

ҚАЗАҚ ТІЛІНІҢ ҰЛТТЫҚ КОРПУСЫНЫҢ КОНЦЕПТУАЛДЫҚ НЕГІЗДЕРІ

Анар Фазылжан

Ахмет Байтұрсынұлы атындағы Тіл білімі институтының директоры,
филология ғылымдарының кандидаты,

ҚХА Ғылыми-сарапшылық кеңесінің мүшесі

Қазақстан, Алматы

e-mail: tilbilimi@bk.ru

Аңдатпа. Бұл мақалада Қазақ тілінің ұлттық корпусының концептуалды негіздері, ҚТҰК қажеттілігі, тіл тарихы, қазақ тілін оқыту, ҚТҰК және тілдік мұра, ҚТҰК бірінші кезең басымдықтары, ҚТҰК екінші кезең басымдықтары, тілтанымдық аннотациялармен қамсыздандырылу принциптері қарастырылған.

Түйін сөздер: Қазақ тілінің ұлттық корпусы, тіл білімі институты, тілдік мұра, тілтаным, тіл мәртебесі.

Аннотация. В данной статье рассмотрены концептуальные основы Национального корпуса казахского языка (НККК), необходимость НККК, история языка, преподавание казахского языка, НККК и языковое наследие, приоритеты первого этапа НККК, приоритеты второго этапа НККК, принципы обеспечения языковыми аннотациями.

Ключевые слова: Национальный корпус казахского языка, Институт языкознания, языковое наследие, языкознание, статус языка.

Abstract. This article discusses the conceptual foundations of the National Corpus of the Kazakh language, the need for the NCCC, the history of the language, the teaching of the Kazakh language, the NCKL and the linguistic heritage, the priorities of the first stage of the NCKL, the priorities of the second stage of the NCKL, the principles of providing language annotations.

Keywords: National Corpus of the Kazakh language, Institute of Linguistics, linguistic heritage, linguistics, language status.

Қазақ тілінің ұлттық корпусы (*әрі қарай - ҚТҰК*) – үздіксіз толықтырылып, жетілдіріліп отыратын ашық жоба. Оны әзірлеуге қазақ тілін зерттейтін әлемдегі жалғыз ғылыми мекеме Ахмет Байтұрсынұлы атындағы Тіл білімі институты ғалымдарының әлеуеті толық жеткенмен, корпус мәтіндердің ауқымды электрондық көлемін қажет етеді. Ол тілтанушы мен ІТ-маманының тандемі негізінде жүзеге асады және оған тілтанушылардың ғана қамқорлығы

жеткіліксіз, сондықтан бірнеше қаржы көздерінің қолдауы аса қажет. Соған қарамастан, Институт өз қызметі шеңберінде ҚТҰК мәтіндік базасын ұлғайтты, ішкорпустарын кеңейтті, метабелгіленімі мен тілтанымдық белгіленімін, сайты, іздеу жүйесін – корпус-менеджментін жетілдірді. ҚТҰК – мемлекеттік тілдің лингвистикалық базасы ретіндегі инновациялық ақпараттық-анықтамалық жүйе, оның мәтіндері арнайы белгіленім арқылы өңделіп, іздеу тетіктерімен қамтамасыз етіліп, статистикалық дәл ақпаратты барынша қысқа уақытта алуға мүмкіндік береді.

Теңгерімделген және көлемі ауқымды (бірнеше миллион сөзқолданыстан басталатын), метабелгіленіммен, лингвистикалық белгіленімдермен терең аннотацияланған, пайдаланушылардың қазақ тілінің синхронды және диахронды сипаты туралы, оны оқыту, үйренуге қатысты алуан түрлі сұранысына жауап беретін, әлемдік деңгейдегі талаптарға сәйкес келетін инновациялық-ақпараттық ашық жүйе ретіндегі Қазақ тілінің ұлттық корпусын әзірлеу және жасау идеясы ең алдымен Ахмет Байтұрсынұлы атындағы тіл білімі институтының қабырғасында айтылды.

