

А.С. Барменкулова^{1*} , Р.Ж. Сисенбаева² ,

Г.О. Сейдалиева³ 

¹А. Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан

²М. Өтемісов атындағы Батыс Қазақстан университеті, Орал, Қазақстан

³Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы, Қазақстан

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

ҚАЗАҚ ТІЛІНІҢ ҰЛТТЫҚ КОРПУСЫНДАҒЫ ЭТНОМӘДЕНИ БІРЛІКТЕРДІҢ РЕПРЕЗЕНТАЦИЯСЫ

Қазақ халқының ұлттық болмысы мен мәдениетін, тыныс-тіршілігін білдіретін, рухан мәдениетін танытушы этномәдени бірліктерді зерттеу өзекті мәселе. Қазіргі цифрландыру кезеңінде өмір сүріп жатқан жас ұрпаққа этномәдени бірліктердің мән-мағынасын жеткізуші ұлттық корпусның құрылуы тіліміздің дамуына өз үлесін қосады. Осы бағытта, зерттеу нысаны – қазақ тілінің ұлттық корпусындағы этномәдени бірліктер таңдалып алынды. Қазақ тілінің ұлттық корпусындағы этномәдени бірліктер жүйесіндегі заттық мәндегі мәдени атаулар репрезентациясына құрылымдық-семантикалық талдау жасау аса өзекті мәселе.

Зерттеудің мақсаты – ұлттық корпустың этномәдени бірліктердің репрезентациясын талдау арқылы заттық мәндегі тұрмыстық реалиялардың мағынасын ажыратушы ұлттық дүниетаным мен ұлттық болмыстың көрінісіне сипаттама беру. Қазақ тілінің ұлттық корпусындағы заттық мәндегі ұлттық мәдени бірліктердің сандық-сапалық көрінісіне талдау жасалып, ұлттық корпусқа этномәдени бірліктерді енгізудің жолдары мен әдістері сипатталады. Ұлттық корпусқа этномәдени бірліктерді енгізуде модельдеу, кодтау, алгоритмдік-программалық тәсілдердің қолданысына сипаттама берілді. Ұлттық корпус базасын құрудағы заттық мәндегі этномәдени бірліктердің жинақталуы мен олардың репрезентациясын талдау негізінде реалиялардың лингвомәдени-семантикалық, сөзжасамдық топтары жіктеліп, топтасу принциптері анықталды. Зерттеудің практикалық нәтижелерін қазақ тілінің ұлттық корпусының базасында этномәдени бірліктерді жинақтауда, олардың лексикалық, сөзжасамдық, семантикалық белгіленімдерін енгізуде тіл мамандары мен лексикографтарға қолдануға мүмкіндік бар.

Түйін сөздер: қазақ тілі, ұлттық корпус, этномәдени бірліктер, сөзжасам, белгі-код, этномәдени талдау, лексика-семантикалық талдау.

A.S. Barmenkulova^{1*}, R. Sisenbayeva², G. Seidaliyeva³

¹A. Baitursynov Institute of Linguistics, Almaty, Kazakhstan

²M. Utemisov West Kazakhstan University, Uralsk, Kazakhstan

³Abai Kazakh National Pedagogical University, Almaty, Kazakhstan

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

Representation of Ethnocultural Units in the National Corpus of the Kazakh Language

The study of ethnocultural units representing the national identity and culture, lifestyle and spiritual culture of the Kazakh people is a topical issue. The creation of the national corpus contributed to the development of the Kazakh language, which conveys the importance of ethno-cultural units to the young generation living in the modern era of digitalization. In this direction, ethnocultural units of the national corpus of the Kazakh language were chosen as the object of research. The problem of structural-semantic analysis of the representation of cultural names in material values in the system of ethno-cultural units of the national corpus of the Kazakh language is very relevant.

The purpose of the research is to describe the national worldview and national identity, which separate the meanings of everyday realia in the objective meaning, by analyzing the representation of ethnocultural units in the national corpus. The quantitative and qualitative representation of ethnocultural units in the national corpus of the Kazakh language has been analyzed, the ways and methods of introducing ethno-cultural units into the national corpus were described. Methods of modeling, coding and application of algorithmic-programming methods of ethnocultural units for inclusion in the national corpus were described. On the basis of the analysis of the collection and representation of ethnocultural units, when creating the national corpus base, linguistic-semantic, word-formative groups were classi-

fied according to the principle of classification. The practical results of the research can be applied by linguists and lexicographers when compiling ethnocultural units and determining their lexical, word-formative and semantic meanings based on the national corpus of the Kazakh language.

Key words: Kazakh language, national corpus, ethnocultural units, word formation, ethnocultural analysis, lexical-semantic analysis.

А.С. Барменкулова^{1*}, Р.Ж. Сисенбаева², Г.О. Сейдалиева³

¹ Институт языкознания имени А. Байтұрсынова, Алматы, Казахстан

² Западно-Казахстанский университет имени М. Утемисова, Уральск, Казахстан

³ Казахский национальный педагогический университет имени Абая, Алматы, Казахстан

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

Репрезентация этнокультурных единиц в национальном корпусе казахского языка

Изучение этнокультурных единиц, представляющих национальную самобытность и культуру, быт и духовную культуру казахского народа, является актуальным вопросом. Создание национального корпуса способствовало развитию казахского языка, доносящего значение этнокультурных единиц молодому поколению, живущему в современный период цифровизации. В этом направлении в качестве объекта исследования были выбраны этнокультурные единицы национального корпуса казахского языка. Весьма актуальной является проблема структурно-семантического анализа репрезентации культурных названий в материальных ценностях в системе этнокультурных единиц национального корпуса казахского языка.

Цель исследования – описание национального мировоззрения и национальной идентичности, разделяющие смыслы бытовых реалий в материальном значении, путем анализа репрезентации этнокультурных единиц в национальном корпусе. Проанализирована количественная и качественная представленность материальных национально-культурных единиц в национальном корпусе казахского языка, описаны пути и методы внедрения этнокультурных единиц в национальный корпус. Описаны методы моделирования, кодирования и применение алгоритмо-программных методов этнокультурных единиц для включения в национальный корпус. На основе анализа сбора и репрезентации этнокультурных единиц при создании национальной корпусной базы были классифицированы лингво-семантические, словообразовательные группы реалий и определены принципы классификации. Практические результаты исследования могут быть применены лингвистами и лексикографами при составлении этнокультурных единиц, при определении их лексических, словообразовательных и семантических значений на базе национального корпуса казахского языка.

Ключевые слова: казахский язык, национальный корпус, этнокультурные единицы, словообразование, этнокультурный анализ, лексико-семантический анализ.

