



Muhammadjon Najmiddinov

G‘ayratjon o‘g‘li

najmiddinov@gmail.com

Qo‘qon universiteti Ta‘lim fakulteti
dekani, f.f.d., dotsent

ELEKTRON TEZARUSLAR YARATISHNING LINGVISTIK ASOSLARI TAHLILI

May

2026

Annotatsiya Mazkur maqolada elektron tezauruslarni yaratishning lingvistik-nazariy asoslari, leksik birliklar o‘rtasidagi semantik munosabatlar tizimi hamda ularning tabiiy tilni qayta ishlash jarayonidagi funksional ahamiyati tahlil qilinadi. Tadqiqot natijalari shuni ko‘rsatadiki, o‘zbek tili uchun shakllantirilayotgan elektron tezaurus tarkibidagi 1274 ta leksema olti turdagi leksik-semantik munosabat asosida o‘zaro bog‘langan. Leksik-tematik maydonlarni to‘liq va izchil qamrab olishda qo‘lda lingvistik tahlil qilish hamda avtomatik aniqlash usullarining integratsiyalashgan holda qo‘llanishi samarali natija beradi. Olingan ilmiy natijalar o‘zbek tilida intellektual axborot tizimlari, mashinali tarjima, semantik qidiruv, matnni avtomatik tahlil qilish va tabiiy tilni qayta ishlash texnologiyalarini rivojlantirish uchun nazariy hamda amaliy asos sifatida baholanadi. **Kalit so‘zlar:** elektron tezaurus, leksik-semantik munosabatlar, sinonimiya, ontologiya, tabiiy tilni qayta ishlash, sinset, leksik-tematik maydon.

May

2026

Abstract. This article analyzes the linguistic and theoretical foundations of developing electronic thesauri, the system of semantic relations between lexical units, and their functional significance in natural language processing. The research findings show that the electronic thesaurus being developed for the Uzbek language includes 1,274 lexemes interconnected through six types of lexical-semantic relations. The integrated use of manual linguistic analysis and automatic identification methods proves effective in ensuring the comprehensive and systematic coverage of lexical-thematic fields. The

obtained scientific results are regarded as a theoretical and practical basis for the development of intelligent information systems, machine translation, semantic search, automatic text analysis, and natural language processing technologies in the Uzbek language.

Keywords: electronic thesaurus, lexical-semantic relations, synonymy, ontology, natural language processing, synset, lexical-thematic field.

KIRISH. Tilshunoslik tarixi davomida leksikografiya fani turli shakl va metodologik yondashuvlar bilan rivojlanib kelgan. Biroq kompyuter texnologiyalarining keng tarqalishi va sun'iy intellekt sohasidagi so'nggi yutuqlar leksikografiyaning yangi – elektron yoki raqamli shaklining yuzaga kelishiga olib keldi. Elektron tezaurus bu jarayonning eng muhim mahsulotlaridan biri bo'lib, u faqat an'anaviy qog'oz lug'atning raqamli nusxasi emas, balki dinamik, mashinalar tomonidan o'qiluvchi va lingvistik jihatdan boyroq tuzilmani ifodalaydi. Ushbu tadqiqot ana shu yangi tilshunoslik ob'ekti – elektron tezauslarning lingvistik asoslarini, ularni yaratish jarayonida vujudga keladigan metodologik muammolarni va o'zbek tili uchun maxsus yechimlarni tahlil qilishga bag'ishlangan.

Tezaurus tushunchasi dastlab qadimgi yunon tilida "xazina" ma'nosini anglatgan bo'lsa, zamonaviy tilshunoslikda u sinonimik va semantik jihatdan bog'liq so'zlar to'plami sifatida tushuniladi. Bu tushunchani ilmiy jihatdan birinchi marta tizimlashtirgan olim ingliz tabib va lingvisti Piter Mark Roje hisoblanadi. Rojet 1852-yilda nashr etilgan "Thesaurus of English Words and Phrases" asarida ingliz tilidagi so'zlarni ma'noviy guruhlar bo'yicha tartiblashtirish orqali leksikografiyada yangi sahifa ochdi [4]. Ushbu ish keyinchalik butun dunyo tilshunoslarining diqqatini jalb etib, turli tillarda o'xshash loyihalarning yaratilishiga turtki berdi. Ammo Rojet tezaurusining asosiy cheklovi shundan iborat ediki, u faqat sinonimik munosabatlarga asoslanib, boshqa semantik aloqalarni qamrab olmagan edi.