Тұлғалар. Қазақ тілінің ұлттық корпусын жасау мәселесін күн тәртібіне қойған Институттың бас ғылыми қызметкері, филология ғылымдарының докторы, профессор Асқар Жұбанов болды. Профессор А. Жұбанов ұлттық тіл корпусының ғылыми негіздерін әзірлеуге зор үлес қосқан ғалым, оның қазақ тілін автоматтандырудың формалды белгілеріне, статоллингвистика мәселелеріне, математикалық тіл біліміне, корпустық тіл біліміне арналған іргелі әрі қолданбалы зерттеулері корпустық лингвистика бойынша жүргізіліп келе жатқан ізденістің барлығы үшін тоқталмай кетпейтін бестселлерге, тірек еңбектерге айналды [1].

Қазақ корпусын жасау ісіне мұрындық болған ғалымдар қатарында профессор Элеонора Сүлейменованы ерекше атауға болады. Корпустық лингвистикаға қатысты терең ғылыми зерттеулерімен бірге [2], ғалымның Әл-Фараби атындағы Қазақ ұлттық университетінде тұңғыш рет «Қолданбалы лингвистика» мамандығын ашып, профессор Асқар Жұбановты оқытушылыққа шақырып, осы саланың алғашқы кәсіби мамандарын әзірлеген ұстаздық еңбегі бүгінде өз жемісін беріп отыр. Ахмет Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің өзегін осы мамандар құрап, олар отандық корпустық ресурсты әзірлеуші авангард құрам.

Қазақ тілінің ұлттық корпусын жасау ісіне Ахмет Байтұрсынұлы атындағы Тіл білімі институтының ұжымы түгел жұмылды, оның бүгінгі деңгейінің барлық жетістігі, озық тәжірибесі мұнсыз мүмкін болмас еді. Осы істе корпустық лингвистика бойынша іргелі, қолданбалы зерттеулер жүргізе жүріп, корпустық тіл білімі мамандарын даярлау, оқулық жазу ісімен қоса, корпустың мәтіндік базасын жинау, тілтанымдық белгіленімін әзірлеу, интернет сайт параметрлерін қалыптастыру, сайтқа тілдік базаны орналастыру сияқты толып жатқан күрделі ірілі-ұсақты жұмыстарды тынбай атқарып жүрген мықты маман филология ғылымдарының докторы, профессор Айман Жаңабекованың [3] еңбегі орасан зор.

ҚТҰК қажеттілігі. Толық әрі ыңғайлы белгіленіммен қаматамасыз етіліп, үйлестірілген және ауқымды көлемнен тұратындықтан, Қазақ тілінің ұлттық корпусы ғана тілдік нормалардың қалыптасуы мен бекітілу динамикасы туралы толық мәлімет береді, сонымен қатар олардың жүйелі әрі стандартталған түрде жазу ережесіне, сөздіктерге, грамматикаларға, оқулықтарға енгізуге мүмкіндік тудырады. Бұл өз кезегінде тіл жүйесінде болып жатқан неологиялық, грамматикалық және ортологиялық процестерді анықтауға, семантикадағы, тіл бірліктерінің тіркесіміндегі болып жатқан өзгерістерді тануға, яғни узустағы қолданысын кеңейтеді (узус - белгілі бір жағдайда тілдік бірліктердің қолданылуын айтады). ҚТҰК жасау қажеттілігі әлемдік тәжірибеден артта қалушылықтан туындап отыр – бүгінгі күні көптеген үлкен және шағын тілдер өз ұлттық корпусын жасап алған. Профессор Э.Сүлейменова «корпустың мәні микроскоп, телескоп, рентген аппараты, томограф сияқты т.б. ғылыми өнертабыстың жемісімен барабар деп көрсетеді, дамыған елдер зор зияткерлік әрі материалдық шығынды қажет ететініне қарамастан, мұндай жобаны жасап отыр, сондықтан ҚТҰК жасау – қазақ тілтануы ғана емес, тіпті мемлекет алдындағы да міндет. Осы айтылған тілтанушылық мақсатпен қатар корпус төмендегідей ауқымды әлеуметтік-мәдени мәнділікке ие болып отыр.

