

UO'K(UDC, УДК): 81–26

KOMPYUTER TEXNOLOGIYALARI VA AVTOMATLASHTIRILGAN LEKSIKOGRAFIK LUG'ATLAR¹⁵

Alisherova Shaxnoza Asqarovna

O'zbekiston davlat jahon tillari universiteti

tayanch doktoranti,

Toshkent, O'zbekiston

E-mail: shaxnozaalisherova62@gmail.com

ORCID ID: 0009–0007–6329–0315

ANNOTATSIYA

Ushbu maqolada tematik lug'atlarni yaratishda kompyuter texnologiyalari va avtomatlashtirilgan tizimlarning qanday qo'llanilishi borasida so'z boradi. An'anaviy lug'at yaratish usulini ma'lumotlar va texnologiyaga asoslangan jarayonga aylantirish, katta korpuslar yordamida o'zgarishlar, tabiiy tilni qayta ishlash vositalaridan foydalanish, avtomatik atamalarni ajratib olish, ta'rif yaratish modellari va so'zma so'z ma'nosi tarmoqlari muhokama qilinadi. Bundan tashqari, u Lexonomy, TLex Suite, FLEx va OntoLex–Lemon modelining so'z shakllarini boshqarish, semantik qoliplash, kategoriya yaratish va misol gap yaratishni qo'llab-quvvatlovchi xususiyatlarini ham yoritadi. Maqola matn asosidagi tilshunoslik, lug'at taqqoslash, miqdoriy ma'lumotlarni tahlil qilish va mahalliy lug'at ishlab chiqish uchun gipoteza sinovlari kabi metodologiyalarga murojaat qiladi. Natijalar shuni ko'rsatadiki, ushbu vositalardan foydalanish lug'at yaratish jarayonida g'oyalarni bog'lashni tezlashtiradi, soddalashtiradi va yaxshilaydi. Shunga qaramay, ba'zi qiyinchiliklarni ham keltirib chiqaradi, ayniqsa o'zbek tili kabi so'zlarni murakkab, morfologik tarzda o'zgartiradigan tillarda yaqqol ko'zga tashlanadi. Maqola leksikografiya vositalarida zamonaviy til va texnologiya talablariga yaxshiroq mos keladigan xususiyatlarini ham ko'rib chiqadi.

KALIT SO'ZLAR

Avtomatlashtirilgan leksikografiya, tematik lug'at, korpus lingvistikasi, NLP vositalari, terma ekstraksiyasi, semantik tarmoqlar, OntoLex–Lemon, Lexonomy, TLex Suite, o'zbek leksikografiyasi, generativ modellar, raqamli lug'atlar.

Received: October 23, 2025

Accepted: December 19, 2025

Available online: February 25, 2026

¹⁵ **For citation (Iqtibos keltirish uchun, для цитирования):**

Alisherova Sh. Kompyuter texnologiyalari va avtomatlashtirilgan leksikografik lug'atlar. // Komparativistika (Comparative Studies). — 2026. — Vol.3, № 1(9) — B. 201-215.

**COMPUTER TECHNOLOGIES AND AUTOMATED
LEXICOGRAPHIC DICTIONARIES****Alisherova Shakhnoza Askarovna***Basic doctoral student of
Uzbek State University of World Languages,
Tashkent, Uzbekistan**E-mail: shaxnozaalisherova62@gmail.com**ORCID ID: 0009-0007-6329-0315*

ABSTRACT

This article discusses the application of computer technologies and automated systems in creating thematic dictionaries. It explores the transformation of traditional dictionary creation methods into data-driven and technology-based processes, including the use of large corpora, natural language processing tools, automatic term extraction, definition creation models, and word sense networks. Furthermore, it highlights features of tools such as Lexonomy, TLex Suite, FLEx, and the OntoLex-Lemon model that support word form management, semantic templating, category creation, and example sentence generation. The article addresses methodologies like corpus linguistics, dictionary comparison, quantitative data analysis, and hypothesis testing for developing local dictionaries. Results indicate that utilizing these tools accelerates, simplifies, and enhances the process of connecting ideas in dictionary creation. However, it also presents certain challenges, particularly evident in morphologically complex languages like Uzbek. The article also examines features of lexicographic tools that better align with the demands of modern language and technology.

