of T. Yusupova, K. Mavlonova, Sh. Naraliyeva, in the 32nd paragraph named “Lexical-semantic relations” and in the 46th paragraph “Word categories and principles of their classification”. (Reference No. 04 / 1-2341 of November 26, 2021, TSUULL). As a result, it has been possible to explain the factors, such as developing oral competence and linguistic competencies, revealing the relationships of synonymy and hyponymy with examples, showing the importance of lexical-semantic relations in language enrichment, highlight the necessity of word groups classifications;

– the outcomes and materials taken from the conducted research regarding matters of tagging word categories on the basis of semantic relations in lexical databases and the concept of developing UzNet linguistic ontology were beneficial to use in the textbook called “The modern Uzbek language” printed by the authors R.Sayfullayeva, B.Mengliyev, L.Raupova, M.Kurbanova, M.Abuzalova, D.Yuldasheva” (Tashkent: Innovation-Ziya, 2021). (Reference No. 04 / 1-2342 of November 26, 2021, TSUULL). As a result, synonymous and hyponymic cases with examples, clarification of their types, semantic classification of words and explanation of lexicalization have been clearly understood.

– by using theoretical and analytical data from research results, recommendations for creating lexical data and the author's monograph “Linguistic modules of editing and analysis programs”, the worker and study programs have been developed for students and specialist masters on this educational sphere in order to develop qualification requirements for the specialty 70230801 – Computer Linguistics. (Reference No. 04 / 1-2343 of November 26, 2021, TSUULL). As a result, the tasks that computer linguistic performs in the section of professional competences of qualification requirements have been developed, the themes including the use of lexical information, development of linguistic ontologies, tagging words groups, establishment of semantic relationships in subjects such as

“Linguistic bases of machine translation”, “Natural language processing / NLP”, “Ontologies and semantic systems” have been covered, and the range of basic literature and information sources has been enriched;

– borrowed words in the Uzbek language, their part of speech, their explanation(s), information about their original language, synonym(s) of the borrowed words, information about their contradictory meanings were used in the stressed base of the borrowed words that was intended to develop speech recognition and speech competence, to eliminate speech ambiguities called “Dictionary with the stress of borrowed words of the Uzbek language” (Dictionary with the stress of borrowed words of the Uzbek language [Text]: educational-methodical dictionary. – Tashkent: Nodirabegim, 2021. – 988 p. ISBN 978-9943-6940-9-5). As a result, a base of words stresses of borrowed words have been developed, which served as a linguistic support for the creation of the Uzbek speech synthesizer.

Approbation of research results. The results of this research have been announced in 14 international and 22 national scientific-practical conferences.

Scientific works of the author have been discussed in the following profiles of the scientific platform.

<https://scholar.google.com/citations?user=ZSEZYo8AAAAJ&hl=ru>,
<https://www.researchgate.net/profile/Manzura-Abjalova>,
<https://www.linkedin.com/in/manzura-abjalova-0125b21ba/>,
<https://orcid.org/0000-0002-1927-2669>

Announcement of research results. 48 scientific works on the topic of the dissertation, including 14 articles in scientific publications recommended for publication of the main scientific results of doctoral dissertations of the Higher Attestation Commission of the Republic of Uzbekistan (6 of them in foreign journals), 5 copyright certificates, 1 at prestigious international conference indexed to Scopes base and 25 scientific articles and theses have been announced in Republican and foreign conferences. The results include 2 dictionary and 2 monograph.

The structure and scope of the dissertation. The dissertation consists of an introduction, four main chapters, a conclusion, a glossary and a list of references and the volume of which is 266 pages. An appendix of 33 pages has been attached to the dissertation.

MAIN CONTENT OF THE DISSERTATION

The relevance and necessity of the topic has been explained in the introductory part, as well as the relevance of the research to the priorities of the development of science and technologies of the republic has been showed. Reviews of foreign scientific examinations on the dissertation topic, analysis of the level of the problem being investigated, goals and objectives have been identified, its object and subject has been described, its scientific novelty and practical results have been illustrated, scientific and practical significance of the results have been revealed, and information on their introduction, approbation, published works and dissertation structure has been included.

The first chapter of the dissertation is called “Linguistic ontology – lexical database” which consists of three sections. In the first section, that is called “Linguistic Ontology and Thesaurus Analysis” the concepts of ontology and thesaurus, the first matter stated in the process of the research, have been analyzed, WordNet – the English lexical information system, and the Russian ontological systems have been analyzed, and their structure, key elements, and capabilities have been explained.

Recently, specialized forms of information retrieval, such as medical, scientific, banking and financial, and political search have become increasingly important, and the role of knowledge in the fields of science is important in ensuring the quality of such information systems. In general, it is a difficult task to incorporate knowledge of language and the world into software systems using modern methods of automatic word processing. The solution lies in the fact that knowledge of language and the world is reflected in specially created sources (thesaurus, ontologies), in which such sources describe tens of thousands of words and phrases and include the opportunities of entering into semantic relationships with other words and units, and drawing logical conclusions. When they are used, the ambiguity, homonymy, and polyfunctionality of words are usually resolved automatically.

Ontology (Greek *ontos* [óntos] – being and *logy* [lógos] – doctrine) is actually a branch of philosophy, the doctrine of all beings. It examines the general foundations, principles of existence, its forms and laws. Although the term ontology was introduced to science by the German philosopher R. Goklenius¹²⁵ in 1613 and later used by H. Wolf (1679 - 1754) in his textbook (1730), at first, Greek philosophers entered¹²⁶ its various interpretations and the term “metaphysics” was also used as a synonym of “ontology”. Unlike the Greek philosophers, Central Asian intellectuals such as Kindi, Zakaria, Razi, Farobi, and IbnSina raised the doctrine of ontology to a whole new level. For example, Farobi approached¹²⁷ ontology as a doctrine that revealed the essence of a single being.

Connections and relationships are primary sources in ontology. This feature of the term has led to its widespread use in other spheres. Thus, the term “ontology” is used in many spheres and has two meanings: 1) a philosophical concept that embodies “being” and “essence”; 2) a system that describes the content of the elements, a network relationship made between them. Views on ontology as a semantic network began in the late 1990s. Linguistic ontology (LO) is rarely mentioned as a science of the existence and essence of language. LO serves to reveal the essence of language through the analysis of linguistic being. The basic spheres of being include¹²⁸ nature, society, and consciousness. Linguistic ontologies also cover the richness of natural language, its usability, and language.

¹²⁵Goklenius R. *Lexicon philosophicum*. Francofurti, 1613.

¹²⁶<https://uz.wikipedia.org/wiki/Ontologiya>

¹²⁷To‘rayev B.O. *Borliq: mohiyati, shakllari, xususiyati: monografiya/ B.O.To‘rayev; maxs. muharrir M.N.Abdullayeva, O‘zR FA I.Mo‘minov nomidagi Falsafa va huquq instituti. – Toshkent: Falsafa va huquq instituti nashriyoti (FHIN), 2011. – 128 b.*

¹²⁸The same source. – B. 5.

experiments, a WordNet lexical database for English is being developed as an imitation of human memory and brain neural networks. However, WordNet has aroused the interest of computer linguists more than psycholinguists.

J. Miller summarized the factors of the development of WordNet in the following 3 hypotheses ¹³²(three hypotheses): 1) Separability hypothesis: the description of the lexical component of natural language can be distinguished and studied separately. Such a large lexical database can be used to separate all bases of categories in order to tag words categories for a certain direction, for instance, a machine translation linguistic database; 2) Patterning hypothesis: there are words in the language that have their own formal interpretation, and such explanations can be applied to most words in the language. Such descriptions are generally consistent with synonyms. That's why WordNet is based on synonyms; 3) Comprehensiveness hypothesis: a comprehensive electronic dictionary with a wide range of lexical units. So as to effectively use computer dictionaries in automatic texts processing programs, the dictionary must be very large and of great importance.

Nouns, verbs, adjectives and adverbs of WordNet base are grouped into sets of cognitive synonyms called “synsets”, each of which has the following semantic relations: synonym, hypernym-hyponym are *in noun*, synonym, hypernym-hyponym *in verb*, antonym *in adjective*, semantic groups and valences *in adjective* have been reflected. In synsets, the unit in each class expresses a distinct concept. For example, the word "book" has meanings like *rule book / record book / volume / Book (a proper noun) / record / reserve*. They, in turn, have structural meanings, such as: *accumulation, aggregation, assemblage, collection* - a noun group; *section, subdivision* - a noun group; *product, production* - a noun group; *schedule* - a noun group; *reserve, hold* - a verb group; *put down, enter* - a verb group; *record - put down (enter), enter* - a verb group; *record book - record* - a noun group; *account book, book of account, ledger, leger* - a noun group; *playscript, script-publication (manuscript)* - a noun group. In this network, it is clear that the word "record" belongs to both the noun and the verb groups, and is the node that connects the synonymous series of nouns and verbs.

From March 1996 to September 1999, a multilingual version of WordNet, EuroWordNet, was created as a result of its funding by the European Commission¹³³. The system is a formal ontology and includes WordNet dictionaries in English, Danish, Spanish, Italian, German, French, Czech and Estonian, and Princeton WordNet 1.5. EuroWordNet dictionaries are commercial products.

In the second section of the first chapter, “*RuThes Lexical Database Software*”, the resources of RuThes and RussNet linguists have been analysed. RuThes is a thesaurus in Russian (ontology¹³⁴ in some sources). It has been developed by the Center for Information Research as an automated indexing tool

¹³²http://db4.sbras.ru/elbib/data/show_page.phtml?20+1531 _Тезаурус WordNet

¹³³P.Vossen. Building a multilingual database with wordnets for several European languages. <http://www.ilc.uva.nl/EuroWordNet/>

¹³⁴<https://nlpub.ru/PyTcz>; <https://new.labinform.ru/pub/ruthes/index.htm>

since 1994 and its structure continues to be developed to this day. The development of the RuThes thesaurus began¹¹ with the development of the socio-political thesaurus¹³⁵. It contains 45,000 concepts, 107,000 words and phrases, and 177,000 synonymous relations.

RuThes is based on the following four principles and therefore consists of four XML files¹³⁶: 1) concept – concepts.xml; 2) relations between concepts – relations.xml; 3) text input element – text_entry.xml; 4) The relationship between concepts and textual entries -

The second chapter of the dissertation, entitled “**Technology and criteria for creating linguistic ontologies**”, describes the model of linguistic ontology, formal schemes and specific resources produced on the basis of this model, which can be used to process texts in a wide range of knowledge fields. The model takes into account three paradigms aimed at describing knowledge in a wide range of subject areas: information retrieval thesauruses, thesauruses belonging to the WordNet-type, and ontologies. Special attention is paid to the system of relations between concepts. In addition, as a result of the structure and content of the KeNet resource in the Turkish WordNet system, highlighting problems the technology and criteria for creating an ontology of the Uzbek language are specified.

