



XҒТАР 16.31.51
Ғылыми мақала

<https://doi.org/10.32523/2616-678X-2025-152-3-89-101>

Мәдени-репрезентативті ішкорпус: ұлттық мәдени-танымдық бірліктерді цифрлық кеңістікте сақтаудың маңыздылығы

Э.Ә. Өтебаева^{*1}, Л.Р. Қойшығұлова², А.Б. Нұстаева³

^{1,2,3}А.Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан

(E-mail: ¹elmira_abdigali@mail.ru*, ³ardana.nustaeva@bk.ru)

Аңдатпа. Мақалада мәдени-репрезентативті ішкорпустың құрылымы мен маңыздылығы, ұлттық мәдени тілдік бірліктерді цифрлық кеңістікте сақтаудың жолдары мен танытудағы рөлі қарастырылды. Қазақ халқының мәдени мұрасын цифрлық кеңістікте сақтап, оны қазіргі ақпараттық қоғамда кеңінен танылуына және таралуына ықпал етіп отырған электрондық Қазақ тілінің Ұлттық корпусы – ұлттық тіл мен мәдениетті дамытудың маңызды әрі сенімді ақпараттық базасы. Корпус құрамындағы мәдени-репрезентативті ішкорпус – қазақ этносының этномәдени даму жолын, рухани және заттық құндылықтарын, этнографиялық, мәдени-танымдық ұғымдарын тереңірек танудың бір құралы ғана емес, тілдік қазынаның елді тұтастырушылық қызметін сипаттайтын маңызды таным тетігі ретінде қарастырылды. Мақаланың мақсаты – қазақ этнолингвистикасының ғылыми негіздеріне сүйене отырып, ҚТҰК (Қазақ тілінің ұлттық корпусы) құрамындағы мәдени-репрезентативті ішкорпустың жасалуы және оның қазақ тіліндегі тілдік және мәдени құндылықтарды цифрлық платформада жинақтау арқылы ұлттық мәдениетті, дәстүрлерді, тарихи және әлеуметтік аспектілерді зерттеу мен сақтау жолдарын анықтау. Сонымен қатар, ішкорпустың құрамын, құрылымын, қолдану әдістерін көрсету және оның білім беру мен ғылыми зерттеулердегі маңыздылығын дәлелдеу. Мақаланың зерттеу нысаны – ҚТҰК құрамындағы мәдени-репрезентативті ішкорпус пен онда жинақталған қазақ этнолингвистикасының тілдік және мәдени материалдары.

Түйін сөздер: қазақ тілінің ұлттық корпусы, мәдени-репрезентативті ішкорпус, цифрлық кеңістік, мәдени-семантикалық ақпарат.

Түсті: 26.06.2025; Жөнделді: 22.08.2025; Мақұлданды: 15.09.2025; Онлайн қолжетімді: 25.09.2025

*хат-хабар үшін автор

Кіріспе

Қазақстан үкіметі UNESCO-мен ынтымақтастық аясында еліміздің материалдық және материалдық емес мәдени мұраларын халықаралық тізімдерге енгізуге ерекше мән беріп отыр. «Әлем жады» (Memory of the World) бағдарламасы – UNESCO-ның маңызды бастамаларының бірі. Оның мақсаты – дүниежүзілік мәдени және тарихи құндылықтарға жататын құжаттық мұраны сақтап, қорғау, сондай-ақ оларды кеңінен қолжетімді ету, цифрландыру. Материалдық емес мұра – ұрпақтан ұрпаққа ауысып, халықтың рухани өмірін білдіретін мәдени құндылықтар. Оған әдет-ғұрыптар мен салт-дәстүрлер, дәстүрлі өнер, қолөнер, мерекелер мен салтанатты рәсімдер т.б. жатады. Әдет-ғұрыптар мен салт-дәстүрлерге қазақтың Наурыз тойы, қонақ қабылдау дәстүрлері, той мен құдалық рәсімдері сияқты ұлттық мәдени ерекшеліктер кіреді. Дәстүрлі өнерге ән, би, музыка және халықтық аспаптарда орындалатын шығармалар жатады. Қолөнер түрлеріне дәстүрлі киім тігу, кестелеу, сырмақ тоқу, қыш құмыра жасау сияқты халықтық қол еңбек түрлері жатады. Сонымен қатар, жыл сайын өткізіліп тұратын ұлттық мерекелер мен салтанатты рәсімдер де материалдық емес мұраға жатады. Материалдық емес мұра – ұлттың рухани бейнесі мен мәдени ерекшелігін сақтайтын, тарихи сабақтастық пен халықтың бірегейлігін көрсететін құндылық. Ол қоғамның бірлігін нығайтып, ұрпақтан ұрпаққа мәдениет пен дәстүрді жеткізуге қызмет етеді. UNESCO бұл мұраны қорғау, зерттеу және насихаттау, цифрландыру арқылы халықтардың мәдени мұрасын сақтап қалуға үлкен көңіл бөліп отыр. Сондықтан жаңа заманның талабына сай Қазақстан алдында мәдениетті жаңғырту және дәстүрлер мен инновацияларды үйлестіру сияқты маңызды міндеттер тұр. Тарихтан көріп-білгеніміздей, көптеген елдер өздерінің ұлттық салт-дәстүрлерімен қатар, қазіргі заманғы инновацияларды тиімді біріктіру арқылы дамудың жолын таңдады. Сол сияқты, біз де төл мәдениетімізді ғаламдық әлемдегі шынайы жағдайларға бейімдеуіміз қажет.