Кіріспе

Қазіргі цифрландыру заманында басқа да ғылым салалары сияқты лингвистика ғылымы да көптеген жетістіктерге қол жеткізді. Жаңа ақпараттық коммуникативтік технологиялар жетістіктерінің әсерінен XX ғасырдың аяғы мен XXI ғасырдың басында пайда болған тілдік корпус немесе корпусық лингвистика тақырыптары лингвистика саласына кешенді өзгерістер алып келді. Тіл корпусы – белгілі бір тілдегі сөздер мен мәтіндердің электрондық жинағын білдіреді. Тілдік корпус тілдегі барлық лексиканы, тұтас мәтіндерді электронды форматқа көшіруге және тілдер бойынша топтастыруға жағдай туғызды.

«Қазақ тілінің ұлттық корпусы» туралы тұжырымдамада «қазақ тілі негізгі тілдердің бірі ретінде ақпараттық ресурстардың көмегімен

жаһандық бәсекелестікке, оның ішінде, ұлттық-мемлекеттік бірегейлік күреске өз үлесін қосуда» делінген. Дегенмен, «Қазақстан Республикасының мемлекеттік тілі ретінде жоғары мәртебеге ие қазақ тілі ақпараттық мүмкіндіктер жарысына біршама кешігіп қалды, ал оның қызмет етуінің өзгерту жағдайлары шұғыл және барабар шаралар қабылдауды талап етеді» деп жарияланды (Тұжырымдама).

Шын мәнінде, қазақ тілінің ұлттық корпусын құрастыру ісі кеш басталғанымен, осы уақытқа дейін көптеген шаралар істелді. Корпусың автоматты нұсқаларын көпшілік пайдаланып келе жатыр. Арнайы бағдарламалар көмегіне жүгіне отырып, лингвист-лексикографтар мәтіндерді өңдеу арқылы арнайы базаға автоматты түрде енгізуде. Қазақ тілінің корпусы да дүниежүзілік принциптерге негізделі отырып құрастырылды және кең ауқымды материалдардың жинағын

құрайды. Корпус арнайы бағдарламалар көмегіне жүгіне отырып, арнайы өңделген мәтіндер мен аннотацияланған мәтіндер корпусын жинақтап келеді.

Қазіргі әлемде 100 миллионнан асатын мәтіндер мен миллиардтаған сөздік қордан құралған корпустар аз емес. Бұндай жетістіктер корпус пайдаланушыларына қажетті материалдарды оңай, әрі тез табуға көмегін тигізеді. Дегенмен, корпусты құрушылар үшін көптеген мәселелер мен міндеттер туындайды. Олар: корпустың құрамына қандай материалдар қарастырылуы керек, қандай лексиканы енгізу қажет, қандай материалдар сұранысқа ие, оқырмандарға пайдалану үшін тиімді формат қандай болуы тиіс, мәтіндердің көлемі мен мазмұны қандай болуы шарт, іріктеу және енгізу формалары қандай болуы керек және т.б. толып жатқан сұрақтар туындайды. Сондай мәселелердің бірі – ұлттық корпусқа қазақтың ұлттық мәдени лексикасын енгізу мәселесі. Осы мәселе төңірегінде жасалатын іс-шаралар әлі де көп. Қазақ тілінің корпусында көптеген мәдени лексиканың мағынасы тіркелмеуі, мәтіндердің болмауы, аймақтық лексиканың енбеуі сияқты мәселелер әлі де болса шешімін тапқан жоқ. Біз қарастырып отырған заттық мәндес этномәдени лексиканың зерттелмей келе жатқан қабатын талдау бойынша, сөздіктерге енгізу сияқты мәселелерді шешуде, оларды корпусқа енгізу жолдары мен семантикалық талдау жұмыстарын кешенді орындау бағыттарын табуда зерттеу жұмыстарының тапшы екендігі анықталды.

Мәселен, А. Барменкулованың зерттеу жұмысындағы талдаулары аймақтық қолданыста сақталған, не шетелдік балама атаулар табылатын *саба, күбі, көсе* сияқты қолданыста жоқ этномәдени атаулар мен этнолексикалық бірліктердің түсініксіз мағынасын анықтауға арналған. Зерттеуші *адалбақан мен кебеженің, тай мен құнанның* айырмашылығын корпусты талдау арқылы көрсетеді. Осы бағытта, «Қазақ әлеміндегі» этномәдени бірліктер жүйесінің ғылыми-әдістемелік үлгісін көрсетуге тырысты. Дегенмен, бұндай көлемді жұмыс корпусты құраудың бір мәселелесін ғана талдауға арналған (Барменкулова, 2024).

Қазақтың ұлттық тіліндегі лексиканың аумақтық ерешеліктерге байланысты әртүрлі сипатта келуі қазақ жерінің, олардың елдің әр түрлі аймақтарында әртүрлі атаулармен берілуі ұлттық мәндегі лексиканың ерекше қабатын түзеді. Осындай, «заттық мәндегі этномәдени лек-

сика» немесе «ұлттық мәдени лексика» – деп қазақ этносының мәдениеті мен тұрмыс-тіршілігінде қалыптасқан, құнды төл деректерді бойында сақтаған, ғасырдан-ғасырға мұра ретінде келіп жеткен байлығы мен жәдігерлерінің атауларын айтамыз. Қазақ тілінің өзгеше сипатын құрайтын жергілікті тіл ерекшеліктеріне қарай түрленіп, қорланып отыратын ұлттық тілдің қабаты этномәдени лексиканың ерекше тобын қалыптастырды. Олардың дамып қалыптасуына ішкі, сыртқы факторлар да әсер берді. Ұлттық тілде өте кең ұғымдарды анықтаушылар ретінде, заттық мәндегі этномәдени лексика өзінің тарихи дамуында күрделі өткелдер мен сан қилы өзгерістерге ұшырай отырып, құрамы мен мағынасы жағынан бірыңғай бола бермейді. Осы бағытта, қазақ халқының бай мұрасы мен тарихи құндылықтарын анықтаушы этномәдени бірліктерді корпусқа енгізудің жолдары мен әдістерін талдау зерттеуіміздің маңыздылығын көрсетеді.

Материалдар мен әдістер

Зерттеуде теориялық әдебиеттерге талдау, салыстыру, диахронды, синхронды талдау әдістері басшылыққа алынды. Зерттеу нысаны заттық мәндегі этномәдени лексиканы этнолингвистикалық талдау барысында лексикалық талдау, компоненттік талдау, тақырыптық жіктеу әдістері қолданылды. Қазақ тілінің ұлттық корпусына (<https://qazcorpus.kz/>) талдау жасалды, оның құрамындағы заттық мәндегі этномәдени лексикасының мәдени-репрезентативті ішкі корпусына талдау жасалды. 1 миллионға жуық ішкі корпустың арасынан заттық мәндегі этномәдени лексиканың құрылымдық-мағыналық сипатына сандық-сапалық көрсеткіштері бағаланып, нәтижесі ұсынылды.

Әдебиеттерге шолу

Қазақтың ұлттық мәдени лексикасына ментальді стереотиптер жиынтығын қамтитын сөздер мен сөз тіркестері жатады. Бұл салада қазақ тіл білімінде Қ. Жұбанов, І. Кеңесбаев, Ғ. Мұсабаев, Т. Қордабаев, М. Оразов және т.б. көптеген ғалымдардың зерттеу еңбектері кездеседі.