The first section of this chapter, entitled “*KeNet – the principles of creation of the ontology of the Turkish language*”, illustrates and analyzes in detail the comprehensive WordNet for the Turkish language, that is KeNet¹³⁷ and its creation. KeNet has also intrinsic semantic relationships, with 76,757 synsets. It is connected with PWN (Princeton WordNet) through interlingual communication. This section explains that due to the main problem in the semantic relationship of synsets components in the creation of synsets, two processes were performed to manage synonymous relations in synsets – *the combination process* and *the separation process*. In KeNet, during the combination process of different sets of synsets that need to be combined have been specified, identified, and synsets are grouped as a single set. Three things are very important when combining a set of synsets¹³⁸: 1) that each set of data has a unique and distinctive interpretation/description; 2) the presence of real synonyms as a synonym member in each set of data; 3) The first component of each data set should be represented (that represents, depicts linguistic data in itself, and of grammatical character. There are three main factors of incorrect combining: 1) heterogeneity of the meanings and interpretations; 2) POS, a problem with tagging word groups; 3) morphological problems¹³⁹. These factors are explained in the dissertation. The problem with meaning is that words that are semantically close to each other but cannot be synonymous (not exactly or not used interchangeably) are given as a

¹³⁵ Лукашевич Н.В., Салий А.Д., Тезаурус для автоматического рубрицирования индексирования: разработка, структура, ведение // НТИ. Сер.2. - 1996. - N 1. – С.1-6.

¹³⁶ <https://nlpub.ru/PyTez>

¹³⁷ KeNet qisqartma nomidagi Ke qismi turkcha “kelime” (kalam, so‘z) so‘zining birinchi bo‘g‘ini hisoblanadi.

¹³⁸ ÖzgeBakay and others. Turkish WordNet KeNet. Global Wordnet Virtual Conference. 2021. January. – P. 166. https://www.researchgate.net/publication/348264475_Turkish_WordNet_KeNet

¹³⁹ Bakay O., Ergelen O., and Yildiz O.T. 2019. Problems caused by semantic drift in wordnetsynset construction. InUBMK.

component of the synsets. For example, nouns with similar meanings, such as *dere* (*stream*) and *irmaknehir* (*river*), are ideographically put into an appropriate place, but are incorrectly united regarding semantic correspondence to each other. The POS problem occurs when synset components that are semantically compatible but belong to different word classes. For example, the words *güç* (strength, power) and *güç* (resistance) belonging to the group of nouns, the word *güç* (difficult) belonging to the group of adjectives are combined into one root and explanations for it are provided in a general term. This results in incorrect combination of the elements of synsets. The problem relating morphology happens when different morphological forms of a word are incorrectly united into a single component of synset. For example, the word *sopalamak* (to beat) having the same basis is a pure verb, in an active voice, in the indefinite form of the verb, and the word *sopalanmak* (to be hit, to be beaten), that is the verb in the passive voice, having various performers expresses different meanings, although its morphological indexes are different, they have been united in the same group. In KeNet, such forms also have been provided and a special data set has been developed for them.

The second chapter, entitled “Classification of Lexical Databases”, analyzes the 11 definitions given to linguistic ontology with approaches in the sources and, as a result, we define an acceptable definition: “**Ontology is a sectoral lexical database that bases on language and world knowledge, embraces terms among fields and that has been formed on the basis of the relationships between them**”. This section also analyzes the classification of ontologies and describes its following types, which has emerged on the basis of two approaches: *meta-ontology* – describe the most general concepts that are not related to subject areas; *a field ontology* – a formal description of the field; it is generally used to define specified concepts in meta-ontology (if used) and / or to define the general terminological base of a subject area; *an ontology of exact tasks* – an ontology that defines the general terminological base connected with a task or problem; *network ontologies* are often used to describe the final results of actions performed by objects in the field or task.

The section of the second chapter, called “*Ontological Network – Lexical Information System Models*”, discusses three main paradigms of resources that include knowledge of the world and language currently used in information retrieval and information analysis systems. These are: thesauruses, WordNet (thesaurus), and formal ontologies.¹⁴⁰

Formal ontologies are a modern paradigm¹⁴¹ of computer resources for information retrieval applications. This system has appeared as a result of advancing the concept of the Semantic Web¹⁴², which is based on constructing a wide range of ontological resources¹⁴³.

¹⁴⁰ Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – Москва: МГУ, 2011. – 512

¹⁴¹ Qarang: <http://globalwordnet.org/resources/wordnets-in-the-world/>

¹⁴² Berners-Lee, T. The Semantic Web / T. Berners-Lee, J. Handler, O. Lassila // Scientific American - 2001. – V. 284. – No 5. – P. 28-37.

¹⁴³ Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования, том 5, №1(15)/2015. – С. 48.

However, contrary to the views of proponents of formal ontologies, S. Nirenberg expresses his opinion that it is difficult to carry out automatic processing of unstructured texts in natural language with the phenomena of polysemy, homonymy and polyfunctionality in them with the help of axiomatic theories¹⁴⁴. Therefore, special types of ontology (terminological or special ontologies) are being developed for automatic processing of texts¹⁴⁵.

On the one hand, concepts are not fully reflected in formal ontologies. On the other hand, textual analysis often requires a formal logical conclusion based on ontology, as the volume of the information in a coherent text is not clearly stated.¹⁴⁶ In addition, due to the usage of ontologies in the automatic processing of texts, a concept called *linguistic ontology* has emerged, which seriously concerns the concepts of language units and terms of the field of knowledge¹⁴⁷.

N.V. Lukashovich describes the official definition of ontology in the following formula:

$$O = \langle C, E, At, R, A \rangle,$$

where: C – concepts (classes) in ontology, E - copies in ontology (registration of proper nouns as common nouns), At – concepts of ontology and attributes of copies, R – relationships between concepts, A – axioms of ontology¹⁴⁸.

One of the popular official models of ontology is represented as follows¹⁴⁹:

$$O = \langle L, C, F, G, H, R, A \rangle,$$

where: $L = L_C \cup L_R$ is an ontology dictionary. It contains a set of lexical units (characters) for the concept of L_C and a set of characters for the concept of L_R ; C is a set of concepts in ontology; F and G – connect the sets of lexical units $\{l_i\}$ belonging to L with this set of ontological concepts and relations; H – defines the taxonomic nature of relations (relationships), in which case the concept of ontology is associated with non-reflexive, acyclic, transitive relations in the form $H \subset C \times C$; R means a non-taxonomic relationship between the concepts of ontology; A is a set of ontological axioms. In creating a natural language ontology, relying on this model is considered appropriate.

The third chapter of the research, “**Semantic relationships in lexical databases**”, explains a set of synonyms – synsets, i.e., the process of creating the WordNet system of a particular language and gives information about results of the

¹⁴⁴Nirenburg, S. What’s in a symbol: Ontology, representation, and language / S. Nirenburg, Y. Wilks // Journal of Experimental and Theoretical Artificial Intelligence. - 2001. - V. 13(1). - P. 9-23.

¹⁴⁵Ландэ, Д.В. Подход к созданию терминологических онтологий / Д.В. Ландэ, А.А. Снарский // Онтология проектирования. 2014. № 2(12). – С. 83-91.; Sowa, J. Building, Sharing and Merging Ontologies. – <http://www.jfsowa.com/ontology/ontoshar.htm>

¹⁴⁶Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования, том 5, №1(15)/2015. – С. 48-49.

¹⁴⁷Magnini, B. Merging Global and Specialized Linguistic Ontologies / B. Magnini, M. Speranza // Proceedings of OntoLex. – 2002. – P. 43-48.; Veale, T. A context-sensitive framework for lexical ontologies / T. Veale, Y. Hao // Knowledge Engineering Review. 2007. Vol. 23(1). – P. 101-115.

¹⁴⁸Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования, том 5, №1(15)/2015. – С. 49.

¹⁴⁹Лукашевич Н.В., Добров Б.В. Проектирование лингвистических онтологий для информационных систем в широких предметных областях // Онтология проектирования, том 5, №1(15)/2015. – С. 50.

examinations of the concept of holonymy and hypernymy relationships, principles and importance of their structural modeling and their types.

In WordNet.princeton.edu the following categories are semantically branched: **noun** (synonym, hypernym-hyponym), **verb** (synonym, hypernym-hyponym), **adjective** (antonym), **adverb** (semantic group, valences).

The first section of this chapter, “*Synset collections – preliminary elements of the lexical information system*”, describes the synset collection and the stages involved in its creation. Words with a line of synonyms and groups are the most important element / attribute (in the language of information technology) in linguistic ontology. In LMB, the line of synonymous words is a “set” and because they are connected to each other by a network, it is called “*synset*”.

The following resources are required to create synsets, including a set of synonyms of the ontology of the Uzbek language: 1) a dictionary of synonyms - to create a set of synonyms and / or to fill the structure of synset; 2) Annotated dictionary – to explain words and give a hypernym relationship. For ex.: a chair is defined, it is determined that it is an object of furniture; 3) National corpus – to find lexical synonyms, quasi-synonyms, textual synonyms that are not included in the dictionary; 4) parallel corpuses – to find a reliable lexical, textual translation equivalent.

It is noteworthy that initially in PWN a relationship between word groups were not established. Due to the serious problems in these software applications, the EuroWordNet project introduced additional relationships between word groups: 1) XPOS-synonymy – synonymy of word groups; 2) XPOS-antonymy – antonymy of word groups; 3) XPOS-hyponymy – hyponymy of word groups. Thus, in Section 2.5.1, the relationship between the synsets *adornment2* (decoration process) and *adorn1* (decoration) can be explained by the synonymy of word groups¹⁵⁰. This relationship is a type of synonymy called quasi-synonymy.

In Chapter III, “*The Relationship of the Holonim-Meronim: Principles of Modeling*”, the holonymy (ancient Greek *hólos* = “whole” + *ὄνομα* = “name”) is the whole state (appearance)¹⁵¹ of a particular concept (s) – whole and meronymy (ancient Greek. *méros* = “part” and *ὄνομα* = “name”) is a component of another concept¹⁵², meronymy has been analyzed in some literatures as partonym¹⁵³ (lat. *pars*, from – *partis* = “part”) – part relationship and their types.

According to M.V. Nikitin meronymic relations are the most important factor in the semantic ordering of a dictionary, so it appears as an integral hierarchical structure. M.V. Nikitin describes meronymy as follows: “Undoubtedly, whole-part

¹⁵⁰ Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – Москва: МГУ, 2010. – С. 72.

