

ЎОЖ 81.373:001.4

ЭЛЕКТРОНДЫ СӨЗДІКТЕР: ҚАЗАҚ ЛЕКСИКОГРАФИЯСЫНЫҢ ЖАҢА МҮМКІНДІКТЕРІ

Дәулетші Дильназ Рахатқызы,
dilnazrahatkizi0505@gmail.com

Қ.Жұбанов атындағы Ақтөбе өңірлік университеті

Ғылыми жетекшісі: филология ғылымдарының
кандидаты, доцент **Б.С.Карагулова**

Андатпа: Бұл мақалада компьютерлік лексикографияның теориялық негіздері мен қазақ тіліндегі электронды сөздіктердің даму бағыттары қарастырылады. Зерттеу барысында отандық және шетелдік электронды сөздік ресурстары салыстырылып, олардың құрылымдық, функционалдық және әдістемелік ерекшеліктері талданады. Сонымен қатар, қазақ тіліндегі заманауи электрондық платформалар – sozdik.kz, qazcorpus.kz, termincom.kz сияқты ресурстардың лексикографиядағы рөлі мен әлеуеті сипатталады. Мақалада қазіргі электронды сөздіктердің артықшылықтары мен кемшіліктері сараланып, болашақтағы даму бағыттары мен жетілдіру жолдары бойынша нақты ұсыныстар беріледі.

Abstract: This article explores the theoretical foundations of computational lexicography and the development trends of electronic dictionaries in the Kazakh language. The study compares domestic and foreign electronic dictionary resources, analyzing their structural, functional, and methodological features. Additionally, the article discusses the role and potential of modern Kazakh electronic platforms, such as sozdik.kz, qazcorpus.kz, and termincom.kz, in lexicography. The advantages and disadvantages of contemporary electronic dictionaries are examined, and specific recommendations for future development and improvement are provided.

Аннотация: В статье рассматриваются теоретические основы компьютерной лексикографии и направления развития электронных словарей казахского языка. В ходе исследования проводится сравнительный анализ отечественных и зарубежных электронных словарных ресурсов, их структурных, функциональных и методологических особенностей. Также рассматривается роль и потенциал современных казахских электронных платформ, таких как sozdik.kz, qazcorpus.kz, termincom.kz, в лексикографии. Анализируются преимущества и недостатки современных электронных словарей, даются конкретные рекомендации по направлениям их дальнейшего развития и улучшения.

***Мақаланың негізгі мақсаты** – компьютерлік лексикографияның теориялық негіздерін зерттеп, қазақ тіліндегі электронды сөздіктердің даму бағыттарын анықтау.*

***Мақаланың өзектілігі** - компьютерлік лексикографияның дамуы ақпараттық технологиялар мен тіл ғылымы арасындағы байланысқа жаңа көзқарас ұсынады. Қазақ тіліндегі электронды сөздіктердің сапасы мен қолжетімділігі тілдің сандық дәуірдегі өзара байланысы мен дамуын анықтайды. Сонымен қатар, электронды платформалардың лексикографиядағы рөлі мен әлеуетін зерттеу қазақ тілін цифрлық кеңістікке енгізуге ықпал етеді. Мақала тілдің қазіргі жағдайы мен даму бағыттарын, сондай-ақ, болашақтағы жаңа технологиялар мен ресурстарды жетілдіру жолдарын айқындауға мүмкіндік береді.*

***Тірек сөздер:** компьютерлік лексикография, электронды сөздіктер, қазақ тілі, лексикографиялық ресурстар, ақпараттық технологиялар, заманауи платформалар, функционалдық ерекшеліктер, лексикографияның дамуы.*

Қоғам мен ғылым мен техникалық прогресс дамыған сайын қоғамда ақпаратпен алмасу, жаңа ақпарат іздеу оңай бола бастады. Технологиялардың, соның ішінде ақпараттық технологиялардың адам өміріне берер пайдасы зор. Олар адамның әртүрлі қызмет саласындағы мүмкіндіктері мен қызметтерін кеңейтуге, жеңілдетуге бағытталған. Соның ішінде, ақпаратты өңдеуден автоматтануы қазақ тіл біліміне елеулі үлес қосты. Ақпараттық автоматтандыру қазақ лексикографиясында компьютерлік бағдарламаларды тиімді түрде пайдалануға мүмкіндік бар екенін дәлелдейді. Осы орайда қазақ тіл білімін дамытуға септігін тигізіп отырған сала - компьютерлік лексикография саласы.