Қазіргі таңда мәдениет пен идеология саласы жасанды интеллектінің ықпалымен түбегейлі өзгерістерге ұшырауда. Дегенмен, бұл технологияларға толықтай сенуге де болмайды, өйткені адам факторы әлі де маңызды болып қала береді. Мысалы, танымал сервистерге Қонаққа ұсынылатын бастың тісін қаға ма әлде қақпай ма деген сауалға «Бас ұсыну қазақтың ұлттық мәдениеті мен тағамға деген құрметінен туындаған әдеп-ғұрып. Сол себепті бастың тісін қақпайды, абайлап, сыпайы түрде тамақтану ұсынылады» деген жалған ақпараттар ұсынады. Қазақтың дәстүрінде жиын-тойларда, үлкен аста, қонақ күткенде сойылған малдың басы – ерекше қадірлі әрі құрметті тағам ретінде есептеледі. Басты табаққа салып, қонаққа ұсыну – сыйлы адамның мәртебесін көрсетеді. Міне, осы рәсімге қатысты дәстүрлі тағам мәдениетінде қонаққа қойдың тісін қағып әкелу дәстүрі бар. Ертеде дәстүрлі қазақ қоғамында, егер реніші болса, оны тікелей бетіне айтпай, назын тұспалмен, ыммен білдіру мақсатында қонағына тартқан бастың тісін қақпай ұсынған. Біреуге ренжісе де, бетінен алмаған, өйткені өзгенің көңіліне қаяу, жүрегіне қылау салармын деп сақтықпен қараған. Міне интернетте осындай мәдени ақпараттар, әсіресе әлемде кең тараған тілдердегі, оның ішінде ағылшын тіліндегі ақпараттардың жеткіліксіздігінен де туындайтын мәселе.

Осындай кейбір жалған ақпараттарды болдырмау үшін мемлекетіміз бұл бағытта үлкен жұмысқа кірісті: қолда бар мұрағаттық құжаттар, анықтамалық мәліметтер, ғылыми зерттеулер мен мәдени ақпараттарды цифрлық форматқа көшіру жұмыстары жүріп жатыр. Сонымен қатар, ұлттық цифрлық мұрағат құру арқылы еліміздегі және шетелдердегі нейрожүйе жасаушылардың бұл деректерді еркін пайдалануы үшін жағдай жасау көзделуде. Осы бағытта жүргізіліп жатқан зерттеулер мен жобалар ел мәдениетін сақтау мен дамытудың негізі болары сөзсіз. Сондықтан бүгінгі күні ұлттық дәстүрлерімізді, тілдік мұрамызды және этнолингвистикалық ерекшеліктерімізді цифрлық платформаға енгізу – ең басты міндеттердің бірі болып табылады. Осы жағдайды ескере отырып, қазақ халқының ұлттық құндылықтарын сандық форматта сақтау мақсатында Ахмет Байтұрсынұлы атындағы Тіл білімі институтының мамандары электрондық Қазақ тілінің ұлттық корпусын (<https://qazcorpus.kz/>) әзірледі. Қазіргі таңда ҚТҰК ауқымды корпустық ресурс ретінде көпшілікке қолжетімді форматта үздіксіз жұмыс жасап тұр. 18 ішкорпустың өзіндік функционалдық, тілтанымдық ерекшелігі бар және үздіксіз толықтырылып, жетілдіріліп отырады. Ішкорпустың бірегейі Мәдени-репрезентативті ішкорпус. Мәдени-репрезентативті ішкорпус базасындағы ақпараттар отандық этнолингвистика ғылымының сан жылғы дамуының нәтижесі деуге болады. Ішкорпус қазақы ортадағы адам, қоғам және қоршаған табиғи ортаның арасындағы сан-салалы қатынастардың субъектілерін, объектілерін және осы қатынастардың жүзеге асырылу жолдары мен механизмін бейнелейтін мәдени бірліктерді, ұғым-түсініктерді, атауларды қамтиды. Кейбір мағынасы солғындаған, тіпті тілімізде өзі сақталса да, халықтың есінде, жадында оның бір замандағы айқын-анық лексикалық мағынасы ұмыт болған тіпті айтылудан қалған көптеген этномәдени бірліктер де ғылыми айналымға енгізіліп отыр. Мәдени ұғым-түсініктерді цифрлық кеңістікке жинақтап, оларды ғылыми зерттеулерде, білім беру саласында және жалпы халық арасында тиімді пайдалануда әлемдік тәжірибеде тұңғыш рет дала академиясының білімдер жүйесін цифрлық кеңістікте дайындаған электрондық база құрылды. Ұлттық мұраны цифрландыруға бағытталған мұндай жұмыстар мәдениет пен тілдің әртүрлі аспектілерін сақтау мен зерттеуде жаңа мүмкіндіктер ашып, болашақтағы зерттеулер үшін тың дереккөздер ұсынады. Ұлттық мәдениет пен тілдің цифрлық сақтау әдістері бойынша ғылыми зерттеулердің жаңа кезеңіне бастама болады. Қазақ халқының ғасырлар бойы сақталып, атадан балаға, ұрпақтан ұрпаққа рухани мұра болып келген сөз байлығын жинақтап, сұрыптап, топтастырып, мән-мағынасын айқындап, анықтамасын беріп, болашақ ұрпақтың игілігіне айналдыру аталмыш ішкорпустың негізгі мақсаты болып табылады.

Зерттеу әдістері

Зерттеу жұмысында сипаттау, жүйелеу, топтастыру, жинақтау, сандық мәліметтер мен түсіндіру сынды әдіс-тәсілдер қолданылды. Үлкен көлемдегі тілдік материалдарды тиімді өңдеу және талдау үшін компьютерлік-бағдарламалық тәсіл де белсенді түрде пайдаланылды. Осы әдістер мен тәсілдердің үйлесімі зерттеудің толық әрі жан-жақты жүргізілуіне мүмкіндік берді.