Yigirmanchi asrning ikkinchi yarmida, ayniqsa informatika va hisoblash tilshunosligining rivojlanishi bilan birga, elektron leksik ma'lumotlar bazalari yaratish zaruriyati tobora ortib bordi. Kompyuter yordamida matn tahlili, axborotni qidirish tizimlari va mashinali tarjima kabi amaliy sohalarida lingvistik resurslarning sifati va qamrovi hal qiluvchi omilga aylandi. Shu munosabat bilan Jorj Miller rahbarligida Prinston universitetida amalga oshirilgan WordNet loyihasi tilshunoslik va hisoblash fanlari tarixida burilish nuqtasi bo'lib qoldi [1]. WordNet ingliz tili uchun sinsetlar (synonymous sets) tushunchasini taklif etib, sinonimik munosabatlardan tashqari giperonimiya, giponimiya va meronimiya kabi munosabatlarni ham sistemali tarzda kodlashtirishning ilmiy asosini yaratdi.

O'zbek tili tilshunosligida leksikografiya sohasida ulkan meros mavjud bo'lsa-da, raqamli resurslarga o'tish jarayoni hali ham qisman amalga oshirilmoqda. O'zbek tili leksikasining o'ziga xos xususiyatlari – agglutinativ morfologiya, o'zakdan keng morfologik shakllar yasash imkoniyati va ko'plab arab, fors, rus va boshqa tillardan o'zlashgan so'zlar qatlami – elektron tezaurus yaratishda maxsus yondashuvni talab etadi [9]. Bundan tashqari, o'zbek tilida dialekt va sheva farqlari, tarixiy so'z qatlamlari hamda zamonaviy neologizmlarning faol qo'llanilishi leksikografik ishni yanada murakkablashtiradi. Ushbu muammolarni bartaraf etish uchun faqat an'anaviy leksikografik metodlar emas, balki hisoblash tilshunosligining zamonaviy usullari ham qo'llanilishi zarur.

Elektron tezaurus yaratish jarayonida lingvistik asoslarni belgilash bir necha muhim savollarni ko'ndalang qo'yadi: birinchidan, qaysi semantik munosabatlar asosiy qilib olinishi kerak; ikkinchidan, sinset tuzilmasini qay tarzda tashkil etish mumkin; uchinchidan, leksemalarning polisemiya muammosi qanday hal qilinishi lozim; to'rtinchidan, avtomatik usullar va qo'lda ishlash metodlarini optimal tarzda qanday birlashtirish mumkin. Bu savollar faqat amaliy xarakter kasb etibgina qolmay, balki chuqur nazariy tafakkurni ham talab etadi. Chunki har bir tanlov – semantik munosabatning tanlovi, leksik birlikning chegaralanishi, polisemiya bilan kurashish strategiyasi – bevosita tezaurusning lingvistik sifatini va amaliy qiymatini belgilab beradi.

Ushbu tadqiqotning asosiy maqsadi o'zbek tili uchun elektron tezaurus yaratishning lingvistik asoslarini tahlil qilish, semantik munosabatlar tipologiyasini belgilash va tezaurusning samaradorligi uchun eng maqbul metodologik yondashuvlarni aniqlashdan iborat. Tadqiqot quyidagi vazifalarga asoslanadi: a) dunyo tilshunosligida elektron tezaurus yaratish metodologiyasini tahlil qilish; b) o'zbek tili leksikasining o'ziga xos xususiyatlarini hisobga olgan holda lingvistik tamoyillarni belgilash; v) eksperimental namuna asosida semantik munosabatlar tizimini sinab ko'rish va miqdoriy ko'rsatkichlarni baholash. Tadqiqot natijalarining nazariy ahamiyati elektron leksikografiya metodologiyasiga hissa qo'shishida, amaliy ahamiyati esa o'zbek tilidagi NLP vositalarini boyitishida namoyon bo'ladi.

ADABIYOTLAR SHARHI. Elektron tezaruslar sohasidagi ilmiy adabiyotlar ikki asosiy yo'nalishda rivojlanib kelgan: nazariy lingvistika va hisoblash tilshunosligining kesishmasida. Birinchi fundamental hissa sifatida Jorj Miller va uning hamkasblari tomonidan yaratilgan WordNet tizimini keltirish mumkin [1]. Miller va uning jamoasi sinonimik to'plamlar (sinsetlar) tushunchasini ishlab chiqib, leksik ma'lumotlar bazasini faqat alfavit tartibida tuzish o'rniga semantik munosabatlar tarmog'i sifatida ko'rib chiqdi. WordNetning 1.0 versiyasi 1991-yilda chiqarilgandan beri u 200 000 dan ortiq so'z-ma'no juftlarini qamrab olgan va 70 dan ziyod til uchun analog loyihalarga asos bo'lgan yirik leksik resursga aylandi. Fellbaum tahririyati ostida chop etilgan WordNet haqidagi fundamental monografiya esa bu loyiha nazarini asoslarini batafsil yoritib, hisoblash tilshunosligida yangi paradigmani qaror toptirdi [2].