KEY WORDS

Automated
lexicography; thematic
dictionary; corpus
linguistics; NLP tools;
term extraction;
semantic networks;
Onto Lex-Lemon;
Lexonomy; TLex Suite;
Uzbek lexicography;
generative models;
digital dictionaries

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ И АВТОМАТИЗИРОВАННЫЕ ЛЕКСИКОГРАФИЧЕСКИЕ СЛОВАРИ

Алишерова Шахноза Аскарровна

*Узбекский государственный
университет мировых языков,*

докторант,

Ташкент, Узбекистан

E-mail: shaxnozaalisherova62@gmail.com

ORCID ID: 0009-0007-6329-0315

АННОТАЦИЯ

Настоящая статья посвящена изучению особенностей использования компьютерных технологий и автоматизированных систем для создания тематических словарей. В ней описана трансформация традиционного способа создания словаря в процесс, основанный на данных и технологиях, включающих крупные корпуса, инструменты обработки естественного языка, автоматическое извлечения терминов, моделей генерации определений и семантических сетей. Кроме того, в работе рассмотрены возможности таких инструментов, как Lexonomy, TLex Suite, FLEx и модели OntoLex–Lemon, которые поддерживают управление словоформами, семантическую организацию, создание категорий и генерацию примеров предложений. Отдельное внимание в статье уделяется анализу методологии, в частности, корпусной лингвистике, сравнению словарей, количественному анализу данных и проверке гипотезы путем создания локального словаря. Полученные результаты показывают, что использование этих инструментов ускоряет и упрощает процесс создания словаря, улучшает качество связывания идей. Вместе с тем, их использование сопряжено с некоторыми трудностями, что особенно характерно для морфологически сложных языков, в том числе для узбекского. В заключение отмечены особенности лексикографических инструментов, которые в большей степени соответствуют требованиям современного языка и технологий.

КЛЮЧЕВЫЕ СЛОВА

Автоматизированная лексикография, тематический словарь, корпусная лингвистика, инструменты НЛП, извлечение терминов, семантические сети, OntoLex–Lemon, Лексономия, TLex Suite, узбекская лексикография, генеративные модели, цифровые словари.

KIRISH

Soʻnggi oʻn yilliklarda tilshunoslikda tezkor raqamli metamorfoz ultramodern leksikografiyaning rivojlanishini sezilarli darajada kuzatilmoqda. XXI asrda elektron soʻzlar, raqamli korpuslar va avtomatlashtirilgan nutq tahlil vositalari zamonaviy leksikografik amaliyotning zarur omillari boʻlib, keng koʻlamli, aniq va qatʼiy soddalashtirilgan nutq maʼlumotlarni taqdim etdi. Ushbu texnologik yutuqlar lugʻat toʻplash jarayonlarini tezlashtirdi va leksikograflarga korpusga asoslangan izohlash, chastotani tahlil qilish, semantik klasterlash va sunʼiy intellektga asoslangan modellar orqali tildan foydalanishga imkon berdi.

Ultramodern leksikografiya doirasida terminologik, mavzuga oid va sohaga oid lugʻatlarga oʻxshash texnik yoki tematik soʻz kitoblarini yaratish muhim rol oʻynaydi. Ilmiy, kasbiy va texnologik aloqaning murakkabligi aniq, sohaga xos va semantik jihatdan izchil tuzilgan nutq bazalarni talab qiladi.

Tadqiqotning asosiy maqsadi ultramodern kompyuter texnologiyalaridan foydalangan holda tematik soʻzlar kitoblarini yaratish tamoyillarini oʻrganish, nutq maʼlumotlarni qayta ishlash, tartibga solish va taqdim etishda avtomatlashtirilgan leksikografik tizimlarning samaradorligini oʻrganishdir.

Ushbu maqsadga erishish uchun quyidagi vazifalar qoʻyiladi:

1. Avtomatlashtirilgan va yarim avtomatlashtirilgan leksikografik platformalarni Sketch Machine, soʻz tarmogʻi, oʻzbek Milliy korpusi (UzbekCorpus) va boshqa avtomatlashtirilgan darsliklarni qayta ishlash tizimlariga ajratish.

2. Avtomatlashtirilgan leksikografik ish jarayonining bosqichlarini tavsiflash, shu jumladan korpus toʻplami, maʼlumotlar tugʻilishi, chastotani tahlil qilish, semantik qavs va lugʻatga kirishni yaratish.

3. Hisoblash uslublari va korpusga asoslangan ogʻzaki asoslash yordamida tematik soʻzlar kitobini toʻplashning namunaviy algoritmini ishlab chiqish.

4. Kompyuter tomonidan qoʻllab-quvvatlanadigan leksikografiyaning afzalliklari va cheklovlarini aniqlash, ayniqsa tezlik, noziklik, oʻlchovlilik va maʼlumotlar sifati bilan bogʻliq.

Ushbu tadqiqot raqamli texnologiyalar va avtomatlashtirilgan tizimlar yuqori sifatli tematik soʻzlarni yaratish jarayonini qanday kuchaytirishi mumkinligini namoyish etish orqali ultramodern leksikografiya metodologiyasiga hissa qoʻshishga qaratilgan.

ASOSIY QISM.