Қазақ тіліндегі тілдік корпусты құру мәселесі басталған кезде ғалымдар алдында шешілуі тиіс көптеген міндеттер мен іс-шаралар кездесті. Олардың ең біріншісі болып, А.Қ. Жұбанов тілдік корпусты құрастырудың негізгі қиын-

шылықтары мен мәселелерін саралап берген болатын. Ғалым қазақ тілінің ұлттық корпусын құруды орталықтандыру мәселесіне ерекше мән берді. 2014 жылы «Ана тілі» газетінде мемлекеттік дәрежеде ұлттық тілде мәтіндерді компьютерлік корпуста құру жобасының аз уақытта және белгілі бір ғылыми ұжымның шешетін мәселесі еместігіне тоқталды. Жалпы, тілдік корпусты құрастыру өте күрделі жұмыс екендігін айта келе, А.Қ. Жұбанов ғалымдар тобын құруды ұсынған болатын. Оған қоса, тілдік лексиканы таңдау және жинақтау, тілдік белгілерді модельдеу, газет, журнал мәтіндерін топтастыру сияқты толып жатқан мәселелердің шешімін қарастыру ұсынылды (Жұбанов, 2014: 9). Қазақ тілінің ұлттық корпустарын еліміздің әр аумағында дербес атқарылып келе жатқан ортақ корпусты жасау ісінде орталықтандыру керектігі маңызды мәселе.

Жалпы, корпус дегеніміз – әр тілдегі электронды пішінге келтірілген, яғни, бір басқару орталығынан автоматты түрде жұмыс істейтін, лингвистикалық ақпараттарды қамтитын мәтіндер жинағы» (Жұбанов, 2016: 13). В.В. Рыков осы бағыттағы зерттеуінде келесі реттілікті ұсынады:

- шынайы сөйлеу тілінің құрылымын көрсету (ұсыну) реті;
- корпусты құрастыруға қатысты материалдық шектеулерді айқындау;
- мәтіндерді іздеп, таңдау;
- тілдік корпусты құрастыру (2002: 35).

Корпуста болатын белгіленім түрлері лингвистикалық және экстралингвистикалық деп екі топқа бөлінеді. SGML құжаттарын белгілеу мен Text Encoding Initiative (Tei) сияқты арнайы халықаралық стандартталған компьютерлік белгілеу құралдарын әзірлеуде XML форматы қолданылады. Тілдік корпустағы кездесетін сөздер мен мәтінді пішімдеуде белгіленімдер бар, олар: тақырыптар, автор туралы белгіленімдер, кодтар, шегіністер, абзацтар және т.б. Осындай ақпараттар мен белгіленімдер тілдік мәлімет базасында іздеуге қажет құжатты сәйкестендіру үшін маңызды. Осы бағытта, корпустық лингвистика қазақ тіл білімінің жаңа бағыты ретінде қалыптасып, тіл мамандарына қажет саналатын тілдік деректерді тез табуға, кең көлемді тәжірибелік материалды пайдалануға, оларға тиісті өңдеулер жүргізуге мүмкіндік тудырады (Фазылжанова, Серикбаева, 2020). Тілдік зерттеулердің шынайылыққа жетуіндегі эмпирикалық тәсілдерге жаңа ғылыми айналымда маңызды тілдік материалдарды енгізу мен көзқараспен қарауға жағдай жасайды.

Корпустық лингвистикада оның негізін құрайтын тіл болып табылады. Тіл әлеуметтік құбылыс десек, оны тәжірибе жүзінде негіздеу үшін мәліметтерге жүгінеміз. Корпустық лингвистиканы құрастыруда компьютерлік технологияны қолданудағы жалпы принциптерге:

- тілдік корпус ішінен тілге байланысты ақпаратты автоматты түрде таңдап алу;
- алынған ақпаратты өңдеу және;
- өңделген ақпаратты тексеру және оларға түсініктеме беру жатады (Барменкулова, Жаңабекова, Манкеева, 2023).

Алдыңғы тәсілдер толық автоматтандырылғанымен, өңделген ақпаратты тексеру және оларға түсініктеме беру адамның қатысуымен орындалатыны белгілі. Бұндағы тиімді саналатын белгіленімдер (кодтар) – морфологиялық белгіленімдер. Компьютерлік әдістерді қолдану, оның ішінде іздеу және белгілеу жүйесін әзірлеу қажетті мәліметтерді жылдам табу және талдау процесін жеңілдетеді. Бұл әдістер үлкен көлемдегі ақпаратты жинақтауға және жүйелеуге ғана емес, сонымен қатар, зерттеудің сапалы нәтижелерін беретін терең мағыналық және лингвистикалық талдау жүргізуге де мүмкіндік берді (Bakytkaevna et al., 2024).

Тілдік корпусты алдына қойған мақсатына және классификациялау белгілеріне орай түрлі типтерге бөліп көрсету маңызды. Корпусты құрастырудағы негізгі шарттардың бірі – ауқымдылық немесе көлемділік және теңгерімділік (репрезентативность). Репрезентативтілік дегеніміз – әртүрлі кезеңдерде мәтіндердің стильдерге, жанрларға, авторларға және т.б. қатысты пропорционалдық шамамен алынған көлемін, яғни тілдік зерттеу аясындағы жан-жақты ашатын мәтін көлемін айтамыз. Корпустарды құрастыру ісі оның түріне қарай келесі топтарға бөлінеді. Олар: ұлттық, стильдік, кезеңдік мәтіндер бойынша таңдалып алынады және корпустың көлемі сөзқолданыс санымен де есептеледі (Жұбанов, 2016: 306).

Корпуста әрбір құжат UTF8 кодтауында мәтін пішімінде сақталады. Құжаттар мазмұнды да, ақпаратты да қамтиды, мета-деректерді бір файлда және қарапайым құрылымы төмендегідей:

- TITLE – құжаттың тақырыбы;
- SOURCE – құжаттың көзі;

- АВТОР – құжаттың авторы(лар);
- DATE – құжат жарияланған күн;
- META – қосымша ақпарат;
- TEXT – құжаттың мазмұны.

Сәйкес ақпарат дереккөзде болған жағдайда, <META> тегінде құжат жататын корпус бөлімінің атауы да, әдеби шығарманың түрі сияқты одан әрі категориялық бөлімше де болады, мысалы: өлең жолдары немесе мақала. Яғни, мүмкіндігінше мұндай санаттар автоматты түрде тағайындалады, мысалы, кейбір веб-сайттар бұндай ақпаратты бере алады. Цифрланған кітаптар, диссертациялар және ғылыми жұмыстар сияқты метадеректері жоқ дереккөздер үшін сәйкес санаттар (информатика, биология, химия және т.б.) жыл сайын тағайындалады (Makhambetov, et all, 2013).

