

Қ. Рысберген¹ , З. Садырбаева¹ , Д. Пашан¹ , Ә. Сейтбатқал^{1*} , Д. Садық¹ 

¹Ахмет Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан

*e-mail: arlen_98@mail.ru

ҚАЗАҚ ТІЛІНІҢ ҰЛТТЫҚ КОРПУСЫНЫҢ ОНОМАСТИКАЛЫҚ ІШКОРПУСЫН ӘЗІРЛЕУ ТӘЖІРИБЕСІ

Аннотация. Ономастикалық ішкорпус тіл білімі, география және когнитивистика ғылымдары үшін өзекті құрал болады. Ішкорпусты құрастыруға қазақ ономастикасы бойынша қорғалған диссертациялар мен ғылыми монографиялар арқау болды. Мақалада қазақ ономастикасының дамуына шолу жасалып, Ономастикалық ішкорпус құрастырудың маңызы мен әдістемелік технологиялары көрсетілді. Ішкорпусқа енгізілген оним түрлері мен оларға берілген сипаттамалар талқыланады. Ономастикалық ішкорпусқа енгізілген жалқы есімдерге берілген лингвистикалық ақпарат пен метаақпараттар сөз болады. Сондай-ақ ішкорпусты ғылыми және практикалық мақсатта да пайдалануға болатындығы негізделді. Ең алдымен түрлі жанрдағы мәтіндерде кездескен жалқы есімдердің мәні мен мазмұны мен қызметі ашылады. Екінші жағынан, ішкорпус үлкен сөздік ретінде пайдаланушыға өзінің туған жері, есімі, өсімдіктер мен жануарлар әлемі және т.б. туралы толыққанды мағлұмат береді. Ономастикалық атаудың бойында сақталған ақпараттарды жүйелеп беру – ономастикалық ішкорпустың негізгі мақсаты. Жалқы есімдер – қоғам өмірімен тығыз байланысты әлеуметтік құбылыс, идеологиялық қызметі басым тілдік бірлік. Әрбір оним белгілі бір кезеңге қатысты болып, тарихилық қасиетін сақтайды. Сол себептен де оның бойында тілден тыс географиялық, этномәдени, тарихи ақпараттар жинақталады. Осы мағлұматтарды жинақтап, жүйелеп, электронды түрде көпшілікке қолжетімді ету ономастикалық ішкорпус енісінде. Ішкорпусқа енгізілген жалқы есімдер сөзжасамдық, грамматикалық заңдылықтары тұрғысынан, сондай-ақ, семантикалық, этимологиялық, дұрыс жазылуы және т.б. қырынан сипатталады.

Тірек сөздер: Ономастикалық ішкорпус, корпуслық лингвистика, корпуслық ономастика, топонимдер, антропонимдер.

**Мақала ҚР БҒМ қаржыландырып отырған BR11765619 «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы» іргелі-зерттеу бағдарламасы аясында орындалды.*