¹⁵¹ https://ru.wikipedia.org/wiki/Мероним_и_холоним

¹⁵² <https://kartaslov.ru/значение-слова/мероним>

¹⁵³ Глобина Л.В. Лексико-семантическое поле партитивной лексики в современном русском языке: дис. ... канд. филол. наук. – Воронеж, 1995. – 205 с.; Коннова М.Н. Введение в когнитивную лингвистику : учебное пособие. Изд. 2-е, перераб. – Калининград : Изд-во БФУ им. И. Канта, 2012. – 313 с.; Материнская О.В. Система меронимов в немецком и английском языках: дис. ... д-ра филол. наук. – Донецк, 2013. – 403 с.; Колодько Д.А. К вопросу о классификации меронимов // Научные записки Национального университета «Острожская академия». Серия «Филология»: сборник научных трудов. – Острог, 2015. Вып.51. – С. 226-228;

relations embrace the whole world from the bottom up, from micro to macrocosm, from elementary particles to galaxies. They include objects of varying degrees of complexity, bringing them into different levels of the hierarchy of parts-whole, element-systems. These relations are general and global”.¹⁵⁴

It should be noted that synecdoche is also based on a whole-particle relationship. In contrast to holonymy, synecdoche is a type of semantic transfer, while holo-meronymy literally has a part-of relationships between functional and physiological or physical structure.

The study of the phenomenon of paronymy is important from the point of view of unraveling the “mysteries” of the vocabulary system and interpreting the meanings of words clearly and correctly¹⁵⁵. This phenomenon is to some extent reflected in Uzbek linguistics in the works of such linguists as H. Nematov et al¹⁵⁶. The works of E. Lysi, D. Cruz, R. Chaffin, E. Winston, D. German, V. Storey from foreign experts in the study of whole-unit relations in the creation of linguistic ontology; Russian scientists M. Nikitin, N. Lukashevich, Ye. Materinskaya, Yu. Rusina, and D. Kolodko became the fundamental base.

In classical mereology, there are three axioms in whole-part (part) relation¹⁵⁷. In the formulas P – whole, x, y, z – parts.

1. Reflexivity. Everything is an integral part of itself. (P.S. or simply $P = x, y$)
2. Antisymmetry: nothing can be share of its components. (P.S. $P \neq x \leftarrow p, y \leftarrow p$)
3. Transitivity: parts of parts are also parts of the whole. (P.S. $x \leftarrow y, y \leftarrow z = P \leftarrow z$)

This system of axioms of whole-part relations is usually called basic mereology (base mereology, ground mereology)¹⁵⁸.

Many authors argue that linguistic analysis poses serious problems with the transitivity of whole relationships. For example, *the hand is a part of the conductor (member), and the conductor is a part of the orchestra*, but according to the abovementioned transitivity (3) axiom, it is confusing to say that *the hand is a part of the orchestra*¹⁵⁹. N.V. Lukashevich emphasizes that such problems with transitivity are associated with the interference of different types of objects in the whole-part relationship¹⁶⁰. In the works of Winston et al. (Winston et al) and others explains transitivity problems are explained as follows: “When one type of

¹⁵⁴Никитин М.В. Курс лингвистической семантики: учебное пособие для студентов, аспирантов и преподавателей лингвистических дисциплин в школах, лицеях, колледжах и вузах. СПб.: Научный центр проблем диалога, 1996. – С. 442.

¹⁵⁵Jamolxonov H. Hozirgi o‘zbek adabiy tili: Darslik. – Toshkent: Talqin, 2005. – B. 147.

¹⁵⁶Begmatov E., Ne‘matov H., Rasulov R. Leksik makrosistema va uning tadqiq metodikasi (Sistem leksikologiya tezislari) // O‘zbek tili va adabiyoti. 1989, № 6. – B.35-40.; Qilichev B. O‘zbek tilida partonimiya. Filol.fan.nomz....dis. – Toshkent, 1997.; Jamolxonov H. Hozirgi o‘zbek adabiy tili: Darslik. – Toshkent: Talqin, 2005. – 260 b.

¹⁵⁷ Simons P. (1987). Parts.A study in Ontology. Oxford University Press. – 390 p.; Varzi A. (2006). A Note on Transitivity of Parthood // Applied Ontology, 1:2, pp. 141-146.

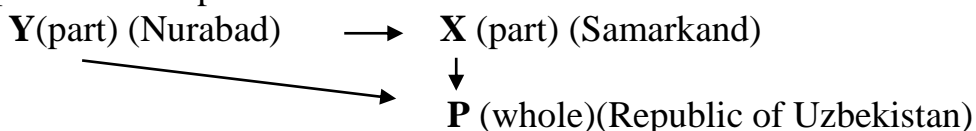
¹⁵⁸Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – Москва: МГУ, 2011. – 512 с.

¹⁵⁹ Лукашевич Н.В. Отношения часть-целое: теория и практика. // «Нейрокомпьютеры: разработка, применение». – Москва: Радиотехника, 2013. – С. 9.

¹⁶⁰This source.

relationship is used, the whole part is always transient. But when different aspects of meronymy are mixed, there is a problem with transitivity”¹⁶¹. Another example: *a leaf is a part of a tree, a tree is a part of a forest*, but it is strange to say that a leaf is a part of a forest.

In his work, D. Cruz argues that a well-formed hierarchy consists of elements of the same type¹⁶². We explain D. Cruz's idea with an example: in a whole-part relationship, if one element is a geographical name or region, then other elements also must belong to the same type. For example, Nurabad district is a part of Samarkand region, Samarkand is a part of the Republic of Uzbekistan, so Nurabad is a part of the Republic of Uzbekistan.



Thus, if an element of meronymy is a physical object, then all other elements of meronymy must be physically the same. If one element is an abstract noun, then the others must be of the same type. In our opinion, it would be expedient to group the whole-part relationship in transitivity as “complete transitivity” and “partial transitivity”. In this case, the units that are found in the text partial transitivity, but which are considered to be partial in the part-whole relationship. In general, this is due to the narrowing of the concept of “part” in everyday life.

Now let us go back to the example aforementioned: *the conductor's hand* - the conductor – orchestra. We see that the mass of the hand is part of the mass of the orchestra, and that the hand of the conductor is in one part of the space occupied by the orchestra. If the conductor's hand is injured, it can cause problems with the orchestra's performance (even a serious tragedy for the orchestra). The *leaf-tree-forest* relationships can be explained in the same way. Commenting on this issue, N.V. Lukashevich argues that it is possible to set additional conditions for the interpretation of the concept of “part”, that is, the additional requirement that the part should be functional, etc., of course, it may not create permeability¹⁶³. In this case, the functions of the whole itself and the functions that complement the functions of the part as a whole may be factors that do not cause transitivity or lead to partiality. Based on this distinction, it is possible to group the two categories we have proposed above in the lexical database.

1. *Functional parts* are bounded by their function in space and time condition. For example, if the handle of a bowl with bundle (cup) performs the function of holding, it can be located in a limited part of the space.

2. *Homeo-dimensional parts* represent the species to which the whole belongs, that is, the parts are exactly the same as the whole or belong directly to the whole to which it belongs. For example, *a piece of bread – bread, part – cake*.

¹⁶¹Winston M., Chaffin R, Herrmann D. 1987. A Taxonomy of Part-Whole Relations // Cognitive Science, 11, – pp. 417-444.

¹⁶²Cruse D. 1986. Lexical Semantics. Cambridge. University Press. – 310 p.; Winston M., Chaffin R, Herrmann D. 1987. A Taxonomy of Part-Whole Relations // Cognitive Science, 11, – pp. 417-444.

¹⁶³Лукашевич Н.В. Отношения часть-целое: теория и практика. // «Нейрокомпьютеры: разработка, применение». – Москва: Радиотехника, 2013.– С. 11.

Homomeric or non-homeo-dimensional parts are completely different, for example, a tree – a forest, a table is furniture.

3. *The individual parts* are relatively separate from the whole. For example, a drawer – a table (the drawer can be separated from the table), a handle is a cup (inseparable parts).

On the basis of these 3 combinations of signs, six types of occurrences of part and whole relations are distinguished (the dissertation describes in detail the types of holonymy with examples).

The Wordnet database identified three types of meronymic relationships associated with a particular concept¹⁶⁴: 1) a partial meronym: a “wheel” is part of a “machine”; 2) the meronym of the participant: “car” – the participant of “congestion”; 3) Meronym of matter: “wheel” is made of “rubber”.

In the process of studying the bases of different ontologies and research on them, it was observed that the principles and interpretations of the establishment of whole-part relations are different in all lexical bases and there is no common norm in the studied sources. However, in a meronymic (paronymic) relationship, it should be noted that the sign connecting the whole and the part is determined by the fact that they perform the same function.

The third chapter, “*Hypo-hypernym relations in ontological dictionaries*”, examines the relationship between genus and species. In the study of the lexical richness of language as a system, it is important to rely on the hypo-hypernymic relationship of lexemes: it expresses meanings of lexemes in the language which are the names of things and events in the nature and society and allows generalization and differentiation of ideas about happenings in the being through these meanings¹⁶⁵. Moreover, hyponymic connections of words formed as a result of lexical classification are considered to be the most important way of hierarchical organization of a dictionary¹⁶⁶.

The phenomenon of hyponymy was extensively studied by French linguist Dj.Layonz, Sh.Balli, V.G.Gak; from Russian researchers Yu.N.Karaulov, A.Vejbiskaya, D.N.Shmelev, A.A.Ufimseva, L.A.Novikov, M.V.Nikitin, Ye.E.Kotsov¹⁶⁷. The separation of hyponymic relations in Uzbek linguistics and its reference to the general public is connected with Rohatoy Safarova's research.

¹⁶⁴<https://www.greelane.com/ru/what-is-a-meronym-1691308/>

¹⁶⁵Look at: Новиков Л.А. Семантика русского языка. – М.: Высшая школа, 1982, – С.136-142.; Ne'matov H, Rasulov R. O'zbek tili sistem leksikologiyasi. – T.: O'qituvchi, 1995, 111-123-b.; Qo'chqortoyev I. So'zlarning leksik-semantik to'dalari haqida. // ToshDU ilmiy asarlari, 359-chiqishi. – Toshkent, 1969.; Расулов Р. Лексико-семантические группы глаголов состояния и их валентность. – Тошкент, 1991.; Сафарова Р. Гипонимия в узбекском языке. АКД. – Тошкент, 1990.

¹⁶⁶Котцова Е.Е. Гипонимические связи глаголов и существительных в лексической системе русского языка. // Вестник Нижегородского университета им. Н.И.Лобачевского, 2011. № 6 (2), – С. 324-327

¹⁶⁷Уфимцева А.А. Семантика слова // Аспекты семантических исследований. М., 1980. С. 5-80; Степанов Ю.С. Имена, предикаты, предложения (Семиологическая грамматика). М., 1981. –360 с.; Новиков Л.А. Семантика русского языка. – Москва: Высшая школа. 1982. – С.241-243.; Никитин М.В. Основы лингвистической теории значения. – Москва: Высшая школа, 1988. – С.73-87.; Котцова Е.Е. Гипонимия в лексической системе русского языка (на материале глагола): Автор.дисс...д.ф.н. – Архангельск, 2010. – С. 10.

