

А.С. Барменқұлова^{1*} , А.Ә. Жаңабекова² , Ж.А. Манкеева² 

¹Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан;

²А. Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан

*e-mail: bs.aida@mail.ru

ЭТНОМӘДЕНИ БІРЛІКТЕРДІҢ КОРПУСТА БЕРІЛУ ЖОЛДАРЫ

Аннотация. Мақалада Қазақ тіліндегі этномәдени бірліктер жүйесіндегі заттық мәдениет атауларына жасалған сипаттамаларды қазақ тілінің ұлттық корпусына енгізу және корпуста белгілеу, берілу жолдары айтылады.

Этномәдени бірліктер қазақ тілінің ұлттық болмысын танытатын, ұлт тілін, дәстүрін, дүниетанымын, рухани әлемін сақтаушы кумулятивтік қызметі зор тілдік таңбалар болып табылады. Осы ретте бүгінгі ұрпаққа тілде бейнеленген этномәдени бірліктер арқылы дүниенің қазақи бейнесін жеткізу – қазіргі жаһандану заманында жас ұрпақтың білім-білігін арттыру үшін оңтайлы да тиімді құрал ретінде өзекті мәселе. Бүгінгі жас ұрпақтың кітаптан емес, ақпараттық құрылғылар арқылы оқуды, білім аруды дағдыға айналдырып бара жатқаны белгілі. Осымен байланысты тіл арқылы бүгінгі ұрпақты ұлттық құндылықтарымызбен сусындатып, ұлттық білім қорын келешек ұрпаққа жеткізу – аса маңызды. Заттық мәдениеттің ұлттық сипаттық жүйесі халық тұрмысының ең негізгі көрсеткіші тіл арқылы сақталған тіл иесі санасында жалғасып, бүгінгі өміршеңдігімен құнды. Сонымен қатар ол басқа халықтардың мәдениетімен ұштаса отырып дамиды. Бірақ кез келген ұлттың дүниетанымы, ұлттық болмысы, ұрпақтан ұрпаққа беріліп келе жатқан ғасырлар бойы жасаған ұлттық мәдениеті оның тілінде көрініс табады. Сондықтан ұлттық мәдениеттің тірегі де тілде.

Қазіргі таңда антропозектік бағытта тілге зерттеу жүргізу қарқынды жүруде. Осымен байланысты аталған бағытта үдерісті тереңдетіп, кеңейтіп Ұлттық корпустар құрастырудың өзектілігіне айқындалуда. Қазақ тілінің ұлттық корпусы осындай ұлттық мәдени бірліктерді кең арнада сақтаушы, ұрпаққа танытушы қызметін атқаратын заманауи инновациялық-ақпараттық құрал болып табылады.

Мақаламызда Қазақ тілінің ұлттық корпусына этномәдени бірліктерді енгізудің жолдары сипатталып түсіндіріледі: этномәдени бірліктерді Ұлттық корпуста енгізуде *кодтау, модельдеу, алгоритмдік-программалық* тәсілдер пайдалану: ең алдымен корпус базасын құруға негіз болатын этномәдени бірліктер жинау, олардың мәдени-семантикалық топтарын анықтап, сол топқа жататын сөздердің тізімін жасау. Олардың әрқайсысына белгі-код қойылғаннан кейін ғана программист мамандарға беру. Программист мамандар этномәдени бірліктерді белгілі бір алгоритмге салып, әрі қарай корпуста ашық пайдалануға қолжетімділікті туғызу үшін бағдарламалық өңдеу жұмыстарын қарастыру үстінде.

Ғылыми-зерттеу жұмысымызда Қазақ тілінің ұлттық корпус базасында айқындалған белгіленімдерді (морфологиялық, сөзжасамдық, семантикалық, лексикалық, фонетика-фонологиялық белгіленімдері) қазақ тілінің корпусына мәдени-семантикалық белгіленімдер ретінде енгізу мәселелерін қарастыруға арналған.

Тірек сөздер: корпус, этномәдени бірліктер, белгіленім, лексикографиялау, белгі-код, тег.

А.С. Барменқұлова^{1*}, А.А. Жаңабекова², Ж.А. Манкеева²

¹Казахский национальный университет имени аль-Фараби, Алматы, Казахстан;

²Институт языкознания имени А.Байтұрсынұлы, Алматы, Казахстан

*e-mail: bs.aida@mail.ru

ПУТИ ПЕРЕДАЧИ ЭТНОКУЛЬТУРНЫХ ЕДИНИЦ В КОРПУСЕ

Аннотация. В статье обосновывается необходимость включения в национальный корпус казахского языка наименований материальной культуры в системе этнокультурных единиц казахского языка посредством их описания, а также необходимость их разметки в корпусе.

Этнокультурные единицы являются языковыми символами с большой кумулятивной функцией, отражающими национальную идентичность казахского языка, сохраняющими язык, традиции, мировоззрение, духовный мир нации. Известно, что сегодняшнее молодое поколение приобретает навыки чтения, обучения не из книг, а с помощью информационных устройств. В связи с этим очень важно через язык наполнить сегодняшнее поколение нашими национальными ценностями и донести национальную базу знаний до будущих поколений. Национальная характерная система материальной культуры, являющаяся важнейшим показателем жизни народа,

продолжается в сознании носителя языка и ценна своей жизнеспособностью и сегодня. Она также развивается в сочетании с культурой других народов. Но мировоззрение, национальная идентичность любой нации, национальная культура, которую она создавала на протяжении веков, передавалась из поколения в поколение, отражаются в ее языке. Поэтому опора национальной культуры есть и в языке.

В настоящее время ведется интенсивное изучение языка в антропоцентричном направлении. В связи с этим углубление и расширение процесса в данном направлении определяется актуальностью сборки национальных корпусов. Национальный корпус казахского языка представляет собой современный инновационный информационный инструмент, сохраняющий такие национально-культурные единицы в широком русле и представляющий их следующему поколению.

В нашей статье описаны способы внедрения этнокультурных единиц в Национальный корпус казахского языка: кодирование, моделирование, использование алгоритмико-программных подходов при внедрении этнокультурных единиц в Национальный корпус: прежде всего, сбор этнокультурных единиц, которые являются основой для создания базы корпуса, определение их культурно-семантических групп и составление списка слов, относящихся к этой группе. Передача данных профессионалам-программистам происходит только после проставления разметки-кода. Специалисты-программисты работают над разработкой программного обеспечения для построения этнокультурных единиц на определенном алгоритме и создания в дальнейшем доступа к открытому использованию в корпусе.

Научно-исследовательская работа посвящена рассмотрению вопросов включения в корпус казахского языка разметок, определенных на базе Национального корпуса казахского языка (морфологические, словообразовательные, семантические, лексические, фонетико-фонологические разметки) в качестве культурно-семантических разметок.