Әдебиетке шолу

Мәдени-репрезентативті ішкорпустың жасалу жолын, құрылымын зерттеуге арналған бірнеше ғылыми мақалалар бар. А.Жұбановтың (2016), А.Фазылжанованың (2016) еңбектерінде алғаш рет ұлттық корпус базасына мәдени-семантикалық белгіленімнің енгізу қажеттілігі айтылады, Э.Өтебаева, Г.Исаева және М.Серікқызы (2023) мәдени-репрезентативті ішкорпусты әзірлеудегі мәдени-семантикалық белгіленімнің берілу жолдарын анықтайды; мәдени-репрезентативті ішкорпусты әзірлеудегі мәдени бірліктерді аннотациялау проблемалары Ж.Манкеева, Н.Шүленбаев, А.Бисенғалидың (2023) еңбегінде баяндалады, лингвомәдени білімдер жүйесін ұлттық корпусқа енгізуде көне атауларды зерттеу қажеттілігін Н.Әшімбаева мен А.Жаңабекова (2023) ұсынады, А.Аманбаева мен Ж.Жұмабаева (2024) мәдени-репрезентативті ішкорпуста берілген діни сөздердің мәдени семантикасын қарастырады.

Материалдық мәдениет атауларының, прецеденттік есімдердің, мифологиялық атаулар мен ұғымдардың, зооморфтық және фитонимдік атаулардың этномаркерленуі қазақ этностық дүниетанымын құрайтын негізгі элементтер ретінде қарастырылған зерттеу еңбектері (Talgatqyzy et al, 2024) талданды. Сонымен қатар кейбір ономастикалық тілдік бірліктердің прагматикасы мен мәдени семантикасын, соның ішінде олардың фразеологизмдер құрамындағы қолданысын салыстыра зерттеген еңбектер (Tlegenova et al, 2024) де негізгі дереккөз ретінде пайдаланылды.

Зерттеу жұмысына «Қазақ тілінің мәдени-танымдық сөздігі» (Шойбеков және т.б., 2023) мен «Лингвомәдени сөздік» (Уәли және т.б., 2024) еңбектерінде берілген тілдік бірліктердің сандық көрсеткіштері негізге алынып, олардың мәдени ақпараттарымен құндылығы ескеріле отырып талдау жүргізілді. Қазақ тіліндегі мәдениетке қатысты зерттеулер көбіне мәтіндер мен сөзқолданыстың мәдени және семантикалық ерекшеліктеріне негізделген. Алайда, топтастырылған мәдени бірліктердің цифрлық кеңістікте сақталуы мен танытуы туралы арнайы зерттеулер әлі де аз. Сондықтан осы бағытта тереңірек зерттеулер жүргізіп, әдістемелік негіздерді кеңейту қажет. Мұндай жұмыстар этнолингвистикалық зерттеулердің сапасын арттырып, тіл мен мәдениеттің өзара байланысын жан-жақты ашуға мүмкіндік береді.

Зерттеу нәтижелері мен талқылау

Ұлт пен ұлтты бір-біріне жақындастыратын олардың ортақ қасиеттері ғана емес, негізгі ерекшеліктері. Сол ерекшеліктерді яғни құндылықтар жүйесінің аттарын мәдени реалилерді, салт-дәстүрді, ұлттық киім-кешегі мен тағамы, қолөнерге қатысты этнобірліктер цифрландырылды. Ұмытылмай үнемі елдің қолданысында, есінде қалу үшін этномаркерлерге арналған бейнебаян де әзірленді. Цифрлық платформада сақтаудың маңызы зор, себебі қазақ халқының дәстүрлі мәдениетімен танысқысы келетін адамға әлемнің кез келген бұрышынан қолжетімді.

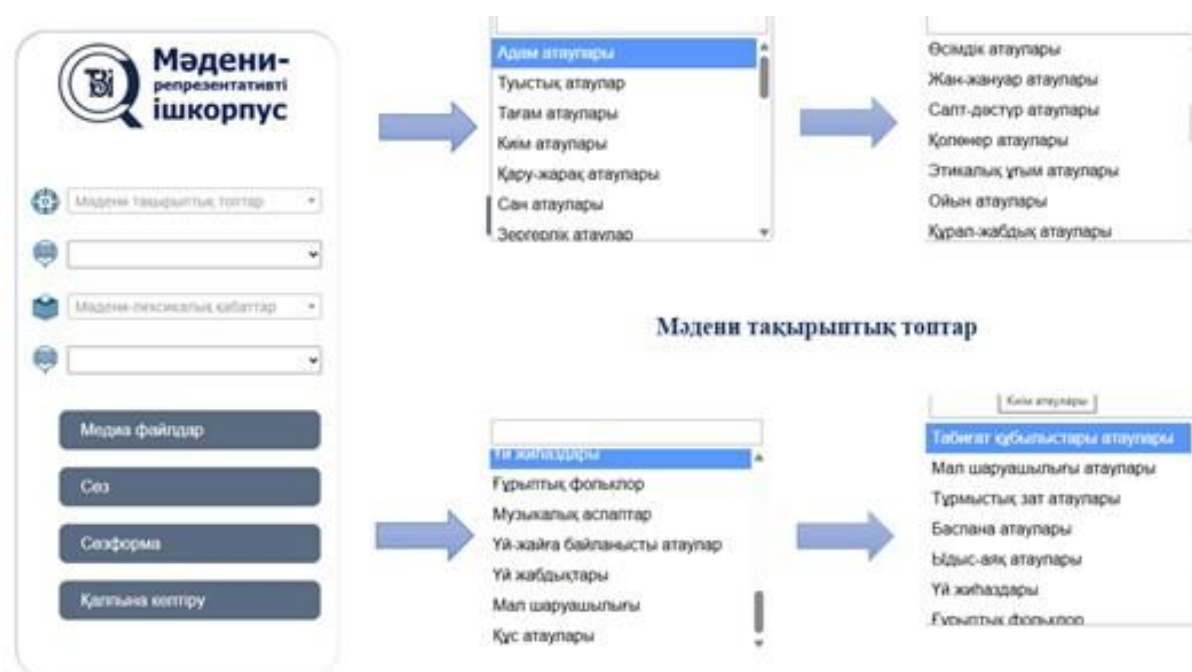
Мәдени-репрезентативті ішкорпустың құрылымы мен мазмұны, сондай-ақ оның қолданыс аясы зерттелді. Ішкорпус құрамындағы 15 миллион сөзқолданыс қазақ мәдениетінің түрлі аспектілерін қамтиды, оның ішінде әдебиет, тарих, этнография, халықтық дәстүрлер мен әдет-ғұрыптар бар. Ішкорпус мәтіндерінде қазақ халқының

тарихи мұралары мен мәдени құндылықтары толық әрі нақты көрсетілген. Сонымен қатар, 26 мәдени тақырыптық топ (1-кесте) пен 6 лексикалық қабат (2-кесте) арқылы іздеу мүмкіндігі пайдаланушыларға қазақ мәдениетін жан-жақты зерттеуге мүмкіндік береді.