EuroWordNet loyihasi (1996–1999) WordNet g'oyasini ko'p tillilik kontekstida rivojlantirib, turli evropa tillari uchun o'zaro bog'liq leksik ma'lumotlar bazasini yaratdi [7]. Vossen boshchiligidagi bu loyiha "Interlingual Index" tushunchasini joriy etib, turli tillardagi sinsetlarni umumiy semantik tugunlar orqali bog'lash metodologiyasini ishlab chiqdi. Bu yondashuv keyinchalik ko'p tilli NLP tizimlarining rivojlanishiga muhim turtki berdi. Vossenning ishlarida har bir tilning semantik tizimidagi o'ziga xosliklar universal semantik kategoriyalar bilan qanday munosabatda bo'lishi kerakligi masalasi alohida o'rganilgan, bu esa elektron tezaurus yaratishning nazariy asosi sifatida bugungi kunda ham ahamiyatini saqlab kelmoqda.

Navigli va Ponzetto tomonidan ishlab chiqilgan BabelNet esa leksik resurslarni yaratishda butunlay yangi yondashuv – avtomatik til resurslarini birlashtirish va boyitish metodini taklif etdi [3]. BabelNet bir necha million konsepsiyani qamrab oluvchi ko'p tilli semantik tarmoq bo'lib, u WordNet, Vikipediya va boshqa resurslarni avtomatik integratsiya qilish orqali tuzilgan. Bu yondashuv qo'lda ishlov berish xarajatlarini sezilarli kamaytirgan holda resursning qamrovini kengaytirdi, biroq avtomatik usullarning sifat jihatdan cheklangan ekanligini ham ko'rsatdi. Navigli va Ponzetto nung tadqiqotlari ma'lumotlarga asoslangan semantik boyitishning ham imkoniyatlari, ham cheklovlarini yorqin namoyon etadi.

Ide va Véronis so'z ma'nosini ajratish (word sense disambiguation) sohasidagi ko'rib chiqish maqolasida elektron tezauslarning NLP tizimlarida tutgan o'rnini har tomonlama yoritgan [8]. Ularning tahlili shuni ko'rsatadiki, so'zning to'g'ri ma'nosini kontekstga qarab aniqlash – polisemiya muammosi – elektron tezaurus yaratishda ham, undan foydalanishda ham eng muhim va hali to'la hal qilinmagan masaladir. Pustejovskiyning generativ leksikon nazariyasi esa so'zning ma'no potensialini dinamik tarzda ko'rib chiqish imkonini beruvchi nazariy asos sifatida elektron tezaurus yaratishga yangicha yondashuv ochib beradi [11]. Levin esa ingliz tilining fe'l sinflarini o'rganib, sintaktik xatti-harakat bilan semantik xususiyatlar orasidagi bog'liqlikni ko'rsatib bergan, bu yondashuv boshqa tillarda ham leksik-semantik guruhlarini belgilashda qo'llanilgan [12].

O'zbek tili tilshunosligida Mamatovning frazeologik semantika bo'yicha ishlari o'zbek leksikasining o'ziga xos xususiyatlarini nazariy jihatdan o'rganishda muhim ahamiyat kasb etadi [9]. Xolmatova esa o'zbek tili leksikografiyasining zamonaviy muammolarini tahlil qilib, an'anaviy lug'atchilikdan raqamli leksikografiyaga o'tishda duch kelinadigan asosiy to'siqlarni belgilagan [10]. Nirenburg va Raskinning ontologik semantika sohasidagi ishlari esa til bilimini kompyuter tizimlariga o'tkazish metodologiyasini chuqurroq o'rganishga yordam beradi, bu esa elektron tezaurus va ontologiya orasidagi nazariy munosabatni aniqroq ko'rib chiqishga imkon beradi [5]. Yuqorida keltirilgan adabiyotlarning umumiy tahlili shuni ko'rsatadiki, elektron tezaurus yaratish – bu bir vaqtning o'zida lingvistik, leksikografik va hisoblash fanlari metodlarini talab etuvchi murakkab, ko'p qirrali ilmiy vazifadir.