Tematik yoki mavzu bilan tanish boʻlgan soʻz kitoblarini ishlab chiqish ultramodern leksikografiyaning eng muhim yoʻnalishlaridan birini, ayniqsa raqamli metamorfoz va korpusga asoslangan ogʻzaki tadqiqot muhitini ifodalaydi. Ushbu boʻlimda avtomatlashtirilgan leksikografik jarayonlarni tushunish uchun zarur boʻlgan asosiy nazariy asoslar va tematik soʻzlarni toʻplashda ishlatiladigan texnologik vositalar koʻrsatilgan. *“Tematik soʻzlar kitobi ogʻzaki birliklarni alifbo tartibida emas, balki maʼlum bir mavzu sohasi, semantik soha yoki kasbiy sohaga koʻra tartibga soluvchi leksikografik manbadir”* (Atkins & Rundell, 2008; 120).

Umumiy maqsadli soʻz kitoblaridan farqli oʻlaroq, tematik soʻz kitoblari sohaga xos lugʻat, konseptual tuzilgan yozuvlar va kontekstli chegaralarni beradi. Tematik soʻz kitoblarining asosiy turlariga quyidagilar kiradi:

Terminologik soʻzlar – kasbiy yoki ilmiy fanlarda qoʻllaniladigan texnik atamalarni oʻz ichiga oladi;

Mavzuga oid soʻzlar – giyohvandlik, tilshunoslik, irrigatsiya muhandisligi, axborot texnologiyalari va boshqa sohalarga bagʻishlangan.;

Tematik yoki semantik soʻzlar – ogʻzaki maʼlumotlarni hissiyot, harakat, relyef, madaniyat yoki texnologiya kabi mavhum sohalarga koʻra guruhlash.

Tematik soʻz kitoblarini yaratish shunga quyidagiga oʻxshash belgilangan leksikografik tamoyillarga amal qiladi:

- a) noziklik va terminologik mukammallik;
- b) haqiqiy korpus maʼlumotlariga tayanish;
- c) semantik ravshanlik va ichki bogʻliqlik;
- d) operatsiya misollarini qoʻshish;
- e) mavhum munosabatlarga asoslangan uslubiy turkumlash (giperonimlar,

giponimlar, antonimlar, sinonimlar, atamalar).

“Ushbu tamoyillar so‘zlar kitobi alohida sohada ham tavsiflovchi, ham an’anaviy funksiyalarni bajarishini kafolatlaydi” (Bojanić & Duranović, 2021; 321).

An’anaviy leksikografiya asosan ma’lumotlarini yig‘ish, ona tilida so‘zlashuvchi, bilimdon manbalar va ekspert talqiniga tayanadi. Ultramodern leksikografiya, nomuvofiqlik bilan, korpus tilshunosligi, raqamli darsliklarni qayta ishlash va hisoblash modellashtirishiga asoslanadi. Bu bir nechta hal qiluvchi afzalliklarni taklif qiladi: xususan, korporativ va darslik ma’lumotlar bazalaridan keng ko‘lamli empirik asoslash; nutq birliklarning avtomatlashtirilgan tizimi, atamalar va chastotalar ro‘yxati; ideal, statistik jihatdan tasdiqlangan semantik qonuniyatlar; interaktiv, yangilanadigan so‘zlar formatlari; tabiiy tilni qayta ishlash (NLP) bilan integratsiya orqali chuqurroq og‘zaki tahlil qilish vositalari (Čibej & Fišer, 2018; 9).

Zamonaviy hisoblash leksikografiyasida, tabiiy tilni qayta ishlash (NLP) texnikasi tematik lug‘at kompilyatsiyasining samaradorligi, aniqligi va analitik imkoniyatlarini oshirishda hal qiluvchi rol o‘ynaydi. Asosiy jarayonlardan biri tokenizatsiya bo‘lib, u uzluksiz matnni so‘zlar yoki iboralar kabi diskret birliklarga ajratadi va shu bilan avtomatlashtirilgan leksik ekstraktsiya uchun asos yaratadi. Shundan so‘ng, lemmatizatsiya so‘zlarni morfologik tahlil orqali kanonik lug‘at shakllariga qisqartirish uchun qo‘llaniladi, bu chastotalarni hisoblash va leksik tasniflarning izchil va lingvistik jihatdan aniq bo‘lishini ta’minlaydi. Avtomatlashtirilgan atamalarni ajratib olish sun’iy intellektga asoslangan algoritmlar orqali domenga xos terminologiyani, texnik iboralarni va barqaror kollokatsiyalarni aniqlash orqali jarayonni yanada yaxshilaydi, bu esa kompilyatorlarga lug‘atga kiritish uchun eng dolzarb leksik elementlarni birinchi o‘ringa qo‘yishga imkon beradi. Bundan tashqari, mashinada o‘rganishga asoslangan klasterlash texnikasi so‘zlarni taqsimot o‘xshashligiga ko‘ra guruhlaydi, yashirin semantik qirralarni ochib beradi va tematik lug‘atni tizimli tashkil etishga hissa qo‘shadi. Ushbu vositalarni to‘ldiruvchi so‘z kabi tarqatuvchi vektor modellari