Ключевые слова: корпус, этнокультурные единицы, разметка, лексикографирование, разметка-код, тег.

A.S. Barmenkulova^{1*}, A.A. Zhanabekova², Z.A. Mankeeva²

¹Kazakh National University Al Farabi, Almaty, Kazakhstan;

²A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan

*e-mail: bs.aida@mail.ru

WAYS OF TRANSMISSION OF ETHNOCULTURAL UNITS IN THE CORPUS

Abstract. The paper substantiates the need to include in the national corpus of the Kazakh language the names of material culture in the system of ethnocultural units of the Kazakh language through their description, as well as the need for their marking in the corpus.

Ethnocultural units are linguistic symbols with a large cumulative function, reflecting the national identity of the Kazakh language, preserving the language, traditions, worldview, and spiritual world of the nation. Therefore, the language is supported by national culture.

Currently, intensive language study is being conducted in an anthropocentric direction. In this regard, the deepening and expansion of the process in this direction are determined by the relevance of the assembly of national buildings. The national corpus of the Kazakh language is a modern innovative and informational tool that serves as a guardian of such national cultural units in a broad way, a carrier of them for generations.

The research paper describes the ways of introducing ethnocultural units into the National Corpus of the Kazakh language: coding, modeling, and using algorithmic and programmatic approaches when introducing ethnocultural units into the National Corpus. First of all, collecting ethnocultural units that are the basis for creating the corpus base, determining their cultural and semantic groups and compiling a list of words related to this the group. Transfer each of them to professional programmers only after applying the markup code. Programmers must work on a certain algorithm and carry out further processing. Works to be done in the future are considered as well.

The research work is dedicated to the consideration of the issues of inclusion in the corpus of the Kazakh language of markings defined based on the National corpus of the Kazakh language (morphological, word-formation, semantic, lexical, phonetic, and phonological markings) as cultural and semantic markings.

Keywords: corpus, ethnocultural units, markup, lexicography, sign-code, tag.

Кіріспе

Қазіргі кезде жаһандануға байланысты жаңа ақпараттарды игеру мақсатында орыс тіл білімінде немесе шетел тіл білімінде үлкен жетістіктерге жетіп отырған тілдік корпусар бар. Шетелдік ғалым Э. Финеган корпусық лингвистикаға табиғи тілдің қолданысы зерттеуге бағытталған корпусы құрастыру мен қолдануды қажет ететін қызмет түрі деген анықтама

береді (Finegan, 2004). Заманауи Қазақстанның да рухани-танымдық жаңа деңгейі Қазақ тілінің ұлттық корпус жасау қажеттілігін көрсетіп отыр. Өйткені, Ұлттық корпусарды құру мәселесі шетел тіл білімінде ХХ ғ. 60-жылдары басталса, орыс тіл білімінде ХХ ғ. 80-жылдары, ал қазақ тіл білімінде осы соңғы жылдарда ғана қолға алынды. Қазақ тілінің корпусық лингвистика мәселесін профессор А.Қ. Жұбанов көтерді. Ғалымның пікірінше, корпусық лингвистика қазақ тіл білімінің де ерекше саласы ретінде қалыптасатын болса, қазақ тілшілеріне көлемді тәжірбиелік материалдарды пайдалануға, қажетті деген тілдік деректерді тез тауып алуға және оларға тиісті деген өңдеулер жүргізуге мүмкіндік тудырады. Осының бәрі қазақ тіліне қатысты зерттеулердің шынайылыққа (ақиқаттыққа) жетудің эмперикалық тәсілдеріне жаңаша көзқараспен қарауға және ғылыми айналым аясына аса маңызды тілдік материалдарды енгізуге жағдай жасайды (Жұбанов, 2017: 4).

Тіл корпусы және онымен байланысты ғылым, корпусық лингвистика тақырыбы ХХ ғасырдың аяғы мен ХХІ ғасырдың басында лингвистердің өміріне өте тез енген сала. Егер біз лингвистиканың анықтамасы бойынша өте заманауи саланы атағымыз келсе, онда бірінші кезекте ойымызға корпусық лингвистика келеді. Бірақ, аталған сала бойынша лингвистерге толық түсінік әлі келген жоқ, дей тұрғанымен корпусық тіл білімі мен ғылыми зерттеулердегі рөлін мойындау қазір өте қарқынды өсуде. Корпус деген не? Тіл корпусы – берілген тілдегі мәтіндердің электрондық жинағы. Егер сіз мәтіндерді әртүрлі бөлшектерден жиналған күрделі ойыншық деп елестетсеңіз, онда бала ойыншықпен ойнай алады, көңіл көтере алады, ал лингвист мәтінді бөлшектейтін жаман бала сияқты әрекет етеді, оған басқа ештеңе қажет емес. Себебі мәтіннен қарапайым адамға мағына қажет, оған не айтылғанын түсіну керек, ал лингвист оны бірден бөлшектей бастайды. Осы мүмкіндікті лингвистерге тіл корпусары береді.

Лингвистер мәтіндерді арнайы бағдарламалардың көмегімен өңдейді. Бұл, әрине, автоматты түрде жасалады. Корпус неғұрлым үлкен болса, соғұрлым жақсы. 100 миллион сөзқолданыстан тұратын корпус болды, бұл ұзақ уақыт бойы көлемі мен саны жағынан үлкен корпус болып саналды. Ал қазір миллиардтаған базадан тұратын корпус аз емес. Көлемі тіпті көптеген миллиардтардан асады – оған интернет үлкен мүмкіндік береді. Мәтінді қолмен белгілемейтініңіз анық. Мәтінді өңдеу бағдарламалары бар. Бұл салада қолмен араласуды қажет етпейтін бағдарламалар бағаланады. Осылайша, өңделген мәтіндер, аннотацияланған мәтіндер корпусы құрайды. Бұл бізге зерттеу жұмысын жүргізуде мысалдарды тез әрі жан-жақты іздеу үшін қажет. Мен, мысалы, септіктерді зерттеумен айналысамын делік. Дәстүр бойынша, лингвистер кітапты ашып, септіктердің түрлі формаларын жинайды, содан кейін карталар пайда болды – бұл көп уақытты алатын жұмыс, бірақ көп тілдік бірлік пен материал жинау ғылыми жұмыстың маңызды бөлігі деп саналады. Ғылыми тұрғыдан алғанда ақсайтын, бірақ жинаған материалдары құнды жұмыстар бар. Материалдарды да жинауға айлар емес, жылдар да кететін. Қазір Корпус материалды қолдап жинау жұмысын өте қысқартып, уақытты үнемдеуге септігін тигізуде. Бұрын материалды жинауға жылдар кетсе, қазір материалды бірнеше минут ішінде жинауға болады.

