

Серікжан Фаридә Жарасқызы
«Қазақ тілі мен әдебиеті» БББ 2-курс магистранты
Л.Н. Гумилев атындағы Еуразия ұлттық университеті

*Ғылыми жетекші:
ф.ғ.к., доцент м.а.
Шоқабәева С.С.*

«ЭЛЕКТРОНДЫ СӨЗДІК» АУДАРМА ІСІН ЖЕТІЛДІРУДЕГІ КОМПЬЮТЕРЛІК ЛИНГВИСТИКА МҮМКІНДІГІ РЕТІНДЕ

Ц

Цифрлық дәуірде технологияның қарыштап дамуы, ақпараттандыру жүйесінің жаһандануы ғылымның қай саласына болсын үлкен өзгерістер әкелді. Тіл білімі де бұл қатардан қалыс қалған жоқ. Оның бір дәлелі – ақпараттық технология мен лингвистиканың тоғысуынан пайда болған пәнаралық ғылым – компьютерлік лингвистика. Бұл ғылымның пайда болуына жасанды интеллект жүйесін, жасанды тілдер үлгілерін құру қажеттілігінің алғышарттары зор ықпал жасады. Тарихи жүйе ретінде қалыптасқан табиғи тіл – адамдардың өзара қарым-қатынасы үшін, ал жасанды тіл «адам – машина – адам» байланысы үшін қызмет етеді. «Адам – машина – адам» жүйесі бойынша тілді автоматты өңдеудің түрлі жолдары жүзеге асады. Соның аса маңызды бір түрі – машиналық аударма.

Машиналық аударма – толығымен дерлік компьютер арқылы жүзеге асатын, мәтінді бір табиғи тілден екіншісіне аудару үдерісі. Автоматты жүйе көлемді мәтіндердің тез аударылуына қол жеткізуге мүмкіндік береді. Толассыз ақпарат ағыны заманында уақыт үнемдеу – басты қажеттіліктің бірі. Десе де жылдамдыққа сай жоғары сападағы аудармаға қол жеткізу үлкен еңбекті талап етеді. Оның басты себебі, әрине, машинаның, біздің жағдайымызда аударма үдерісін жүзеге асыратын компьютердің, өзіндік ойлау қабілеті жоқ, ол аударма «мәнін» толық түсіне алмайды. Машиналық аударманы қажетті деңгейге дейін жетілдіру үшін сапалы, ауқымды компьютерлік қор енгізілуі тиіс. Мұндай қордың негізгі әрі басым бөлігін «электронды сөздіктер» құрайтыны белгілі.

«Электронды сөздік» терминіне дейін тіл білімінде «компьютерлік лексикография» термині қалыптасқан. ХХ ғасырдың екінші жартысынан бергі еңбектерінде ғалымдар оны түрлі терминмен атайды, мысалы: Ю.Н. Марчук – «есептеуіш лексикография» (1976 ж.), А.Казакевич – «автоматтандырылған лексикография» (1985 ж.), ал В. Марковкин компьютерлік лексикография (1990 ж.) деп берген. Аталған терминнің бәрі де компьютер жадына енгізілген, белгілі бір заңдылыққа негізделген, жүйелі сөздер тобын білдіреді. Уақыт өте келе техниканың қарыштап дамуы компьютерлік лексикография мүмкіндіктерінің кеңеюіне жол ашты. Соның негізінде бұл терминге берілетін анықтамалар аясы да кеңейді. Мысалы, ғалым Е.Ю. Чепик «компьютерлік лексикография – сөздіктерді құрастыру теориясы мен тәжірибесінде компьютерлік технологияны қолданудың әдістерін, технологиясын зерттейтін тіл біліміндегі қолданбалы ғылыми пән. Ол арқылы арнайы мәліметтер базасы, компьютерлік файлдар шкафы, мәтінді өңдеу бағдарламалары сөздік жазбаларын автоматты түрде жасауға мүмкіндік береді» деген тұжырым жасайды [1]. О.П.Фесенко компьютерлік лексикографияны дәстүрлі (қағаз) сөздіктерден олардың электрондық нұсқалары мен модификацияларына көшуді жүзеге асыруға арналған лексикографияның қолданбалы саласы ретінде түсіндіреді. Сонымен қатар «арнайы компьютерлік бағдарламалар арқылы электронды сөздіктерді, мәліметтер базасын дайындайды» деп пікірін толықтырады [2].