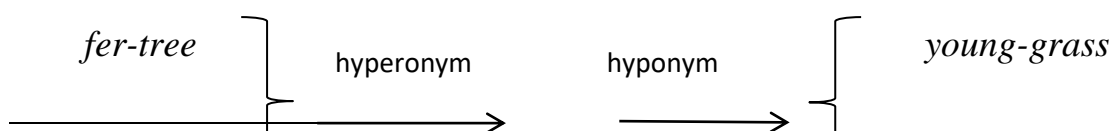
This is the first study of genus-species relations in Uzbek¹⁶⁸. R.Safarova divided the names of about 1000 animals in Uzbek into ten semantic groups and revealed hyponymic (genus-species) relations between them. In her research, D.Akhmedova also spoke about the role of lexical-semantic relations in the tagging of adjectives. She analyzed the research done in this regard.¹⁶⁹

The hyponymic sequence of words consists of a hypernym and a hyponym. The broader word is called a hypernym, and the narrower word is called a hyponym: dishes (hypernym), plate, cup (hyponym). More precisely, the hypernym is genus, i.e., dominant, or, as A.Nurmanov says, a homonym; a hyponym is a lexical unit that is semantically richer than a hyperonym, but in its place is dependent on a word that represents the names of certain species of genus and includes a semantic word in its semantic structure.

According to A. Sobirov, genus-species (hyper-hyponymic) relationship is the basis of semantic fields at the lexical level. Genus-species relations continue to emerge when members are grouped into cells, cells into larger groups, larger groups into groups, and groups into semantic fields¹⁷⁰. Obviously, the smaller things go into the bigger things. Each lexeme in the paradigm has a hyponymic status. Each hyponym, in turn, can combine several lexemes, and it becomes a hyperonym relative to others in the group.

Hypernym (genus / homonym)	Hyponyms(species/homonymous)
Prosepoetry	Drop, feuilleton, sketch, novella, story, povest, novel
Poetry	Ghazal, tuyuk, rubai, fard, masnavi, ode, white poem, poem, epic
Drama	Drama, comedy, tragedy, tragicomedy
Literature	Prosepoetry, poetry, drama
Art	Literature, sculpture, painting, cinema, theater

On the one hand, the concept of *a tree* is generally associated with the words *bush* and *grass* as a whole, clear, real concept. *The tree* hypernym is associated with words that represent species of the same genus as a word that expresses the concept of genus. On the other hand, it can be used instead of a lexical unit denoting a species, where the word *tree* is a hyponym. However, the word hypernym, which refers to the concept of genus, does not fully and accurately express the meaning of the species when the lexical units that represent the name of the species are used instead of hyponyms. Therefore, the concept of species that exists in our minds needs to be clearly expressed in language.



¹⁶⁸Safarova R. O'zbek tilida giponimiya. // O'zbek tili va adabiyoti. – T., №1. 1987.; Сафарова Р. Гипонимия в узбекском языке (на материале общеупотребительных зоонимов): Автор-дисс.учен.ст.канд.ф.н. – Ташкент, 1990. – 20 с.; Safarova R. Leksik-semantik munosabat turlari. – Toshkent: O'qituvchi, 1996.

¹⁶⁹Ahmedova D. Atov birliklarini o'zbek tili korpuslari uchun leksik-semantik teglashning lingvistik asos va modellari: Filol.fan.bo'yicha falsafa doktori (PhD)...diss. – Buxoro, 2020. – 145 b.

¹⁷⁰Sobirov A. O'zbek tilining leksik sathini sistemalar sistemasi tamoyili asosida tadqiq etish. – Toshkent: Ma'naviyat, 2004. – 104 b.

willow
elm
poplar

tree

bush
flower
branch

Each phrase in the WordNet model has its own set of relationships. It should be noted that in various computer applications, in most cases, synonyms, antonyms, hyponymy (hypernymy), meronymy (whole-part) are used interchangeably. The main relationship between noun synsets is the genus-species relationship¹⁷¹ that species synset is called hyponym and genus synset as hypernym. These are transitive hierarchical relationships, also known as *is A* relationships. If English speakers consider sentences of the type “*An X is a (kind of) Y*” to be the norm, then the X – synset is said to be the hyponym of the Y-synset¹⁷². In this case, the relationship between the synsets forms a hierarchical structure. In the creation of hierarchical systems based on genus-species relations, it is generally assumed that the characteristics of higher concepts are inherited, i.e., the hereditary nature. The horses in the WordNet model are organized in the form of a hierarchical system with the following inheritance: for each synset, its own unique concept, systematic connections are made in order to find its hypernym.

In the third chapter, entitled “*Types of antonymic relations in the lexical database*”, counter correlates, vector correlates, conversions, pragmatic antonyms, enantiosemia, enanthema according to the expression of concepts of antonyms; different and monolithic in structure; linguistic (usual) and textual (contextual, verbal, occasional) in terms of language and speech; proportional and disproportionate in terms of action; types of words according to form and meaning were studied.

The fourth chapter of the dissertation is entitled “**Principles of creating a linguistic ontology UzNet**”. It outlines the concept of creating the UzNet linguistic ontology, discusses the problem of tagging word groups in the UzNet database, and discusses in detail the possibilities of a syn set database and synonymizer in UzNet.

In the section *The Concept of Designing the creation of UzNet Linguistic Ontology* the UzWordNet linguistic resource referenced in the international WordNet system was analyzed, its non-compliance with ontology requirements was found out.

Problems related to the social nature of language arise in shaping the relationships between natural language concepts, words, and expressions. In addition, establishing a clear and stable relationship between ontological concepts is also a difficult issue for the pragmatic aspect of language, especially when it comes to establishing a lexical unit-concept (or its interpretation). In this case, it is necessary to form a complete and complete explanation of each lexical unit (LB) in the natural language. To do this, you must first (1) summarize the entire

¹⁷¹Miller, G. Nouns in WordNet / G. Miller // WordNet – An Electronic Lexical Database. – The MIT Press, 1998. – P. 23-47.

¹⁷²Miller, G. 1998.

vocabulary, (2) define the category of each LB, and (3) formulate a clear and concise explanation (s) of that LB in a hierarchical (or numbered) order. This principle is common, and the formation of the ontology of the Uzbek language on the basis of this principle gives the expected results.

In the course of our research, the requirements for the creation of a linguistic ontology, the principles of creating an information retrieval thesaurus were studied. If the thesaurus creates a hierarchy of all lexical riches in one field, then in linguistic ontology it is necessary to create a whole network of lexicons of languages. Therefore, the creation of thesauruses is relatively successful and is done in a relatively short time compared to ontology. Creating a network of words from scratch or expanding an existing one is a laborious process that involves several steps and requires extensive use of human labor and automated systems. Therefore, a concept was developed to create an ontology of the Uzbek language – UzNet.

“CONCEPT OF CREATION OF THE UZBEK LANGUAGE ONTOLOGY”

Contents:

- 1. The study of the conceptual basis of the creation of linguistic ontology**
- 2. Creating a lexicographic database for the UzNet system**
- 3. Tagging of word groups**
- 4. Defining the structural and semantic relationship of words**
- 5. Synsets (lexical, of meaning and textual synonyms)**

Each stage has several structural stages.

In this study, steps 1.1, 1.2, and 1.3 of the first phase of the concept are reflected. In practice, steps 2.1, 2.2, 2.3, and 2.4 were formed as a lexicographic base. Creating a linguistic ontology is a multi-step process that requires a great deal of research. Great team work gives great results.

Uzbek language ontology – the field of application of UzNet, its types, language, significance, users are as follows: **I. Language** – formalized Uzbek language¹⁷³. **II. Type:** 1) by degree of generalization: a) general ontology; b) field ontologies; c) application ontologies; 2) in depth of development: a) glossary; b) thesaurus; c) a set of taxonomies; e) a set of logical constraints. 3) by subject area: a) ontology of official texts; b) ontology of fields of knowledge; c) industry ontology; e) ontology of material and intangible concepts. **III. Application status:** 1) general semantic system; 2) knowledge management system; 3) educational technology; 4) expert systems; 5) a means of processing natural language; 6) application integration; 7) the basis of modeling and design. **IV. Significance:** 1) the basis of language learning; 2) the primary source of artificial intelligence; 3) source of knowledge; 4) integration of different areas. **V. User:** 1) person; 2) group; 3) the organization; 4) the state.

WordNet – type linguistic resources are designed to describe the vocabulary of a language in accordance with linguistic traditions. However, any information

¹⁷³view: Abjalova M. Linguistic modules of edit and analysis programs. [Text]: Monography / M.A. Abjalova. – Tashkent: Nodirabegim, 2020. – 176 p.

system deals not only with a general dictionary, but also with specific disciplines and their terminology. An analysis of attempts to create terminological resources based on WordNet concluded that the structure of WordNet was not adapted to describe terminology. Specific descriptions of word groups, a very large set of unrelated meanings, and insufficiently developed principles for entering multi-syllable expressions all lead to problems in the development and use of terminological resources based on the WordNet model. Therefore, the development of a concept based on the nature of each language is considered expedient.

In the section of the chapter “*The matter of tagging word groups in the UzNet database*” the necessity of defining a group of words in the processing of natural language was proven, which is currently aimed at creating an artificial intelligence system. From the world of computer linguistics, it is known that this linguistic-technical analysis is a stage of word group identification – POS-tagger, i.e., tagging of word groups, as well as the process of automatic processing of texts. Early tagging efforts to create language corpora have addressed many of the text's most pressing issues to date. In NLP, the process of natural language processing, word grouping is also a primary task, which results in the analysis of various linguistic ambiguities in the text, such as the definition of homonymy and the semantics of polysemous words.

In the “Explanatory Dictionary of the Uzbek language” there are four explanations for the word *long*: 1) the distance is large, relatively far; distant; 2) continuous large, many over time; 3) friendship, kinship, acquaintance, etc. the connection is not so close, not so close; 4) (mobile) no direct connection; unrelated; without connections¹⁷⁴. The color of expression in works of art, in accordance with the situation of realization in everyday life, gives rise to additional meanings of the word *long*: 1) the past (history); 2) thinking about the future, thinking about the consequences (future); 3) quantity, many; 4) length measurement; 5) superficial approach to work (superficiality); 6) not continuous, ends quickly (short time); 7) a distant place.

Identifying parts of speech is a difficult process¹⁷⁵. This is because it is not possible to tag all Uzbek words universally within 12 categories. A word can be polyfunctional depending on the state of its realization in the sentence and the semantic valence of the N-gram¹⁷⁶ words. For example: In the sentences “*The patient was brought to the hospital*” and “*The patient man was brought to the hospital*”, the word *patient* is a category of noun according to the sign of its case (the word in accusative case that answers the question “Who?”). and (answers the question how?) is a word in the function of the category of quality. Out of 11,000 borrowed words in the “Dictionary with the stress of borrowed words of the Uzbek language”¹⁷⁷, 66 similar polyfunctional words were identified¹⁷⁸.

¹⁷⁴Ўзбек тилининг изоҳли луғати: 80 000 дан ортиқ сўз ва сўз бирикмаси (А. Мадвалиев таҳрири остида). 4-жилдли. – Тошкент: Ўзбекистон миллий энциклопедияси, 2006. – Б. 268.