Қазақ тілінің корпусында барлық ақпараттар осы тектес беріледі. Біз қарастырып отырған қазақ тіліндегі этномәдени мәндегі зат атауларының ерекшелігі – олардың тарихи сипатта болуы мен тұрмыс-тіршілікті бейнелеуші қасиеті (Манкеева, 2008).

Көне мәдениетте заттардың таңбалық, символдық мәндері басым болды және олар белгілі бір ақпаратты білдірді. Көптеген киім-кешек бұйымдары мен атауларында айқын беріледі. Киім-кешек атауларының өзіндік мәртебесіне қарай символдық мәні де бар. Олар ер, әйел, ұл, қыз, апа, ата сияқты жынысы мен жасына қарай бөлініп жіктеледі. Заттық мәндегі этномәдени атаулардың негізгі ерекшелігі – олар материалдық мәдениетті қана көрсетпейді, олар халық тұрмысы мен адамдардың статусын да көрсетеді. Материалдық мәдениет пен рухани мәдениет бірінен-бірі туындайтын егіз ұғымда деп қарауға да болады (Ахметжанова, Оналбаева, Умирзакова, 2014). Б.М. Қайдарова ұлттық стереотиптерді анықтаушы лексиканы сипаттай келе, «халық мұралары, ауызша әдебиет нұсқалары және жазба әдебиетте сөздің мағынасына ерекше мән беріледі, әрі ол әрқашан лингвистердің басты назарында болып келеді» деп тұжырымдайды (2023: 77).

Қазақ халқының байырғы сөздері түркі тілдерінен қалыптасып дамып келді. Түркі тілдерінің ішіндегі ең бай және көркем тілі ретінде қазақ тілі бұрынғы қалпын барынша сақтап қалған (Қайдар, Оразов, 2004).

Адам өмір сүріп жатқан тіршілік әлемін қоршайтын, адамның өз қолымен жасаған заттарын,

киім-кешек, тамақ атаулары мен құрылыс материалдарының атаулары да халық мәдениетінен сыр шертеді. Басқа халықтар сияқты, қазақ халқы да материалдар ретінде ағаш, тас, саз, мақта, тері, жүн, ет, сүт сияқты өнімдері тұтынғандығы белгілі. Мысалы: киім-кешектің ежелгі, көне түрлері тері мен жүннен жасалды. Ал тамақ өнімдері арпа, бидай, тары, жүгері, сүт, ет т.б. өнімдерімен тікелей байланысты. Мысалы: *жекей құндыз, сәукеле, бөрік* – бас киім түрлерін атаса, *жекей табан, байтаба, байпақ, жорап, бестақа, шепшер* сияқты атаулар аяқ киім түрлерін сипаттайды.

Осы бағытта, қазақ тілінің этномәдени сипаттағы заттық мәндес лексикасы қазақ тілі корпусының едәуір бөлігін қамтиды. Сондықтан, аталған лексиканы зерттеу жұмысымыздың нысаны ретінде қарастырып, талдау жаасалуы зерттеуіміздің ерекшелігін көрсетеді.

Нәтижелер мен талқылау

Қазақ тілінің ұлттық корпусына талдау жасау барысында, оның құрамындағы заттық мәндегі этномәдени лексиканың келесі топтары жіктелді. Шамамен 1 миллионға жуық ішкі корпусның арасынан заттық мәндегі этномәдени лексиканы құрылымдық-мағыналық сипатына сандық-сапалық талдау жасалды. «Мәдени тақырыптық топтардың»: «тағамдық атаулар» мен «киім атаулары» құрамындағы этномәдени атауларға талдау кезінде мәдени-репрезентативті ішкі корпусының құрамындағы келесі тақырыптар талданды. Ұлттық тіл корпусындағы «мәдени бірліктер» деп аталатын тобындағы кейбір этномәдени бірліктерге компоненттік талдау тапты. Осы бағытта, қазақ тілінің корпусындағы «мәдени-лексикалық қабаттарындағы»: мәдени онимдер, этнографизмдер, көнерген сөздер, кірсе сөздерге ішінара талдау жасап көрелік. Мысалы, осы топтағы «Балдыз көрімдік» сөзіне – «күйеу жігітке балдызын көрсетіп, таныстырғанда қыз жеңгесіне беретін күйеудің кәдесі» деп анықтама берілген. Бұл лексика бойынша көптеген мәндес мәтіндер сөздің мағынасын бір-бірінен алшақтатпайды.

Келесі кезекте, кәдуілгі «саптама етік» сөзіне талдау жасап көрелік. Осы тақырыпта табылған 33 мәтінде «саптама» етіктің энциклопедиялық сөздіктер бойынша анықтамалары берілген (1-сурет).

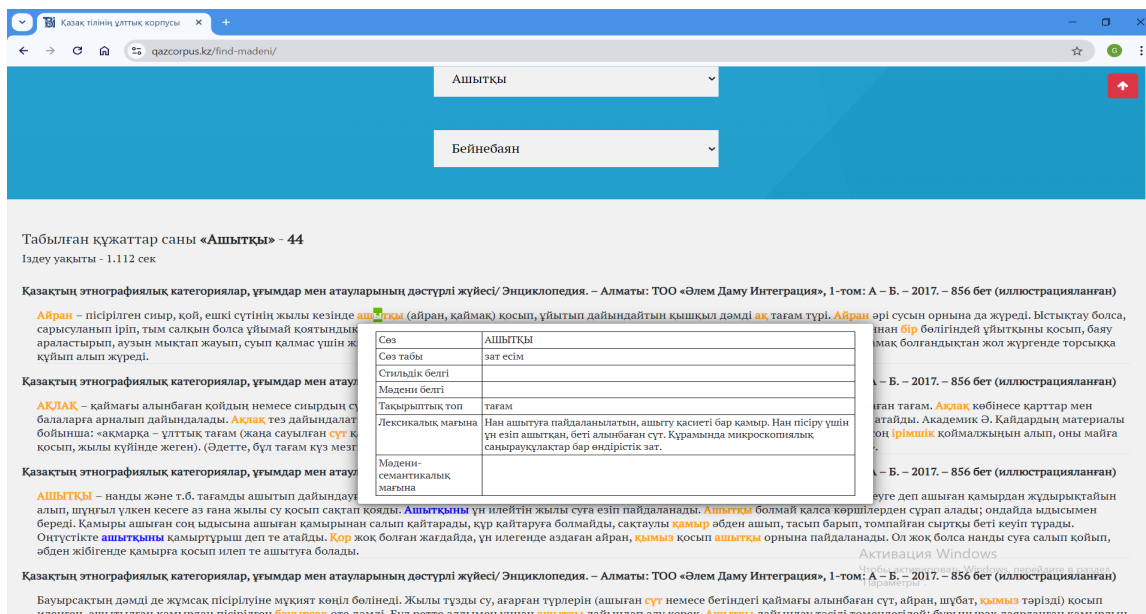


ғарыдан тігілетін жылы аяқ киім». Сонымен қатар, келесі парақшаны ашсақ, сөздің мәдени-семантикалық мағынасы қоса берілген (2-сурет).