Қ. Рысберген¹, З. Садырбаева¹, Д. Пашан¹, А. Сейтбатқал¹, Д. Садық¹

¹Институт языкознания имени А.Байтұрсынова, Алматы, Казахстан

*e-mail: arlen_98@mail.ru

ОПЫТ РАЗРАБОТКИ ОНОМАСТИЧЕСКОГО ПОДКОРПУСА НАЦИОНАЛЬНОГО КОРПУСА КАЗАХСКОГО ЯЗЫКА

Аннотация. Ономастический подкорпус актуален для ряда наук, особенно для лингвистики, географии и когнитивной науки. Базовой основой для разработки подкорпуса послужили защищенные диссертации и научные монографии, труды по казахской ономастике. В статье представлен обзор истории развития казахской ономастики, также рассматриваются роль и значимость создания ономастического подкорпуса. Обсуждаются типы онимов, включенных в подкорпус, и характерные черты, приписываемые им. В статье указывается, что в ономастический подкорпус включается лингвистическая информация и метаинформация имен собственных. Ономастический подкорпус может быть использован как в научных, так и в практических целях. Во-первых, раскрываются содержание и функции имен собственных, встречающихся в текстах разных жанров. С другой стороны, данный подкорпус в качестве объемного словаря предоставляет пользователю всестороннюю информацию о географии топонимического объекта, месте рождения, имени, флоре и фауне и т.д. Основной целью ономастического подкорпуса является систематизация разнообразной научной, культурологической информации, которая содержится в ономастическом названии. Имена собственные – социальное явление, тесно связанное с жизнью общества, в связи с чем в его семантике часто преобладает идеологическая нагрузка. Каждый оним относится к определенному периоду и сохраняет свой исторический характер. Поэтому он в себе хранит неязыковую информацию, а именно географическую, этнокультурную, историческую. Задача ономастического корпуса – собрать, систематизировать и сделать эту информацию общедоступной в электронном виде. Имена собственные, включенные в подкорпус, характеризуются с точки зрения словообразования, грамматических закономерностей, а также с точки зрения семантической, этимологической, правописания и т.д.

Ключевые слова. Ономастический подкорпус, корпусная лингвистика, корпусная ономастика, топонимы, антропонимы.

K. Rysbergen¹, Z. Sadyrbayeva¹, D. Pashan¹, A. Seitbatqal¹, D. Sadyk¹

A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan

*e-mail: arlen_98@mail.ru

EXPERIENCE OF CREATING THE ONOMASTIC SUBCORPUS OF THE NATIONAL CORPUS OF THE KAZAKH LANGUAGE

Abstract. The onomastic subcorpus is relevant for a number of sciences, especially for linguistics, geography, and cognitive science. The basic basis for the development of the subcorpus was the defended dissertations and scientific monographs and works on Kazakh onomastics. The article provides an overview of the history of the development of Kazakh onomastics and discusses the role and significance of the creation of the onomastic subcorpus. The types of onoms included in the subcorpus and the characteristic features attributed to them are discussed. The article points out that the onomastic subcorpus includes linguistic information and metainformation of proper names. The onomastic subcorpus can be used both for scientific and practical purposes. First, the content and functions of proper names found in texts of different genres are revealed. On the other hand, this subcorpus as a comprehensive dictionary provides the user with comprehensive information about the geolocation of the toponymic object, place of birth, name, flora and fauna, etc. The main purpose of the onomastic subcorpus is to systematize a variety of scientific, cultural information that is contained in the onomastic name. Proper names are a social phenomenon, closely connected with the life of society, due to which its semantics is often dominated by ideological load. Each onym refers to a certain period and preserves its historical character. Therefore, it contains non-linguistic information, namely geographical, ethno-cultural, and historical information. The task of the onomastic corpus is to collect, systematize, and make this information publicly available in electronic form. The proper names included in the corpus are characterized in terms of word formation, grammatical regularities, as well as in terms of semantic, etymological, orthographic, etc.

Keywords. Onomastic subcorpus, corpus linguistics, corpus onomastics, toponyms, anthroponyms.

Кіріспе

Бүгінде ғылым мен білімнің барлық салаларында ақпараттық технологияларды барынша молынан қолдану басты мәселелердің біріне айналып жатыр.

Цифрлық технологияларды ғылым саласына енгізуге мемлекет те мүдделі болып, қарқынды қолға алды. Осыған орай қазақ тілінің ҚР мемлекеттік тілі ретіндегі мәртебесін арттыру, оның әлеуметтік функцияларын күшейту және қолдану аясын кеңейту үшін оны әлемдік ақпараттық технологиялар кеңістігіне енгізу қажет.