¹⁷⁵https://en.wikipedia.org/wiki/Hidden_Markov_model

¹⁷⁶View: Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – B. 73-77.

¹⁷⁷Ўзбек тилининг изоҳли луғати: 80 000 дан ортиқ со‘з ва со‘з бирикмаси (А. Мадвалиев таҳрири остида). 5 жилдли. – Тошкент: Ўзбекистон миллий энциклопедияси, 2006.

The tagging of word groups is necessary in the following processes: 1) in grammatical tagging of word formations in the corpus; 2) in linguistic ontology in the correct and complete formation of word interpretation and intergroup relations; 3) in determining the synonymousness and / or homonymy of a word; 4) is important in the syntactic analysis of a sentence. Most importantly, WG tags are the most essential linguistic element for Natural Language Processing (NLP), so WG tagging is performed as a preliminary demand to simplify various problems in NLP.

There are words that do not have a categorical sign or that the contextual meaning of the sentence confuses the reader. Look at the following examples: “... *test sinovlaridan o‘tkazish yuzasidan shaxsan javobgarligi belgilab qo‘yilsin*”¹⁷⁹, (determine **personal** responsibility regarding the conduction of the test), “*Shaxsan o‘zim keldim*”, (I came **in person**), “*Shaxsan bajardim*”, (I did it **in person**), “*Bular hammasi lotincha yoki lotinchaga yaqin so‘zlar. Men, shaxsan, shunday deb bilaman*”(These are all Latin words or words close to Latin. I **personally** consider this that way). (A. Kahhor, Literature Teacher). It is difficult to identify the word group of the word formation *personal* in those sentences. In some places, a lexeme (pronoun) is used instead of a personal pronoun, and in some cases it is used as an adverb lexeme. In this case, the categorical features of the groups are used to determine the group of the word. There are four characteristics of it¹⁸⁰: semantic, syntactic, morphological, and word-formation.

It is known that as a result of the adding word-forming suffixes to 12 word groups in the Uzbek language (independent word groups: noun, verb, adjective, adverb, numerals, pronoun; dependent word groups: conjunction, auxiliary, particle; separate word groups: modal, onomatopoeic, exclamatory) 4 word groups are formed: noun, verb, adjective, adverb. Among the identified forming suffixes (337: 114 noun forming suffixes, 58 verb forming suffixes, 117 adjective forming suffixes, 48 adverb suffixes)¹⁸¹, the -an is an adverb affix. Based on this parameter, it can be concluded that because of uniting the suffix – an with the word “*person*” belonging to the group of nouns an artificial adverb is produced: *shaxs(person) (Noun) ∪ {-an} => shaxsan(personal)*.

In most cases, word groups are tagged using the following methods (methods, algorithms)¹⁸²: 1) rule-based method; 2) stochastic (or statistical) method.

Rule-based POS tags. One of the oldest tagging methods is POS-tagging based on these rules. The Brill method is mainly useful in this case¹⁸³. Rule-based taggers use a dictionary or lexis to tag each word. If a word (here multifunctional,

¹⁷⁸View: Qurbonova M., Abjalova M. va boshq. O‘zbek tili o‘zlashma so‘zlarining urg‘uli lug‘ati. [Matn]: o‘quv-uslubiy lug‘at / M.Qurbonova, M.Abjalova, N.Axmedova, R.To‘laboyeva. – Toshkent: Nodirabegim, 2021. – 988 b.

¹⁷⁹<https://lex.uz/acts/-4276890>

¹⁸⁰https://ru.wikipedia.org/wiki/Часть_речи

¹⁸¹view: Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – B. 122-123.

¹⁸²<https://www.freecodecamp.org/news/an-introduction-to-part-of-speech-tagging-and-the-hidden-markov-model-953d45338f24/>; <https://coderlessons.com/tutorials/akademicheskii/obrabotka-estestvennogo-iazyka/pometka-chasti-rechi-pos>; <https://habr.com/ru/post/125988/>

¹⁸³ Brill E. 1992. A simple rule-based part of speech tagger //Proceedings of ANLC. – P. 154.

homonymous, synonymous words are meant) has a number of tags, then rule-based taggers use handwritten rules to correctly identify the categorical tag of a word in a sentence. In order to give more precise tags, the linguistic features of the word are defined on the basis of rules by analyzing the words that have come before and after it. For example, a linguistic unit that comes after a word related to the name in possessive case is a word in noun group that has a possessive suffix. For example, *my book, my brother's house, Salima's dress*. Thus, in this case, the fact that the word is in the group of nouns is defined by the name in coming before itself. Let's look at an example from English: if the first word is an article, then the word following it is a lexical unit of the noun group. For example, *an egg, a book, the train, the windows*.

Such cases in POS tags are coded in the form of rules. These rules may include: 1. *Rules based on linguistic norms*. Hundreds of rules based on the spelling rules of the language are formed in the form of a base of general, private and exceptional rules¹⁸⁴. 2. *Contextual template rules*, that is, a regular connotative use of a word with a figurative sense in a sentence is kept in the program memory, as a result of which inaccuracies associated with that connotative word are eliminated in subsequent processes.

The stochastic tagging method is based on frequency or probability (statistics). Therefore, in some sources it is explained as a statistical or probabilistic method¹⁸⁵. *In the frequency approach*, stochastic taggers eliminate grammatical inaccuracies based on the probability that a word will meet a particular tag in the text. The probability of a sequence of tags, or the n-gram method, calculates the probability of using a given sequence of tags.

The lexical units of the Uzbek language ontology contain 84,094 language units, including 47817 nouns, 17081 verbs, 14727 adjectives, 2644 adverbs, 276 numerals, 240 pronouns, 65 conjunctions, 111 auxiliaries, 18 particles, 115 modals, 174 exclamatory and 822 imitation words are tagged in a rule-based manner.

The fourth chapter, entitled “*A base of synsets and synonymize capabilities in UzNet*”, describes the types of synonyms, the problem of their study, the capabilities of the synonymize program in educational corpus of the Uzbek language and, most importantly, provides information about the study of quasi-synonyms.

By the principle of creating Princeton WordNet, all synonyms in a sentence should be reflected in the lexical database. So as to do it, of course, the relationship between the concepts that are possibly synonymous should be defined. One of these kinds of phenomena is quasi-synonyms, which occur in the course of speech

¹⁸⁴view: Abjalova M. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – 176 b.

¹⁸⁵<https://www.freecodecamp.org/news/an-introduction-to-part-of-speech-tagging-and-the-hidden-markov-model-953d45338f24/>; <https://habr.com/ru/post/125988/>; https://ru.wikipedia.org/wiki/Частеречная_разметка; https://en.wikipedia.org/wiki/Part-of-speech_tagging#:~:text=In%20corpus%20linguistics%2C%20part%20of,its%20definition%20and%20its%20context.

and, in some cases, they can be mentioned in the dictionary as synonyms and currently it is one of the important issues that needs to be studied.

The element of the term “**quasi-**” is derived from the Latin word meaning “*imaginary, not real, false, imaginary*”. Quasi-synonyms (fake synonyms, partial synonyms¹⁸⁶) are words that are relatively close in meaning, but not in all cases they are interchangeable.

There are the following types of quasi-synonyms¹⁸⁷: 1) quasi-synonymous terms that have similar meanings, for example: *home - building, talent – genius*; 2) partial synonyms in which the meaning and scope of one word apply to another field, for example: *metal – iron*; 3) synonymous words (such as quantitative characteristics of a single phenomenon) used as a form of synonyms in the process of speech occurrence of words having opposite semantics, for example: *hardness – softness, transparency – darkness*.

It is known that lexical synonyms are mainly lexical units with similar meanings belonging to the same word group. Quasi-synonyms are also formed by the synonymousness of words belonging to different word groups. For example: *lemmatization* (analysis of the word up to the base morpheme; specification of the base) - (*N – noun*) ↔ *lemmatizing-* (*V_h – verb_{name of the action}*); *regeneration* (*N–noun*) ↔ *regenerating* (*V_h – verb_{name of the action}*); *horror* (*N – noun*) ↔ *great* (*Adj – adjective*) (verbally). To conclude, although quasi-synonyms (partial synonyms) in the process of speech occurrence are close in meaning, their categorization varies. Quasi-synonyms include combination or intersection semantics.

CONCLUSION

1. The creation of a linguistic ontology is a promising area of modern research in the field of natural information processing, a system that reflects the potential of natural language. With the formation of linguistic ontology, a number of problems have been successfully begun to solve using automated systems. That is why, at the current time the number and quality of applications that include successful usage of many linguistic ontologies developed by a particular algorithm is expanding.

2. Ontology is a network-based lexical database based on knowledge of language and the world, embracing terms among fields and formed grounded on relations between them. Synonymous lines are the primary components of lexical database interactions. Hyponymy and holonymy are also important for search engines. Thesauruses and ontologies are also necessary sources in the processing of natural language.

3. The terms *linguistic ontology* or *language ontology* are more commonly used in the field of information technology than in linguistics, which mainly means information retrieval thesaurus specialized for automatic processing of texts, i.e., various special linguistic software gathering language vocabulary in itself and establishing (or having a branch of words) semantic relations of the words.

¹⁸⁶ <https://ru.wiktionary.org/wiki/quasi>

¹⁸⁷ [Квазисинонимы – Википедия \(wikipedia.org\)](https://ru.wikipedia.org/wiki/Квазисинонимы)

Linguistic ontologies cover most part of the words related to a language or subject area; at the same time, it is an ontological structure in which there is a relationship between concepts. Therefore, linguistic ontologies can be considered as a special type of lexical database and a separate type of ontology.

4. Formally, an ontology is a system that consists of a set of concepts and a set of assertions about concepts which can be used to construct classes, objects, relationships, functions, and theories.

5. Thesauruses are lexicographical sources that cover concepts in a collection of texts or language corpuses, concepts, definitions, and terms of a special field of knowledge or field of activity. While linguistic ontology encompasses the full range of linguistic capabilities, thesauruses are limited to a relationship of concepts to concerning a certain set or direction or field

6. With the development of linguistic ontology, it is possible to use them as construction blocks of knowledge base components for large software systems, as well as construction blocks of object diagrams in object-oriented systems and conceptual schemes of databases. Moreover, they can be used in automatic semantic and pragmatic analysis of texts, natural language processing, to further develop the artificial intelligence system and increase opportunities in the field of robotics. Ontologies create an opportunity for the formation of knowledge bases for solving problems in various intellectual, in particular, expert systems.

7. In studying the lexical richness of language as a system, it is important to rely on the hypo-hypernymy relations of lexemes: it provides with the meanings of lexemes that are the names of things and subjects, phenomena in a language occurring in nature and society, and by these meanings allow generalization and differentiation of concepts about things and phenomena themselves in being. Hyponymy is a semantic relationship that results from a hierarchical relationship of lexemes in the vocabulary. The essence of such relationships is that lexemes that express a narrower concept or meaning are in a relationship of species (hyponym) and genus (hypernymy) with lexemes that represent a broader concept or meaning. This type of relationship is accomplished through combining (integral) semantics.