Алдымен, лексикография дегеніміз не? Oxford Dictionary сөздігінде бұл терминге “сөздіктерді жазудың теориясы мен практикасы” деген анықтама берілсе [1], Қазақ әдеби тілінің он бес томдық сөздігінде: “сөздік кітап болып, басылып шығуы да, тарауы да мүмкін. Мұндай сөздіктердің жинақты аты лексикография деп те аталады” делінген[2].

Ал компьютерлік лексикография болса - лексикография ғылымының заманауи техникалық мүмкіндіктерге бейімделіп, электронды форматтағы сөздіктер жасау, жаңарту және қолдану үрдістерін қамтитын бағыт.

А. Жұбановтың еңбектерінде лексикография және тілдің ақпараттық өңделуі жайлы пікірлері кездеседі:

Автоматты сөздіктің міндеті ғылыми-техникалық мәтіндердің барлық түрлерімен әрекет жасау болғандықтан, оның құрамына барлық терминдер

түрлері енулері қажет. Мысалы, оларға жататындар: стандарттанған және ұсынылған, сол сияқты ұсынылмаған, тұрақтанған немесе жаңадан қолданысқа енген терминдер. Сонымен қатар А.Жұбанов автоматты сөздіктің құрылымын анықтау жолында мынадай негізгі үш мәселенің шешімі табылғаны жөн дейді:

- 1) сөздік мақаланың басқы бірлігінің тұрпатын анықтау;
- 2) тілдік бірлікті сипаттауға қатысты ақпарат жиынтығын анықтау;

3) автоматты сөздік базасындағы сөздік мақаланы құрастырудың ұтымды тәсілін таңдау[3]. Сол сияқты Г.Б. Мәденова, С.Б.Бектемірова, Ж.Ж. Күзембекова электрондық сөздіктерді – қажетті лексикалық бірліктерді жедел аударуға, іздеуге және аударма бағытын икемдеуге мүмкіндік беретін, құрылымы мен қызметтік ерекшеліктеріне қарай әртүрлі мақсатта пайдаланылатын заманауи компьютерлік лексикографиялық құрал ретінде сипаттайды[4].

Қазақстанда соңғы онжылдықта қазақ тілін цифрландыру және компьютерлік лексикография бағытында нақты ілгерілеушіліктер байқалды. ҚР Ұлттық Ғылым академиясы жанындағы Тіл білімі институты тарапынан «Қазақ сөздігі» жобасы жасалып, 15 томдық электрондық база құрылды. Қазақ тіліндегі электронды сөздіктер мен ресурстар қазақ тілінің дамуында, оның қолданылуын жеңілдетуде және тіл үйренушілер мен зерттеушілер үшін өте маңызды құралдар болып табылады. Әрбір электронды сөздік өз ерекшеліктерімен және мақсатымен ерекшеленеді, бұл пайдаланушыларға қажетті ақпаратты тез әрі тиімді табуға мүмкіндік береді. Бүгінде халық қолданысы аясында кездесетін электронды сөздіктер қорлары:

❖ Emle.kz – қазақ тілінің орфографиялық электрондық базасы. Бұл сайт орфографиялық ережелерді үйрену және тексеру үшін тиімді құрал болып табылады. Пайдаланушылар бұл ресурсты сөздердің дұрыс жазылуын тексеру, онлайн диктанттар жазу және орфографиялық ережелермен танысу үшін қолдана алады. Базада 100 000-нан астам сөз қамтылған, бұл тіл үйренушілер мен қазақ тіліндегі жазу мәдениетін жетілдіруге ұмтылғандарға пайдалы ресурс.