Тілдік корпусарды құрастырудың қиындығын айқындап, саралап берген профессор А.Қ. Жұбанов қазақ тілінің ұлттық корпусын құрастыруды орталықтандыру, мемлекеттік дәрежеде барлығы бірдей жұмыла кірісу керектігі туралы 2014 жылғы «Ана тілі» газетінің бетінде осы мәселеге қатысты бағдарламалық мақала жариялады. Ғалым ол туралы былай дейді: «Ұлттық тіл мәтіндерінің компьютерлік корпусын құру жобасы бір ғана ғылыми ұйымның шешетін мәселесі емес және ол зерттеу жұмысы 3-5 жылда аяқтала қояды деуге де болмайды. Себебі, бұл аса күрделі және оның нәтижелері әлемдік дәрежедегі аса маңызды ғылыми жұмыс болып саналады. Зерттеу жұмысының мақсатына сай орындалатын міндеттері мен сала-салаға кезең-кезеңге бөлініп, тек қана ғалымдар қауымдастығын құру арқылы ғана ауқымды нәтижеге жетуге болары сөзсіз. Қазақ тілінің ұлттық корпусарын жасау – көп болып жұмылып атқаратын ұлттық құндылығымыз. Сондықтан Қазақстанның әр жерінде бірінен-бірі дербес атқарылып жатқан корпус жасау ісін орталықтандыру керек. Оған Қазақстанның әр

жерінде ғылыми-педагогикалық қызмет атқарып жүрген ғалымдардан арнайы лингвистикалық топ құру керек. Өйткені, тілдік бірліктерді модельдеу – өте күрделі мәселе Сондай-ақ тілдік корпустар құрастырудың өзі ең алдымен лингвистикалық белгіленім талдамасын жасауға тіреледі. Осы кезге дейін жасалып жатқан корпус жасау тәжірибесіндегі нәтижелерді бір орталыққа жинақтау керек. Кітап, газет-журнал шығаратын баспалармен шартқа отырып, мәтіндердің электронды нұсқаларын алу қажет» – дейді (Жұбанов, 2014: 9).

Қазіргі уақытта мәтіндерде лингвистикалық және басқа ақпарат түрлерін ұсынудың жалпыға бірдей танылған стандарттары жоқ. Text Encoding Initiative (Tei) арнайы халықаралық жобасы стандартталған белгілеу құралдарын әзірлеуге арналған. Ол үшін SGML құжаттарын белгілеудің жалпыға бірдей танылған халықаралық тілі немесе оның XML ішкі жиыны қолданылады. Корпуста болуы мүмкін белгілеу түрлерін шартты түрде лингвистикалық және сыртқы лингвистикалық деп бөлуге болады. Соңғыларына мыналар жатады:

– мәтінді пішімдеу ерекшеліктерін көрсететін белгіленім (тақырыптар, абзацтар, шегіністер және т. б.);

– автор мен мәтін туралы мәліметтерге қатысты белгіленім. Сонымен қатар, автор туралы ақпарат оның атын ғана емес, сонымен қатар жасын, жынысын, өмір сүрген жылдарын және тағы басқаларын қамтуы мүмкін, ал мәтін туралы ақпарат, әдетте, тақырыптан басқа, ол жазылған тілді, басылымның жылы мен орнын және т. б.

Мұндай ақпараттың болуы мәтіндік мәліметтер базасында іздеуді едәуір егжей-тегжейлі көрсетуге мүмкіндік береді, сонымен қатар тиісті құжатты сәйкестендіру құралдарын ұсынады.

Қазіргі кезде қазақ тіл білімінде тілдік корпустарды құрастыруға қатысты қазақ тілінің тілдік деңгейінің зерттеу нысанына сәйкес лингвистикалық белгіленімдер, морфологиялық деңгейде белгіленім қою – аннотациялау жұмыстары жүргізілуде. Орыс тіл білімінде лингвистикалық ақпаратты «разметка» деп атайды. Ал, бұл терминді профессор А.Қ. Жұбанов қазақ тілінде «белгіленім» немесе «белгі-код» терминдерімен атауды ұсынды. Тілдік корпустарды жасауда осы терминдер белсенді қолданылып жүр. Корпуста этномәдени бірліктерге қатысты белгіленімдердің сипатына қарай оларға тегтар пайдаланылады. Мұндағы белгі-код «тег» сөзі (ағыл. tag – ярлық. Анотациялау дегеніміз – қазақ әдеби тілінен сұрыпталып алынған таңдама мәтіндердегі сөзформаларға лингвистикалық мәліметтер беріп, ақпараттандыру. Лингвистикалық белгі-код қою арқылы тілдік мәліметтерді бейнелесе, белгілер семантикалық (лексика-семантикалық), морфологиялық, синтаксистік сипатта болады. Ұлттық корпуста сөз табына, септелуіне, тегіне, жіктелуіне, шырай, жанды-жансыздығына қарай оларға морфологиялық, синтаксистік белгіленімдер қойылады. Ал, семантикалық талдаулар арқылы мәтін ішінде қажетті белгіленімдер қойылады. Корпус жасағанда белгіленімдердің барлығын бір уақытта енгізу қиындық туғызды. Корпустарда лингвистикалық белгіленімдер кезең-кезеңімен асырылып отырады. Осы аталған белгіленімдерді «терең анотацияланған» (глобоко аннотированный) деп аталады. Лингвистикалық белгіленімдерді автоматтандыратын арнайы бағдарламалар жасалады. Корпустарға белгіленімді қою ең күрделі жұмыс деуге болады. Корпустардың көлемінің үлкен болуына байланысты жұмысты күрделендіре түседі.

Сондықтан да, корпуста этномәдени белгіленімдер қою шартты түрде арнайы тәгтерді «тіркеп» жазу арқылы белгіленетін болады.

Бүгінгі ақпараттық технология заманында кез келген ғылым саласын цифрландыру мәселесі гуманитарлық ғылым салаларына да қажеттілік туғызды. Өйткені қазір барлық қоғамдық қызмет түрлері мен оқу-білім алу жүйесі компьютерлік техника арқылы орындалуда. Ғылым дамыған сайын кез келген қызмет түрін жеңілдету, оңтайландыру қажеттік тудырып отыр. Қазақ тілінің ұлттық корпусын жасау да лингвистикалық зерттеулер жүргізудің мүмкіндігін арттыру, мемлекеттік тілді корпустық ресурстар арқылы жедел меңгерту мақсатын көздейді. Қазақ тілінің ұлттық корпусы негізінен, этнос тіліндегі

материалдық және танымдық мәдени ақпараттарды жеткізуші де қызмет атқаратын Ұлттық корпусқа этномәдени бірліктерді енгізудің маңызы зор.