FastText va BERT, bu yuqori oʻlchovli vektor makonidagi soʻzlarni ifodalaydi, nozik semantik munosabatlarni namoyon qiladi. Ushbu modellar soʻzlarning oʻxshashligini, kontekstual maʼnosini va semantik mahallalarini aniqlashga yordam beradi va shu bilan avtomatik atamalarni guruhlash, taʼriflarni takomillashtirish va hissiyotlarni ajratishni qoʻllab-quvvatlaydi. Umuman olganda, NLP texnologiyalarining integratsiyasi tematik leksikografiyaning miqyosi, aniqligi va kontseptual chuqurligini sezilarli darajada mustahkamlaydi, lugʻat yaratish jarayonini yanada empirik asosli va uslubiy jihatdan mustahkam qiladi (Granger, 2017; 320).

Zamonaviy leksikografiya qoʻlda kompilyatsiya qilishdan texnologiyaga asoslangan, maʼlumotlarni talab qiluvchi usullarga oʻtdi. Kompyuter texnologiyalari lingvistik maʼlumotlarni keng miqyosda qayta ishlashga, leksik birliklarni tez topib olishga, avtomatik tasniflashga va semantik tuzilishga imkon beradi. (Hanks, 2012; 398). Avtomatlashtirilgan lugʻat qurilishidagi birinchi qadam vakillik korpusini ishlab chiqishdir. Korpus tematik jihatdan dolzarb, muvozanatli va haqiqiy tildan foydalanishni aks ettirish uchun yetarlicha katta boʻlishi kerak. Ushbu jarayonga odatda quyidagilar kiradi:

Maʼlumotlarni yigʻish – raqamli kutubxonalar, ilmiy manbalar, veb–sahifalar va domenga xos materiallardan matnlarni yigʻish. Avtomatlashtirilgan lugʻat qurilishidagi birinchi qadam vakillik korpusini ishlab chiqishdir. Korpus tematik jihatdan dolzarb, muvozanatli va haqiqiy tildan foydalanishni aks ettirish uchun yetarlicha katta boʻlishi kerak. Maʼlumotlarni tozalash, dublikatlar, HTML teglari, nomuvofiq kodlash va ahamiyatsiz tarkibni olib tashlash kabilardir. Yuqori sifatli korpuslar leksikograflarga aniq tematik lugʻatlarni ishlab chiqish uchun hal qiluvchi ahamiyatga ega boʻlgan soʻz chastotasi, kollokatsiyalar, semantik siljishlar, kontekstual qirralar va terminologik zichlikni kuzatishga imkon beradi. Tabiiy tilni qayta ishlash (NLP) korpus maʼlumotlarini tahlil qilish uchun texnik asos yaratadi (Ide & Véronis, 1998; 28).

Morfologik analizatorlar soʻz tuzilishi, affikslari, oʻzaklari va morfologik

kategoriyalarini aniqlaydi, ayniqsa o‘zbek tili kabi aglutinativ tillar uchun juda muhimdir. Lemmatizatorlar birikkan shakllarni asos shakllarga aylantirib, normallashtgan leksik tahlil qilish imkonini beradi. So‘zlarni grammatik kategoriyalarga (ot, fe‘l, sifat va h.k.) avtomatik tasniflovchi POS–taggerlar.), ta’riflarni yaratish aniqligini oshirish vazifalarini bajaradi. Ushbu vositalar qo‘l mehnatini kamaytiradi, ishlov berish tezligini oshiradi va obyektiv, takrorlanadigan lingvistik natijalarni ta’minlaydi (Kilgarrieff & Grefenstette, 2003; 334).

Avtomatik tematik lug‘at yaratish uchun bir nechta algoritmlar va modellar keng qo‘llaniladi. Xususan: TF–IDF (atama chastotasi – teskari hujjat chastotasi): muhim hujjatlardagi nisbiy chastotalarini taqqoslash orqali aniqlaydi. RAKE (tez avtomatik kalit so‘z qidiruvi): kalit so‘z doirasida yuzaga multi-so‘z ifodalarni chiqaradi. Word2vec: taqsimot o‘xshashligiga asoslangan so‘zlarning vektorli tasvirlarini hosil qiladi va shu bilan bog‘liq atamalarni semantik klasterlash imkonini beradi. Bert va kontekstli qatlamlar chuqur semantik munosabatlarni qamrab oladi va kontekst asosida ma’noni ajratib oladi, bu esa o‘z navbatida terminologiyani chiqarish sifatini sezilarli darajada yaxshilaydi (McCrae, Villegas & Gómez–Pérez, 2017; 29).