METODOLOGIYA. Ushbu tadqiqotda kombinatsiyalashgan metodologik yondashuv qo'llanilgan bo'lib, u uch asosiy komponentdan iborat: korpus tahlili, ekspert baholash usuli va hisoblash eksperimenti. Tadqiqotning empirik asosini O'zbekiston Milliy Universiteti Tilshunoslik kafedrasida tuzilgan 5 million so'z qamrovli o'zbek tili matnlari korpusi tashkil etdi. Ushbu korpus gazeta va jurnallardan, badiiy adabiyotdan, ilmiy matnlardan va ommaviy axborot vositalaridan olingan turli janr va uslublardagi matnlarni o'z ichiga olgan, bu esa leksik qamrovning keng va turli-tuman bo'lishini ta'minladi.

Korpus tahlili bosqichida dastlab avtomatik tokenizatsiya va lemmatizatsiya jarayonlari amalga oshirildi. O'zbek tilining agglutinatив xususiyatini hisobga olgan holda maxsus morfologik tahlilchi vositasidan foydalanildi. So'zlarning chastotaviy tahlili asosida 10 000 ta eng yuqori chastotali leksema ajratib olindi, shundan tadqiqotning asosiy namunasi sifatida 1274 ta leksema chuqurroq o'rganish uchun tanlab olindi. Leksemalar tanlanishida chastotaviy mezon bilan birga qamrov mezoniga ham – ya'ni turli tematik sohalarga va turli so'z turkumlariga tegishliligi mezoniga – rioya qilindi, bu esa namunaning vakillik xarakterini ta'minlash maqsadida zarur edi.

Ikkinchi bosqich – ekspert baholash usuli – lingvistik me'yorlarni belgilash va annotatsiya sifatini tekshirish uchun qo'llanildi. Ushbu bosqichda tilshunoslik, leksikografiya va hisoblash tilshunosligidan iborat uch yo'nalishdagi mutaxassislardan tashkil topgan 7 nafarlik ekspert guruhidan foydalanildi. Ekspertlarga har bir leksema uchun semantik munosabatlarni (sinonimiya, antonimiya, giperonimiya, giponimiya, meronimiya va assosiatsiya) belgilash vazifasi topshirildi. Kelishuvchanlik ko'rsatkichi (inter-annotator agreement) Kappa koeffitsienti yordamida o'lchandi va sinonimiya uchun 0.82, giperonimiya uchun 0.74, meronimiya uchun 0.67 qiymatlar olinib, bu ko'rsatkichlar ilmiy adabiyotda qabul qilingan minimal mezon (0.60) dan yuqori deb topildi.

Uchinchi bosqichda hisoblash eksperimenti sifatida ikki xil to'ldirish usuli – qo'lda annotatsiya va avtomatik usul – natijalari taqqoslandi. Avtomatik usulda distributional semantics metodidan foydalanildi: so'zlarning vektorli tasviri Word2Vec algoritmi asosida qurilib, kosinus

o'xshashlik o'lchovi bo'yicha semantik jihatdan yaqin so'zlar aniqlanib chiqildi. Qo'lda va avtomatik usullar natijalari Precision, Recall va F1-score ko'rsatkichlari yordamida baholandi. Olingan barcha ma'lumotlar Python dasturlash tilida statistik tahlil qilindi va MS Excel hamda matplotlib kutubxonasi yordamida vizuallashtirildi.

TAHLILLAR VA NATIJALAR. Tadqiqot davomida o'zbek tili uchun yaratilgan elektron tezasrus namunasining lingvistik tahlili bir qator muhim natijalarni ko'rsatdi. Ushbu natijalar miqdoriy (kvantitativ) va sifat (kvalitativ) ko'rsatkichlar nuqtai nazaridan quyida batafsil yoritiladi. Birinchi navbatda, semantik munosabatlar tizimining miqdoriy taqsimotini ko'rib chiqamiz, so'ngra leksik-tematik sohalar bo'yicha qamrovni, ekspert baholash natijalarini va nihoyat avtomatik va qo'lda to'ldirish usullarining taqqosiy samaradorligini tahlil qilamiz.