Жеті тақырыптық топқа жататын тілдік бірліктердің лексикалық және мәдени-семантикалық мағыналары түсіндіріліп, ұжымдық «Қазақ тілінің мәдени-танымдық сөздігі» жарық көрді. Тақырыптық топтарға сәйкес сөздер мен сөз тіркестерінің саны төмендегідей:

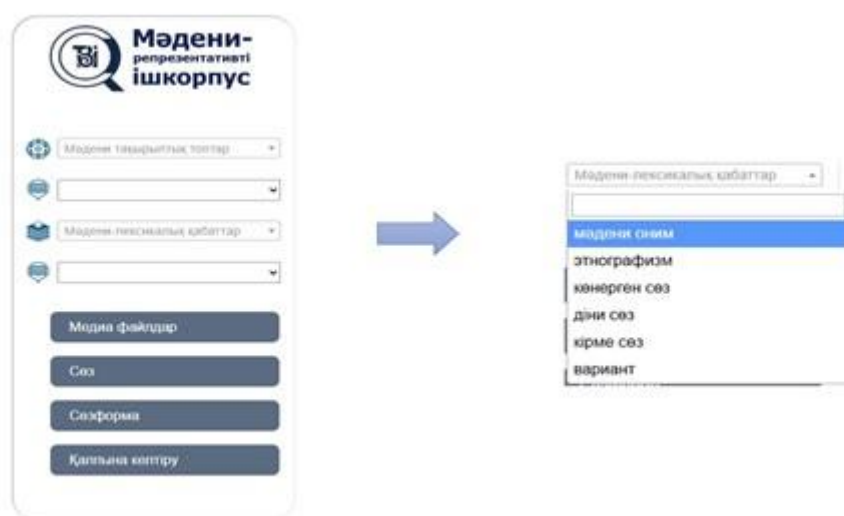
Адам атаулары – 59 сөз, 54 сөз тіркесі; Туыстық атаулар – 47 сөз, 73 сөз тіркесі; Тағам атаулары – 68 сөз, 139 сөз тіркесі; Киім атаулары – 65 сөз, 241 сөз тіркесі; Ыдыс-аяқ атаулары – 27 сөз, 21 сөз тіркесі; Қару-жарақ атаулары – 27 сөз, 32 сөз тіркесі; Сан атаулары – 9 сөз, 70 сөз тіркесі; Зергерлік атаулар – 28 сөз, 170 сөз тіркесі.

1-кесте. Мәдени тақырыптық топ



Сонымен қатар, мәдени-лексикалық қабатқа енген тілдік бірліктердің – мәдени онимдер (86 сөз), этнографизмдер (122 сөз), көнерген сөздер (111 сөз), діни сөздер (117 сөз), кірме сөздер (155 сөз) және вариант сөздер (162 сөз) – лексикалық және мәдени-семантикалық мағыналары түсіндіріліп, ұжымдық «Лингвомәдени сөздік» түрінде жарияланды.

2-кесте. Мәдени-лексикалық қабат



Ішкорпустың тағы бір ерекшелігі – оның визуалды мазмұны. Базада 145 танымдық бейнеролик және 700 сурет бар (3-кесте), олар қазақ халқының мәдениеті мен этнографиясын визуалды түрде танытуға арналған.

3-кесте – Медиа файл



Мультимедиялық жасақтамамен қамтамасыз етілген бейнероликтер қазақ тіліндегі этнографизмдердің лексикалық мағынасы және ашық/жабық кодта сақталған мәдени-семантикалық белгіленімі қойылған мәтіндер арқылы дайындалған. Бейнероликке

алынған мәтіндер жай ғана мәтін емес, мәдени-танымдақ ақпараттармен толыққан, бейнелік мәні зор, ақпаратты жылдам есте сақтауға көмектесетін және мәтінге қоса оның дыбыстық сапасының әуезділігі, құлаққа жағымдылығымен ерекшеленеді. Этномәдени бірліктердің иллюстрациялық базасын жинақтау жұмысы бойынша алдағы уақытта мағыналық құрамында мәдени-танымдық ақпараты бар этномәдени атаулар туралы белгілі этнолог-ғалымдардан аудио ақпараттар мен бейнероликтер түсіру жұмыстары жүргізіледі. Бұл мультимедиялық база мектеп оқушыларына, студенттерге, қазақ тілін ана тілі ретінде білетін және оны жаңадан үйренуші тілұстармандарға этномәдени құндылықтарды бейнероликтер арқылы тез қабылдауына, ұлттық сана-сезімін оятуға, мәдени құзыреттілігін арттыруға мүмкіндік береді.

Мәдени ақпарат жан-жақты қамтылған мәтіндер базасының IT-әзірлемесі пайдаланушының түрлі сұранысын қанағаттандыруға ыңғайланған іздеу жүйесімен қамтамасыз етілген. Әсіресе, бейнероликтердің білім беру және ғылыми зерттеулерде қолданылуы өте тиімді, өйткені олар этномәдени құндылықтарды тез әрі тиімді қабылдауға ықпал етеді. Ол мектеп оқушылары мен студенттер үшін этномәдени құндылықтарды бейнероликтер арқылы тез қабылдауға көмектеседі. Ішкорпус тек ғылыми зерттеушілер мен этнографтар үшін ғана емес, сондай-ақ мектеп оқушылары мен студенттер үшін де пайдалы ресурс. Оқушылар мен студенттер бұл ішкорпус арқылы қазақ халқының мәдениеті мен салт-дәстүрлерін көрнекі түрде зерттеп, тануға мүмкіндік алады.