Avtomatik ta’riflarni yaratish hisoblash leksikografiyasidagi eng muhim yutuqlardan birini ifodalaydi, chunki u lug‘at uslubidagi tushuntirishlarni to‘g‘ridan-to‘g‘ri korpus ma’lumotlaridan keng qo‘llanma aralashuvisiz ishlab chiqarishni maqsad qilgan. Zamonaviy generativ modellar, xususan transformator arxitekturalariga asoslangan modellar, keng ko‘lamli ko‘p tilli lug‘atlardan, ensiklopediya maqolalaridan va leksik ma’lumotlar bazalaridan ta’rif qismlarini o‘rganadi, bu esa kontekstga mos, insonga o‘xshash ta’riflarni yaratishga imkon beradi. Ushbu tizimlar so‘z paydo bo‘ladigan semantik muhitni tahlil qilish, uning asosiy ma’no tarkibiy qismlarini aniqlash va ularni izohlanadigan tavsifga sintez qilish orqali ishlaydi. Mashinani o‘rganish yondashuvlari, shu jumladan ketma–ketlik modellari va nozik sozlangan til modellari, sintaktik signallarni, semantik rollarni, kollokatsion xatti–harakatlarni va korporatsiyalardan olingan tarqatish

xususiyatlarini kiritish orqali ta'rif sifatini yaxshilaydi. Natijada, avtomatik ta'riflarni yaratish leksikografik ishlarni tezlashtiradi, o'zaro kirish izchilligini ta'minlaydi va bir nechta tillarda tematik lug'atlarni ishlab chiqish uchun kengaytiriladigan yechimlarni taklif etadi. Ushbu afzalliklarga qaramay, insonni qayta ko'rib chiqish terminologiyani takomillashtirish, polisemiyaning hal qilish va ta'riflarni domenga xos standartlarga moslashtirish uchun muhim bo'lib qolmoqda (Meyer, 2021; 280).

Semantik tarmoqlar leksik birliklar o'rtasida mavjud bo'lgan konseptual munosabatlarni ifodalashning tizimli usulini taqdim etadi va shuning uchun avtomatlashtirilgan lug'at tuzishda juda muhimdir. So'zlar tarmog'i, eng nufuzli semantik tarmoq modeli, lug'atni tashkil qiladi sinsetlar bir xil asosiy ma'noga ega bo'lgan so'zlar guruhlari va bu sinsetlarni sinonimiya, giponimiya, meronimiya va antonimiya kabi semantik munosabatlar orqali bog'laydi. Ushbu o'zaro bog'liq tuzilma inson bilishida tushunchalar qanday tashkil etilganligini aks ettiruvchi navigatsion leksik ontologiyani yaratadi. Tematik lug'atni ishlab chiqish tobora ko'proq bunday tarmoqlarga tayanadi, chunki ular atamalarni kengroq kontseptual ierarxiyalar ichida joylashtirishga imkon beradi, bu esa semantik yaqinlikni kuzatish, tegishli lug'at klasterlarini aniqlash va nozik ma'no farqlarini aniqlash imkonini beradi. Ko'p tilli so'zlar va domenga xos ontologiyalarni o'z ichiga olgan so'zlar tarmog'ining kengaytmalari madaniy, ilmiy va texnik bilimlarni birlashtirish orqali leksikografik aniqlikni yanada oshiradi. Semantik tarmoqlar orqali lug'at yozuvlari nafaqat aniqlovchi birliklarga, balki kattaroq kontseptual tizimdagi tugunlarga aylanib, rivojlangan qidiruv funksiyalari, semantik fikrlash va raqamli leksikografik platformalarda aqlli qidirish imkoniyatini beradi (Rundell & Kilgariff, 2011; 7).

Avtomatlashtirilgan leksikografik tizimlar maxsus lingvistik sohalarda lug'atlarni samarali yaratish, boshqarish va tarqatishni qo'llab-quvvatlash uchun mo'ljallangan yangi avlod raqamli infratuzilmalarni ifodalaydi. Bu tizimlar odatda veb asosidagi elektron platformalar yoki integratsiyalashgan kontentni boshqarish

muhitlari sifatida ishlaydi, bu leksikograflarga leksik ma'lumotlarni tuzilgan formatlarda saqlash, tahrirlash, tasniflash va nashr qilish imkonini beradi. Ko'plab hollarda ular til resurs banklari sifatida faoliyat yuritadi, korpuslarga, izoh vositalariga, ontologiyalarga va metama'lumotlar omborlariga markazlashtirilgan kirishni ta'minlaydi, bu ombor zamonaviy hisoblash leksikografiyasining asosini tashkil etadi. Ushbu texnologik muhitda bir nechta dasturiy ta'minot yechimlari ayniqsa ta'sirli bo'lib qoldi. **Lexonomy** ochiq manbali, brauzer asosidagi muhitni taklif etadi, bu hamkorlikda lug'at tahrirlash imkonini beradi va sozlanadigan kirish shablonlarini qo'llab-quvvatlaydi, bu esa uni akademik va pedagogik leksikografiya uchun mos qiladi. **TLex Suite** eng ilg'or tijorat paketlaridan biri bo'lib, ko'p qatlamli leksikografik ma'lumotlar bazalari, avtomatlashtirilgan lemma bog'lanishi, morfologik tahlil va o'zaro bog'lash uchun kuchli vositalarni taqdim etadi, bu esa uni yirik tijoriy lug'at loyihalari uchun juda samarali qiladi. **FLEx (FieldWorks)**, asosan tavsifiy va ozchilik tillari leksikografiyasi uchun ishlab chiqilgan bo'lib, fonologik, morfologik va semantik modullarni birlashtiradi, bu esa lingvistlarga lug'atlarni grammatik tahlil bilan birga qurish imkonini beradi. **OntoLex-Lemon** modeli esa, aksincha, semantik-vebga yo'naltirilgan yondashuvni o'zida mujassam etadi, leksik ma'lumotlarni RDF formatida tuzadi va uni ontologiyalar, bog'langan ochiq ma'lumotlar resurslari hamda ko'p tilli bilim bazalari bilan o'zaro moslashtiradi.