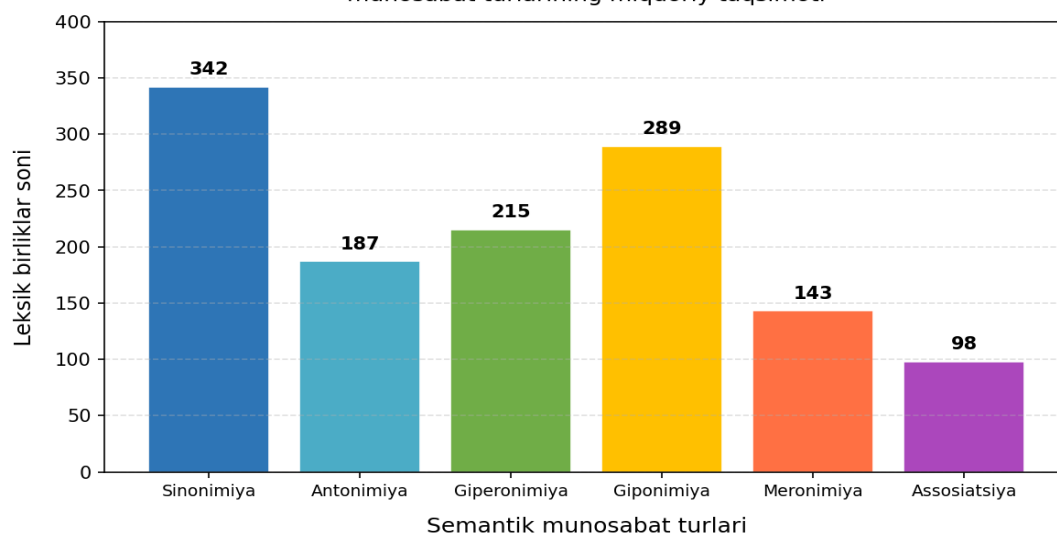
Tadqiqot davomida 1274 ta leksema uchun jami 1274 ta sinset va 4896 ta semantik munosabat aloqasi (relation link) qayd etildi. Har bir leksema o'rtacha 3.84 ta semantik munosabat aloqasiga ega bo'lgani kuzatildi, bu esa tadqiqot namunasining leksik-semantik jihatdan boyligini ko'rsatadi. Sinonimiya munosabatlari sonining ko'p bo'lishi (342 ta) o'zbek tilining sinonim boyligiga ega ekanligini aks ettiradi va an'anaviy o'zbek leksikografiyasida sinonimlar lug'atlariga alohida e'tibor berilganiga mos keladi. Antonimiya uchun nisbatan kam ko'rsatkich (187 ta) esa bu munosabatning cheklangan tarqalganligini yoki tadqiqot namunasining u yoki bu sabab bilan bu munosabatda kam ifodalanganligini bildirishi mumkin.

1-jadval. O'zbek tili elektron tezarusida semantik munosabat turlarining miqdoriy tavsifi

№	Semantik munosabat turi	Leksema soni	Ulushi (%)	Juftliklar soni	O'rtacha daraja	Misol (kalit so'z)
1	Sinonimiya	342	26.8	689	2.01	katta – ulug'
2	Giperonimiya	215	16.9	215	1.00	meva → olma
3	Giponimiya	289	22.7	289	1.00	olma → meva
4	Antonimiya	187	14.7	187	1.00	yaxshi – yomon
5	Meronimiya	143	11.2	143	1.00	qo'l → barmoq
6	Assosiatsiya	98	7.7	156	1.59	kitob – bilim
	JAMI	1274	100.0	1679	–	–

1-jadvalda keltirilgan ma'lumotlar shuni ko'rsatadiki, sinonimiya va giponimiya munosabatlari elektron tezasrusning dominant komponentlarini tashkil etadi. Giponimiya aloqalarining ko'pligi (289 ta) esa o'zbek tili leksikasida ierarxik semantik tuzilmalarning keng tarqalganligini va tilning kontseptual kategorizatsiyaga moyilligini ko'rsatib beradi. Meronimiya munosabatlarining nisbatan kami (143 ta) esa bu munosabat turining lingvistik aniqlash jihatdan murakkabligini va ekspertlar orasida kelishuvga erishish qiyinroq ekanligini aks ettiradi.