ҚТҰК және ұлттық бірегейлік. Қазақ тілі ірі тілдердің бірі ретінде виртуалды кеңістіктегі бірегейлік үшін жүріп жатқан әлемдік күреске тартылып отыр. Саяси-әлеуметтік жақтан мемлекеттік тіл ретінде жоғары мәртебеге ие болғанымен, ол ақпараттық мүмкіндіктер жарысында әлі де қалып келеді. Қазақ тілінің ұлттық корпусы – мемлекеттік қазақ тілін барлық көркемдік құралдары мен мүмкіндіктерін толықтай қамти отырып сипаттайтын және оның мәртебесін арттыра түсетін заманауи тиімді құралдардың бірі ретінде ұлттық мәдениетті дамытуға, ұлттық-мемлекеттік бірегейлікті нығайтуға, ортақ мәдениетті қалыптастыруға, сол арқылы Қазақстан қоғамының ұйысуына мүмкіндік жасайды.

ҚТҰК және тілдік мұра. Корпус қазақ тілінің бүкіл байлығын жинап, бір жерге шоғырландыруға және оның узустағы бойынша графикалық бөлінісі салдарын жоюға көмектеседі. Қазақ тілінің бай мұрасы кирилл, латын және араб графикаларымен бөлініп, басы бірікпей жатыр. ҚТҰК осы тілдік жазба мұраны түгендеп, жинап, әлемнің барлық тілтанушыларының зерттеу нысанына қолжетімді түрде ұсына алады. Жазба мұралардың Қазақстандағы және басқа да елдердегі қызығушылық білдірген көпшілікке қолжетімді болуын қаматамасыз етеді

ҚТҰК және тіл тарихы. Қазақ тілінің тарихын жаңғырту мен стандартталуы мәселесін шешуде ҚТҰК тапсырмайтын құрал. ҚТҰК тіл дамуынан барынша көлемді ақпарат беретін көне және заманауи мәтіндер жиынтығы ретінде тілде үздіксіз болып жататын құбылыстарды тіркеп отыруға, әрбір тілдік құбылысты нақты және тұтастай сипаттауға мүмкіндік тудырады.

ҚТҰК жән қазақ тілін оқыту. ҚТҰК тіл оқыту оқулық, оқу құралдарын, түрлі типте жаттығу, онлайн-мультимедиялық сервистер әзірлеуде кең қолданыс табады, өйткені оқыту мәтіні үшін аса маңызды сөздердің жиілігі мен контексті,

қолданыс табатын мәтін типі мен жанры туралы ақпаратты ҚТҰК жүйелі түрде беріп отыра алады.

ҚТҰК және ұлттық сөздік қор. ҚТҰК тілдің эмпирикалық базасын кеңейтіп, ол туралы барынша терең ақпарат ауқымын ұсынады. Бұл түрлі типтегі қазақ сөздіктерін жасау сияқты зор сұранысқа ие мұқтаждықты өтеуге, сол сияқты қазақ тілінің сөздік қорының бүгінгі санын анықтауға септігін тигізеді (орыс тілінің үлкен академиялық сөздігінде 150 000, Оксфорд сөздігінде – 600 000, үлкен «Қазақ сөздігінде» 106 000 сөз бар), терминологиялық жүйесін реттеп, шет тілінен енген сөздерді игеру, стандарттау жолдарын көрсетеді. Бұл тілтанушылар үшін қазақ тілінің барлық саласы мен нақты тілдік құбылыстары (мысалы, кәсіби сөздер, жаргондар, социолектілер, сән сөздер мен неологизмдер т.б.) туралы дәл әрі дәйекті мәліметтер алып, жаңа теориялық міндеттерді алға тарту арқылы қолданбалы лексикография бағытындағы көптеген міндетті шешуге септігін тигізеді: мәселен, әртүрлі жиілік сөздіктер, мектеп оқушыларына арналған ортологиялық құралдар, заманауи түсіндірме сөздіктер, тарихи сөздіктер, диалектологиялық атластар мен аймақтық сөздіктер, ономастикалық анықтамалық пен сөздіктер, морфемдік, сөзжасамдық т.б. грамматикалық сөздіктер, валенттілік сөздігі мен жарыспа сөздер сөздігін, антоним, синоним, омоним, пароним сөздіктерін, ресми іскери, кәсіби тіл, жаргон сөздер сөздігін, орфографиялық, орфоэпиялық сөздіктер мен құралдар шығаруға база болады.