Қазақ тілі түсіндірме сөздіктерінде этномәдени бірліктер беруде бірізділік сақтала бермейді, соның негізінде сөздікті пайдаланушыға ұлттық құндылықтарды жеткізу жағы әлі де тереңдетуді, жетілдіруді қажет етеді. Осы орайды қазақ тіліндегі этномәдени бірліктерді лексикографиялаудың әдіс-тәсілдерін айқындаудың мәні зор, ол мәселе этномәдени бірліктерді Қазақ тілінің ұлттық корпусына этномәдени белгіленімдер енгізу әзірлемесін жасау міндетін алға шығарады. Осымен байланысты әлемдік Ұлттық корпусардағы этномәдени белгіленімдерді енгізудің әдіс-тәсілдерін зерттеп, саралау қажеттілігі туындайды. Соның барысында қазақ тіліндегі мағыналық құрылымында этномәдени семантикасы бар тілдік бірліктерді жинақтау арқылы этномәдени бірліктерді түсіндірме сөздікте лексикографиялаудың жолдары айқындалады.

Қазақ тілінің ұлттық корпусында этномәдени бірліктерді кодтаудың технологиясын жасау және Қазақ тілінің ұлттық корпусына этнобелгіленімдерді енгізу жұмыстарын да қарастырудың маңызы зор.

Материал және әдістер

Ғылыми-зерттеу жұмысымызда тіл білімінде лексикалық бірліктерді зерттеуде қалыптасқан әдіс-тәсілдерді, оның ішінде қазақ тіліндегі этномәдени бірліктерді жинауда жинақтау, топтастыру (классификациялау) тәсілдері қолданылады. Этномәдени бірліктердің түсіндірме сөздіктердегі лексикографиялану жолдарын айқындауда сипаттамалы, синонимдік, суреттемелі, иллюстрациялық т.б. дефиницияны беру тәсілдері негізге алынады. Этномәдени бірліктерді Ұлттық корпусқа енгізуде кодтау, модельдеу, алгоритмдік-программалық әдістері пайдаланылады.

Әдебиетке шолу

Тіл тек қана коммуникативтік қызмет емес, өз бойында этномәдени ақпараттарды жинақтап, сақтаушы, оларды ұрпақтан-ұрпаққа жеткізуші және ұлттарды бір бүтін етіп тұтастырушы кумулятивтік те қызмет атқарады. Ғаламның тілдік бейнесі тіл мен мәдениеттің бір-бірімен сабақтастығында қалыптасып, тіл арқылы тіл иесінің ұлттық болмысының сипаты, ұлттық рухы танылады. Ғасырлар бойы мазмұнына сіңірілген тілдің осы құдіретті қызметі мен қасиетін қағида еткен тіл білімінің когнитивті лингвистика, лингвомәдениеттану, этнолингвистика т.б салалары кең қанат жайып дамып келе жатыр.

Ш. Уәлиханов, Қ. Жұбанов, Ә. Марғұлан, М. Әуезов өз еңбектерінде тіл мен мәдениеттің бір-бірімен тығыз байланыстылығына мән бергені белгілі. Мысалы, Қ. Жұбанов тіл арқылы халықтың тағылымдық мұраты мәдени әрі тілдік мұра ретінде сақталатынын және этностың әр дәуірдегі ұрпақ өкілдеріне тіл мен мәдениетті табыстыратын жаңа саланың дүниеге келетіні туралы сол кезеңдердің өзінде-ақ болжам айтқан. Ғалымның «халықтардың тұрмыс қалпы түрлі-түрлі болған соң, олардың әр затқа қоятын аттары да түрлі-түрлі. Тіл ұзақ заман жасалады, бірден өзгермейді, оның өзгерісі тұрмысқа байланысты» – деген пікірі ұлт мәдениетінің ұрпақ жалғастығын сипаттаса, ал, мәдениет жай өмірдің туындысы ғана емес, белгілі бір қоғамда өмір сүрген рухани әрі материалдық қоры бар іргелі елдің өмір тіршілігінің қажетті туындысы деп қарауын: «Мәдениет – қазанның күйесінше жұғатын нәрсе емес, мәдениеттің белгілі бір сатысы көксегендей жағдай тумаған болса, жоғары мәдениеттің өзі келіп ауызға түспейді» деген пікілерінен көруімізге болады (Жұбанов, 1999: 7). Ертедегі мәдениеттің көріністері, олардың салт-дәстүрілері тіл арқылы сақталып, бізге жеткендігін көреміз.

Қазіргі тіл білімінде тілді мәдени-танымдық тұрғыдан арнайы зерттеу лингвомәдениеттану саласының еншісінде. Тіл мен мәдениеттің байланысын лингвомәдениеттану саласы құрайды. Лингвомәдениеттанушы ғалым В.И. Телия лингвомәдениеттану саласының басты мақсатын былай түсіндіреді: «Тіл мен мәдениеттің өзара

қатынасын қазіргі заман мәдени ұлттық сана-сезім тұрғысынан қарастырып, оның тілдегі көрінісін зертеу мен сипаттау» (Телия, 1987: 28). Демек, Лингвомәдениеттану мәдениеттің тілдегі синхрондық жағын сипаттап, түрлі тарихи кезеңдердегі тіл тұтынушының айналадағы тылсым дүниенің сырларын, мифтік пайымдарын, наным-сенімдерін бейнелеген тілдік деректері, көне этномәдени сөз қолданыстары арқылы халықтың бай ауыз әдебиеті мен көркем әдеби мәтініндегі ұлттық-мәдени компоненттері мен ескіліктері келіп жетті. Соның негізінде тілде сақталған халқымыздың этномәдени даму жолдарын көрсететін материалдық және рухани құндылықтарымыз танылады. Осы арада материалдық мәдениет атаулары да рухани мәдениетпен қатар өз бойында ұжымдық стереотипті сіңіріп, қалыптасады, бір этностың ұжымда тұрмыс кешіп жатқан адамдардың күнделікті тұрмыс барысындағы тіршілік мұқтаждықтарынан туындап қалыптасқан. Материалдық мәдениет пен рухани мәдениеттің бір-бірімен тығыз байланыстылығы екеуін бір бүтіннің екі жағы деп қарастырыруға негіз болады. Өткені, адамды қоршаған табиғат пен өмір сүруші субъектінің қатынасы барысында айналадағы дүниені танып, өз ақыл-ойының елегінен өткізіп, табиғаттың затын өз затына өңдеп айналдыру – зор еңбектің көрінісі. Бұл – түптеп келгенде, заттық мәдениеттің дүниеге келуінің негізгі шарты. Ол өз кезеңінде сол қоғамның эстетикалық символдық тәжірибиесін көрсететін заттық атаулардың рухани өмірмен байланысын білдіретін «материалдық мәдениет өнімдері рухани қажеттіліктерден туындайды» деген пікірді нақтылай түседі. Ол туралы ресейлік этнолингвист ғалым, академик Н.И. Толстой: «Халықтың рухани мәдениеті оның ана тілінде, говорларда, салт-дәстүрлерде материалдық мәдениет арқылы да танылады. Оларды біртұтас деп қарау қажет» деп көрсетеді (Толстой, 1995: 21).