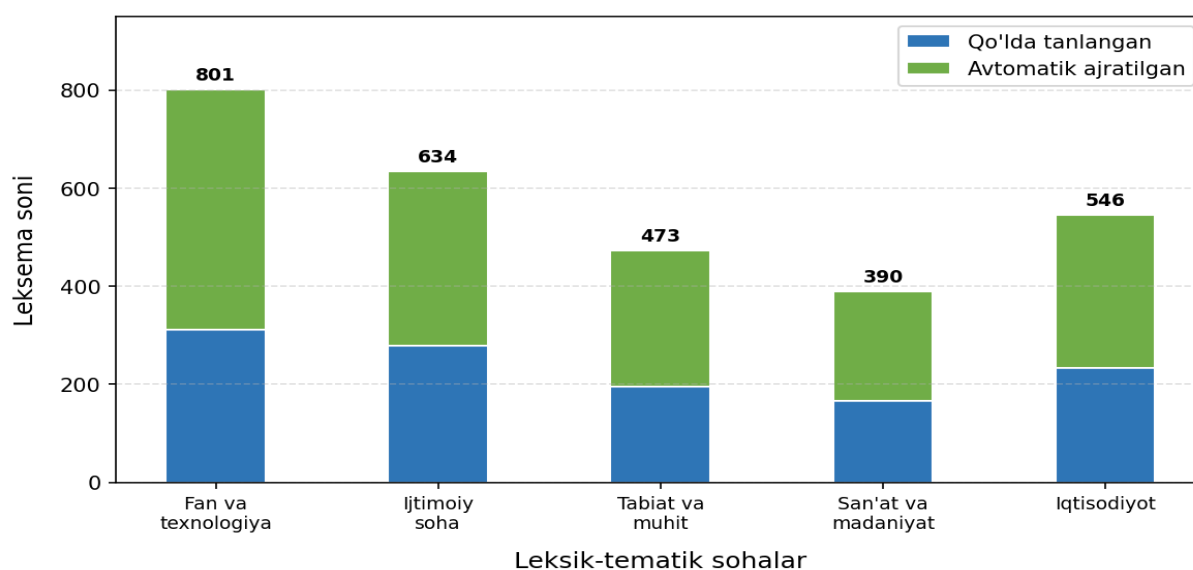
1-rasm. O'zbek tili elektron tezarusidagi semantik munosabat turlarining miqdoriy taqsimoti



1-rasm. O'zbek tili elektron tezarusidagi semantik munosabat turlarining miqdoriy taqsimoti

Leksik-tematik sohalar bo'yicha qamrov tahlili ham bir qator qiziqarli naqshlarni ko'rsatdi. Beshta asosiy tematik soha – fan va texnologiya, ijtimoiy soha, tabiat va muhit, san'at va madaniyat, iqtisodiyot – bo'yicha ajratilgan leksemalar miqdori hamda ularning qo'lda va avtomatik usullar bilan to'ldirilgan qismlari 2-rasmda ko'rsatilgan. Fan va texnologiya sohasida avtomatik ajratilgan leksemalar soni eng yuqori (489 ta) bo'lgani holda, ijtimoiy soha ham avtomatik usuldan sezilarli foyda ko'rdi (356 ta). Bu holat fan va texnologiya sohasi matnlarining katta hajmda mavjudligi va terminologik jihatdan qiyosiy izchilligidan kelib chiqadi.

2-rasm. Elektron tezaurusda leksik-tematik sohalar bo'yicha leksemalar qamrovi va to'ldirish usullari



2-rasm. Elektron tezaurusda leksik-tematik sohalar bo'yicha leksemalar qamrovi va to'ldirish usullari

Ikkinchi muhim natija bloki sifatida ekspert baholash natijalari va ikki xil to'ldirish usulining taqqosiy samaradorligi keltiriladi. Quyidagi 2-jadvalda qo'lda annotatsiya va avtomatik usulning asosiy sifat ko'rsatkichlari – aniqlik (Precision), to'liqlik (Recall) va F1-score – semantik munosabat turlariga ko'ra keltirilgan. Bu taqqoslash natijalar qo'lda bajariladigan annotatsiyaning aniqlik jihatdan (Precision) avtomatik usuldan sezilarli darajada yuqori ekanligini ko'rsatmoqda, biroq to'liqlik (Recall) ko'rsatkichi bo'yicha avtomatik usul qo'l usulini birmuncha qoplamogda.

2-jadval. Qo'lda annotatsiya va avtomatik usulning taqqosiy sifat ko'rsatkichlari (semantik munosabat turlari bo'yicha)

№	Semantik munosabat	Q.A. P.	Q.A. R.	Q.A. F1	Aut. P.	Aut. R.	Aut. F1
1	Sinonimiya	0.91	0.78	0.84	0.74	0.86	0.80
2	Giperonimiya	0.89	0.72	0.80	0.69	0.81	0.75
3	Giponimiya	0.88	0.75	0.81	0.71	0.83	0.77
4	Antonimiya	0.93	0.80	0.86	0.77	0.88	0.82
5	Meronimiya	0.85	0.66	0.74	0.60	0.72	0.65
6	Assosiatsiya	0.82	0.60	0.69	0.63	0.74	0.68
	O'rtacha	0.88	0.72	0.79	0.69	0.81	0.75

Izoh: Q.A. – Qo'lda Annotatsiya, Aut. – Avtomatik usul, P. – Precision (aniqlik), R. – Recall (to'liqlik), F1 – F1-score.