ҚТҰК және тілдік норма. Корпусқа жазба және ауызша мәтіндердің ауқымды көлемі енгізілетіндіктен, ол тілді кодификациялау мен стандарттаудың негізгі бола алады. Қазақ тілінің кодификациялануы мен стандартталуы оның мемлекеттік тіл ретінде сапалы қолданыс табуына жағдай жасайды, сөйтіп, қазақ тіліндегі коммуникацияны жеңілдетуге мүмкіндік береді.

ҚТҰК маңыздылығын ескере отырып, Ахмет Байтұрсынұлы атындағы Тіл білімі институты ғалымдары оны әзірлеуді алдымен гранттық зерттеу жобалары аясында бастап, кейіннен ірі ғылыми жобаларда жалғастырды. ҚТҰК жасалу тарихын осы ұжымдық монографияның «Қазақ тілінің ұлттық корпусын жасау тәжірибесі» атты тарауында профессор А.Жаңабекова толық әрі жан-жақты сипаттап көрсетіп отыр.

ҚТҰК әзірлеудің алғашқы кезеңінде ауқымды корпустарға тән мындай негізгі қағидалар мүмкіндігінше ескеріліп отырды:

Ауқымдылық немесе репрезентаттық қағидасы. Бұл қағида қазақ мәтіндерінің мейлінше барлық (тілдің көркем, публицистикалық, оқу, ғылыми, іскерлік, эпистолярлық және т.б. стильдері мен әртүрлі жанрындағы) түрін қамтуды білдіреді. Мұны толықтай жүзеге асыру корпустың мәтіндік базасы әрі ішкорпустық жүйесі неғұрлым ауқымды болған сайын жеңілдей бермек.

Теңгерімділік қағидасы. Бұл қағида әртүрлі кезеңдегі, жанрдағы, стильдегі, автордағы және т.б. тіл фактілерінің көрініс табатын қолданыс салалары мен жағдаяттарындағы мәтіндерді корпуста пропорционалды бейнелеуді, тиісті тарихи кезеңге тән мәтіндерді сол кезеңдегі әрбір жанр,

стильдің үлесіне сәйкес түрде корпусқа енгізуді білдіреді. Мәселен, ҚТҰК-да XIX ғасырдағы мәтіндер ішінде көркем әдебиет, ал қазіргі заманғы мәтіндер арасында публицистикалық, содан соң ғылыми мәтіндер стильдегі үлгілері басым болуы тиіс. ҚТҰК-ның қазіргі деңгейінде бұл принцип негізгі корпуста ғана жақсы сақталып тұр. Ал ауызша мәтіндер, тарихи мәтіндер ішкорпусында мәтіндік базаның әлі де толығыуы қажеттілігінен ішінара ғана жүзеге асып отыр. Әрине, нақты тілге мінсіз ұқсастық жасау мүмкін емес, бірақ әр уақыт кезеңі үшін корпустағы жанрлардың үлесін анықтауды шешуші түрткі жайт тілдің табиғи тіршілігіне мүмкіндігінше жақындыққа ұмтылу болып табылады.