Мәдени-репрезентативті ішкорпусқа берілген «қонақ» сөзінің лексикалық, мәдени-семантикалық мағынасы және тіркесімділік өрісінің берілу нұсқасы:

Мәтіндік базада қонақ сөзі – 4238 рет қолданылған.

ҚОНАҚ – жолаушылып келе жатып, үйге түскен таныс/бейтаныс жан; Сый-сияпат көрсету үшін арнайы шақырылған сыйлы адам; Ауыл арасын қыдыра келген адамдардың бірі деген ұғымның атауы. Адамдарды тіліне, дініне, ұлтына бөліп-жармай, достық ықылас көрсетіп, ас-дәм беру қонақжайлықтың негізгі белгісі. Дәстүрлі қазақ қоғамында әкенің балаға қалдыратын мұрасының белгілі бір бөлігі міндетті түрде қонаққа тиесілі енші деп есептелінген. Әрбір үй иесі келген қонақты барынша сән-салтанатымен, пейілімен қарсы алып, ақ тілеумен шығарып салуды өмірдің бұлжымайтын моральдық-этикалық қалып ретінде қараған. Қонақтың қазақ салтында құдайы қонақ, құтты қонақ, арнайы қонақ, ерулікті қонақ, қыдырма қонақ деп аталатын түрлері бар. Қазақ халқы – ежелден қонақ десе ішкен асын жерге қоятын халық. Үй иесі бұрын танысын, танымасын «Құдайы қонақпын» десе болды, жылы шыраймен қарсы алып отырған. Сапарда жүрген жолаушы «бөлінбеген еншісі бар» деп, ауылдың кез келген үйіне түсе берген. Қазақтың танымында «Қонақты құдай жібереді», «Қонақты қондырмасаң құт кетеді», «Құтты қонақ келсе, қой егіз табады, құтсыз қонақ келсе, қойға қасқыр шабады», «Қырықтың бірі Қыдыр келеді» делінген. Сондықтан халқымыз қонақты қарсы алу, шығарып салуға ерекше мән берген. Қонақты үй иесі, қарсы алып, есікті өзі ашып, үйге енгізіп, соңынан өзі кіріп есікті жабады. Бұл – қонақпен еріп келген «құт» бірге кірсін дегені. Ал қонақты шығарып саларда, есікті қонақтарға аштырып, соңынан өзі жауып шығатын болған. Келген қонақтармен жөн сұрасқаннан кейін үй иесі қонағына ең жақсы тамақтарын беріп, риза етуге тырысады (Қ.Ғабитханұлы).

Қонған шығар бұрын да талай қонақ, Бәріне де жетеді, асың мол-ақ. Кейін тағы келетін керуен бар, Жер босатам, бұл үйге кірсін сол-ақ (Ә.Найманбаев, Шығармалары). Аздан

соң қайтатұғын уақыт жетті, Бізге де қонақ бол деп тілек етті. Марияны қасына алып Троекуров, Қонаққа үш күннен соң, о да кетті (Ш.Құдайбердиев, Шығармалары). Балалар шулап бақырды, Ынтасы даусын білдірмек. Қонақтың күйін қашырды, Басты ауыртты көп дүрмек (Абай, Толық жинақ).

Атты қонақ. Атпен жүрген қонақ. Ертеде қазақтың қонағының көбі атпен жүргені белгілі. «Атты қонақ келсе, шөптің соры» деген тәмсіл сөз бар. Сондықтан өз малына жететін шөп жинап қана қоймай, қыста келетін қонақтардың ат-көлігі үшін де шөп жинайды. Қонақтың өзіне ғана емес, астындағы атының бабына да ерекше көңіл бөлген. Қонақтың аты аш қалса, жоғалып кетсе, ол қонақ иесі үшін ұятты құбылыс саналады.

Арнайы қонақ. Әдейі шақырылған ағайын-туыс, дос-жаран, көрші-қолаң, құда, сыйлы, құрметті адамдар. Арнайы қонаққа «көкқасқа тай», «ақсарбас қой» сойылады, мол дастарқан жайылып, түрлі тағамдар әзірлеп, күтіп алады. Беделі бар, әңгімешіл адамдарға қонақтармен бірге отыруға ұсыныс етіледі. Ауылдың өнерпаздары өнер көрсетіп қонақтардың көңілін көтереді.

Ерулікті қонақ. Тұрақты қоныста отырған ауылдың үстіне көшіп келген отбасын құрметтеп, ерулік асқа шақырылған қонақ. Жаңадан көшіп келген, жапсарлас қонған отбасына бұрыннан отырған көршілері қонақасы береді, дәм-тұз татырып, ықылас-пейіл білдіреді.

Құдайы қонақ. Шалғай елден, алыс жерден жолаушылап жүріп, сапарында көлігі болдырып немесе жолдан адасып тынығуға қонған тосын, танымайтын бөгде қонақ. Дәстүрлі қазақ қоғамында «Құдайы қонақымыз» деп жолда кездескен ауылға түссе, құдайы қонақты үйге түсіріп күтуге міндетті болған. Қолынан келген көмегін көрсетіп, тамағын беріп шығарып салып отырған. Егер үй иесі құдайы қонақты қондырмаса, дұрыс қарсы алмаса, қонақ ауыл ақсақалына, биге шағымданып, үй иесіне айып салдыра алған.

Құтты қонақ. Келген кезінде түскен отбасында бір жақсылық, қуаныш болған қонақ. Қазақ халқында: «Құтты қонақ келсе, қой егіз табады», – деген нақыл бар. Осылай деген ырымшыл қазақ қашанда барын дастарқанға салып, қабағын шытпастан қонағын қарсы алған. Жақын білісе келе таныстықтары тамырлыққа, одан кейін туысқандыққа немесе құдалыққа ұласқан. Осылайша тамырласқан адамымен ешқашан байланысын үзбеген.