Келесі тағам категориясы бойынша, «ашытқы» сөзін таңдаған кезде, 44 құжат табылып, «ашытқы» сөзі бойынша мағыналас сөздер то-

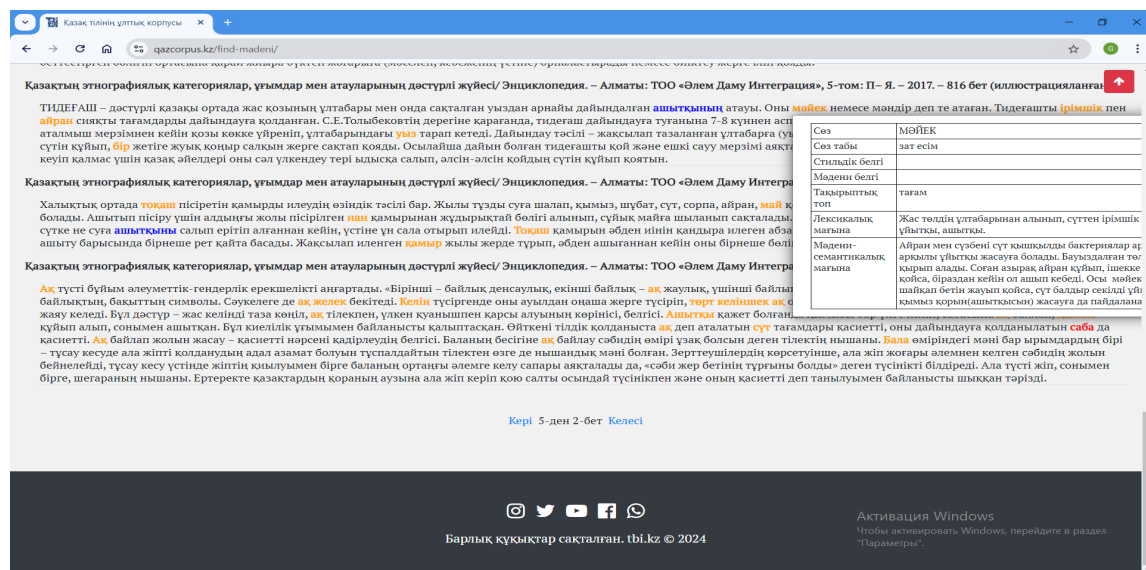
бының мәтіндердегі мағыналық сипаты мен туыстас заттық мәндегі лексиканың қатары мен субкорпустар қатары көрініс берді.



3-сурет – «Ашытқы» сөзінің ұлттық корпустағы көрінісі

«Ашытқы» сөзінің мағынасымен қатар, оған туыс айран, ақлақ, шалап, жанторсық, қымыз, тоқаш және т.с.с. лексикалар тізбегі көрініс береді. Осы топтағы «мәйек» сөзіне келер болсақ,

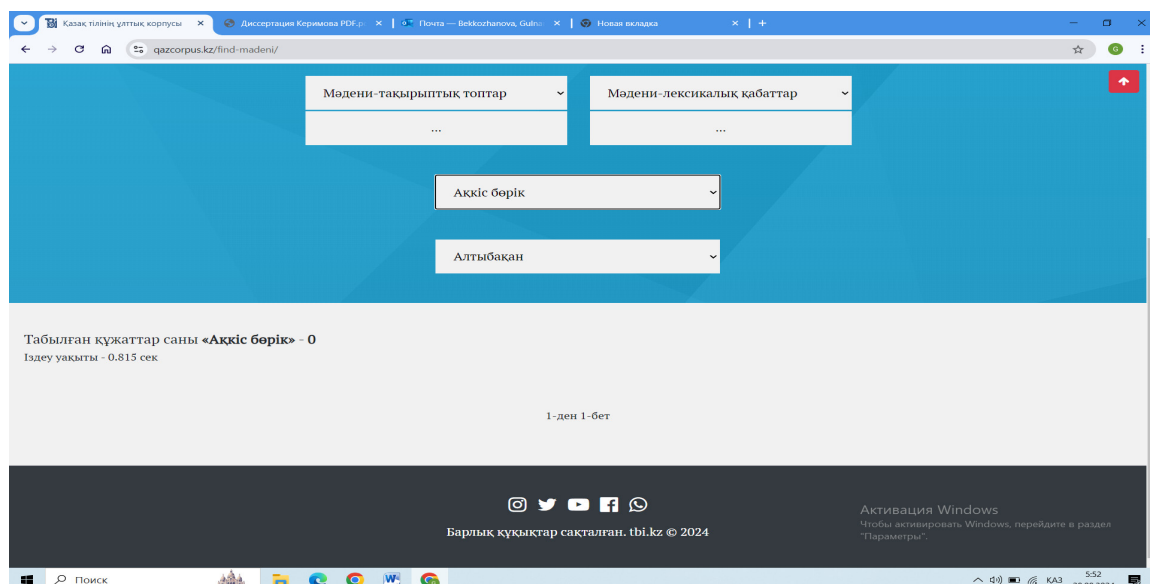
«айран мен сүзбені сүт қышқылды бактериялар арқылы ашықты дайындауға болады ...» деп анықтай келе, «мәйек» сөзінің ерекшелігі мен қолданысы туралы сипаттама берілген.



4-сурет – «Мәйек» сөзінің мағыналық репрезентациясы

Осылайша, этномәдени сипаттағы сөздердің атауы, сипаттамасы, анықтамасын тізбектеп көре отырып, олардың мәтіндердегі функцияларын анықтауға мүмкіндік бар. Дегенмен, кейбір атаулар сөздікте көрінгенімен, олардың категориялары бойынша анықтамалары мен

мәтіндер мүлде жоқ. Мысалы: *ақкіс бөрік, шар сорпа, қатық сорпа, сілекей сірне, үкпе қуырдақ* және т.б. көптеген сөздер туралы ақпараттар мен мәтіндер берілмеген. Бұл қазақ тілінің корпусының әлі де болса жасалу үстінде екендігін білдіреді.



5-сурет – Қазақ тілі корпусында берілмеген сөздердің көрінісі

Қазақ тілінің ұлттық корпусындағы заттық мәндегі ұлттық мәдени бірліктерге қазақ мәдениетіндегі, тұрмыс-тіршілігіндегі жалпытүркілік қабатқа жататын сөздерді құрайды. Этномәдени бірліктердің модельдеу кезінде, алдымен оларды тақырыптық топтарға бөлуіміз маңызды. Заттық мәндегі ұлттық мәдени лексиканы төмендегі топтарға бөліп жіктеледі (Бисенбаева, 2019: 89):

- үй-жай, қора-қопсы атаулары: *сарай, абзар, қора, жай* және т.б.
- ыдыс-аяқ, құрал-жабдық атаулары: *ырғақ, дере, ожау* және т.б.
- көрпе-төсек, көрпе-жастық атауларына қатысты атаулар өте көптеп кездеді, мысалы: *жуырқан (көрпе), жамылғы, сырмақ, жастық, көпшік, сандық, алаша* және т.б. жатады.
- киім-кешек атаулары: *күпі, шапан, малақай* және т.б.
- өсімдік, оның түрлеріне қатысты атаулар: *алғы, асқабақ, тары, жүгері, жусан* және т.б.
- туыстық, туыстық жақындықты білдіретін атаулар: *аға, нағашы, жиен, құда-жесекжат* және т.б.