Тілтанымдық аннотациялармен қамсыздандырылу қағидасы. Терең аннотация корпусқа тән барлық тілтанымдық белгіленіммен қамсыздандыруды білдіреді, тілтанымдық аннотацияның ең танымалы әрі ауқымдысы морфологиялық белгіленім болып табылады. Корпусқа сонымен бірге фонетикалық, просодикалық, лексика-семантикалық, фонологиялық, синтаксистік, сөзжасамдық т.б. белгіленімдер тән. ҚТҰК белгіленім жүйесінің басқа корпустардан үлкен өзгешелігі – мәдени-семантикалық және идеографиялық (фразеологиялық) белгіленімдердің әзірленуі. Қазақ тілінің этномаркерлік лексика-фразеологиясын қамтитын мәдени-семантикалық белгіленім ҚТҰК-ын рухани және материалдық ұлттық мәдениет туралы біліммен қамтамасыз етіп, пайдаланушыларды қазақ мәдениеті негізінде бірыңғай ұлттық-мемлекеттік бірегейлікке ұйытуға жағдай жасайды. Белгіленімнің мұндай түрі ҚТҰК-ның басқа тіл корпусынан айырмашылығын мен өзіндік ерекшелігін қамтамасыз етеді, өйткені мәдени-семантикалық белгіленім қою тілдік корпустар жасаудың әлемдік тәжірибесінде алғаш рет қазақ тілінің корпусында қолға алынып отыр. Бүгінгі корпуста синтаксистік белгіленім ғана іске қосылмады.

Зерттеуші мен пайдаланушы сұраулары бойынша оңтайландыру қағидасы. Әдетте алғашқы корпустар тілтанушының зерттеуіне қажет ауқымды эмпирикалық материал жинақтау, жүйелеу мақсатында әзірленді. Кейіннен кез келген пайдаланушы сұрауын өтеу керек деген қағида да ескеріле бастады, ал қазақ тілінің ұлттық корпусы басынан бастан осы екі мұқтаждықты бірге өтеу мақсатында жасалды. Бір жағынан, тілтанушы зерттеулерін жақсарту, екінші жағынан, қалың жұртшылықтың мемлекеттік тіл мәртебесіндегі қазақ тілі бойынша жан-жақты ақпарат ала алу мүмкіндігі ескеріліп, корпустың базасы, белгіленімі мен іздеу жүйесі жетілдіріліп отырды. Іздеуді жеке сөзбен ғана емес, сөзформамен және әртүрлі параметрлер (авторы, шығарма стилі, мерзімі, дереккөзі, т.б.) бойынша жүргізуге болатындай іздеу жүйесі ұсынылды, әрі ол әрбір ішкорпус ерекшелігіне қарай түрленіп берілді. Мұның бәрі келесі кезеңде үздіксіз өзгертіліп, жақсартылып, жетіліп отырады.

Онлайн режимде қолжетімділік қағидасы. ҚТҰК көпшілікке онлайн режимде қолжетімділігі 2020 жылдан бастап жүзеге асырылды. Корпус qazcorpus.kz сайтында орналасты. Басқа корпустарға қарағанда ҚТҰК ерекшелігі оның барлық қызметінің тегін болуында жатыр.

Ашықтығы, құрамы, көлемі мен тереңдігі бойынша үздіксіз толығып отыру қағидасы. Аталған қағиданы жүзеге асыру барлық корпус жасаушылардың міндеті, алайда бұл оның орындалуы көбіне корпус жасаушы мамандардың мүмкіндік аясында үнемі бола бермейтін қаржыландыру мәселесімен тығыз байланысты. Мұндайда ұлттық корпустар жасаудың маңызын қалың жұртшылыққа ғана емес, қаржы көздеріне жауапты шешім шығарушы органдар мен құзырлы тұлғаларға түсіндіріп, менеджерлік іскерлік таныта алатын жетекшінің де қажеттілігі шындық. Алдағы уақытта Институт ҚТҰК-ын әзірлеу, жетілдіру, толықтыру ісін өз қызметімен үздіксіз байланыстыруды, барлық зерттеу жобаларының бір міндеті ретінде корпустық ресурс жасауды немесе оны жетілдіре алуды көздеп отыр.