8. In information searching and ontological systems, it is desirable to establish a strict hierarchical relationship to obtain accurate information. We can only include a textual whole-part relationship into a “partial whole-part relationship”.

9. The proposed approaches (definitions) in defining the concept of ontology help to create a single system of terms that are convenient for discussing ontologies and their characteristics and to provide a single definition reflecting the function, structure, relationships and importance of ontology. Thus, ontology is a network-based lexical database that bases on knowledge of language and the world, embraces terms among fields and is formed on the basis of the relationship between them.

10. The structure of the thesaurus, the principles of describing its units and relations are important in improving the quality of solving information searching problems. RuThes thesaurus (ontology according to its structure) differs from both WordNet-type ontology (thesaurus in some sources) and traditional data retrieval

thesaurus in a number of important developmental principles. Qualitative coordination of knowledge and models in RuThes improves the quality of this ontology by 10-15%, knowledge gained in RuThes enables to tackle the problem faster and better than using statistical and machine teaching methods in finding solutions for a number of problems.

11. Hierarchy is the entire structure in its general form, in which the elements are interconnected with its relevancy relations from top to bottom, from general to specific, from primary to secondary, from genus to species. The construction of the hierarchy is mainly based on the hierarchical principle, which assumes the distribution of elements in parallel levels according to their value. The synsets of linguistic ontology are based on this hierarchy, and in turn the concepts are branched and interconnected.

12. Synsets are the basic elements of linguistic ontology, a collection of synonyms. The semantic relationships in synsets lead to the formation of a network of concepts, and the existence of synonymy among word groups has given a rise to the emergence of species called quasi-synonyms. Such networking is important for anti-plagiarism programs, expert systems, speech analysis programs, semantic analysis programs, information retrieval systems.

13. While in thesauruses a hierarchy of all lexical wealth is created in one field, it is necessary to create a whole network of language vocabulary in linguistic ontology. Therefore, the creation of thesauruses is relatively successful and is done in a relatively short time compared to ontology. Creating a network of words from scratch or expanding an existing one is a very difficult process that involves several steps and requires extensive use of human labor and automated systems.

14. In computer linguistics, it is an important task to organize an effective search on the Internet on the basis of the ontology of the Uzbek language, to provide the necessary information related to the linguistic object that is being sought through its parts, whole, type, gender or synonyms, and contradictory meanings.

15. In the process of examining linguistic ontologies, thesauruses, information searching thesauruses, technologies for creating WordNet-type linguistic resources, especially, it has become clear that the creation of language ontology is a very difficult and arduous working process. Therefore, systematic work on the basis of the “Ontology of the Uzbek language – the concept of creating the UzNet system” will ensure the success of the result. In this study, steps 1.1, 1.2, and 1.3 of the first phase of the concept are reflected. Practically, steps 2.1, 2.2, 2.3, and 2.4 were formed as a lexicographic base. Since creating a linguistic ontology is a multi-step complicated process that requires a great deal of further research. Large team work gives great results.

16. It is significant to create an ontology of the Uzbek language for the purpose of information retrieval, natural language processing, machine translation, formalization of the Uzbek language for artificial intelligence and transformation of the Uzbek language into the Internet language. In this case, WordNet cannot be the base. The reason behind it is that the automatic performance of many stages in

the creation of WordNet resources can cause major problems. This is because the features of one natural language (English) are not compatible with the features of another natural language (such as Uzbek). Therefore, the initial steps in creating UzNet have to be done manually. This process is carried out on a rule-based method, in which after most of the data is processed, the database is enhanced using a stochastic method.

17. For linguistic ontology, tagging of word groups play an important role. Its work process also continues with the conduction of morpho-analysis after graphemic analysis. In this process no synonymousness or synonymousness, homonymousness or polyfunctionality of words is defined. Once a word's group and state of manifestation are determined, its semantic relations are established. In order for these principles to be accurate, it is important that each group of words must be tagged correctly.

18. Linguistic ontologies play a significant role in solving problems regarding machine translation, question and answer systems, information searching, knowledge gaining systems, communicating systems between a computer and a person, language comprehension systems, as well as knowledge demonstration, artificial intelligence, and computer data processing. It is important in solving many problems. In linguistics, in particular, ontologies are used in the semantic annotation of the corpus of a text, machine translation, as well as in automatic tackling of synonymousness and defining of context-based homonymy, and the creation of low-level ontological resources, dictionaries, and thesauruses. In addition, multilingual ontologies are utilized by translators as sources of information, including knowledge and relevant vocabulary.

**НАУЧНЫЙ СОВЕТ DSc.03/30.12.2019.Fil.05.02 ПО ПРИСУЖДЕНИЮ
УЧЁНОЙ СТЕПЕНИ ПРИ ФЕРГАНСКОГО ГОСУДАРСТВЕННОГО
УНИВЕРСИТЕТА**

**ТАШКЕНТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
УЗБЕКСКОГО ЯЗЫКА И ЛИТЕРАТУРЫ ИМЕНИ АЛИШЕРА НАВОИ**

АБЖАЛОВА МАНЗУРА АБДУРАШЕТОВНА

ПРИНЦИПЫ СОЗДАНИЯ ОНТОЛОГИИ УЗБЕКСКОГО ЯЗЫКА

10.00.11 – Теория языка. Прикладная и компьютерная лингвистика

**АВТОРЕФЕРАТ ДИССЕРТАЦИИ ДОКТОРА
ФИЛОЛОГИЧЕСКИХ НАУК (DSc)**

Ташкент – 2022

Тема диссертации доктора наук (DSc) зарегистрирована в Высшей аттестационной комиссии при Кабинете Министров Республики Узбекистан под номером B2021.4DSc/Fil364.

Диссертация выполнена в Ташкентском государственном университете узбекского языка и литературы имени Алишера Навои.

Автореферат диссертации доступен на трех языках (узбекском, английском, русском (резюме)) на веб-странице Ферганского государственного университета (www.fdu.uz) и на Информационно-образовательном портале «ZiyoNet» (www.ziynet.uz).

Научный руководитель:	Дадабоев Хамидулла Арипович доктор филологических наук, профессор
Официальные оппоненты:	Хошимов Ганижон Мирзаахметович доктор филологических наук, профессор Уринбаева Дильбар Базаровна доктор филологических наук, доцент Эшреф Адали доктор технических наук, профессор
Ведущая организация:	Институт узбекского языка, литературы и фольклора

Защита диссертации состоится на заседании Научного совета DSc.03/30.12.2019.Fil.05.02 при Ферганском государственном университете « 5 октября 2022 г. в ___ часов. (Адрес: 100100, 100151, г. Фергана, ул. Мураббайлар, 19. Тел.: (99873) 244-66-02; факс: (99873) 244-44-01; e-mail: info@fdu.uz).

С диссертацией можно ознакомиться в Информационно-ресурсном центре Ферганского государственного университета (зарегистрирован под номером 159) (Адрес: 100100, 100151, г. Фергана, ул. Мураббайлар, 19. Тел.: (99873) 244-71-28.

Автореферат диссертации разослан май «24» 2022 года
(Реестр протокола рассылки № 44 от 2022 года «27» май).


Ш.Искандарова
Председатель научного совета
по присуждению ученых степеней,
д.ф.н., профессор
М.Зокиров
Ученый секретарь научного совета
по присуждению ученых степеней,
к.ф.н., доцент
А.Мамажонов
Председатель семинара при научном совете
по присуждению ученых степеней,
ф.ф.д., профессор

ВВЕДЕНИЕ (автореферат докторской диссертации)

Актуальность и необходимость темы исследования. Важно отметить, что развитие современных отраслей человеческой деятельности, требующих знаний, связано с возрастанием роли компьютерных технологий во всем мире. В настоящее время, поскольку информационные потоки значительно возрастают, возникает необходимость их защиты, оформления и упорядочивания, а также поиска новых способов автоматической обработки. В связи с этим возрастает интерес к широкому спектру информационных баз, которые можно использовать в практическом плане. Тем более, спрос на системы на основе нейросетей, выводящие из текста любую информацию без учета человеческого фактора, огромен. В половине 20-го века появились семантические сети вместе с глобальной веткой, которая была снабжена дополнительными тегами, несущими информацию о семантике элементов гипертекстовых страниц. Неотъемлемой частью семантической сети является понятие онтологии, которая представляет собой лексическую базу данных, состоящую из ветвей слов.

Цель исследования заключается в исследовании структуры лексических баз данных и принципов разработки лингвистических онтологий для цифровых технологий и искусственного интеллекта, а также в формулировании концепции создания UzNet – онтологии узбекского языка.

Объектом исследования выбраны «Толковый словарь узбекского языка», синонимы в узбекском языке и лингвистическая информация на основе узбекскоязычного учебного корпуса.

Предмет исследования составляют разработка лексической базы данных WordNet, технология разработки лингвистических онтологий, семантические связи в лексических базах данных WordNet, KeNet, RuTez и концепция создания лингвистической онтологии UzNet.

Научная новизна исследования заключается в следующем:

доказано, что на основе философско-лингвистических принципов успешно определяются дифференциальные аспекты междисциплинарного разветвления интегральных и лексических единиц с точки зрения охвата ими семантических отношений лингвистической онтологии и тезауруса;

определены требования концептуализации к онтологии как базы лексической данных и математические модели, в которых отражается, что в лингвистической онтологии широкомасштабная междисциплинарная терминологическая база основывается на семантическом разветвлении;

на основе английской онтологической системы Princeton WordNet обоснованы системная технология создания онтологии узбекского языка (UzNet) основанной на семантических отношениях в синсете (синонимических сборниках), принципы предоставления в ней лингвистической и экстралингвистической информации;

определены первичные элементы лингвистической онтологии: формальные принципы образования набора слов-синонимов (синсетов), асимметрия, рефлексивность, транзитивность и наследственность в

отношениях вида и рода (гипоним-гипероним), целое-часть (холоним-мероним), доказано, что она варьируется на основе функциональных, гомеомерных и отдельных частей, порождающих отношения целое-часть;

обоснованы национально-лингвистические особенности концепции создания онтологии узбекского языка на основе сопоставлении со структурной технологией специфической таксономической и семантической веб-системы турецкого языка KeNet;

обоснованы иерархический порядок интерпретации значений слов в UzNet`е, стохастические (теоретико-вероятностных и статистических) и основанные на правилах методы пометки частей речи, актуальные аспекты создания набора квазисинонимов в онтологии узбекского языка и определения лингвистического синкретизма, а также принципы построения базы данных о формальных отношениях.