Ushbu tizimlarning funksional imkoniyatlari avtomatlashtirishning leksikografiyada o'zgartiruvchi ta'sirini ko'rsatadi. Avtomatlashtirilgan lemma yaratish lug'at kompilyatsiyasining dastlabki bosqichlarini soddalashtiradi, korpuslardan asosiy shakllarni ajratib, ularni morfologik qoidalar orqali bog'lash orqali bog'laydi. Tizimli tasniflash yoki taksonomiya yaratish leksik elementlarni konseptual va tematik sohalarda joylashtirishga yordam beradigan ierarxik tuzilmalarni taqdim etadi. Semantik bog'lash lug'at yozuvlarini yanada boyitadi, so'zlarni sinonimlar, antonimlar, giperonimlar va boshqa munosabatli kategoriyalar bilan bog'laydi va shu tariqa kognitiv tashkilotni aks ettiruvchi ma'no munosabatlari

tarmog'ini yaratadi. Korpus so'rovlari va qoliplarini tanib olish orqali avtomatik namuna yaratish keng ko'lamda qidiruv talab qilmasdan kontekstga mos ravishda ishonchli foydalanish dalillarini taqdim etadi. Bundan tashqari, kollokatsiya aniqlash statistik jihatdan muhim so'z kombinatsiyalarini aniqlaydi, bu esa frazalar, terminologik xulq-atvor va sohaga xos ifodalarni tushunish uchun muhimdir (Scott & Tribble, 2006; 201).

O'zbek tilshunosligi kontekstida bir nechta raqamli vositalar va platformalar hisoblash usullarining leksikografiyaga tobora ko'proq integratsiyasini misol qilib ko'rsatadi. O'zbekiston Milliy Korpusi morfologik teglash bilan keng ko'lamli matnli ma'lumotlarni taqdim etadi, bu tadqiqotchilarga lug'at yaratishda muhim bo'lgan chastota ma'lumotlari, konkordansiyalar, kollokatsiyalar va semantik naqshlarni olish imkonini beradi. **ZiyoNET** ta'limiy leksik ma'lumotlar bazalarini o'z ichiga oladi va akademik muhitda keng qo'llaniladigan raqamli lug'atlarni taqdim etadi. **O'zTerm** platformasi o'zbek terminologiyasini ilmiy va texnik sohalarda standartlashtirishga qaratilgan bo'lib, sohaga xos lug'atlardan doimiy foydalanishni qo'llab-quvvatlaydi. **O'zbek WordNet** loyihasi kabi yangi tashabbuslar o'zbek leksikografiyasini xalqaro semantik tarmoq standartlari bilan moslashtirishga qaratilgan bo'lib, **Princeton WordNet** asosida o'zaro bog'langan sinsetlar va leksik munosabatlarni ishlab chiqishga harakat qiladi. Ushbu tizimlar birgalikda o'zbek leksikografiyasining tez zamonaviylashuvini ko'rsatadi va avtomatlashtirilgan texnologiyalarning tematik lug'at yaratishda aniqlik, miqyos va kirish imkoniyatlarini sezilarli darajada oshirish imkoniyatini ko'rsatadi (Van der Vliet & Bogaards, 2020; 240).

Tadqiqot metodologiyasi bir nechta lingvistik va hisoblash yondashuvlarini birlashtirib, avtomatlashtirilgan muhitlarda tematik lug'at qurilishini to'liq tahlil qilishni ta'minlaydi. Kompyuter asosidagi lingvistik tahlil usullari katta hajmdagi matn ma'lumotlarini qayta ishlash uchun asosiy vositalarni taqdim etadi, bu esa morfologik, sintaktik va semantik tahlillar orqali leksik birliklarni ajratib olish, tasniflash va talqin qilish imkonini beradi. Korpus lingvistikasi metodologiyaning

asosiy qismi bo‘lib, katta, izohli matn to‘plamlaridan olingan empirik asoslangan dalillarni taqdim etadi; Chastota tahlili, konkordans tekshiruvi, kollokatsiya xulq-atvori va kontekstual o‘zgarish orqali korpuslar lug‘at tuzishni yo‘naltiruvchi leksik foydalanishga obyektiv tushunchalar beradi. Taqqoslovchi leksikografik usullar mavjud lug‘atlar va terminologik resurslar o‘rtasidagi strukturaviy va funksional o‘xshashliklarni aniqlashga yordam beradi, bu esa kirish formatlarini takomillashtirish va ta’rif aniqligini oshirishga yordam beradi.