Демек, тірнек (заттық) мәдениеттің мәні дегеніміз – адам әрекетінің нәтижесінде заттандыру. Оның ішкі мәні соны тұтынушы қоғам өкілдерінің санасында, рухани дүниесінде.

Лингвомәдениеттанымдық зерттеулерге сүйеніп, этномәдени атауларды талдаудың негізінде мәдениет ұғымының тілге қатысты мәнін былай айтуға болады: ұлттық мәдениеттің негізі – ортақ тілде сөйлеп, ортақ бір мүддеге келу, этноұжымда қалыптасқан адамның дүниетанымының тұтастығы мен этностың материалдық және рухани кеңістігі.

Материалдық мәдениетке тілдегі заттық ұғымдар қатарына: тұрақ, киім, тамақ, тұрмыс бұйымдары, еңбек құралдары жатады. Ұлттық тіл арқылы ұрпақтан-ұрпаққа жеткізілетін рухани мәдениетке салт-дәстүр, миф, өнер, тәлім-тәрбие, көркем мәтін т.б жатады. Осыдан келіп ғылыми-зерттеу барысында мәдени лексиканы Ю.М. Лотман, Е.М. Верещагин, В.Г. Костамаров, Ә.Т. Қайдар, Е.Н. Жанпейісов, Р.С. Сыздық, Е.Қ. Жұбанов, Н.М. Уәлиұлы, Ж. Манкеева, Р. Шойбеков сияқты ғалымдарымыздың зерттеулерінен көрдік.

Тіл мен мәдениет сабақтастығын диахрондық аспектіде қарап, Этнолингвистика этнос болмысын тіл арқылы ашудың ретінде де тілде таңбаланып отырғанын дәлелдеу мақсатында академик Ә. Қайдардың жоғарыда көрсетілгендей, тілдің кумулятивтік немесе «құжаттық» (Э. Сепир) қызметімен байланыстыруын қарастырдық: «Этнос және оның тілін біртұтас, өзара тығыз байланысты құбылыс деп қарасақ, тілдің қоғамда тек сол этностың қарым-қатынас құралы ретінде ғана емес, сонымен қатар оның (этностың) бүкіл рухани, мәдени байлығының куәгері іспетті барлық болмысы мен өмір тіршілігінің, дүниетанымы мен әдет-ғұрпын бойына сіңіріп, ата мұрасы, асыл қазына ретінде ұрпақтан-ұрпаққа жеткізіп отыратын қызметі тағы бар» (Қайдаров, 1998: 34).

Этномәдени мұралардың тілдің өз бойындағы сақтаушылық қызметіне ғалым Н.М. Уәли де ерекше назар аударады: «Тіл – өмір шындығы. Сол шындықты уақыт қалдырған қатпарлардан аршып ала білсек, шежіредей шыр шертіп тұрады. Өмірде болған нәрсе тілде де із тастамай кетпейді. Сондай іздің сорабын асыл мұрамыздың тілінен кездестіреміз» (Уәли, 1998: 13).

Осындай пікірлерге ден қойсақ, мәдениеттің негізгі арқауы тілде сақтала отырып, сол арқылы ұрпақтан-ұрпаққа жетіп отыратынын тұжырымдадық.

Қазіргі қазақ тіл білімінде материалдық лексиканы арнайы зерттеу нысан болуы – профессор Ж. Манкееваның «Қазақ тіліндегі этномәдени атаулардың танымдық негіздері» монографиясында талданады (Манкеева, 2008: 5).

Зерттеуіміздің барысында аталған әдебиеттер бойынша материалдық лексиканың бұл түрі күнделікті өмірге қажетті ұғымдарды қамтитын бөлшегі тұрмыстық лексика туралы мәліметтер аталған еңбектерде барынша жинақталып берілгеніне көз жеткізіп, шолу жасадық. Тұрмыстық лексика сөздік қордағы негізгі бөлікті құрайтын сөздерден тұруы – кез-келген тілге тән құбылыс. Өйткені олар әр халықтың, әрбір ұлттың өмірі үшін, тіршілік етуіне керекті, аса қажетті, маңызды ұғымдардың жиынтығынан тұрады. Тұрмыстық лексиканың семантикалық аттаулары адамның күнделікті өміріне байланысты, адамның тұрмыс-тіршілігінің түрлі қырларын білдіретін сөздерді жинақтайды.

Нәтижелер және талқылау

Тұрмыстық лексиканың сөздік қордағы негізгі бөлікті құрайтын сөздерден тұруы – кез келген тілге тән құбылыс. Ал «белгіленім» корпуста издеген сөзіңіз туралы толық ақпараттарды көрсететін термин. Қазақ тілінің ұлттық корпус базасында бес түрлі белгіленімдер бар: морфологиялық, сөзжасамдық, семантикалық, лексикалық, фонетика-фонологиялық белгіленімдер. Біздің алға қойған мақсатымыз – қазақ тілінің ұлттық корпусына этномәдени бірліктердің белгіленімдер жүйесін енгізу. Оларды түрлеріне қарай топтастырып, өңдеп, одан әрі жетілдіру этномәдени бірліктерді корпусқа енгізіп, оларға шартты белгілер қою арқылы жүзеге асырғаннан кейін мына міндеттер орындалуы тиіс:

Этномәдени бірліктердің корпуста берілу жолдары:

1. Қазақ тілінің ұлттық корпусына этномәдени бірліктердің белгіленімдерін беру үшін Excell бағдарламасында бастапқы мәтіндер базасы жиналады;
2. Excell бағдарламасына жиналған этномәдени бірліктердің мәтіндерін негізгі базаға енгізу үшін өңдеу жұмыстары жүргізіледі;
3. Excell бағдарламасына жиналған этномәдени бірліктер негізгі базаға енгізіледі және мәдени-семантикалық корпуспен жалғанады;
4. Корпуста этномәдени бірліктерге белгіленім қойылады. Мысалы; (этно) белгіленімі;
5. Корпустардың мәдени-семантикасы бойынша іздеу (поиск) жасалады;
6. Әрбір белгіленімдер бойынша іздеу жұмысы жетілдіріледі.