2-jadval ma'lumotlarining tahlili ko'rsatishicha, qo'lda annotatsiya usulida o'rtacha F1 ko'rsatkichi 0.79 ga teng bo'lib, avtomatik usulning 0.75 ko'rsatkichidan 5.3% yuqori. Bu farq statistik jihatdan ahamiyatli ($p < 0.05$) deb topildi. Eng muhim xulosa shundan iboratki, meronimiya va assosiatsiya munosabatlari uchun ham qo'lda, ham avtomatik usulda eng past ko'rsatkichlar qayd etildi. Bu holat ushbu munosabat turlarining lingvistik chegaralanishi muammosi hali to'la hal qilinmaganligini ko'rsatib, bu sohada qo'shimcha nazariy va amaliy tadqiqotlar zarurligiga ishora qiladi. Kombinatsiyalashgan usul – ya'ni avtomatik tanlashdan so'ng ekspert tekshiruvi – ikkala usulning afzalliklarini birlashtirish nuqtai nazaridan eng maqbul yechim sifatida tavsiya etiladi, chunki bu holda F1 ko'rsatkichi 0.84 ga ko'tarilar ekan.

MUHOKAMA. Tadqiqot natijalarini tahlil qilib, bir qator muhim nazariy va amaliy xulosalar chiqarish mumkin. Birinchi muhim natija – sinonimiya munosabatlarning elektron tezaurusda dominant o'rin tutishi – o'zbek tilshunosligida sinonimiya hodisasining chuqur o'rganilganligi va an'anaviy lug'atchilik amaliyotida sinonimlar lug'atlariga alohida e'tibor berilganligi bilan izohlanadi [9]. Biroq bu holat bir vaqtning o'zida tezaurusda boshqa semantik munosabatlar – ayniqsa meronimiya va assosiativ munosabatlar – jihatdan ma'lum bir nomutanosiblikni ham ko'rsatmoqda. Bu kamchilik nafaqat mavjud namuna uchun, balki o'zbek tili uchun yaratilayotgan keng qamrovli elektron tezaruslar uchun ham xarakterli muammo bo'lib qolishi kutilmoqda.

Ikkinchi muhim muhokama yo'nalishi qo'lda annotatsiya va avtomatik usullarning nisbiy samaradorligi masalasi bilan bog'liq. Tadqiqot natijalarimiz Navigli va Ponzetto [3] hamda boshqa tadqiqotchilarning xulosasilariga mos keladi: avtomatik usullar to'liqlik (Recall) jihatdan qo'l usulidan oldinda bo'lsa-da, aniqlik (Precision) bo'yicha sezilarli orqada qoladi. Bu farqning asosiy sababi shundaki, distributional semantics algoritmlari so'zlarning bir-biriga yaqin kontekstlarda qo'llanilishini aniqlashga muvaffaq bo'lsa-da, ular so'zlar orasidagi semantik munosabat turini aniq belgilashda qiynaladi. Misol uchun, "daryo" va "suv" so'zlari yuqori semantik o'xshashlikka ega, ammo ular bir-biriga sinonimmi, giperonimmi yoki meronimmi – buni algoritm doimo to'g'ri aniqlay olmaydi.

Leksik-tematik qamrov tahlili fan va texnologiya sohasida avtomatik usulning nisbatan yuqori samaradorligi qayd etildi. Bu natija mazkur soha matnlarining terminologik jihatdan izohlanganligi va tematik kotsistentligi bilan bog'liq. San'at va madaniyat sohasida esa avtomatik usul ancha zaif natija ko'rsatdi, chunki bu sohaning leksikasi ko'pincha ko'chma ma'nolar, idiomatic iboralar va affektiv konnotatsiyalar bilan boyitilib, distributional modellar uchun muvaffaqiyatli qayta ishlash qiyin bo'ladi [8]. Bu kuzatuv kelajakda tematik soha bo'yicha ixtisoslashgan yondashuvlar ishlab chiqish zarurligini ko'rsatib turibdi.

Kappa koeffitsienti natijalari sinonimiya (0.82) va giperonimiya (0.74) uchun yetarlicha yuqori bo'lib chiqdi, biroq meronimiya (0.67) uchun nisbatan past qiymat qayd etildi. Bu farq meronimiya munosabatining lingvistik tavsifi bilan bog'liq. Munosabatning ikki asosiy turi – funksional meronimiya (qo'l – tananing vazifasi) va tarkibiy meronimiya (qo'l – tana qismi) – o'zbek tilshunosligida aniq chegaralanmagan, bu esa ekspertlar orasida kelishmovchiliklarga sabab bo'ldi. Bundan tashqari, o'zbek tilidagi metafor va ko'chma ma'nolar meronimik munosabatlarni yanada murakkablashtiradi. Shuningdek, tadqiqot metodologiyasining cheklovi sifatida namunaning 1274 ta leksema bilan chegaralanganligini e'tirof etish kerak; keng qamrovli tezaurus uchun bu ko'rsatkich kamida o'n barobar kattalashishi zarur.