Қазіргі заманауи ақпараттық технологиялар мен техникалық құралдар тілдік корпустың базасы негізінде лингвистикалық зерттеулерге жаңа мүмкіндіктер ашуда. Корпустық лингвистика – цифрлық кеңістікте пайда болып, дамып жатқан ғылым. Ахмет Байтұрсынұлы атындағы Тіл білімі институтының «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы» атты зерттеу бағдарламасы аясында қазақ тілінің ономастикалық ішкорпусы алғаш рет жасалып отыр. Корпуста қажетті ақпаратты іздеу жүйесі арқылы табуға болады. Мамандар мен әрбір қоғам мүшесі корпусты әртүрлі мақсатта пайдалана алады. Корпустың базасы әрдайым жетілдіріліп, толықтырылып тұрады.

Ономастикалық атаулар қоршаған ортаны игеру процесінде этностың білімі мен алуан түрлі ақпаратын жинаған мәдени сананың нәтижесі болып табылады. Сол онимдерді жинақтап, жүйелеп, цифрландырып көрсететін құрал – Ономастикалық ішкорпус (бұдан кейін ОІК).

Материал және әдістер

Корпустық лингвистика және географиялық ақпараттық жүйелер (ГАЗ) – тілді және тілді қолдануды, үйренуді, сондай-ақ географиялық дерекқорлардағы кеңістік моделдерін қолдануға арналған компьютерлік әдістемелерді пайдаланатын тәсілдер болып есептеледі. Корпус пен ішкорпуста ономастикалық атаулардың моделдерін табу үшін сөзді тегтеу, іздеу арқылы топоним, мифоним, антропоним, фитоним, зоонимдермен, олардың бірқатары прецеденттік атаулар т.б. шектеу үшін метабелгіленімдер технологиясын қолдану арқылы көрнекі түрде көрсетуге және талдауға болады.

Ономастикалық ішкорпусты зерттеу әдісі бірнеше қадамдардан тұрады. Бірінші кезекте корпуста сұраныс жасалады. Іздеу таңдалған параметрлерге байланысты жүзеге асырылады. Келесі қадам – ақпараттар базасын құру: алынған ақпараттар Excel форматындағы кестеге орналастырылады. Ол ақпараттар қолмен өңделеді. Деректер базасын қалыптастыру бойынша жұмыстар редакцияланады: олардың дұрыстығы тексеріледі. Өз кезегінде, сапалы талдау

зерттеушінің өзіндік ақыл-ой еңбегінің нәтижесі болып табылады. Корпусты зерттеу әдісі пайдаланылған корпусқа және оның техникалық іздеу ерекшеліктеріне байланысты. Қазіргі уақытта ономастикалық ішкорпусты жасаушылардың алдында іздеу нәтижелеріндегі белгілеудің дәлдігін арттыру міндеті тұр. Бұл жағдайда жұмыс істейтін әдістер салыстырмалы түрде қарапайым болуы керек.

Лексикограф-ономаст В.Н.Подольскаяның классификациясында ғаламның ономастикалық бейнесі тілдік тұрғыдан қарастырылса, Н.Ю.Шведованың жансыз табиғатты лексика-семантикалық, идеографиялық аймақтарға бөліп қарауы, жалпы «жансыз табиғат» концептосферасының құрылымын көрсетуінен – таза тілдік жіктеу мен концептуалдық аймақтарға бөлшектеу сынды екі тәсілдің ұштасуын көруге болады. Дегенмен, концептуалдық тұрғыдан қарастыру анағұрлым кеңірек, оның ішіне ономастикалық концептосфера да енеді.