Тұрмыстық лексикаға байланысты «саба» атауының дәстүрлік және оның құрылымдық жағынан мағыналарының ұлттық корпуста қалай ашылатынын көрсетуге болады:

1. Этномәдени бірліктерді корпуста жинаған кезде ең алдымен *саба* туралы анықтамалық ақпарат жиналады;
2. *Саба* – кептірілген шикі теріні әдемі құрамға салып, құм кептеп қатырып, содан кейін ысқа салып төрт бойдан құрастырылған тері ыдыс;
3. Сабаның ең басты белгісі туралы ақпараттарды жинақталады: сабасыз іркіт жиналып, қымыз құйылмайды;
4. Сабаға (этно) белгіленімі қойылады;
5. Сабаның түрлері туралы ақпарат жиналады. *Саба* – жылқы терісінен жасалып, терінің түгі жидітіліп, терісі сыпырылып, көкетін алып тазартылды. Дайын болған теріні талдың қабығымен боялады. Ол шамамен 100 литр болатындай түбі кең, аузы тарлау болатындай етіп пішіп, қосылған жерлерін тарамыс немесе шуда жіппен жөрмелеп тігеледі. Сабаның ішінде қымыз жарық түскенде түбі көрініп тұрады. Ысталған ыдыс берік болады, құрт-құмырсқа түспейді, әрі ыстың дәмі мен иісі шауып, оның иісін арттыра түседі. Сабаны іркіт жинап, қымыз пісу үшін пайдаланады. Сабаның *бұлқын*, *қарақұла* сияқты түрлері бар:

Бұлқын түрі – сабаны піскенде бүлкілдеп, бұлқынып тұратын, сабаның бүйірінен сәл жоғары бөлігінің атауы;

Қарақұла – сабаның жүні мен қылы жидітілген теріден жасалады. Жылқының жал тұсынан сабаның ортаңғы бөлігіне тігінен келтіріп тігетін болған. Сабаның қарақұласы көзге

әдемі көрінетін болған. Сабаны жерге қоймаған. Оның астына пішініне лайықтап *сәкі* (орындық) жасайды, немесе оны *саба орындық* деп атайды. Сабаның астыңғы жағына *күнтулақ* немесе *киіз* салып пайдаланады. Сабаны үйдің кіре берісіне қойып, іркіт жинап отырған. Сондықтан да, қымыз сабасы іркіт сабасынан жоғары тұрады. Сабаның аузын керегенің басына байлап қояды. Қымызды пісу үшін сабада піспек пайдаланылады. С. Мұқанов «*Тайжүзген*» деп аталатын саба туралы былай дейді: «*Біздің елде бес мың жылқысы болыпты деген Есенеі сабаны алты айғырдың терісінен тіктіріп, оның атын «тайжүзген» қойған екен... Тайжүзгеннің піспегін арқанға кереді де, екі қарулы жігіт әрең көтеріп піседі екен...*». Осындай үлкен сабаның басы керегенің басына байланып, ондағы піспектің басы жабдықтан шығарылып, үй сыртынан адамдар келіп, аттың үстінен піскен. Үлкендігі соншалық адамдар піспекті жай қолымен пісе алмаған. Сүттен жасалған тағам, сусын түрлерін әзірлейтін ыдыстың түрі болғандықтан саба үйдің мүлкі ретінде саналады. Қазақ халқының танымдық ұғымында *сабадай* сөзі *ірілікті, молдықты, жуандықты, үлкендікті* ұғымын білдіреді. Сондықтан да бес биенің *сабасындай* деген ұғымы ірілікті, семіздікті білдіретін тұрақты сөз тіркестер де кездесіп отырады. Келісті әйелдерді *бес биенің сабасындай* деп мадақтайды (Қазақтың этнографиялық категориялар, ұғымдар мен атаулардың дәстүрлі жүйесі энциклопедия, 2011: 115).

Міне, осындай келтірілген мысалдардан сабаның жылқы терісінен, өгіз терісінен жасалатындығын көреміз. Қазақ халқы сабаны ырымдап үйдің жоғарғы жағына сәндік үшін де іліп қойған. Ғажайып саба жасайтын шеберлер сабаны сәндік бұйым ретінде жасаған.

Этномәдени бірліктерді ұлттық корпусқа енгізуге қатысты маңызды жұмыстың бірі – корпустарда этномәдени бірліктерді бере отырып, олардың дефинициясын анықтап, мәдени-семантикасын ашу. Мысалы, *қобди* сөзі – ұсақ-түйек салатын кішкене сандықшаның атауы. Оның ұстайтын тұтқасы болады. Қобдиға әшекей бұйымдар мен ине-жіп, т.б керекті заттар салуға таптырмас құрал. Оның сыртқы жағы қаңылтырмен қапталады. Ұзындығы 28,4 см, ені 18,8 см, биіктігі 9 см. Қақпағының ортаңғы бөлігінде темірден жасалған үлкен тұтқа орналасады. Алдыңғы бетінде құлып салуға арналған бастырма болады. Кейбір әйелдер қобдиды *бөгіре* деп атайды. Осындай қобдиға шекіме әдісімен өрнектеу, заттық бұйымның беріктігі мен оның эстетикалық танымдылығын арттырады. Осыдан көретініміз – корпуста *қобди* сөзінің мағыналық дефинициясының ашылуы.

Корпустарда заттық атаулардың олардың қолданысына қарай мағыналық динамикасы да тілде көрініс табады. Мысалы: *тегене* – ағаштан жасалған шұңғыл *тік жиек* ыдыс. Тегене қатты ағаштан тұтас шабылып, ойып жасалады. Тегененің шұңғыл жасалған түрі де кездеседі. Тегененің екі құлағы болады. Құлағы осы ағаштан шауып жасалатын болса, кейде темірден немесе жезден де жасалады. *Тегене* қымыз, шұбат түрлі сусындар құюға арналған қолайлы ыдыс. Тегененің сыртын ағаштан оюлап безендіреді. Қымызға арналған тегенелерді сүйектен асыл тастармен әшекейлеп те қолданған. Оюдан ойылып жасалған қақпағы болған. Өйткені, тегененің көлемі үлкен болғандықтан, күнделікті өмірде көп қолданылмайды, тек той жиындарда ғана қолданылады. XX ғ. екінші жартысында қазақ халқының арасында талай жылдан бері өткізген *В. Плотниковтың* айтуы бойынша: тегене көлемі *5-15 литрге* дейін жеткен. Ағашы қаңсып шытынап кетпес үшін, күн көзіне тигізбей оны сүртіп, майлап отырған (Қазақтың этнографиялық категориялар, ұғымдар мен атаулардың дәстүрлі жүйесі энциклопедия, 2011: 385).