Практические результаты исследования следующие:

– разработана «База данных слов-синонимов узбекского языка» для коллекций синсетов (синонимов) лингвистической онтологии UzNet, получено авторское свидетельство;

– созданы базы данных омонимов, антонимов и паронимов для узбекской лексической базы данных UzNet и получены авторские свидетельства;

– разработана база заимствованных слов с ударением для теггирования частей речи заимствованных слов и дан толкования 11 000 заимствованных слов на узбекском языке и получено авторское свидетельство;

– разработана морфологическая база данных узбекского языка и система орфографических правил для процесса лемматизации онтологии узбекского языка;

– созданные формальные орфографические правила, база данных словаря словообразования, база паронимов использованы в практическом проекте «Формирование образовательного корпуса узбекского языка» № АМ-ФЗ-201908172, данные раздела «Фонетика», разработанные для универсальной грамматике узбекского языка и комплекс заданий, направленных на развитие устной компетенции использованы в проекте учёных женщин № ПЗ-2020042022 «Создание лингводидактической электронной платформы тюркских языков».

Внедрение результатов исследования. На основе полученных научных результатов по концепции онтологии узбекского языка:

результаты по определению того, что на основе философско-лингвистических принципов успешно определяются дифференциальные аспекты междисциплинарного разветвления интегральных и лексических единиц с точки зрения охвата ими семантических отношений лингвистической онтологии и тезауруса, использованы в фундаментальном проекте за № И-ОТ-2019-42 «Создание электронного поэтического словаря узбекского и английского языков (изображение облика человека, поведения, природы и национальной символики)», выполненном в Ташкентском

государственном университете узбекского языка и литературы имени Алишера Навои (справка № 04/1-2341 Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои от 26 ноября 2021 года). В результате освещены роль и значение слов холонимов (целое), меронимов (часть), гиперонимов (род), гипонимов (вид) в описании внешности человека, характера, природы и национальных символов; выявлены смысловые соотношения многозначных слов в узбекском и английском языках;

материалы и результаты исследования по определению первичных элементов лингвистической онтологии как формальные принципы образования набора слов-синонимов (синсетов), асимметрия, рефлексивность, транзитивность и наследственность в отношениях вида и рода (гипоним-гипероним), целое-часть (холоним-мероним), о том, что они варьируется на основе функциональных, гомеомерных и отдельных частей, порождающих отношения целое-часть, используется в параграфах учебника «Методика обучения узбекскому языку» как «Принципы формирования компетентности у учащихся», «Лексико-семантические отношения» и «Принципы словосочетаний и их классификации» (справка № 04/1-2341 Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои от 26 ноября 2021 года). В результате освещаются факторы, формирующие речевую компетенцию и языковые компетенции, на примерах раскрывается связь синонимии и гипонимии, показано значение лексико-семантических отношений в языковом обогащении, выделена необходимость классификации частей речи;

результаты и материалы исследования по иерархическому порядку интерпретации значений слов в UzNete, стохастическим (теоретико-вероятностных и статистических) и основанным на правилах методам пометки частей речи, актуальным аспектам создания набора квазисинонимов в онтологии узбекского языка и определения лингвистического синкретизма, а также принципам построения базы данных о формальных отношениях использованы в учебнике «Современный узбекский язык» (Разрешение № 285-078 в соответствии с Приказом Министерства высшего и среднего специального образования № 285 от 4 мая 2020 г.). В результате в учебнике иллюстрируется на примерах явления синонимии и гипонимии, разъясняются их виды, освещаются семантическая классификация слов и явление лексиколизаия;

требования концептуализации к онтологии как базы лексической данных и математические модели, в которых отражается, что в лингвистической онтологии широкомасштабная междисциплинарная терминологическая база основывается на семантическом разветвлении, используются в практическом проекте «Формирование образовательного корпуса узбекского языка» № АМ-ФЗ-201908172 (справка № 04/1-2341 Ташкентского государственного университета узбекского языка и литературы имени Алишера Навои от 26 ноября 2021 года). В результате достигаются результаты систематического

поиска на основе синонимов, антонимов, омонимов и паронимов в учебном корпусе узбекского языка;

в целях создания лексикографической базы паронимов при формировании семантических отношений в онтологии узбекского языка в UzNete издан «Словарь паронимов узбекского языка» (ISBN 978-9943-7870-6-3). В результате создана база паронимов узбекского языка и сформировано лингвистическое обеспечение онтологии узбекского языка;

заимствованные слова в узбекском языке, их ряды, комментарий(и), сведения о языке оригинале, синоним(ы) заимствований, сведения об их противоположных значениях, база ударных заимствований, которая определяется в целях развития распознавания речи и речевой компетенции, устранения речевой неспособности, использовалась в издании «Словаря ударных заимствований в узбекском языке» (Словарь ударных заимствований в узбекском языке [Текст]: учебно-методический словарь. – Ташкент: Нодирабегим, 2021. – 988 с. ISBN 978-9943-6940-9-5). В результате сформировалась акцентированная словесная база, которая послужила лингвистической опорой для создания синтезатора узбекской речи.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения, списка использованной литературы. Общий объем диссертации составляет 228 страниц.

E'LON QILINGAN ISHLAR RO'YXATI
СПИСОК ОПУБЛИКОВАННЫХ РАБОТ
LIST OF PUBLISHED WORKS
I bo'lim (I часть; I part)

1. Abjalova M.A. Tahrir va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2020. – 176 b. ISBN 978-9943-6939-0-6
2. Abjalova M.A. O'zbek tili ontologiyasi: yaratish texnologiyasi va konsepsiyasi. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2021. – 210 b. ISBN 978-9943-7804-5-3.
3. Qurbonova M., Abjalova M., Axmedova N., To'laboyeva R. O'zbek tili o'zlashma so'zlarining urg'uli lug'ati [Matn]: o'quv-uslubiy lug'at. – Toshkent: Nodirabegim, 2021. – 988 b. ISBN 978-9943-6940-9-5
4. Abjalova M. O'zbek tilidagi paronim so'zlar lug'ati [Matn]: o'quv-uslubiy lug'at. – Termiz, 2022. – 64 b. ISBN 978-9943-7870-6-3
5. Abjalova M. The linguistic norms of the grapheme editing module //International Journal of Science and Research (IJSR), India. ISSN: 2319-7064. Volume 8 Issue 2, 2019. – P. 1005-1007. www.ijsr.net. (№ 40; Research Gate: 0.28)
6. Abjalova M. Milliy korpusi mavjud bo'lmagan tillarning lingvistik dasturlarida omonimlarni tahlil qilish texnologiyasi. // So'z san'ati. Xalqaro jurnal. – Samarqand, 1/2020. – B. 117-128. ISSN 2181-9297 DOI <http://dx.doi.org/10.26739/2181-9297>. (10.00.00 № 31)
7. Abjalova M., Iskandarov O. Matnlarni morfologik tahlil qilish moduli va uning algoritmi. – T.: O'zMU xabarleri // Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti ilmiy jurnali. – Toshkent, 2/2020. – B. 160-163. (10.00.00 № 15)
8. Axmedova X., Yusupova D., Abjalova M. Algorithm Based on Linguistic Models in Machine Translation Between English and Uzbek. // Global Scientific Journal Publications Volume 8, Issue 12. December 2020. – P. 1498-1502. ISSN 2320-9186. http://www.globalscientificjournal.com/researchpaper/algorithm_based_on_linguistic_models_in_machine_translation_between_english_and_uzbek.pdf. (№ 2, № 35; ImpF.:1.2).
9. Abjalova M., Erkinov F. Elektron lug'at va kiberleksikografiya. // O'zMU xabarleri. Mirzo Ulug'bek nomidagi O'zbekiston Milliy universiteti ilmiy jurnali. – Toshkent, 1/2021. – B. 160-163. (10.00.00 № 15)
10. Abjalova M., Elova D. Tabiiy tilni qayta ishlash (NLP)da so'z turkumlarini teglash masalasi. // O'zbekistonda til va madaniyat, – Toshkent: ToshDo'TAU, 1/2021. – B. 6-20. (10.00.00 № 33)
11. Abjalova M., Yuldashev A. Methods for Determining Homonyms In Homonymy And Linguistic Systems. ACADEMICIA: An International Multidisciplinary Research Journal. Vol. 11, Issue 2, February 2021. Impact Factor: SJIF 2021 = 7.492 (<https://saarj.com>). ISSN: 2249-7137 (№ 2, № 35).
12. Abjalova M. Lingvistik ontologiya – leksik ma'lumotlar bazasi. // O'zbekistonda til va madaniyat, – Toshkent: ToshDo'TAU, 4/2021. – B. 6-12. (10.00.00 № 33)
13. Abjalova M. Leksik ma'lumotlar bazalar tasnifi. // O'zbekistonda til va madaniyat: Amaliy filologiya, – Toshkent: ToshDo'TAU, 4/2021. – B. 95-107. (10.00.00 № 33)
14. Abjalova M. Ta'lim tizimida Shanxay hamkorlik tashkiloti davlatlari. // Academic Research In Educational Sciences Volume 2 | Issue 9 | 2021 ISSN: 2181-1385 Scientific Journal Impact Factor (SJIF) 2021: 5.723 Directory Indexing of International Research Journals- CiteFactor 2020-21: 0.89 DOI: 10.24412/2181-1385-2021-9-67-72. – P. 67-72. (№ 16).