ҚТҰК бірінші кезең басымдықтары туралы. Осыдан сегіз жыл бұрын жариялаған мақаламызда қазақ тілдік корпусының концептуалды параметрлері туралы ойларымыз «орынсыз көрінуі мүмкін, өйткені қазақ тілінің ұлттық корпусын әзірлеу мен жасау мәселесі теориялық та, тәжірибелік тұрғыдан да толықтай шешімін таппаған» [4] деп көрсеткен едік. Қазіргі кезде бұл ой әбден ескіргені, ақиқатқа жанаспайтыны анық. Осы уақытқа дейінгі аралықта бұл мәселе қазақ тілтанушылары шоғырланған Ахмет Байтұрсынұлы атындағы Тіл білімі институты ұжымының зор еңбегінің арқасында шешімін тапты. Алдағы уақытта көтерілетін мәселелер құрылған, біршама жасақталған «жас» ресурсты – Қазақ тілінің ұлттық корпусы мен оның сайтын заман талабына қарай жетілдіру туралы ғана болмақ.

Тілдік корпус жасаудың жалғыз ғана басымдығы болады, ол – терең аннотациялар берілген мәтіндердің ауқымды массивтерін жасау арқылы лингвистикалық зерттеулердің сапасы мен олардың эмпирикалық базасын жақсарту екені мәлім. Дегенмен дамыған елдердің тілдік корпус жасаудағы тәжірибесі мен тілдік жағдаяттары отандық жағдайдан ерекшеленеді. Бүгінгі таңда ең ірі тілдік корпустар инфрақұрылымы және ақпараттық технологиялары дамыған елдерде жасалғаны белгілі, сонымен қатар тағы бір айта кетер жайт, бұлар - монотілдік жағдаятпен ерекшеленетін және толыққанды дамыған мемлекеттік тілі бар елдер (Жапония, Ұлыбритания, Ресей, Финляндия, т.б.). Бұл елдермен салыстырғанда, Қазақстан көптілді жағдаятымен және өз мемлекеттік тілін дамытуға әлі көп уақыт болмағанымен ерекшеленеді. Сонымен қатар экономикамыздың қарқынды даму жетістіктерін ерекше атай отырып, дегенмен де елімізде лингвистикалық инновациялармен байланысты, соның ішінде қазақ тілін бағдарламалауға (корпустық лингвистердің жетіспеуі, қазақ тілін жетік меңгерген компьютерлік бағдарламалар жасаушы IT-мамандардың болмауы) қатысты ақпараттық технологиялардың (ұлттық лингвистикалық корпустың инфрақұрылымы) даму деңгейі төмен екенін мойындау қажет. Осыған байланысты қазақ тілінің корпусын жасауда тілтанушылар таза ғылыми сипаттағы қажеттіліктерді, яғни зерттеушілердің корпуста мұқтаждығын ғана емес, сонымен қатар қарапайым тұтынушының, корпусты қолданушылардың - Қазақстан Республикасы азаматтарының қажеттіліктерін де жіті ескеру қажет. Сондықтан қазақ тілінің

корпусын жасаудың басымдықтарын дұрыс айқындау немқұрайлы қарайтын жайт емес. Қазіргі таңда қазақ тілінің қолданылу аясын кеңейту және қызмет ету сапасын арттыруға қатысты туындап отырған мәселелерге байланысты тілдік жағдайды талдай отырып, қазақ тілінің ұлттық корпусын жасаудағы бірінші басымдық - мемлекеттік тілдің ұлттық-мемлекеттік бірегейлігінің көрсеткіші ретіндегі рөлін орнықтыру, яғни көпмәдениетті қазақстандық қоғамды біріктіруші түрткі жайт ретіндегі қазақ тілінің базасын нығайту деп санаймыз.