Осындай қазақ халқының тұрмыстық бұйымдарды қолдану кезіндегі асқан ұқыптылықпен пайдаланылғаны жайлы көрініп тұр. Дәстүрлі ортада тегенелердің оюларына қарап, үйдің әл-ауқатын байқайтын болған. Міне, содан қазақ халқының тұтынған заттарының өрнек байқампаздығын, даналығын, көрегендігін көреміз.

Торсық – түйенің, бұғының, өгіздің мойын терісінен жасалған қымыз құятын ыдыс.

Торсық – түйе, жылқы, сиыр терісінен, түйенің бітеу сойылған мойын терісінен жасалады. Торсық тігілуіне қарай *өркешторсық, мүйізді торсық, түйенің* мойын терісінен жасалады.

Үнемі өзімен бірге малшылар сусын құйып алып жүргендіктен, *жанторсық* деп аталады. Мүйізді торсықтың сыртына түрлі ою-өрнектер салынады. Оның оюланып, әдеміленіп жасалған түрі *кеужерін* деп аталады. Пішіліп алынған терілер бір-бірімен беттестіріліп қойылып, олардың шеттері *тарамыспен* немесе *шуда жіппен* тігілген. Оның екі бүйіріне *батырғы* (*батырма*) құрал ұшымен өрнек салынады. Мойын тұсынан кішкене «кұлақ» шығарыла пішіліп, оған тұйықтап темір шегелеп, дөңгелек етіп кішкене ғана тесік тесіліп, оған бүтін бір қайыс таспа тағылған. Ол таспаны өріп бау таққан. Торсық жасалып біткеннен кейін кептіріліп, ол кепкеннен кейін оның ішіндегі құмы төгіліп, жуып тазартылады. Осыдан кейін торсық кептіріліп, маймен майланып, тобылғымен ысталып, сусын қоюға қолданылған (Қазақтың этнографиялық категориялар, ұғымдар мен атаулардың дәстүрлі жүйесі энциклопедия, 2011: 454).

Көріп отырғандарымыздай, торсықтың қандай мақсатта қолданатындығы жайлы, оның қандай материалдардан жасалатындығына көңіл бөлініп, оларды талдап, топтастырғаннан кейін ғана корпус тар базасына енгізіледі.

Сүйретпе – жылқының бас терісін шикі күйінде жібітіп, түтігін алып, бітеулей тігіп, ысқа 7-8 күн қойып дайындалған қымыз сақтайтын ыдыс, сүттен жасалған сусын түрлерін сақтауға, оларды тасымалдауға арналған теріден жасалған ыдыс. Ол көлемі жағынан сабадан кіші ыдыс. Сабаны дайындаудың тәсілі: саба, бұлқыншақ, торсық сияқты ыдыстармен ұқсас болып келеді. Олар жылқы, өгіздің терісінен жасалады. Сүйретпе – ұзынша, түбі сопақтау келген мойыны қысқаша. Сүйреткіні құралға салып алып тасымалдауға болады. Сүйретпенің негізгі қызметі – сабаны босатып алу үшін сабадағы қымызды қою. Қымызды сүйретпеде уақытша сақтауға, онымен тасымалдауға болады. М.Ж. Көпейұлының айтуынша: *Абылай Жарылғантың аулына келгенде Маңқай бәйбіше жүктің астында сүйретпеде байдың сыбаға осы түнемел тұр еді*, – дейді (Қазақтың этнографиялық категориялар, ұғымдар мен атаулардың дәстүрлі жүйесі энциклопедия, 2011: 358).

Осы мысалдан көретініміз сыйлы қонақтарға сүйретпедегі қымыздан құйып берген сияқты.

Сонымен, корпус базасын құру үшін ең алдымен этномәдени бірліктер жинақталады. Олардың мәдени-семантикалық топтары анықталып, сол топқа жататын сөздердің тізімі жасалады, одан кейін оларға *белгі-код* қойылады. Әрқайсысына *белгі-код* қойылғаннан кейін ғана программист мамандарына беріледі. Содан кейін программист мамандары белгілі бір алгоритмге салып, әрі қарай өңдеу жұмыстары жүргізіледі.

Ұлттық корпус ең бірінші кезекте тілші ғалымдарға сол тілдің мәдени-семантикасын жан-жақты зерттеуге көптеген мүмкіндіктер тудырады. Бүгінгі заманауи техникалардың, компьютерлердің қарқынды дамуына байланысты корпус қол еңбектерін жеңілдетіп, корпусты пайдаланушы іздеп отырған сұраныстар бойынша жан жақты лингвистикалық ақпарат алуға мүмкіндік береді. Сондықтан да тілдің тарихын, даму бағытын зерттеуді мақсат ететін зерттеушілерге корпустың берер ақпараты мол.

Ұлттық корпус әртүрлі ғылыми зерттеулерді жүргізуді қамтамасыз ету мүмкіндігіне ие. Ұлттық корпус жүйесінде жүргізілген осы жұмыстардың негізінде қазақ халқының тұрмысында асқан шеберлікпен қолданылған материалдық мәдениетінің жарқын көріністерін жан-жақты талдауға мүмкіндік береді.

Қорытынды

Тіл корпустарының тіл білімінің барлық салаларында қолданылу мүмкіндігі бар. Нақтырақ айтсақ, әртүрлі сөздіктерді құрастыруда (жиілік, түсіндірме, аймақтық, фразеологиялық, этнографиялық) құрастыруда, оқулықтар құрастыруда, аудармалар жасауда қолданылады.

Тіл корпустарында этномәдени бірліктер компьютерлік базаға салынып, басқару жүйесі бойынша жұмыс істейтіндіктен, этномәдени бағыттағы тілдік зерттеулерді корпус материалдары бойынша зерттеуге болады. Этномәдени бірліктерді корпуста берудің маңызы

ерекше. Этномәдени бірліктерді корпустарын құрастыру ісі – кең көлемдегі мәтіндерді қамтитын күрделі жұмыстардың бірі болып отыр.

Ең бастысы ешбір корпуста этномәдени бірліктер берілмеген. Осымен байланысты зерттеу жұмысымыздың барысында жинақталып, өңдеуден өткен этномәдени бірліктерге шартты белгілер арқылы лингвистикалық белгіленімдер беріп, корпуста мағынасын ашуымыз қазақ тілін әлемнің кез келген жерінен ғаламтор арқылы зерттеуге жағдай жасайды.