Қыдырма қонақ. Туыс-туған, дос-жаран, құда-жегжаттарын арнайы іздеп барып аунап-кунап жататын қонақтар. Қыдырма қонақ – қазақтың көне жолында, қонақ болу салтында көп кезігетін қонақтың бір түрі. Мейрам-мереке, демалыс күндері ауыл азаматтары топ-топ болып, ауыл-ауылға, үй-үйге қыдырады, үлкендерге амандасады, арнайы сәлем береді. Қазақ оны үйіне келген қонақтарды бізді арнайы іздеді, елеп-ескеріп келді деп, құшақ жая қарсы алады, дәмді құжырын дастарқанға жаяды. Қазақтағы қыдырма қонақтың да өзіндік мәні, кісілік сәні бар. Олар жай қыдырып, қарын баққалы, жанын сақтағалы жүрген жоқ. Қайта ауыл-ауылдағы, үй-үйдегі көңілі жақын, ниеті бір, тілеуі қабыл адамдарға жүздесу, сәлемдесу, сырласу, ауырған-сырқағандары болса, көңілін сұрау, өлген-жеткендері болса, құран оқу, туыс-туғандарына көңіл айту, жақсы ниетпен тілдесу, береке бірлікті арттыру. Ол ғана емес, тірлікте, аз өмірде, саналы ғұмырда сапалы өмір сүру, көңіл тауып, мәнді-мағыналы тірлік өткізу болып табылады. Әуелі ауыл-ауылды, үй-үйдің берекеcін жалғайды, бір-біріне құрметін көрсетеді, сый-сыбағасын жалғайды, адамдық құнын, арын, адамгершілік, кісілік кереметін білдіреді. «Қыдырма қонақтан қырық ауыз сөз қалады, үлгі-өнеге, өшпес із қалады» деген сөз содан қалған.

Мұндай қонақтар жұрт сыйлаған беделді кісі немесе сары сүйек құда бастаған адамдар болса, онда қонақтың аз-көптігіне қарай жылқы баласынан – көк қасқа, қой баласынан – ақсарбас сойылса; ал, дос-жаран, алыс-беріс, барыс-келістері жиі адамдар болса, қонақтың аз-көптігіне қарай тай, қой сойып күтеді

Қылқыма қонақ. Қайсы ауыл, қайсы үйге қонақ түсіп, мал сойып, ет асып тамақ пісіргенін аңдып жүріп, дәл үстінен түсіп тамақ асырайтын қонақ. Мұндай қонақты қазақ халқы «шақырылмаған қонақ – жұт» деп жаратпайды. Алайда қонақты үйден қумайтын қазақы дәстүр бойынша, оларға да барын беріп, қондырып аттандырады.

Қорытынды

Мақалада қазақ тілінің ұлттық корпусының маңызды бөлімдерінің бірі – мәдени-репрезентативті ішкорпустың құрылымы мен маңыздылығын, сондай-ақ ұлттық мәдени тілдік бірліктерді цифрлық кеңістікте сақтап, танытудағы рөлі жан-жақты талданды. Қазақ халқының мәдени мұрасын цифрлық форматта жинақтап, ақпараттық қоғам жағдайында кеңінен таратуға бағытталған жұмыс тәжірибесі баяндалып, ҚТҰК-ның мәдени-репрезентативті ішкорпусы қазақ тілінің дамуындағы және жасанды интеллект жүйелерінде қолданылуындағы рөлі көрсетілді. Сонымен бірген, UNESCO-ның «Әлем жады» бағдарламасы аясында Қазақстанның материалдық емес мәдени мұраларын сақтау және цифрландыру жұмыстары қарастырылып, дәстүрлі салт-дәстүрлер, ұлттық өнер, қолөнер, мерекелер және олардың қазіргі заманға үйлесімділігі туралы айтылды. Сонымен қатар, дәстүр мен инновацияның өзара үйлесімін сақтаудың маңызы атап өтілді. Қорыта келгенде, бұл ішкорпус – қазақ халқының ұлттық құндылықтарын, салт-дәстүрін білетін рухани дүниесі бай тілдік тұлғаның қалыптасуына мұрындық болады. Ең маңыздысы қазіргі кітап оқу үрдісі тарылып бара жатқан ақпараттық қоғамда Мәдени-репрезентативті ішкорпус жалпыға қолжетімді цифрлық ресурс болуымен құнды. Ішкорпус – мектеп оқушыларына, студенттерге, магистранттарға этномәдени құндылықтарды бейнебаяндар арқылы тез қабылдауына, сондай-ақ өз халқының мәдениеті мен этнографиясын танып, білгісі келетін зерттеушілер мен этнографтарға арналған. Қазіргі уақытта ішкорпус үздіксіз толығып, жаңарып отырады, бұл оның болашақта қазақ халқының мәдени мұрасын әлемге танытуда үлкен рөл атқаруына ықпал етеді.

Мүдделер қақтығысы, алғыс айту және қаржыландыру туралы ақпарат

Мақала бойынша мүдделер қақтығысы жоқ.

Мақала ЖТН BR24993244 «Қазақ тілі ұлттық корпусын Smart-мәтіндер мегажобасы және қазақтілді жасанды интеллект негізі ретінде жетілдіру, ішкорпустарын әзірлеу» атты бағдарламалық-нысаналы зерттеу аясында жазылды

Авторлардың қосқан үлесі

Э. Өтебаева мақаланың негізгі идеясы мен тұжырымдамасын айқындап, зерттеу материалдарына талдау жүргізді және мақаланың негізгі бөлігін жазды. **Л.Қойшығұлова** тілдік материалдарды жинақтау, сұрыптау және талдау, мақаланы рәсімдеу жұмыстарын атқарды. **А.Нұстаева** тілдік материалдардың сандық деректерін жасауға үлес қосты. Жұмыстың барлық аспектілері бойынша авторлар тең дәрежеде жауапты бо-

лып табылады, зерттеу жұмысының тұтастығы мен деректердің дұрыстығы өзара талқыланып, пікір алмаса отырып жүзеге асырылды.