Екі әдіс-тәсілге ортақ тек топонимдік нысаналарды В.Н. Подольскаяның ізімен ономастикада қалыптасқан төмендегідей негізгі топтарда қарастырамыз (Подольская, 1988: 17):

ороним – жер қыртысының бетіндегі, жер бедері (ландшафт, рельеф) шығыңқы (положительный рельеф) бедері → *тау, қыр, төбе, шоқы, дөңбел, жартаc, шың, адыр* т.б. жер қыртысынан төмен жатқан (отрицательный рельеф) бедері → *сай, шұңқыр, шыңырау, ойпат*, т.б. бейтарап, тегіс жатқан жер бедері → *жазық, шөл, дала* т.б.;

гидроним – *мұхит* немесе «су» концептісі → *теңіз, көл, өзен, қақ, бұлақ, жылға, қайнар, бастау, бөген, саз, батпақ*;

ойконим – елдімекен атаулары, *астионим* қала атауы, *комоним* кез келген ауыл типтес елдімекен;

хороним кез келген табиғи немесе жасанды (әкімшілік) аймақ, аудан, аумақтың, атауы;

Ономастикалық ішкорпус базасына енгізілетін ономастикалық бірліктерді онтологиялық топтастыру әдісіне сай топоним, антропоним, гидроним, ороним, ойконим, мифоним, теоним, агионим, фитоним, зооним т.б. негізгі топтары қамтылады. Олар метабелгіленім арқылы іздестіру жүйесінен негізгі Корпус пен ішкорпуста да табылатын болады. Көрсетілген әр терминнің мазмұнына қарай топоним – жер-су атаулары, антропоним – кісі есімдері, гидроним – су нысаны атаулары, ороним – жер бедеріне қатысты термин, ойконим – елді мекен, мифоним – мифтік мазмұндағы атаулар, теоним – ұлттық пантеондағы Құдай аттары, агионим – әулие, киелі атаулар, фитоним – өсімдік әлеміндегі нысан атаулары, зооним – мал-жануарлар, аң атаулары т.б.

Алдағы уақытта жалқы есімдердің басқа да түрлерін енгізу көзделіп отыр. Корпусқа жаңа оним түрлері қосылған сайын оның аудиториясы да кеңейе түседі. Мысалы, топонимдер тіл білімінің ғана емес, география ғылымының да зерттеу нысаны болып саналады. Сондықтан ОІК-та географ мамандарда пайдалана алады. Ішкорпустағы фитонимдер бөлігін биология ғылымының мамандары керегіне жаратады. Ал экзонимдер, шеттілдік ономастикалық атаулар, енгізілетін болса, ономастикалық ішкорпусты журналистер мен дипломаттардың қолдануына мүмкіндік туады. Бұдан ономастикалық ішкорпустың практикалық та маңызы басым екенін көреміз. Сонымен бірге, ОІК-та ономастикалық лексиканың үлкен қоры жинақталады. Жалқы есімдер жүйеленіп беріледі. Оларды тіркеу, сақтау және стандарттауға да үлкен септігін тигізеді. Сол арқылы ғылыми зерттеулер жүргізу мен танымдық ақпараттар алуға өз үлесін қосады.

Әдебиетке шолу

Корпус – қолданбалы тіл білімінің бір саласы болғанымен, оның пайда болуы тіл білімінің келесі деңгейге, жаңа бір кезеңге аяқ басқанын білдіреді. Өйткені, корпуста мәтіндердің үлкен қоры лингвистикалық ақпараттармен қамтамасыз етіліп, жіктеліп, жүйеленіп беріледі. «Корпус» компьютердің жадына электронды түрде енгізілген мәтіндер жинағы болып табылады.

Ахмет Байтұрсынұлы атындағы Тіл білімі институты әзірлеп отырған Ұлттық корпуста *Ахмет Байтұрсынұлы ішкорпусы, ауызша, диалектологиялық, мәдени, мақал-мәтел, тарихи, параллель, жарнама және ономастикалық ішкорпус* іске қосылды. Әр ішкорпус өзіне тән мазмұнымен, соған сәйкес белгіленім түрлерімен ерекшеленеді. Ономастикалық ішкорпусты басқа корпустардан даралайтын бірқатар өзгешеліктері бар. ОІК осы күнге дейін қалыптасқан академиялық қазақ ономастикасына табан тірейді. Ішкорпустың пайда болуы, қазақ ономастика ғылымының қалыптасып, дамуымен тығыз байланысты. Ономастикалық зерттеу нәтижелері ономастикалық корпусты құрастыруға негіз болады. Яғни қазірге дейін қалыптасқан, дамыған қазақ