Берілген анықтамаларға сүйене отырып компьютерлік лексикографияны тұтас жүйе деп танысақ, оның негізгі нысаны – электронды сөздік. Электронды сөздіктің басты белгілерін санамалай отыра ғалым Л.П. Нелюбин мынадай анықтама береді: «Электронды сөздік – жинақталған, жүйелі, компьютер жадына енгізілген және өңдеу, яғни толықтыру немесе өзгерту мүмкіндігі бар лингвистикалық ақпарат қоры [3]. Я.А. Первановтың да анықтамасы

осыған ұқсас: «электронды сөздік – электронды нұсқада жасалған, электронды түрде пайдалануға ыңғайлы, жетілдіріп отыру мүмкіндігіне ие сөздер жинағы» [4].

Заман талабына сай, қазіргі таңда электронды сөздіктер аса қажеттілікке ие болып отыр. Бұған оның бірқатар артықшылықтары басты негіз болады. Артықшылықтарын санамалап көрсету үшін Н. Сивакова мен Г. Мандрикованың пікіріне сүйене аламыз:

- біріншіден, икемділігі – лингвистикалық бағдарламалы кешен сөздікті пайдалануды жеңілдетеді және пайдалану барысында мүмкіндіктер аясын кеңейтеді;

- екіншіден, тәуелсіздігі – әртүрлі тілдердің лексикалық бірліктері арасында байланыс орната біледі;

- үшіншіден, динамикалылығы – сөздік мақалаларын түзете алу – ескірген сөздерді алып тастау немесе жаңа сөздерді қосу, қамтылған сөз мағынасының аясын кеңейту немесе тарылту мүмкіндігін ұсынады [5; 112].

Электронды сөздіктер құрастырудың аса өзекті мәселе екенін отандық ғалым Г.Б.Мәдиева да өз ғылыми еңбектерінде қарастырып жүр. «Компьютерлік лингвистика кеңістігіндегі электронды сөздіктердің артықшылығы» мақаласында ғалым сөздік мақаласының мазмұнын, оның ішінде әртүрлі өлшемдер бойынша ішінара көрсету мүмкіндігін, морфологиялық және синтаксистік талдау, дыбысты тану және синтездеу сияқты әртүрлі лингвистикалық технологияларға қол жеткізу мүмкіндігін электронды сөздіктердің басты артықшылықтары ретінде атап көрсеткен [6]. Бұдан байқайтынымыз: қазіргі таңда ғылым электронды сөздіктен тек ақпарат емес, тілді автоматты өңдейтін, яғни түрлі талдаулар жасайтын функцияны да талап етеді.

Бүгінде қазақ тілінің қолданыс аясында бірнеше электронды сөздіктің қызмет етіп келе жатқанын білеміз. Оған «lugat.kz», «emle.kz», «termincom.kz», «sozdikqor.kz», «sozdik.kz» сияқты белгілі бір деңгейде өз мақсатына жетіп, нәтиже беріп тұрған электронды сөздіктерді мысалға келтіре аламыз. Кеңірек тоқталар болсақ, «lugat.kz» – жалпы сөздік қоры 530 мыңнан асатын қазақ тіліне арналған сөздік. Бұл сөздікті пайдалану үшін кез келген пайдаланушы желі арқылы тіркеліп, жұмыс істей алады және сөздік мақаланы толықтыруға да мүмкіндік берілген. Бұл кешенді сөздікке қазақша-орысша, қазақша-түрікше, қазақша-ағылшынша, қазақша-арабша сөздіктер мен аймақтық сөздік, орфографиялық, синонимдік сөздіктер кіріктірілген.