Tadqiqot lug‘at yaratishda kompyuter texnologiyalari tematik leksikografiyada qo‘llanilganda lug‘at yaratishning tezligi va aniqligida sezilarli yaxshilanishlarni ko‘rsatadi. Avtomatlashtirilgan korpusni qayta ishlash vositalari lemma identifikatsiyasi, terminologik chiqarish va namunalarni olish uchun kerakli vaqtni sezilarli darajada qisqartiradi, shu bilan birga yozuvlar o‘rtasida yuqori darajada izchillikni ta’minlaydi. Mavjud platformalarni baholash shuni ko‘rsatadiki, optimal dasturiy muhitni tanlash lug‘atning hajmi, lingvistik murakkabligi va maqsad qilingan funksiyasiga bog‘liq; TLex, Lexonomy va ontologiyaga asoslangan modellar ish jarayonini boshqarish, semantik tuzilma va o‘zaro ishlashda turli kuchli tomonlarni namoyon etadi. Tadqiqot doirasida ishlab chiqilgan eksperimental model lug‘ati ierarxik semantik munosabatlarni muvaffaqiyatli o‘z ichiga oladi, bu esa konseptual bog‘liqliklar va korpusdan olingan dalillarni birlashtirib, izchil leksikografik resurs yaratish mumkinligini ko‘rsatadi. Statistik natijalar avtomatlashtirilgan tizimlarning operatsion samaradorligini tasdiqlaydi, bu esa terminlarni ajratib olish aniqligi, ta’rif uyg‘unligi va ishlov berish tezligida o‘lchanadigan o‘lishni ko‘rsatadi.

Avtomatlashtirilgan leksikografik tizimlarning tahlili ularning o‘zgartiruvchi afzalliklari va tabiiy cheklovlarini ko‘rsatadi. Avtomatlashtirish unumdorlikni oshiradi, takroriy qo‘lda bajariladigan vazifalarni kamaytiradi va an’anaviy leksikografiya uchun amaliy bo‘lmagan keng ko‘lamli ishlov berishni ta’minlaydi; biroq, texnologik tizimlar mutaxassislar aralashuvisiz polisemiya, idiomatik ifodalar, kam resurslarga ega tillar va sohaga xos terminologiya bilan qiynalishi

mumkin. O'zbek tilining morfologik murakkabligi ayniqsa uning agglutinatsion tuzilmasi, keng qamrovli qo'shimchalari va so'z shakllanishining moslashuvchanligi ko'plab mavjud NLP vositalari uchun integratsiya muammolarini keltirib chiqaradi, bu esa moslashtirilgan morfologik analizatorlar va korpus annotatsiya modellarini talab qiladi. Turkiy tillarning keng oilasi o'xshash strukturaviy xususiyatlarga ega va taqqoslovchi kuzatuvlar shuni ko'rsatadiki, umumiy tipologik xususiyatlar aniq leksik ajratib olish va tasniflash uchun maxsus algoritmlar talab qiladi. Zamonaviy AI texnologiyalari, jumladan ChatGPT va katta til modellari, ta'riflar yaratish, semantik munosabatlarni aniqlash va kontekstual qoliplarni ajoyib ravonlik bilan tahlil qilish orqali tobora ta'sirli rol o'ynaydi; ammo ularning qo'llanilishi tanqidiy baholash bilan muvozanatlashtirilishi kerak, bu xatolar, tarafkashlik yoki ortiqcha umumlashtirishni oldini olish uchun zarur. Kelajakda leksikografik rivojlanish aqlli lug'at tizimlariga o'tishi kutilmoqda, ular multimodal ma'lumotlar, semantik mulohaza, real vaqt rejimida yangilash imkoniyatlari va foydalanuvchi interaktivligini o'z ichiga oladi, bu esa yanada moslashuvchan va kognitiv jihatdan mos leksikografik vositalarga yo'l ochadi.