Одан бұрын қазақ тіліндегі лексикографиялық және корпустық зерттеулердің теориялық және практикалық мәселелерін толықтырылып, этномәдени бірліктерді лексикографияландырудың әдіс-тәсілдерін айқындайды. Себебі зерттеу нәтижелерін түсіндірме сөздіктерде, этномәдени бірліктерді лексикографиялау кезінде қолдану ұсынылады.

Мақала ҚР ҒЖБМ Ғылым комитеті ЖТН BR18574132 «Мәдени-репрезентативті және жарнама мәтіндерінің ішкорпустарын әзірлеу» атты бағдарламалық-нысаналы қаржыландыру аясында жарияланды.

Әдебиеттер

- Жұбанов А.Қ. (2009) Компьютер қазақ сөзін жатсынбасынбайды. – Алматы, «Ана тілі», №17 30 сәуір-6 мамыр, 2009. – 3-11 б.
- Жұбанов А.Қ. (2013) Компьютерлік лингвистикаға кіріспе // Монография. – Алматы, 2013. – 204 б.
- Жұбанов А.Қ. (2014) Принципы автоматизации морфологияческой разметки текстов Национального корпуса казахского языка (НККЯ) // Материалы Международной научно-практической конференции «контрастивные исследования и прикладная лингвистика» 29-30 октября 2014 года в Минском государственном лингвистическом университете. – Минск. – 4 комп.стр.
- Жұбанов А.Қ., Жаңабекова А.Ә. (2017) Корпустық лингвистика. – Алматы: «Қазақ тілі», 2017. – 336 б.
- Жұбанов Қ. (1999) Қазақ тілі жөніндегі зерттеулер – Алматы: Ғылым, 1999. – 325 б.
- Қазақтың этнографиялық категориялар, ұғымдар мен атаулардың дәстүрлі жүйесі (энциклопедия) Алматы, 2012. 5-том.
- Қайдаров Ә.Т. (1998) Қазақ тілінің өзекті мәселелері. – Алматы: «Ана тілі», 1998. – 308 б.
- Манкеева Ж.А. (2008) Қазақ тіліндегі этномәдени атаулардың танымдық негіздері – Алматы, «Жібек жолы», 2008. – 356 б.
- Телия В.Н. (1987) О специфика отображения мира психики и знания в языке. // Сущность, развития и функция языка. – М., 1987. – 28 стр.
- Толстой Н.И. (1995) Язык и народная культура: очерки по славянской мифологии и этнолингвистике. – М., 1995. – 300 стр.
- Уәлиев Н.М. (1998) Фразеология және тілдік норма. – Алматы, «Мектеп», 1998. – 128 б.
- Finegan E. (2004) Language: its structure and use. – N.Y.: Harcourt Brace College Publishers.

References

- Ginegan E. (2004) Lingua: Eius structura et usus. – NY: Harcourt Brace College Press, MMIV. [Finegan E. (2004) Language: its structure and use. – N.Y.: Harcourt Brace College Publishers.] (in English)
- Kaidarov A.T. (1998) Qazaq tilinin ozekti maseleleri. – Almaty: «Ana Tili», 1998. 308 b. [Kaidarov A.T. (1998) Actual problem of the Kazakh language. – Almaty: «Mother Tongue», 1998. 308 p.] (in Kazakh)
- Mankeeva J.M. (2008) Qazaq tilindegi ethnovadeni ataylardin tanimdik negizderi – Almaty: «Jibek Joli», 2008. 356 b. [J.A. Mankeeva. (2008) Cognitive bases of ethnocultural names in the Kazakh language – Almaty, «Silk Road», 2008 356 p. (in Kazakh)
- Qazaqtin ethografijlik kategorijlar, ygimdar men ataylardin dastyrlı jiesi (ensiklopedij) Almaty 2022. 5-tom. [Kazakh traditional system ethnographic categories, concepts and names (encyclohedias) Almaty 2012. Volume 5. (in Kazakh)
- Telia V.N. (1987) O spesifika otobrojenij mira psihiki i znaniy v jzika. // Suchnost, razvitij i punkcija ijzika. – M., 1987.28 str. [Teliya V.N. (1987) About the specificity of the world of the psyche and knowledge in the language. // Essence, development and function of the language. – M., 1987. 28 p.] (in Russian)
- Tolstoy N.I. (1995) Ijzik i narodnıy kultura: ocherki po slabijnskoj mifologii i etholingvistiki. – M., 1995. 300 str. [Tolstoy N.I. (1995) Language and folk culture essays on Slavic mythology and ethnolinguistics.] – M., 1995. 300 pages. (in Russian)

- Ualiev N.M. (1998) Frazеologij jәne tildik norma. – Almaty, «Mektep», 1998. 128 b. [Valiev N.M. (1998) Phraseology and linguistic norms. – Almaty, «Mektep», 1998. 128 p. (in Kazakh)]
- Zhubanov A.K. (1999) Qazaq tili jonindegi zerteuleri – Almaty: Gulum, 1999. 325 p. [Zhubanov A.K. Computatorium Kazakh verbum non agnoscit. – Almaty, «Mater Lingua», 2 (34), p.] (in Kazakh)
- Zhubanov A.K. (2013) Komputerlik lingvistikaga kirispe // Monografij. – Almaty 2013. – 204 b. [Zhubanov A.K. (2013) Introduction to Computational Linguistics // Monograph. – Almaty, 2013. 204 p.]
- Zhubanov A.K. (2014) Principia automatizacij morfologiziskoi razmetki textov Nacionalnogo Corpora Kazakh // Materiali Megdunarodni nauchno-prakticheskoi konferesii «kontrastivnye izledovanij I prikladnaj linguistica» 29-30 octobrij 2014 coda v Minskom gosudarstvennim lingusticheskom universitete – Minsk – 4 comp. [Zhubanov A.K. (2014) Principles of the morphological markup of texts of the National Corpus of the Kazah Language (NCCL) // Proceedings of the International Scientific and Practical Conference «Contact Studies and Applied Linguistics» October 29-30, 2014 at the Minsk State Linguistic University. – Minsk. -4 comp.] (in Russian)
- Zhubanov A.K. Komputer qazaq sozin zhatsynbasynbaydy. – Almaty, «Ana Tili», No. 2 (34) – 3-11 b. [Zhubanov A.K. Computatorium Kazakh verbum non agnoscit. – Almaty, «Ana Tili», 2 (34), p.] (in Kazakh)
- Zhubanov A.K., Janabekova A.A. (2017) Korpustik lingvistica. – Almaty: «Qazaq tili», 2017. 336 b. [Zhubanov A.K., Zhanabekova A.A. (2017) Corpus Linguistics. – Almaty: «Kazakh language» 2017. 336 p.] (in Kazakh)