- іс-әрекет, қимыл атаулары: *мошқау, жону* және т.б.

- азық-түлік, тағам атаулары: *тоқаш, күлше, көмбе* және т.б.

- жан-жануар, оларға қатысты ұғым атаулары: *қодық, күре, бота, ешкі, қом, жазы* және т.б.

- табиғат құбылыстарына, жер бедеріне қатысты атаулар: *қабыз, шық, жайлау*, және т.б.

- сындық-сапалық ұғым атаулары: *әпет (күшті), дайыр (анық), қары (оңбаған)* және т.б.

Этномәдени лексиканы модельдеудің тағы бір әдісі – оларды құрылымдық-компоненттік талдау әдісі. Құрылымдық жағынан қазақ тілінің ұлттық корпусындағы заттық мәндегі ұлттық мәдени бірліктерге жалпытүркілік лексикалық қабат жатады. Олар түбір және туынды түбір тәрізді жалаң сөздерден құралады. Мысалы: *өре* – «бір салма киіз»; *қарма* – «майға қуырылған балық»; *күп* – «су сақтайтын орын»; *сел* – «күш-қуат, денсаулық»; *кірші* – «кір жуатын ыдыс». Аналитикалық тәсілмен жасалған күрделі сөздерге мысал өте көп, атап айтсақ: *ашық немесе машық «есіктің топсасы» қылдыбақыр – «шелек», тастабақ – «тарелка», қара үй – «киіз*

үй», *taskedi* – «сырты қатты асқабақ», *қолкиіз* – «қазан көтергіш киіз», т.б. түркі тілдерінің дамуындағы қалыптасқан көне сөздердің төл қабатының үлесіне жатады.

Халық тұрмысындағы қолданылатын этномәдени атаулар арасында от жағуға байланысты қалыптасқан заттық атаулар қатары орын алады. Мысалы. *жәмке, шымыуыр, жіптік, мәстемір, шоқ қысқыш, шаттауық, әтішкір, шоқ салғыш* деген заттық атаулар жиі кездеседі. Ұлттық мәдениеттің өзегі – ұлттық дүниетанымы десек, аймақтық лексика құрамында кірме қабат болуы заңды, өйткені, ешқандай ел басқа әлем халықтарынан жеке дара, оқшау өмір сүрмеген (Сарыбаев, Нақысбеков, 1989: 39).

М.С. Атабаева әрбір этностың өзін қоршаған әлемін өзінше көруі, өзіндік ұлттық дүниетанымының деңгейі мен дәрежесінде танып білуін ерекшелеп көрсетеді. Қазақ халқының ұлттық мәндегі лексикасындағы жергілікті ерекшелігін танытатын элементтерге түркітекті тілдік субстраттар жатады. Ал, оларды кірме сөздер ретінде танытатын элементтерге көне дәуірден бастап қарым-қатынастарды анықтайтын кірме элементтер белгілері. Мысалы, евразиялық түрлі кеңістікте мекендеп келген әртүрлі этностар тарихынан, олардың арасындағы қарым-қатынас нәтижесінде тілден-тілге енген лексика қабатын «кірме» деп атаймыз (Атабаева, 2006: 51). А.Қайыржановтың пікірінше, этникалық мәдениеттер арасындағы үйлесімдік болғандықтан, кірме элементтерге жол ашылады және оларды «кез келген этностың рухани эволюциясының шарты, шығармашылық процессі» деп атауға болады (Қайыржанов, 1999: 195).

Ә.Нұрмағанбетов осындай туынды мәндес лексиканы «әдеби негізден», «әдеби негіз бен диалектілік негізден» және «диалектілік негізден» құралып, сөйленістік туынды атау жасайды (Нұрмағанбетов, 1985: 47). Ұлттық мәдени лексиканың осындай қабатын халық өмірі мен түрмс-тіршілігінің барлық саласын қамтитын күрделі туынды сөздер құрайды. Олар: үй-жай атауларын, жан-жануар аттары, тағам атауларын, күнделікті тұрмыстық заттарды, өсімдік, табиғат құбылыстарды, сапалық ұғым атауларын молынан қамтиды. Сонымен қатар, заттық мәндегі ұлттық мәдени бірліктерді біріккен сөз, қос сөз, тіркескен сөз сияқты түрлері кездеседі:

- жан-жануарға қатысты күрделі атау сөздерге мысал беретін болсақ: *айтұяқ, қыдыражал, алмауыт, қылқұйрық (жылқы)* деп бейнелеп айтады. Немесе: *қара тұқай (түйенің бақайын-*

дағы мүйіз тұяқ), күрдебай (кенже қозы) сияқты мысалдар өте көп.

- үй-жай, қора-қопсы атаулары: *ағаш үй (киіз үй), бақауыз (құрал-сайман қоятын сарай), ашпұзыл (мейрамхана), әуліжай, қора-қопсы, қора-жай, асқазан – (ас үй), жертөле* және т.б.

- бау-бақша, өсімдік атауларына қатысты күрделі атау сөздерге мысал: *айқабақ – күнбағыс, ақпорми (жүгері), айылжуа, сарымсақ,* және т.б.

- құрал-жабдық, ыдыс-аяқ атаулары: *кермағаш (шкаф), айырқап (ертоқым түрі), аққұлақ шелек (ағаш ыдыс), жылқауыз, тістеуік, бақша шот (кетпен), қылдыбақыр (елек)* және т.б.;

- киім-кешек атаулары: *аяққап (аяқ киім), күлөпара, жалбағай (масадан қорғануға арналған киім); қаптама етік (киіз етік), жалғантыс–жадағай күлөпара (бас киім)* және т.б.;

- тағам, тамақ атауларына: *таба нан, көмбеқазан, қазанжаппа (нан түрлері), ақтүймеш (құрт), қаусырма нан (жайма нан), қамырбас (ашытқы),* және т.б.;

- табиғат құбылыстарындағы күрделі атаулары: *ақбайтал (үскірік боран), бейбіл жел (қатты жел), ақ баспа (қалың қар); айық (ажуын-шашынсыз), жалт ойнау (найзағай)* және т.б.