ономастикасының бет-бейнесі ОІК-та көрініс табады. Сол себептен, ең алдымен қазақ ономастикасының қалыптасуына, дамуына шолу жасап өткенді жөн көрдік. Қазақ ономастикасы 1949-2023 жж. дейінгі аралықты қамтып, 74 жылдық тарихты басынан өткізді. Ономаст ғалымдар оны 3 кезеңге бөліп қарайды:

1) 1949-1970 жж аралығы. Бұл жылдары ономастикалық мектеп қалыптасты. Қазақ топонимдері мен антропонимдері зерттелді, олардың этимологиясы, құрылымдық және лексика семантикалық топтары анықталды; топонимиялық жүйенің стратиграфиясы айқындалды, географиялық терминдердің топонимдер жасаудағы орны мен рөлі анықталып, зерттелді.

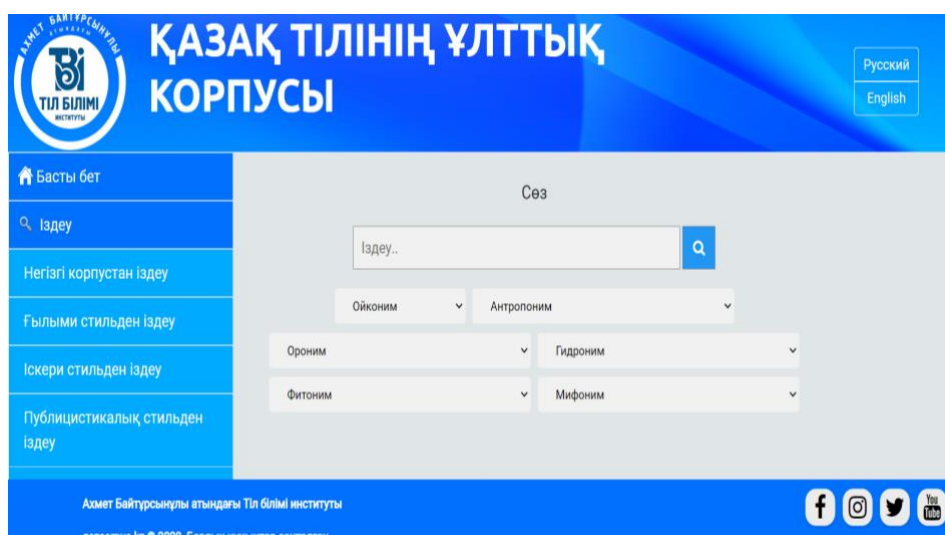
2) 1970-1995 жж. арасы. Бұл кезеңде (ономастикалық) аймақтық топонимиялық зерттеулер дами түсті. Ономастикалық теорияларға назар аударылды. Топонимдер мен антропонимдер этнолингвистикалық бағытта зерттеле бастады. Ономастикалық зерттеулердің ауқымы кеңейді. Космонимдер, зоонимдер, Қазақстан территориясындағы орыс топонимдері зерттелді; ономастикалық қолданбалы аспектілер қарастырыла бастады; транскрипция мен транслитерация мәселесі жетіле түсті. Ономастикалық сөздіктер құрастырылды.

3) 1995-тен қазірге дейін. Бұл кезеңде аймақтық топонимиялық зерттеулер өз жалғасын тапты. Сонымен қатар ұлттық ономастикада антропоэзектік парадигмаға сай когнитивтік зерттеулер қарқынды дами бастады. Ономастикалық зерттеулердің бұрын зерттелмеген, шеткері жатқан аймақтары қамтылып, эргоним, идеоним, мифонимдер (агиотопонимдер, космонимдер, фитонимдер т.б.) қарқынды зерттеле бастады.