❖ Sozdik.kz – қазақша-орысша және орысша-қазақша онлайн сөздік. Бұл сөздік екі тіл арасында сөздер мен сөз тіркестерінің аудармасын ұсынады. Әр бағыт бойынша 60 000-нан астам сөздік мақалалар бар, сондай-ақ қазақ тіліндегі 16 000-нан астам сөздің айтылуы да ұсынылған. Пайдаланушыларға тілдің дыбысталуын нақты тыңдау мүмкіндігі беріледі, бұл әсіресе қазақ тілін үйренушілер үшін маңызды.

❖ Sozdikqor.kz – әртүрлі салалар бойынша қазақ тіліндегі сөздер мен олардың аудармаларын қамтитын платформа. Бұл ресурс синонимайзер, сөздерді жүктеу мүмкіндіктері және басқа да қызметтер ұсынады. Пайдаланушыларға тілдік қорын байыту және нақты бір сала бойынша аударма жасауда көмек көрсетеді.

❖ Termincom.kz – ғылыми, білім, экономика, қоғам және әлеумет салаларындағы терминдер мен ұғымдарды жинақтайтын республикалық терминологиялық электрондық қор. Бұл ресурс әртүрлі мамандық өкілдеріне қажетті терминологиялық ақпаратты ұсынумен қатар, терминдердің дұрыс қолданылуын қамтамасыз етеді. Сайт базасында барлығы 389 000 термин қарастырылған.

❖ Lugat.kz – жалпы сөздік қоры 534мыңнан асатын қазақ тіліне арналған кешенді сөздік. Бұл ресурста қазақша-орысша, қазақша-түрікше, қазақша-ағылшынша, қазақша-арабша сөздіктер мен аймақтық сөздік, орфографиялық, синонимдік, аймақтық сөздіктер бар. Платформа қазақ тілінің байлығын толыққанды зерттеуге және тіл үйренушілерге барынша кеңес беруге бағытталған.

❖ Atau.kz – қазақ тіліндегі ономастикалық атаулар мен терминдерді жүйелеуге арналған платформа. Бұл ресурс ерекше атаулар мен терминдермен жұмыс істейтін мамандарға, сондай-ақ ғылыми зерттеулерде немесе аударма жұмыстарына қажетті ақпаратпен қамтамасыз етеді.

Шетелдік және отандық электрондық сөздіктер: салыстырмалы талдау жасап көрелік.

Компьютерлік лексикография саласы халықаралық деңгейде де, отандық ғылымда да қарқынды дамып келе жатқан бағыттардың бірі. Бұл салада шетелдік тәжірибелердің үлгісі зор, себебі шет елдерде электрондық сөздіктерді құрастыру мен қолдану технологиясы ертерек қалыптасып, жүйелі зерттеулер мен практикалық шешімдер жүзеге асқан. Осы орайда шетелдік және отандық электрондық сөздіктердің ерекшеліктерін салыстыра қарастыру компьютерлік лексикографияның даму тенденцияларын анықтауға мүмкіндік береді. Шетелдік электрондық сөздіктер – мысалы, Oxford English Dictionary (OED), Cambridge Dictionary, Merriam-Webster, Collins Dictionary сынды ресурстар – терең құрылымды, көпдеңгейлі лексикографиялық жүйемен ерекшеленеді. Бұл сөздіктер тек сөздің мағынасы мен аудармасын ғана беріп қоймай, оның этимологиясын, қолдану жиілігін, тіркесімділігін, грамматикалық ерекшеліктерін, стилистикалық реңкін және нақты мәтіндердегі қолданыс үлгілерін де қамтиды. Сонымен қатар, мұндай сөздіктер жиі түрде корпус лингвистикасы негізінде құрастырылып, нақты

деректерге сүйенеді. Қолданушы интерфейсі жоғары деңгейде дамыған, іздеу функциясы күрделі алгоритмдер арқылы жүзеге асады, ал кейбірі жасанды интеллект элементтерін де пайдаланады. Отандық электрондық сөздіктер, мысалы, sozdik.kz, termincom.kz, lugat.kz сынды платформалар, соңғы жылдары айтарлықтай жетістіктерге жетіп келе жатса да, әзірше кейбір техникалық және құрылымдық шектеулерге ие. Алайда отандық лексикографияда тілді сақтау, дамыту және кеңінен қолдану мақсатындағы ұлттық мүдде алдыңғы орында тұр. Қазақ тіліндегі электрондық сөздіктер көбіне екітілді (қазақ-орыс, қазақ-ағылшын) аударма бағытына негізделген және семантикалық өріс пен қолданым аясын көрсету жағынан шетелдік баламалардан қарағанда шектеулі. Сонымен қатар, тілдік деректерді автоматты өңдеу, толық мәтінді корпустармен байланыс жасау секілді мүмкіндіктер әлі де даму сатысында.