Әдебиеттер тізімі

Әшімбаева, Н.М., & Жаңабекова, А.Ә. (2023). Лингвомәдени білімдер жүйесін ұлттық корпусқа енгізуде көне атауларды зерттеу. *Tiltanym*, (3), 25–37 бб. <https://doi.org/10.55491/2411-6076-2023-3-25-37>

Жұбанов, А.Қ. (2016). Этномәдениеттік семантика белгіленімі «қазақ тілінің ұлттық корпусында» ескерілуі қажет. *Tiltanym*, (1), 24–28 бб.

Жұмабаева, Ж.Т., & Аманбаева, А.Ж. (2024). Лингвомәдени ішкорпус: діни сөздердің мәдени семантикасы. *Tiltanym*, (4), 147–154.

Манкеева, Ж.А., Шүленбаев, Н.Қ., & Бисенғали, А.З. (2023). Мәдени-репрезентативті ішкорпустарды әзірлеудегі мәдени бірліктерді аннотациялау проблемалары. *Tiltanym*, (3), 95–102 бб. <https://doi.org/10.55491/2411-6076-2023-3-95-102>

Өтебаева, Э.Ә., Исаева, Г.С., & Серікқызы, М. (2023). Лингвомәдени ішкорпусты әзірлеудегі мәдени-семантикалық. *Tiltanym*, (3), 135–143 бб. <https://doi.org/10.55491/2411-6076-2023-3-135-143>

Уәли, Н.М., Исаева, Г.С., Өтебаева, Э.Ә., Мамырбек, Г.М., Жұмабаева, Ж.Т., Сейітбекова, А.А., Аманбаева, А.Ж., & Қақабаев, Б.С. (2024). Лингвомәдени сөздік. Алматы: Ел-шежіре.

Фазылжанова, А.М. (2016). Национальный корпус казахского языка как инновационно-информационная база государственного языка (об особенностях и задачах первого этапа разработки и создания). *Alatoo Academic Studies*, (3), 103–109.

Шойбеков, Р.Н., Ғабитханұлы, Қ., Өтебаева, Э.Ә., Исаева, Г.С., Солтанбекова, А.Ә., Серікқызы, М., & Барменкулова, А.С. (2023). Қазақ тілінің мәдени-танымдық сөздігі. Алматы: Асыл ЖК.

Talgatqyzy, G., Amirbekova, A., Fazylzhanova, A., Gabitkhanuly, K., & Issayeva, G. (2024). Exploring the semantic significance ‘Camel’ and cultural representation in the Kazakh ethnic worldview as reflected in the national corpus. *Forum for Linguistic Studies*, 6(4), 268–276. <https://doi.org/10.30564/fls.v6i4.6986>

Tlegenova, G., Rysbergen, K., & Zhanabekova, A. (2024). Representativeness of mythonyms and theonyms in language corpora. *International Journal of Society, Culture, & Language*, 12(1), 374–395. <https://doi.org/10.22034/ijscsl.2024.2017262.3288>

Э.А. Утебаева^{*1}, Л.Р. Койшығұлова², А.Б. Нұстаева³

^{1,2,3}Институт языкознания имени А. Байтурсынулы, Алматы, Казахстан

Культурно-репрезентативный подкорпус: значимость сохранения национальных культурно-познавательных единиц в цифровом пространстве

Аннотация. В статье рассмотрены структура и значимость культурно-репрезентативного подкорпуса, а также его роль в сохранении и презентации национальных культурных и языковых единиц в цифровом пространстве. Электронный Национальный корпус казахского языка, сохраняющий культурное наследие казахского народа в цифровой среде и способствующий его широкому признанию и распространению в современном информационном обществе, является важной и надежной информационной базой для развития национального языка и

культуры. Культурно-репрезентативный подкорпус корпуса рассматривается не только как инструмент для глубокого изучения этнокультурного развития казахского этноса, его духовных и материальных ценностей, этнографических и культурно-познавательных представлений, но и как важный познавательный механизм, отражающий объединяющую роль языкового наследия. Цель статьи – на основе научных основ казахской этнолингвистики определить процесс создания культурно-репрезентативного подкорпуса в составе Национального корпуса казахского языка (НККЯ) и пути исследования и сохранения национальной культуры, традиций, исторических и социальных аспектов через цифровую платформу, аккумулирующую языковые и культурные ценности казахского языка. Также показан состав, структура и методы применения подкорпуса, а также доказана его значимость в образовании и научных исследованиях. Объект исследования – культурно-репрезентативный подкорпус в составе НККЯ и накопленные в нем языковые и культурные материалы казахской этнолингвистики.

Ключевые слова: национальный корпус казахского языка, культурно-репрезентативный подкорпус, цифровое пространство, культурно-семантическая информация.

E.A. Utebayeva^{*1}, L.P. Koishygulova², A.B. Nustaeva³
^{1,2,3}A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan

Cultural representative subcorpus: the importance of preserving national cultural and cognitive units in the digital space

Abstract. The article discusses the structure and significance of the cultural-representative subcorpus and its role in preserving and presenting national cultural and linguistic units in the digital space. The electronic National Corpus of the Kazakh Language is an important and reliable informational resource for the development of the national language and culture. The cultural-representative subcorpus within the corpus is considered not only as a tool for a deeper understanding of the ethnocultural development of the Kazakh ethnicity, its spiritual and material values, ethnographic and cultural-cognitive concepts, but also as an important cognitive mechanism reflecting the unifying function of the linguistic heritage. The aim of the article is, based on the scientific foundations of Kazakh ethnolinguistics, to define the creation of the cultural-representative subcorpus within the National Corpus of the Kazakh Language (NCKL) and to determine ways to study and preserve national culture, traditions, historical and social aspects through a digital platform that accumulates linguistic and cultural values of the Kazakh language. The composition, structure, and application methods of the subcorpus are also presented, and its importance in education and scientific research is demonstrated. The research object is the cultural-representative subcorpus within NCKL and the linguistic and cultural materials of Kazakh ethnolinguistics collected therein.