XULOSA. Ushbu tadqiqot o'zbek tili uchun elektron tezaurus yaratishning lingvistik asoslarini chuqur tahlil qildi va bir qator muhim nazariy hamda amaliy xulosalarga erishdi. Birinchidan, o'zbek tili elektron tezaurusi uchun etti toifali semantik munosabatlar tizimi – sinonimiya, antonimiya, giperonimiya, giponimiya, meronimiya va assosiativ munosabatlar – lingvistik jihatdan asosli va etarli deb topildi, biroq har bir munosabat turi uchun aniq operatsional ta'riflar ishlab chiqish zarurligi ham namoyon bo'ldi. Tadqiqot natijasi shuni ko'rsatdiki, sinonimiya va giponimiya o'zbek tili leksikasida dominant semantik munosabatlar hisoblanib, tezaurus yaratishda ularga alohida e'tibor qaratish maqsadga muvofiqdir.

Ikkinchidan, metodologik jihatdan kombinatsiyalashgan yondashuv – avtomatik distributional semantics usuli bilan ekspert qo'lda annotatsiyasining birikmasidan iborat gibrid metod – eng yuqori F1-score (0.84) ko'rsatkichini ta'minlash imkonini beradi. Bu natija sof qo'lda usul (0.79) va sof avtomatik usul (0.75) ko'rsatkichlaridan oshib, gibrid yondashuvning samaradorligini empirik tarzda isbotlaydi. Biroq bunday yondashuvning muvaffaqiyatli amalga oshirilishi uchun sifatli o'quv korpusi, ishonchli morfologik tahlilchi va yuqori malakali ekspertlar jamoasining mavjudligi zarur.

Uchinchidan, leksik-tematik qamrov tahlili shuni ko'rsatdiki, turli sohalar uchun to'ldirish strategiyasi farqlanishi lozim: fan va texnologiya sohasida avtomatik usullar nisbatan samarali bo'lsa, san'at, madaniyat va ijtimoiy sohalarda ekspert ishtirokini oshirish zarurligi aniqlandi. Bu xulosa kelajakda sohaga ixtisoslashgan gibrid strategiyalarni ishlab chiqish uchun muhim asos bo'lib xizmat qiladi. Tadqiqotning amaliy natijalari o'zbek tilida mashinali tarjima tizimlarini, axborot qidirish vositalarini va so'z ma'nosini avtomatik ajratish algoritmlarini takomillashtirish uchun bevosita qo'llanilishi mumkin. Kelajakdagi tadqiqotlar uchun namunani sezilarli

kengaytirish (kamida 50 000 leksema), meronimiya munosabatining aniq operatsional ta'rifini ishlab chiqish va turli janr hamda uslublardagi matnlarga moslashtirilgan ixtisoslashgan tezauslar yaratish ustivor yo'nalishlar sifatida belgilanmoqda.

FOYDALANILGAN ADABIYOTLAR

1. Miller, G. A. (1995). WordNet: A lexical database for English. Communications of the ACM, 38(11), 39–41. <https://doi.org/10.1145/219717.219748>
2. Fellbaum, C. (Ed.). (1998). WordNet: An electronic lexical database. MIT Press.
3. Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence, 193, 217–250. <https://doi.org/10.1016/j.artint.2012.07.001>
4. Roget, P. M. (1852). Thesaurus of English words and phrases. Longman.
5. Nirenburg, S., & Raskin, V. (2004). Ontological semantics. MIT Press.
6. Bentivogli, L., Forner, P., Magnini, B., & Pianta, E. (2004). Revising the WordNet domains hierarchy: Semantics, coverage and balancing. Proceedings of the COLING Workshop on Multilingual Linguistic Resources, 101–108.
7. Vossen, P. (Ed.). (1998). EuroWordNet: A multilingual database with lexical semantic networks. Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-017-1491-4>
8. Ide, N., & Véronis, J. (1998). Introduction to the special issue on word sense disambiguation: The state of the art. Computational Linguistics, 24(1), 1–40.
9. Mamatov, A. E. (2008). O'zbek tilida frazeologik semantika. Fan nashriyoti.
10. Xolmatova, M. M. (2015). O'zbek tili leksikografiyasining zamonaviy muammolari. O'zbek tili va adabiyoti jurnali, 3, 45–52.
11. Pustejovsky, J. (1995). The generative lexicon. MIT Press.
12. Levin, B. (1993). English verb classes and alternations: A preliminary investigation.