*№1 Кесте Шетелдік және отандық электрондық сөздіктердің
салыстырмалы сипаттамасы:*

Критерий	Шетелдік электрондық сөздіктер	Отандық электрондық сөздіктер	Мысал сөздіктер	Ресурстар (сілтемелер)
Құрылымдық деңгейі	Көпдеңгейлі, мағыналық, грамматикалық, стилистикалық сипаттамалар қамтылған.	Қысқа анықтама, аударма және кейде айтылым беріледі.	Oxford English Dictionary, Merriam-Webster / Sozdik.kz, Emle.kz	www.oed.com , www.merriam-webster.com / www.sozdik.kz , www.emle.kz
Корпуспен байланысы	Үлкен тілдік корпустар негізінде жасалған.	Корпустық жұмыс енді қолға алынуда (qazcorpus.kz).	British National Corpus, COCA / Qazcorpus.kz	www.english-corpora.org / www.qazcorpus.kz
Мәліметтер қорының көлемі	Миллиондаған сөздік бірліктер, кең дерекқор.	60 000 – 500 000 аралығында, ресурстарға байланысты.	Cambridge Dictionary / Lugat.kz , Sozdikqor.kz	dictionary.cambridge.org / www.lugat.kz , www.sozdikqor.kz
Функционалды мүмкіндіктері	Айтылым, тіркесім, контекст, транскрипция, аудиомысалдар.	Аударма, айтылым, іздеу, диктант құралдары бар.	Longman Dictionary, Collins / Sozdik.kz , Emle.kz	www.ldoceonline.com , www.collinsdictionary.com / www.sozdik.kz
Интерфейс және қолдану ыңғайлылығы	Заманауи UX/UI, мобильді нұсқалар, пайдаланушыға бейімделген жүйе.	Кейбіреулерінде заманауи интерфейс бар, бірақ толық оңтайландырылмаған.	Macmillan Dictionary / Termincom.kz , Atau.kz	www.macmillandictionary.com / www.termincom.kz , www.atau.kz

Лексикографиялық әдіснама	Корпустық, құрылымдық, семантикалық тәсілдер үйлесімі.	Дәстүрлі лексикографияға негізделген, жаңа әдістер енді еніп жатыр.	Oxford, Collins / Lugat.kz, Sozdikqor.kz	www.collinsdictionary.com / www.lugat.kz
Мақсатты аудитория	Тіл үйренушілер, зерттеушілер, аудармашылар, ғалымдар.	Қазақ тілін үйренушілер, оқушылар, мамандар, мемлекеттік мекемелер.	Merriam-Webster, Cambridge / Sozdik.kz, Emle.kz, Termincom.kz	www.merriam-webster.com / www.termincom.kz
Мемлекеттік қолдау	Көбінесе коммерциялық немесе академиялық жоба ретінде жүзеге асады.	ҚР БҒМ, Тіл комитеті, Тіл білімі институтының ресми қолдауымен жасалған.	– / Qazcorpus.kz, Atau.kz, Termincom.kz	www.qazcorpus.kz, www.atau.kz, www.termincom.kz
Мақсаты	Тілдің жаһандық қолданылуын кеңейту, лингвистикалық зерттеулерді қолдау.	Ұлттық тіл саясатын жүзеге асыру, тіл үйрету, терминологияны жүйелеу.	Oxford, COCA / Sozdik.kz, Lugat.kz, Atau.kz	www.oed.com, www.english-corpora.org / www.lugat.kz, www.atau.kz