Корпустағы этномәдени лексиканың сөзжасамдық қалыптарына компоненттік талдау нәтижелерін қарастырар болсақ, келесі түрлері кездесті:

- *зат есім + зат есім* үлгі қалыбы жатады, мысалы: *жал-жұғын (жал-құйрық), қасқазан (асхана), атқұлақ (саңырауқұлақ), ол шаршы (қол орамал), қойқосақ (кемпірқосақ), ши қасық (шай қасық), түйе табан (үлкен ожау), ауыз там, ауыз үй, кіре беріс* және т.б.;

- *зат есім + етістік* үлгі қалыбы: *тәртіп беру (бұйрық, нұсқау), қирасақал (бұйра сақал), ит кірмес (киіз үйдің есігі), тоң бұзар (сүймен)* және т.б.;

- *сын есім + зат есім* үлгі қалыбы: *қара кеспе (жеңсіз камзол), боз үй, қара қазан, қара үй (киіз үй), қызыл үй (бекет), қарабура (қамыс қора), қара алқа (құлып), терме үй немесе киіз үй, қамкірпіш (күйдірілмеген кірпіш),* және т.б.;

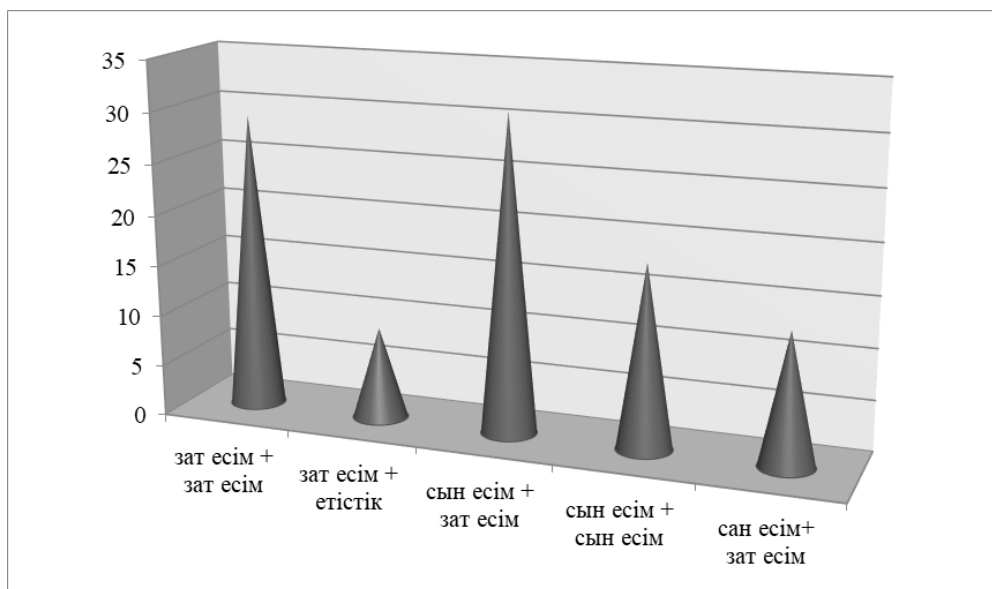
- *сын есім + сын есім* қалыбындағы күрделі туындылар: *қып-қызыл (жап-жас өрімдей), қызыл жүйрік (аудың торын құрт жемес үшін қолданылатын дәрі), көне-көскі (ескі құсқы), ала қиғаш (әлек, әбігер);*

- *сан есім + зат есім* қалыбындағы күрделі туындылар: *бесақа (тармақты айыр), жеті қоңыраулы шаибау (жеті қоңырауы бар шаибау),*

бестақа – (ерлердің өкішесі аласа етігі), бір тақта (ширек гектардай жер) және т.б.

Осы бағытта, қазақ тілінің ұлттық корпусындағы 2000-нан астам заттық мәндегі лексикаға талдау жасалды. Зерттеу барсында олардың мә-

нідегі қолданысы мен мағынасына және құрылымдық ерекшеліктеріне талдау жасалды. Келесі 6-суретте қазақ тілі корпусындағы заттық мәндегі ұлттық мәдени бірліктердің құрылымдық талдау нәтижелерінің статистикасы берілген.



6-сурет – Қазақ тілі корпусындағы ұлттық мәдени бірліктердің құрамындағы күрделі туындылардың өнімділік көрсеткіші

Көріп отырғанымыздай, қазақ тілінің ұлттық корпусындағы заттық мәндегі ұлттық мәдени бірліктердің құрамындағы күрделі туындылар арасында бір сыңарлары зат есім, сын есім және етістік құрылымдағы сөзжасамдық қалыптар өнімді түрлеріне жатады. Олардың қазақ тілі корпусындағы өнімді түрлеріне сандық талдау нәтижесі төмендегі статистикалық нәтижелерді көрсетті. Олардың ішіндегі ең өнімділеріне – зат есім + зат есім – 29% және сын есім + зат есім – 31 көрсетті. Келесі қатарда сын есім + сын есім -18%-ды; сан есім+ зат есім-13%-ды құрайды. Ал, зат есім + етістік үлгі қалыбы – 9%-бен өнімсіз түрге жатады.

Заттық мәндегі этномәдени лексиканың ортақ мағыналары заттар, құбылыстар мен қимыл ұғымдары атаулары болғандықтан, халықтың күнделікті тұрмыс-тіршілігінің көрінісін береді. Олар әр түрлі аймақтарға байланысты айтылуымен ерекшеленеді. Жоғарыда көрсетілгендей, қазақ тілінің жергілікті ерекшеліктері заттық мәндегі ұлттық мәдени бірліктердің құрамындағы күрделі туындылар сөзжасам қалыптарының дамуына септігін тигізеді.

Қорытынды

Қазақ тілінің корпусық лингвистикасының этномәдени бағытын дамыту – мемлекеттік тілдің толыққанды дамуында қоғам қажеттілігін өтейтін күрделі құбылыс. Біз қарастырған қазақ тілінің корпусындағы тілдік ақпараттың мол жинақталған қоры – этномәдени лексиканың базасын толықтыру әлі аяқталмаған сала. Зерттеу барысында заттық мәндегі сөздердің мағынасы мен анықтамасы, оған байланысты мәтіндердің мүлде берілмегендігі анықталды. Олар бүгінгі күннің қолданбалы лингвистика саласындағы тың мәселелердің бірі ретінде ұлттық корпусы дамытуда және оны қолдану мүмкіндіктерін сыннан өткізуде аса маңызды тілдік бірліктер.

Тілдік корпустағы ұлттық мәдени лексиканы талдау үлкен көлемдегі мәтіндердегі лексиканы талдау, зерттеу және нәтижесін анықтауда үнемі дамытылып отыратын күрделі жұмысты қажет етеді. Корпусты құрастыру мен қолдануда ондағы этномәдени лексиканы талдаудың бірыңғай әдістемесі де қарастырылмай келеді.

Сондықтан, тілдік корпусты дамытуда қазақтың ұлттық мәдени бірліктерін әлі де болса талдау мен олар туралы ақпаратты толықтыру бойынша жаңа бағыттағы талдау жұмыстары мен зерттеу жұмыстары қажет деуге толықтай негіз бар.