Бұл басымдықты қалай жүзеге асырдық? Әзірше екі жолмен. Біріншіден, алғашқы кезеңде қамтылатын мәтіндік материалды саралауда көпшілікке танымал мәтіндерге артықшылық берілді. Корпустың стильдер мен жанрларға және мәтін авторы тіліндегі әлеуметтік көрсеткіштерге тәуелсіз тілдің барлық қолданысын түгелдей алаламай қамтуы принципті мәселе, әрі солай болуы тиіс. Алайда бұл жағдайды жалпы қазақ тілдік корпусының негізгі мақсаты етіп ала отырып, алғашқы кезеңде-ақ бірден қолданушыға интернет желісінде қолжетімді болатын корпус мәтіндерінің базасын жасаудың мазмұнына мәдени құнды, оның үстіне отансүгіштік пен елжандылық қасиеттердегі тәрбиелеу әлеуеті күшті мәтіндерге ерекше артықшылық беріліп, оларды көптеп жинау ісі жүзеге асырылды. Д.В.Сичинава, орыс тілінің корпусын жасауға белсенді қатысқан лингвист ғалым ізімен біз де мәдени сипаттағы мәтіндерге «жазушы беделі арқылы артық болатын тілдік мысалды» [5] жатқызамыз. Ол туралы ғалымның пікірі: «Лингвистиканың эстетикалық-мәдени тәсілдемеден тәуелсіздігі қаншама рет жарияланса да, жазба тіл - ол мәдениет тілі. Лингвистикалық жұмыстан, көбінесе, тілдік мысал емес, жазушы беделімен сәулеленген жақсы тілдік мысал талап етіледі»; талаптың пайда болуы барлық лингвистикалық дәстүрлерге тән. Сондықтан, бір тілде мәдени сипаттағы мәтіндерді жинақтау сол тілдің мазмұндама берілген корпусын жасауда өзіндік құндылығы бар» [6], - деп беріледі. Біз бұл пікірмен толықтай келісетінімізді ҚТҰК бірінші кезеңдегі электронды базасының көп бөлігін көркем әдебиет жанры мәтіндері құрағаны көрсетеді.

Мәдени репрезентативті корпус жасау тәжірибесі еуропалық елдерде бар. Мәселен, Венадағы Австриялық ұлттық корпус құрамында ұлттық мәдениетке ықпалын тигізген немістілді мәтіндер жинақталғаны белгілі болса, ҚТҰК-да бұған қоса әлемдік корпус жасау тәжірибесінде тұңғыш рет мәдени-семантикалық белгіленім [7] әзірленіп, іске қосылып отыр. Бұл – сөз болып отырған басымдықты жүзеге асырудың екінші жолы.

Мәдени-семантикалық белгіленім тілтанымдық белгіленімнен тілдік емес мәліметтерді беру арқылы айырмаланады. Бірақ корпус дәстүрінде тілтанымдық емес мәлімет беретін белгіленім аясына да кіре алмайды, корпуста берілу әдісі мен мазмұны бойынша метабелгіленімнен қатты ерекшеленеді. Біздің қараған мәліметтерге сүйенсек, бүгінге дейінгі жасалған тілдік корпустарда мәдени-семантикалық ақпараттарды ұсыну әрекеті болмағаны байқалады. Аталмыш семантиканы беруді көздеген корпустар болған күннің өзінде оның жүйелі түріде аяғына дейін өңделмегені байқалады.

Соған қарамастан, мұндай белгіленім, ҚТҰК-ның біздің анықтауымызша

бірінші болып отырған мемлекеттік, жалпыұлттық бірегейлік қалыптастыруға қызмет ету басымдығын жүзеге асыру құралы болмақ. Бүгінгі күні қазақтілді кеңістік пен қазақстандық қоғамдастық мемлекеттік тілде сөйлеушілер санының артуымен, сөзжасаушылық, терминжасаушылық серпіліс сияқты оң үрдістермен, қоғамдық өмірдің барлық аясы мен арнасында қазақ тілі қолданысының артуымен, сапасының жақсаруымен, тілдік инновациялардың қарқынды түрде көбеюімен ерекшеленіп отырса да, қазақ сөзінің терең семантикаларында тұнып қалған халық санасындағы «мәдени кодтардың өшуі», «рухани сананың көмескіленуі» сияқты қауіпті жағымсыз құбылыстардың күш ала бастауы байқалады. Сол себепті осы мәдени кодтарды ашудың, этнотаңбаларда жасырылған (лингвомәдениеттанымдық, бейбаламалық, лингвокогнитологиялық және т.б.) тілдік бірліктердегі семантиканы жұрт санасына сіңірудің таза ғылыми маңызы ғана жоғары емес, сонымен қатар қазақ этнотілдік қоғамы үшін оның бірегейлігін нығайту, рухани қауіпсіздігін сақтау және ұлттық мәдениеттің батыстандыруға, рухсыздануына, рухани құндылықтардың құнсыздандыруға қарсы иммунитетін нығайтып, бекіту мақсатында тәжірибелік маңызы да өте жоғары болмақ. Себебі мәдени-семантикалық аннотациясы жасалған қазақ мәтіндерінің ауқымды көлемін онлайн-режимде пайдаланушыға бір сәтте ғана қолдануға бере алатын корпус осы тілдік бірліктердің мазмұнында жасырылған құнды лингвомәдениеттанымдық білімдердің әлеуетті және тиімді тасымалдаушысы, насихаттаушысы, ұлттық санадағы жаңғыртушысы бола алар еді. Лингвистикалық аспектіде лингвомәдениеттанымдық анықтаманың берілуі профессор Н.Уәлидің идеялық және ғылыми жетекшілігімен шығарылған «Қазақ сөздігі» атты Үлкен біртөмді түсіндірме сөздігі мен «Қазақ әдеби тілінің сөздігі» атты 15 томдық түсіндірме сөздіктерінің технологияларында өңделіп, жүйелі түрге келтірілді, осылардың негізінде тәжірибе ретінде ҚТҰК-да пайдаланылатын «Лингвомәдениеттанымдық сөздік» ұсынылады.