«Emle.kz» – орфографиялық сөздік. Аталған сөздіктің басты қызметі – сөздердің дұрыс жазылуын тексеру. Электронды сөздік сүйенетін қордың негізі – түрлі ғылыми кітаптар мен мақалалар.

«Termincom.kz» – түрлі саланы қамтитын: ғылым, білім, экономика, қоғам және әлеумет – республикалық терминологиялық электрондық қор. Тіл білімінде өзекті мәселе саналатын терминдердің біріздендірілмеуі бұл электронды сөздіктің қорына да әсер еткені байқалады.

«Sozdikqor.kz» – ақпараттық технология саласында қазақ тілінің өзіне тән ерекшеліктері ескерілетін, соған сәйкес сөз мағынасы жан-жақты түсіндірілетін, аймақтық қолданыс ерекшеліктеріне сай ақпарат та ұсынылатын электронды платформа.

«Sozdik.kz» – қазақша-орысша, орысша-қазақша ең танымал, өте жиі қолданылатын электронды сөздік. Бұл жүйенің ішінде мәтіндерді аударуға болады және қазақша мәтінді кириллицадан латынға көшіруге арналған конвертер функциясы қамтылған. Бір ерекшелігі: пайдаланушы сөздікке жаңа сөз қоса алады.

Аталған электронды сөздіктер негізінен дәстүрлі сөздіктің автоматтанған түрі болып есептеледі. Басты олқылық электронды сөздіктердің тұтас жүйе ретінде мүмкіндіктерінің жетілдірілмеуінде. Жеке сөздердің аудармалары ғана енгізілгендіктен, сөз бен сөз тіркесімі, тіркесу валенттілігі сияқты мәселелер әлі де қажет деңгейде қамтылмай жатыр. Мұның есебінен осы электронды қорға сүйенетін машиналық аударма сапасы да жақсармай тұр. Аудармада сұранысқа ие тілдік ақпарат – мәтін. Мәтіннің сөзден ғана емес, сөз тіркестерінен, одан да күрделі құрылым – сөйлемдерден тұратыны белгілі. Сәйкесінше, мәтінді аударуда сөзбе-сөз аудармадан бөлек, сол сөздердің дұрыс байланысуы, орын тәртібі,

мағыналық тұтастығы, грамматикалық заңдылықтары ескерілсе ғана сапалы аударма аламыз. Бұл мақсатқа қол жеткізу үшін электронды сөздіктер тілді автоматты өңдейтін мүмкіндіктерге ие күрделі жүйе ретінде құрастырылуы тиіс.

Электронды сөздіктердің сапасын жақсарту мәселесінің аса өзекті екенін отандық жас ғалымдардың жүргізген зерттеулерінен де байқауға болады. Мысалы, А.Н. Шормакова қазақ тіліндегі автоматтандырылған синонимдер тізімін құру мақсатында ағылшын электронды сөздіктеріне жүгініп, кейін оны Google аударма қызметі арқылы қазақшаға аударған. Зерттеу нәтижесінде сөздік қоры 3,5 млн. ағылшын синонимдер сөздігі қолданылып, 1000 синоним сөздерден тұратын 6000 қазақ синонимінің тізімі құрылған [7]. Негізге алынған сөздер қарапайым, жеңіл, сондықтан зерттеу нәтижесінде алынған синонимдік қатар шет тілдегі лексикалық қордан жинақталмай, қазақ тіліндегі электронды сөздіктер қорында болуы тиіс еді. Одан бөлек машиналық аударма негізінде синоним деп танылған кей сөздер мағыналық тұрғыда ол қатарға кірмейді, бұл мәселе аударманың ақсап тұрғанының белгісі деп қабылдануы керек.