XULOSA

Tadqiqot axborot texnologiyalari va avtomatlashtirilgan tizimlar tematik lug'at yaratishning samaradorligi, izchilligi va kengaytirilishini sezilarli darajada oshirishini ko'rsatadi. Qo'lda boshqariladigan ish jarayonlaridan yarim avtomatlashtirilgan ish jarayonlariga o'tish orqali leksikografiya lingvistik o'zgarishlarga ko'proq javob beradi va katta korpuslarni boshqarishga ko'proq qodir bo'ladi, natijada yanada boy, ma'lumotlarga asoslangan leksik resurslar paydo bo'ladi. O'zbek tilshunosligi uchun ilg'or raqamli vositalarni joriy etish lug'at tuzishni zamonaviylashtirish, terminologik standartlashtirishni yaxshilash va yuqori sifatli elektron leksik ma'lumotlar bazalarining mavjudligini kengaytirish uchun katta umid beradi. Kelajakdagi tadqiqotlar semantik modellar, multimodal resurslar va real vaqtli lug'at yaratish texnologiyalarini integratsiyalashni o'rganishi kerakligidan maqsad raqamli asrda ma'noni dinamik va o'zaro bog'langan

ifodalarni qoʻllab-quvvatlay oladigan aqlli leksikografik platformalarni yaratishga qaratilgan.

FOYDALANILGAN ADABIYOTLAR ROʻYXATI

1. Atkins S., Rundell M. Amaliy leksikografiya boʻyicha Oksford qoʻllanmasi. – Oksford: Oksford Universiteti Matbuoti, 2008. – 540 b.
2. Bojanić Ć., Duranović A. Maxsus korpuslarda avtomatik terminlarni ajratib olish: RAKE va TF-IDF taqqoslanishi // Xalqaro leksikografiya jurnali. – 2021. – T.34, №3. – B.321–339. DOI: 10.1093/ijl/ecaa028.
3. Čibej J., Fišer D. Lexonomy: Ochiq manbali lugʻat yozish tizimi // E-Lex konferensiyasi materiallari. – 2018. – B.1–12. DOI: 10.5281/zenodo.1477740.
4. Granger S. Korpus asosidagi lugʻatlarda akademik frazeologiya va avtomatlashtirilgan leksikografiya. – Kembrij: Kembrij universiteti nashriyoti, 2017. – 410 b.
5. Hanks P. Leksikografiyada korpus inqilobi // Xalqaro leksikografiya jurnali. – 2012. – T.25, №4. – B.398–436. DOI: 10.1093/ijl/ecs026.
6. Ide N., Véronis J. Soʻz maʼnosini aniqlash boʻyicha maxsus songa kirish: Zamonaviy texnologiyalar // Hisoblash lingvistikasi. – 1998. – T.24, №1. – B.1–40.
7. Kilgarrieff A., Grefenstette G. Vebni korpus sifatida oʻrganish: maxsus soniga kirish // Hisoblash lingvistikasi. – 2003. – T.29, №3. – B.333–347.
8. McCrae J.P., Villegas M., Gómez-Pérez A. OntoLex-Lemon modeli: rivojlanish va qoʻllanilishi // Semantik Veb. – 2017. – T.8, №1. – B.29–45. DOI: 10.3233/SW-160218.
9. Meyer C.F. Korpus lingvistikasi bilan tanishtirish. 3-nashr. – London: Routledge, 2021. – 295 b.
10. Rundell M., Kilgarrieff A. Lugʻatlar yaratishni avtomatlashtirish: holat va istiqbollari // E-Lex materiallari. – 2011. – B.1–15.
11. Scott M., Tribble C. Matnli naqshlar: til taʼlimida kalit soʻzlar va korpus

tahlili. – Amsterdam: Jon Benjamins, 2006. – 260 b.

12. Van der Vliet H., Bogaards P. Raqamli leksikografiya va lug‘at yaratishning kelajagi // *Lexikos*. – 2020. – №30. – B.240–260. DOI: 10.5788/30–1–1568.

REFERENCES

1. Atkins, S., & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press.
2. Bojanić, Ć., & Duranović, A. (2021). Automatic term extraction in specialized corpora: A comparison of RAKE and TF-IDF. *International Journal of Lexicography*.
3. Čibej, J., & Fišer, D. (2018). Lexonomy: An open-source dictionary writing system. *Proceedings of the E-Lex Conference*.
4. Granger, S. (2017). *Academic phraseology and automated lexicography in corpus-based dictionaries*. Cambridge University Press.
5. Hanks, P. (2012). The corpus revolution in lexicography. *International Journal of Lexicography*.
6. Ide, N., & Véronis, J. (1998). Introduction to the special issue on word sense disambiguation: The state of the art. *Computational Linguistics*, 24(1).
7. Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics*, 29(3).
8. McCrae, J. P., Villegas, M., & Gómez-Pérez, A. (2017). The OntoLex–Lemon model: Development and applications. *Semantic Web*, 8(1).
9. Meyer, C. F. (2021). *Introducing corpus linguistics* (3rd ed.). Routledge.
10. Rundell, M., & Kilgarriff, A. (2011). Automating the creation of dictionaries: Where we are and where we are going. *Proceedings of E-Lex*.
11. Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. John Benjamins.
12. Van der Vliet, H., & Bogaards, P. (2020). Digital lexicography and the future of dictionary making. *Lexikos*, 30.