15. Abjalova M., Iskandarov O. Methods of Tagging Part of Speech of Uzbek Language. // IEEE – UBMK – 2021: 6th International Conference on Computer Science and Engineering. – Ankara – Turkey, 15-16-17 September of 2021. – P. 82-85. <https://ieeexplore.ieee.org/document/9558900>. (№ 40; Impakt Faktor 5.5; № 14 ResearchBib).
16. Abjalova M. The Importance of Language Corpus in the Construction of Lexicographic Sources. // Current Research Journal of Philological Sciences (2767-3758), 2(12), – pp. 161–166. <https://doi.org/10.37547/philological-crjps-02-12-31>. (№ 35; CrossRef).
17. Abjalova M. Omonimiya va lingvistik tizimlarda omonimlarni aniqlash usullari. // Oriental Renaissance: Innovative, educational, natural and social sciences. Volume 1 | Issue 10 ISSN 2181-1784 Scientific Journal Impact Factor SJIF 2021: 5.423. – pp. 1016-1021. <http://www.oriens.uz/> (№ 12, № 14, № 23).
18. Абжалова М. Технология анализа омонимов для программ анализа текстов на узбекском языке // Международная научно-практическая конференция “Наука и образование в современном мире: вызовы XXI века”. – Нур-Султан, Казахстан: Молодой ученый, – 2020. 13-17 марта 2020 г. – С. 14-16.
19. Абжалова М. Автоматический анализ фразеологических единиц в лингвистических программах // VIII Международная конференция по компьютерной обработке тюркских языков «TurkLang-2020». (Труды конференции). Уфа: ИИЯЛ УФИТС РАН, 2020. – С. 76-80.
20. Abjalova M., Rashidova U. Lingvistik ta’minotni yaratishda frazeologik birliklar masalasi // “Kompyuter lingvistikasi: muammo va yechimlar”. Xalqaro onlayn ilmiy-amaliy konferensiya materiallari to’plami. – Toshkent, 19-aprel, 2021-yil. – B. 22-26.
21. Abjalova M. O‘zbek tili Milliy korpusida sinonimayzer yoxud sinonimizatorni yaratish masalasi // O‘zbek milliy va ta’limiy korpuslarini yaratishning nazariy va amaliy masalalari / Xalqaro ilmiy-amaliy konferensiya materiallari. 2021 yil 7 may. Elektron nashr / ebook. –Toshkent: ToshDO‘TAU, 2021. – 330 b. – B. 38-40.
22. Elov B., Axmedova X., Abjalova M. Omonim so‘zlarni farqlash usullari // O‘zbek milliy va ta’limiy korpuslarini yaratishning nazariy va amaliy masalalari / Xalqaro ilmiy-amaliy konferensiya materiallari. 2021 yil 7 may. Elektron nashr / ebook. –Toshkent: ToshDO‘TAU, 2021. – 330 b. – B. 41-44.
19. Абжалова М. Синонимайзер (синонимизатор) в образовательной корпусе узбекского языка. // TurkLang – 2021: Turkiy tillarni kompyuterda qayta ishlash IX xalqaro konferensiyasi.
20. Abjalova M., Bozorov S. Kvazisinonimlar nutqiy imkoniyat sifatida. // Актуальные вызовы современной науки, LX Международная научная конференция 26-27 июня 2021 г. – С. 123-126.
21. Raupova L., Normuradova N., Mengliyev B., Elov B., Abjalova M. O‘zbek tilining ta’limiy korpusini yaratish” nomli amaliy loyihasi xususida. // “O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari” mavzusidagi xalqaro konfrensiyasi “Zamonaviy leksikografiya, til korpuslari va turkiy tillar platformalarini yaratish muammolari” nomdagi sho‘ba materiallari, 2021-yil 18-oktabr. – B. 8-14.
22. Normuradova N., Raupova L., Abjalova M. “Turkiy tillarning lingvodidaktik elektron platformasini yaratish” nomli amaliy loyihalar xususida. // “O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari” mavzusidagi xalqaro konfrensiyasi “Zamonaviy leksikografiya, til korpuslari va turkiy tillar platformalarini yaratish muammolari” nomdagi sho‘ba materiallari, 2021-yil 18-oktabr. – B. 15-20.
23. Elov B., Abjalova M., Yuan Min. Прикладная лингвистика: проблемы лингвистических корпусов и создания платформ тюркских языков // “O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari” mavzusidagi xalqaro konfrensiyasi “Zamonaviy leksikografiya, til korpuslari va turkiy tillar platformalarini yaratish muammolari” nomdagi sho‘ba materiallari, 2021-yil 18-oktabr. – B. 94-99.

24. Abjalova M. Lingvistik ontologiyalarni takomillashtirishda til korpuslaridan foydalanish omillari. // “O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari” mavzusidagi xalqaro konferensiyasi “Zamonaviy leksikografiya, til korpuslari va turkiy tillar platformalarini yaratish muammolari” nomdagi sho‘ba materiallari, 2021-yil 18-oktabr. – B. 156-162.

25. Abjalova M., Salomov A. Mashina tarjimasini lingvistik ta‘minoti uchun o‘zbek↔ingliz sinonimlar bazasini yaratish masalasi. // “O‘zbek tili taraqqiyoti va xalqaro hamkorlik masalalari” mavzusidagi xalqaro konferensiyasi “Zamonaviy leksikografiya, til korpuslari va turkiy tillar platformalarini yaratish muammolari” nomdagi sho‘ba materiallari, 2021-yil 18-oktabr. – B. 362-365.

II bo‘lim (II часть; II part)

26. Abjalova M., Iskandarov O., O‘zbek tilidagi matnlarni avtomatik sintez va tahlil qilish dasturi. Guvohnoma № DGU 06645. – Toshkent, 2019. (mualliflik guvohnomasi)

27. Abjalova M., Rashidova U., Rahimov S. O‘zbek tilidagi sinonim so‘zlarning ma‘lumotlar bazasi. Guvohnoma № BGU 00380. – Toshkent, 2019. (mualliflik guvohnomasi)

28. Abjalova M., Iskandarov O. O‘zbek tilidagi omonim so‘zlarning ma‘lumotlar bazasi. Guvohnoma № BGU 00381. – Toshkent, 2019. (mualliflik guvohnomasi)

29. Abjalova M., Rashidova U., Iskandarov O. O‘zbek tilidagi antonim so‘zlarning ma‘lumotlar bazasi. Guvohnoma № BGU 00390. – Toshkent, 2020. (mualliflik guvohnomasi)

30. Qurbonova M., Abjalova M., Axmedova N., To‘laboyeva R., Iskandarov O. O‘zbek tilidagi o‘zlashma so‘zlarning urg‘uli bazasi. Guvohnoma № BGU 00404. – Toshkent, 2020. (mualliflik guvohnomasi).

31. Abjalova M., Xolyorov U., Iskandarov O., Xudoyberganov N., Nafasova V. O‘zbek tilidagi paronim so‘zlarning ma‘lumotlar bazasi. Guvohnoma № BGU 00469. – Toshkent, 2021. (mualliflik guvohnomasi)

32. Abjalova M. Lingvistik dasturlar uchun omonimlarni tahlil qilish texnologiyasi // “Tilshunoslikdagi zamonaviy yo‘nalishlar: muammolar va yechimlar” mavzusidagi Xalqaro ilmiy-amaliy 1-onlayn konferensiya. – Andijon, 2020. – B. 666-668.

33. Abjalova M., Ergasheva N., Azizova S. Uztextanalysis – o‘zbek tilidagi matnlarni avtomatik tahrir va tahlil qilishga mo‘ljallangan dasturiy ta‘minot // “Tilshunoslikdagi zamonaviy yo‘nalishlar: muammolar va yechimlar” mavzusidagi Xalqaro ilmiy-amaliy 1-onlayn konferensiya. – Andijon, 2020. – B. 686-689.

34. Abjalova M. Matnlar bilan ishlashga mo‘ljallangan lingvistik dasturlarni yaratish – davr talabi. O‘zbekistonda ilm-fan va ta‘lim // Oriental Art and Culture. Ilmiy-metodik jurnal. – Qo‘qon, 26.03.2020. – B. 392-395.

35. Abjalova M. Lingvistik kompyuter dasturlari uchun ma‘lumotlar bazasi va lingvistik protsessor. // «O‘zbek tilini dunyo miqyosida keng targ‘ib qilish bo‘yicha hamkorlik istiqbollari» mavzusidagi xalqaro ilmiy-amaliy anjuman materiallari. 2020-yil 19-20-oktabr. – B. 509-512.

36. Abjalova M. Kitob mutolaasi oiladan boshlanadi (tahliliy munosabat) // “Kitobxonlik madaniyatini yuksaltirish: muammo va vazifalar” mavzusida ilmiy-amaliy konferensiya. – Navoiy, 2020.

37. Abjalova M. O‘zbek Milliy korpusining tarixiy asosi. “Filologiyaning dolzarb muammolari // The actual problems of Philology” mavzusidagi Xalqaro ilmiy-amaliy internet konferensiya. – Farg‘ona, 2020. –B. 74-76.

38. Abjalova M., Qurbonova M. Raqamli iqtisodiyot davrida nutq sintezatorining zarurati. O‘zbekistonda ilm-fan va ta‘lim / Oriental Art and Culture. Ilmiy-metodik jurnal. – Qo‘qon, 26.05.2020. – B. 33-36 <http://oac.dsmi-qf.uz>

39. Abjalova M. O‘zbek tili milliy korpusida so‘zshaklni leksikografik baza asosida qidiruv imkoniyatlari // Kompyuter lingvistikasi: muammolar, yechim, istiqbollor / Respublika

ilmiy-texnik konferensiya to‘plami. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 2021. – B. 12-17.

40. Abjalova M., Abdumuminov B. Lingvistik dasturlar uchun o‘zbek tilidagi frazemalar bazasini yaratish tamoyillari // Kompyuter lingvistikasi: muammolar, yechim, istiqbollar / Respublika ilmiy-texnik konferensiya to‘plami. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 2021. – B. 63-67.

41. Abjalova M., Geldiyeva X. Terminlarni tarjima qilish masalalari // Kompyuter lingvistikasi: muammolar, yechim, istiqbollar / Respublika ilmiy-texnik konferensiya to‘plami. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 2021. – B. 85-89.

42. Abjalova M., Sharipov E. O‘zbek tezaurus lug‘ati uchun sifat turkumi bazasini yaratish masalasi // Kompyuter lingvistikasi: muammolar, yechim, istiqbollar / Respublika ilmiy-texnik konferensiya to‘plami. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 2021. – B. 189-193 b.

43. Abjalova M., Ortiqaliyeva G. “Alisher Navoiy asarlarining izohli lug‘ati” mobil ilovasini yaratish – davr talabi // Kompyuter lingvistikasi: muammolar, yechim, istiqbollar / Respublika ilmiy-texnik konferensiya to‘plami. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 2021. – B. 194-199 b. (hammuallif:)

44. Abjalova M., Rahmonaliyeva T. O‘zbek tilini xorijiy mamlakatlarda o‘qitish texnologiyasi va intellektual tizimlari // Kompyuter lingvistikasi: muammolar, yechim, istiqbollar / Respublika ilmiy-texnik konferensiya to‘plami. Elektron nashr / ebook. – Toshkent: ToshDO‘TAU, 2021. – B. 207-211 b.

45. Elov B., Abjalova M. “Sun’iy intellektga moslashtirilgan o‘zbek kompyuter tili yaratildi” (Kompyuter lingvistikasi: taraqqiyot sari qadam). / “Yangi O‘zbekiston”, 26.06.2021. № 129.

46. O‘zbek tili milliy korpusi – muhim madaniy voqelik (davra suhbat) / “Ma’rifat”, 11.08.2021. № 32 (9357).

47. Xolonim – meronim munosabati: modellashtirish tamoyilari. Filologiya fanlari doktori, professor Hamid G‘ulomovich Ne‘matov tavalludining 80 yilligiga bag‘ishlangan “Filologik tadqiqotlar: muammo va yechim” mavzusida xalqaro ilmiy-nazariy konferensiya. – Buxoro, 22.11.2021. – B.72-76.

48. Abjalova M. Ta’limiy korpus – lingvodidaktikada asosiy texnologik vosita. // Xorijiy tillarni o‘qitishning zamonaviy paradigmatlari: Respublika ilmiy-amaliy konferensiya materiallari. 2021-yil 21-dekabr, – Samarqand. – B. 363-369.

Avtoreferat Farg‘ona davlat univeyersiteti “FarDU. ILMIY XABARLAR – Scientific journal of the Fergana State University” jurnali tahririyatida tahrirdan o‘tkazildi (21.01.2022).