Екінші басымдық параллель корпусарды, яғни қазақ-орыс, қазақ-ағылшын, сонымен қатар, қазақ-түрік, қазақ-өзбек, қазақ-қырғыз және республика аумағында өмір сүріп жатқан басқа да этностық топтар тілдеріндегі ішкорпусарды жасау. Бұл басымды жүзеге асыру үшін алғашқы кезеңде қазақ-орыс параллель ішкорпусының бастапқы нұсқасы әзірленді. Осы ұжымдық еңбекте екітілді корпус жасау тәжірибесі бойынша жинақталған ғылыми білімдер ұсынылды. Болашақта параллель ішкорпусты әлемдік, халықаралық, аймақаралық, туыстық және елдегі этностық топтар тілдерінің мәтіндерімен толықтырып, сәйкестендеру қазақ тілінің этносаралық қарым-қатынас тілі мәртебесін нығайтуға, ғылыми-лингвистикалық негіздемесін жасауға мүмкіндік тудырыра отырып, жалпымемлекеттік бірегейлікті нығайтуға да септігін тигізеді.

Дереккөздер мен әдебиеттер тізімі

1. А.Қ. Жұбанов. Корпустық лингвистика – қазақ тілбілімінің жаңа бағыты //

Тілтаным. Языкознание. ҚР БЖҒМ А.Байтұрсынұлы атындағы Тіл білімі институты, сәуір-мамыр-маусым. - Алматы.2009 жыл, №2 (34). - 3-11-б. және т.б.

2. Э.Д.Сүлейменова. Қазақ тілі үшін ұлттық корпус керек пе? А.Байтұрсынұлы атындағы Тіл білімі институтының «Тілтаным» журналы. 2015. №2. – 42-45-б. және т.б.

3. Жаңабекова А.Ә. Қазақ әдеби тілінің аннотацияланған мәтіндер корпусына морфологиялық білімдер жүйесін енгізудің практикалық тәжірибелері және оларды оқыту жүйесіне енгізудің ғылыми-әдістемелік негіздері // «Қазақ тіл білімі: жаңа ғылыми парадигмалар мен оқытудың инновациялық технологиялары» атты халықаралық ғылыми-теориялық конференция материалдары. – Алматы, 1912. – 197-201-бб. және т.б.

4. А.Фазылжанова. Национальный корпус казахского языка как инновационно-информационная база государственного языка (об особенностях и задачах первого этапа разработки и создания) // ALATOO ACADEMIC STUDIES. 2016. № 3. – 103-109 с.

5. Сичинава Д.В. (2002). «К задаче создания корпусов русского языка», НТИ. Серия 2 Информационные процессы и системы. №11. 26 с.

6. Сичинава Д.В. (2002). «К задаче создания корпусов русского языка», НТИ. Серия 2 Информационные процессы и системы. №11. 27 с.