Электронды сөздік негізінде сапалы машиналық аудармаға қол жеткізу мақсатында жасалған «Apertium» платформасына тоқталып кетпеуге болмайды. «Apertium» – арнайы лексикалық қоры бар машиналық аударма жүйесі. Бұл жүйе аясында:

- apertium-kaz.kaz.lexc: қазақ тіліндегі сөздерге морфологиялық талдау жасау функциясы;

- apertium-eng-kaz.eng-kaz.dix: ағылшын-қазақ тілдері арасындағы тікелей аударма жасау функциясы;

- apertium-kaz-rus.kaz-rus.dix: орыс-қазақ тілдері арасындағы тікелей аударма жасау функциясы қамтылған.

Бұл платформаның сапасын арттыру мақсатында тілші ғалымдар мен информатика саласының мамандары бірлесе жұмыс жасаған. Оған бірқатар қазақ ғалымдарының «бұл жүйе жеке сөз деңгейінде сәтті жоба болғанымен, толық сөйлем құрылымы бойынша дұрыс нәтиже бермейді» деген тұжырымы алғышарт болған [8]. Бұл жүйеде септік, көптік, тәуелдік және жақтық категорияларды автоматты ажырату функциясы бар. Бірақ жалғаулардың дыбыстық нұсқасының көп болуы бұл алгоритм жұмысында қиындық туғызады. Бұл жайлы алгоритмді құрастырумен айналысқан информатика саласының мамандары төмендегідей мысал келтіреді: «слово «мектебінен» (с его школы), «мектеп» изменился на «мектеб», но окончание «інен» отличается от суффикса («сыдан»), который относится к «қалта» (карман): қалтасынан (с его кармана)». Яғни, толық сапалы талдау, аударма өте күрделі алгоритм құруды талап етеді. Ол үшін лексикалық қор мен сөздердің түрлену ережелері, олардың дыбыстық нұсқалары математикалық тұрғыда есептеліп, нақты бірнеше формулаға топтастырылуы тиіс. Сонымен, жобаны жетілдіру бойынша жұмыс істеген мамандар сапалы нәтижеге қол жеткізу ұзақ уақытты және ауқымды қор мен сөйлемді автоматты талдау функциясының толық енгізілуін қажет ететінін атап көрсетеді.

Қорытындылай келе, қазіргі таңда лексикография ғылымы өз алдына цифрлық дәуір талаптарына жауап беретін жаңа мақсаттар мен міндеттер қоюына байланысты сапалы электронды сөздіктер құрастыру аса өзекті. Бұл – машиналық аударма ісін компьютерлік лингвистика мүмкіндігі арқылы жетілдірудің бірден-бір жолы.

Пайдаланылған әдебиеттер:

1. Чепик Е.Ю. Компьютерная лексикография как одно из направлений современной прикладной лингвистики [Электронный ресурс] // Ученые записки ТНУ.Т.19 (58). Филология. – 2006. – № 2: Симферополь. – С. 274–280
2. Фесенко О.П. Компьютерная лексикография: программные средства и продукт научной отрасли [Электронный ресурс] // Инновационное образование и экономика. – 2014. – № 15. – С. 28–30
3. Нелюбин Л.П. Перевод и прикладная лингвистика. – М.: Высшая школа, 1983. – 208 с.
4. Перванов Я.А. Заметки по электронной лексикографии. [Электронный ресурс]. URL: <http://arxiv.org/pdf/1107.1753.pdf>
5. Сивакова Н.А. Лексикографическое описание английских и русских фитонимов в электронном глоссарии: дис. канд. филол. наук / Н.А.Сивакова. – Т., 2004. – 162 с.

6. Садық Д.А., Мәдиева Г.Б. Компьютерлік лингвистика кеңістігіндегі электронды сөздіктердің артықшылығы Абай атындағы ҚазҰПУ-дың ХАБАРШЫСЫ, «Филология» сериясы, №3(73), 2020. – Б. 180–184
7. Шормакова А.Н. Қазақ тіліндегі автоматтандырылған синонимдер тізімін құру [Мәтін] / А.Н. Шормакова, У.А. Тукеев // Механика және технологиялар / Ғылыми журнал. – 2022. – № 3 (77). – Б. 44–49
8. Рахимова Д.Р., Қалдашбеков Е.Е., Мұсабекова Ж.Ғ., Абақан М., Кызырканова С., Жамалиева А., Абдуали Б. Arertium платформасы негізінде орыс тілінен қазақ тіліне аудару жүйесін құру. //Материалы международной научной конференции студентов и молодых ученых "Фараби Әлемі», Алматы, Казахстан, 13-16 апреля, 2015 г. – Алматы, 2015. – С. 175

Арыстанғалиқызы Аяулым

Қазақ тілі мен әдебиеті мамандығының 2-курс магистранты

Х.Досмұхамедов атындағы Атырау университеті

arystangalikyzy@bk.ru

Ғылыми жетекші:

Амангазиева М.К.