Keywords: National Corpus of the Kazakh Language, cultural-representative subcorpus, digital space, cultural-semantic information.

References

Ashimbaeva, N.M., & Zhanabekova, A.A. (2023). Lingvomadeni bilimder jujesin ulttyk korpuska engizude kone ataualardy zertteu [Study of ancient names in integrating the linguocultural knowledge

system into the national corpus]. Tiltanym, (3), 25–37. <https://doi.org/10.55491/2411-6076-2023-3-25-37> (in Kazakh)

Zhubanov, A.Q. (2016). Etnomadeniyyetlik semantika belgilenimi "qazaq tilinin ulttyk korpusynda" eskerilui kajet [Ethnocultural semantic markers should be considered in the "National Corpus of the Kazakh Language"]. Tiltanym, (1), 24–28. (in Kazakh)

Zhumabaeva, Zh. T., & Amanbaeva, A. Zh. (2024). Lingvomadeni ishkorpustardy dini sozderdin madeni semantikasy [Linguocultural subcorpus: Cultural semantics of religious vocabulary]. Tiltanym, (4), 147–154. (in Kazakh)

Mankeyeva, Zh.A., Shulenbayev, N.Q., & Bisengali, A.Z. (2023). Mädeni-reprezentativti ishkorpustardy äzirleudegi mädeni birlikterdi annotaciýalaý problemalary [Annotation issues of cultural units in developing culturally representative subcorpora]. Tiltanym, (3), 95–102. <https://doi.org/10.55491/2411-6076-2023-3-95-102> (in Kazakh)

Utebayeva, E.A., Issayeva, G.S., & Serikkizy, M. (2023). Lingvomadeni ishkorpustardy äzirleudegi madeni-semantikalık eresheilikter [Cultural and semantic features in the development of the linguocultural subcorpus]. Tiltanym, (3), 135–143. <https://doi.org/10.55491/2411-6076-2023-3-135-143> (in Kazakh)

Uali, N.M., Issayeva, G.S., Utebayeva, E.A., Mamyrbek, G.M., Zhumabaeva, Zh.T., Seitbekova, A.A., Amanbaeva, A. Zh., & Kakabayev, B. S. (2024). Lingvomadeni sozdik [Linguocultural Dictionary]. Almaty: El-shezhire. (in Kazakh)

Fazylzhanova, A. M. (2016). Nacional'nyj korpus kazahskogo jazyka kak innovacionno-informacionnaja baza gosudarstvennogo jazyka (ob osobennostjah i zadachah pervogo jetapa razrabotki i sozdaniya) [The National Corpus of the Kazakh Language as an innovative and informational basis for the state language]. Alatau Academic Studies, (3), 103–109. (in Russian)

Shoibekov, R. N., Gabitkhanuly, Q., Utebayeva, E.A., Issayeva, G. S., Soltanbekova, A. A., Serikkizy, M., & Barmenkulova, A. S. (2023). Kazak tilinin madeni-tanyndyk sozdigi [Cultural-cognitive dictionary of the Kazakh language]. Almaty: Asyl ZHK. (in Kazakh)

Talgatqyzy, G., Amirbekova, A., Fazylzhanova, A., Gabitkhanuly, K., & Issayeva, G. (2024). Exploring the semantic significance 'camel' and cultural representation in the Kazakh ethnic worldview as reflected in the national corpus. Forum for Linguistic Studies, 6(4), 268–276. <https://doi.org/10.30564/fls.v6i4.6986>

Tlegenova, G., Rysbergen, K., & Zhanabekova, A. (2024). Representativeness of mythonyms and theonyms in language corpora. International Journal of Society, Culture, & Language, 12(1), 374–395. <https://doi.org/10.22034/ijscsl.2024.2017262.3288>

Авторлар туралы мәлімет:

Өтебаева Эльмира Әбдіғалиқызы – хат-хабар үшін автор, филология ғылымдарының кандидаты, А. Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан. E-mail: elmira_abdigali@mail.ru, ORCID: 0000-0002-6500-6062.

Қойшығұлова Лаура Рахимжановна – PhD докторант, А. Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан. ORCID: 0009-0005-4340-0308.

Нұстаева Ардана Батырханқызы – филология магистрі, А. Байтұрсынұлы атындағы Тіл білімі институты, Алматы Қазақстан. E-mail: ardana.nustaeva@bk.ru, ORCID: 0000-0001-8810-3239.

Утебаева Эльмира Абдугалиевна – автор для корреспонденции, кандидат филологических наук, Институт языкознания имени А. Байтұрсынова, Алматы, Казахстан. E-mail: elmira_abdigali@mail.ru, ORCID: 0000-0002-6500-6062.

Койшыгулова Лаура Рахимжановна – PhD докторант, Институт языкознания имени А. Байтұрсынұлы, Алматы, Казахстан. ORCID: 0009-0005-4340-0308.

Нұстаева Ардана Батырханқызы – магистр филологии, Институт языкознания имени А. Байтұрсынова, Алматы, Казахстан. E-mail: ardana.nustaeva@bk.ru, ORCID: 0000-0001-8810-3239.

Utebaeva Elmira Abdugaliyevna – corresponding author, Candidate of Philology, A. Baitursynov Institute of Linguistics, Almaty, Kazakhstan. E-mail: elmira_abdigali@mail.ru, ORCID: 0000-0002-6500-6062.

Koishyghulova Laura Rakhimzhanovna – PhD student, A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan. ORCID: 0009-0005-4340-0308.

Nustaeva Ardana Batyrkhankyzy – Master of Philology, A. Baitursynov Institute of Linguistics, Almaty, Kazakhstan. E-mail: ardana.nustaeva@bk.ru, ORCID: 0000-0001-8810-3239.



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).