References

- Atabaeva, M.S. (2006). Qazaq tili dialektilik leksikasınıñ étnolingvistikalıq negizi. [The Ethnolinguistic Basis of Dialectal Vocabulary in the Kazakh Language]. Almaty. (in Kazakh)
- Akhmetzhanova, Z.K., Onalbayeva, A.T., Umirzakova, Z.A. (2014). Kul'tura na zerkal'nom yazyke [Culture in the Mirror Language]. Almaty. Yeltanym. (in Russian)
- Barmenqulova, A.S., Zhanabekova, A.A., Mankeyeva, ZH.A. (2023). Puti peredachi etnokul'turnykh yedinit v korpuse. [Ways of Transmitting Ethnocultural Units in the Corpus]. Tiltanym. Vol. 89(1), P. 56-66. <https://doi.org/10.55491/2411-6076-2023-1-56-66> (in Russian)
- Barmenqulova, A. (2024). Aymaktyq leksikada saqtalghan etnomádeni birlikterdiń ulttyq korpusta berilwi. [Transfer of ethnocultural units stored in the regional lexicon to the national corpus]. Tiltanym. Vol. (2), P. 193-200. 10.55491/2411-6076-2024-2-193-200
- Bisenbaeva, M.Q. (2019). Étnomadeni leksikanıñ tanımdıq-qatısmdıq qızmeti (aymaqtıq leksika materialdarı negizinde) – Filosofiya doktorı (Phd) darejesin alw ushin dayındalğan dissertaciya. [Cognitive and Communicative Functions of Ethnocultural Vocabulary (Based on Regional Lexical Materials). Dissertation Prepared for the Degree of Doctor of Philosophy (PhD)]. Almaty. (in Kazakh)
- Jubánov, A.Q. (2016). Korpustıq lingvistika [Corpus Linguistics]. Almaty. (in Kazakh)
- Tuzhyrymdama – Kontsepsiya Natsional'nogo korpusa kazakhskogo yazyka (2014). [Kontsepsiya Natsional'nogo korpusa kazakhskogo yazyka]. https://online.zakon.kz/Document/?doc_id=31637787. (in Russian)
- Qazaq tiliniń ulttyq korpwsı. [The National Corpus of the Kazakh Language]. (<https://qazcorpus.kz/>). (in Kazakh)
- Qaydar, A., Orazov, M. (2004). Turkitanwga kirispe. [Introduction to the Study of Turkic Peoples]. Almaty. (in Kazakh)
- Qaydarova, B.M. (2023). Lingvomadeniattanwdagı ulttyq-madeni stereotipterdiń leksika-semantikalıq erekshelekteri [Lexico-Semantic Specifics of National and Cultural Stereotypes in Linguoculturology]. Bulletin of L. N. Gumilyov Eurasian National University. Philology Series. Vol. 1(142), P. 87-96.
- Qairzhanov, A. (1999). PALAEOTURCICA: Mir drevnix tyurkov [PALAEOTURCICA: The World of the Ancient Turks]. Almaty. (in Russian)
- Mankeeva, J.A. (2008). Qazaq tilindegi étnomadeni atawlardıń tanımdıq negizderi. [Cognitive Foundations of Ethnocultural Terms in the Kazakh Language]. Almaty. (in Kazakh)
- Makhambetov, O., Maqajanov A., Esenbaev Zh., Matqarimov B., Sabyrgaliev I., Sharafudinov A. (2013). Kazakh corpus Colloquii. [Assembling the Kazakh Language Corpus]. Asta 2013 Conferentiae de Methodis empiricis in Processing Naturalis Linguae, P. 1022-1031, Seattle, Washington, USA. [Electronic resource]. – <https://aclanthology.org/D13-1104.pdf> (in Kazakh)
- Nurmagambetov, A. (1985) Jergilikti til erekshelekteriniń törkini. [The Etymology of Regional Linguistic Features]. Almaty. 1985. (in Kazakh)
- Rykov, V.V. (2002). Korpus tekstov kak realizatsiya ob'yektno-orientirovannoy paradigmy [Text corpus as an implementation of the object-oriented paradigm]. Trudy Mezhdunarodnogo seminaru Dialog. Moscow. (in Russian)
- Tlegenova, G.B., Rysbergen, K.K., Zhanabekova, A.A. (2024). Repraesentativum Mythonymorum et Theonymorum in Lingua Corpora International Acta Societatis, Culturae, & Language [Representativeness of Mythonyms and Theonyms in Language Corpora International Journal of Society, Culture, & Language]. Vol. 12(1), P. 374-395. <https://doi.org/10.22034/ijsc.2024.2017262.3288>
- Sarybayev, S.H., Nakysbekov, O. (1989). Regional'nyy leksikon kazakhskogo yazyka. Almaty. (in Kazakh)
- Fazilzhanova, A. Serikbayeva, A (2020). O novom neoograficheskom slovare kazakhskogo yazyka. [On the New Neographic Dictionary of the Kazakh Language]. DOI: 10.29013/EJHSS-20-4-138-143: <https://elibrary.ru/item.asp?id=44300573>. (in Russian)

Авторлар туралы мәлімет:

Барменкулова Аида Серікханқызы (корреспондент-автор) – PhD докторант, А. Байтұрсынұлы атындағы Тіл білімі институты (Алматы, Қазақстан, e-mail: bs.aida@mail.ru);

Сисенбаева Раушан Жаңабайқызы – магистр, М. Өтемісов атындағы Батыс Қазақстан университеті (Орал, Қазақстан, e-mail: raushan_sisenbaeva.71@mail.ru);

Сейдалиева Гүлдана Омаровна – PhD, Абай атындағы Қазақ ұлттық педагогикалық университеті (Алматы, Қазақстан, e-mail: g.seidalieva@abaiuniversity.edu.kz).

Свидения об авторах:

Барменкулова Аида Серикхановна (корреспондент-автор) – PhD-докторант, Институт языкознания имени А. Байтұрсынулы (Алматы, Казахстан, e-mail: bs.aida@mail.ru);

Сисенбаева Раушан Жанабаевна – магистр, Западно-Казахстанский университет имени М. Утемисова (Уральск, Казахстан, e-mail: raushan_sisenbaeva.71@mail.ru);

Сейдалиева Гүлдана Омаровна – PhD, Казахский национальный педагогический университет имени Абая (Алматы, Казахстан, e-mail: g.seidalieva@abaiuniversity.edu.kz).

Information about the authors:

Barmenkulova Aida (corresponding author) – PhD student, A. Baitursynov Institute of Linguistics, (Almaty, Kazakhstan, e-mail: bs.aida@mail.ru);

Sisenbayeva Raushan – master, M. Utemisov West Kazakhstan University (Uralsk, Kazakhstan, e-mail: raushan_sisenbaeva.71@mail.ru);

Seidalieva Guldana – PhD, Abai Kazakh National Pedagogical University (Almaty, Kazakhstan, e-mail: g.seidalieva@abaiuniversity.edu.kz).

Келіп түсті: 16 қыркүйек 2025 жыл

Қабылданды: 5 маусым 2025 жыл