ҚАЗАҚ ТІЛІНДЕГІ АЛҒАШҚЫ БАСЫЛЫМДАРДЫҢ ТАҚЫРЫПТЫҚ ЕРЕКШЕЛІКТЕРІ

Алғаш жарық көрген қазақ тілді баспалар тұжырымдылығымен, жоғары көсемсөзділігімен, қазақ әдеби тілінің мағыналылығымен, өз халқының көкейкесті мүдделеріне адал қызмет етуімен ерекшеленді. Газеттер өз беттерінде халық ағарту, кітап шығару мәселелеріне басымдық берді. XIX ғасырдың екінші жартысындағы ұлы ағартушылардың дәстүрлерінде тәрбиеленген қазақ демократияшыл зиялылары Шығыс пен Батыс ағартушыларының тәжірибесін жинақтап, тұтас доктрина жасады. Әр басылымның өзіндік ерекшеліктері болса да, барлығының мүдделері бір болды. Басылым беттеріндегі жарияланымдардың едәуір бөлігі тіл мен әдебиет, этнография, мәдениет пен тәрбие, педагогика мәселелеріне, рыноктың жағдайларға бейімдей отырып, халық санасын жетілдіруге арналды.

Кілт сөздер. Алғашқы басылымдар, тіл мен әдебиет мәселелері, басылымдардың зерттелуі, мақалалар.

«Қазақ» газетінің бірінші нөмірінде басылған бағдарламалық мақаласында А.Байтұрсынов отаршылдық режим жағдайларындағы ұлттың тағдыры туралы мәселені ашық қойды. Ол былай деп жазды: «Өзіміздің өзіндік ерекшелігімізді сақтап қалуымыз үшін біздің бар күш пен құралды жұмсап, оқу – ағартуға және ортақ мәдениетке ұмтылуымыз қажет; бұл үшін біз әдебиетті ана тілінде танытумен айналысуды бірінші борыш деуге міндеттіміз. Өз тілінде сөйлейтін және өз әдебиеті бар халықтың ғана дербес өмірге үміттенуге құқылы екенін ешқашан ұмытпау керек».

«Қазақ» газеті ХХ ғасырдың басындағы азаттық қозғалысының стратегиясын анықтап берді. Ә.Бөкейханов былай деп атап өтті: «Жуық арада далада, бәлкім, қазақ арасында белең алып келе жатқан екі саяси бағытқа сәйкес екі саяси партия ұйымдастырылатын сыңайлы. Олардың бірі ұлттық-діни партия деп аталуы мүмкін және де қазақтарды басқа мұсылман халықтарымен біріктіру оның мұраты болып табылады. Басқа бір батыстық бағыт қырғыз даласының болашағы-осы сөздің кең мағынасында-батыс мәдениетін саналы түрде іс жүзіне асыруда деп біледі».

«Қазақ» газеті екінші бағыттың ой –пиғылын көрсетті. Ол тұжырымдамалық проблемаларды көрсеткен кезде соны негізге ала отырып жазды. Ең алдымен Ә.Бөкейханов, А. Байтұрсынов қазақ халқының өркениет, түрік дүниесі тарихында алатын орнын анықтап берді. Халықты топтастыру, оның ежелгі мәдениетін және демократиялық Ресей құрамындағы қазақ ұлттық мемлекеттігін қайта түлету проблемалары газет қызметінің өзегіне айналды. «Шаруашылықтағы өзгерістер» деген