Жалпылай алғанда, шетелдік сөздіктердің артықшылықтары – мәліметтер қорының ауқымдылығы, интерактивтілік деңгейінің жоғарылығы, жүйелі корпусқа сүйенген құрылымдылығы, айтылым мен тыңдалым мүмкіндіктерінің кеңдігі. Ал отандық сөздіктердің ерекшелігі – ұлттық тілді сақтау мен насихаттауға бағытталуы, мемлекеттік қолдау аясында жүзеге асуы, қазақ тілін үйренушілерге бейімделгені. Дегенмен, қазіргі таңда отандық электрондық лексикографияда да жаңа серпін байқалады. Жоғарыда атап өткен платформалар тілді құрылымдық деңгейде зерттеу және мәтіндер базасын жүйелеу жолында маңызды қадамдар жасауда. Мұның өзі болашақта шетелдік тәжірибелермен тең келетін, тіпті өзіндік ерекшелігімен дараланатын заманауи қазақ электронды сөздіктерінің пайда болуына негіз болады.

Қазіргі таңда компьютерлік лексикография – ақпараттық технологиялармен тығыз байланыста дамып келе жатқан тіл білімінің өзекті саласына айналды. Қазақ тіліндегі электронды сөздіктер тілдің заманауи қолданыс өрісін кеңейтіп, лексикографиялық ресурстардың қолжетімділігін арттыруда үлкен рөл атқаруда. Шетелдік тәжірибемен салыстырғанда, отандық электронды сөздіктердің әлеуеті жоғары болғанымен, оларды

жетілдіру мен сапалық тұрғыдан дамыту қажеттілігі туындап отыр. Осыған байланысты төмендегідей нақты ұсыныстарды атай кетуге болады:

1. Корпустық лингвистика негізінде сөздік базасын жетілдіру: Қазақ тілінің толық мәтіндер корпусы кеңейтіліп, электронды сөздіктермен интеграцияланған жағдайда, сөздердің қолданыс жиілігі, контекстік үлгілері мен стильдік ерекшеліктері ғылыми негізде ұсынылатын болады.

2. Мультимедиялық элементтерді енгізу: Электронды сөздіктерге сөздердің айтылым үлгілері, иллюстрациялық суреттер мен грамматикалық диаграммалар қосылуы олардың практикалық құндылығын арттырады. Бұл әсіресе қазақ тілін үйренушілер үшін тиімді болмақ.

3. Мобильді қосымшалар нұсқаларын жетілдіру: *sozdik.kz*, *emle.kz* секілді платформаларға арналған мобильді қосымшаларға оффлайн режим, дауыспен іздеу және оқу тарихын сақтау сияқты заманауи функциялар енгізілуі қажет.

4. Салалық бағыттағы электронды сөздіктерді көбейту: Медицина, құқық, техника, педагогика, өнер салаларында терминологиялық сөздіктерді электронды форматта әзірлеу мамандарға да, студенттерге де үлкен көмек болмақ.

Жалпы алғанда компьютерлік лексикография – тілдік мұраны сақтау мен жаңғыртудың тиімді жолы. Бұл салада жасалып жатқан әрбір жаңа қадам қазақ тілінің ғылым тіліне, жаһандық ақпараттық кеңістікке бейімделуіне және болашақ ұрпаққа қолжетімді әрі сапалы тілдік ресурстар ұсынуға мүмкіндік береді.

Пайдаланылған әдебиеттер:

1. <https://www.oxfordlearnersdictionaries.com/definition/english/lexicography>
2. Қазақ әдеби тілінің сөздігі. Он бес томдық. 10-том. /Құраст.: М.Малбақов, Н.Оңғарбаева, А.Үдербаев және т.б. - Алматы, 2011. -752 б.
3. Жұбанов А.Қ. 3 Компьютерлік лингвистикаға кіріспе: Оқу құралы. Алматы: Қазақ университеті, 2007.
4. Лингвистикалық зерттеулердегі компьютерлік техно- логиялар: оқу құралы / Г.Б. Мәдиева, С.Б. Бектемірова, Ж.Ж. Кү зем бе ко ва. - Стер. бас. - Алматы: Қазақ универ- ситеті, 2020. - 119 б.