1949-2005 жж. аралығында ономастикадан Қазақстанда 50-ге жуық кандидаттық және докторлық диссертациялар қорғалды (Т.Жанұзақ, А.Әбдірахманов, Е.Қойшыбаев, В.Н.Попова, О.А.Сұлтаньев, Е.Ә.Керімбаев, В.У.Махпиров, Б.Бектасова, М.Қожанов, Г.Б.Мәдиева, Қ.Рысбергген, А.Жартыбаев, Ғ.Сағидолда, Б.М.Тілеубердиев, К.М.Головина, М.А.Диарова, С.Иманбердиева, С.Е.Керімбаева М.Е.Какимова т.б. (Тілеубердиев, 2006: 8).

Нәтижелер және талқылау

Қазақ тілінің ұлттық корпусында жұмыс істеу барасын көрсетсек (1-сурет). Алдымен <https://qazcorpus.kz/news/> сайтына кіресіз. Корпус ұлттық корпус болғандықтан ақпарттар қазақ тілінде берілген. Басты бетте «Іздеу» тармағы тұрады, оны басссаңыз ұлттық корпусының ішіндегі 10 субкорпусы көрсетеді. Солардың қатарында «Ономастикалық ішкорпус» тұрады. Сол паракшаға өтсеңіз, төмендегі бетті көресіз.



1-Сурет – Қазақ тілінің ұлттық корпусындағы Ономастикалық ішкорпусының беткі беті

Ұлттық корпусының бір бөлігі саналатын ОІК-та ономастикалық атаулардың бірнеше түрі енгізілді:

- Топонимдер:
 - ороним,
 - гидроним

- ойконим
- Антропоним;
- Мифоним;
- Фитоним;

Ономастикалық ішкорпустың беткі бетінде осы оним түрлері берілген. Мысалы, оронимнің ішінен «Алатау» деген онимді іздесек, корпус мәтінінен 501 қолданысты көрсетіп отыр. Батырманы оронимнің үстіне апарсақ, лингвистикалық белгіленімі көрсетіледі (2-сурет). Ал шығарма авторының жанына апарғанда, автор мен шығарма туралы толық ақпаратты беріледі (3-сурет).

Табылған құжаттар саны 501 - Алатау

Ақұштап Бақтыгереева "Өлең жазарда"

Бүлт кешіп **Алатаудың** басында анау, **Ақ жауын** Алматыда ау ашылмады.

Кекілбаев С

Алатаудың

Әбділда Тә

Алатау қауы

Әбділда Тә

Тап кеше Ал

Әбділда Тә

Сом балға

Әбділда Тә

Алатау түр

Оним атауы	Алатау
Оним түрі	топоним
Шартты белгісі	t.topon
Топоним түрі	ороним
Шартты белгісі	t.oron
Орфографиялық норма	Алатау
Орфограмма (бейнорма)	жоқ
Сөзжасамы	Біріккен сөз: Ала + тау
Геоақпарат	Алатау – тау, басы Хантәңірі тауынан басталып, аяғы Шымкентке дейін созылып жатқан асқар биік, заңғар тау.
Мәдени семантика	Алатау – Түркілер ежелгі заманнан-ақ биік шыңды, басында мәңгілік мұздар мен аппақ қар жататын тауларды Алатау деп атағаны Кузбас Алатауы, Жоңғар Алатауы, Іле Алатауы, кыргыз Алатауы, Тяньшань Алатауы атауларынан белгілі. Осы атаулардағы ала сөзінің төркіні, қазіргі қазақ тіліндегі "ақ пен көктің аралас келген түсіне /орысша пестрый/ катысты емес, ұлу немесе одан да көнелеу оло тұлғасында артатындай" (Е. Қойшыбаев, 1985, 42). Бұған қарағанда, Алатау "ұлы, аса зор, биік тау" деген мағынаға ие екені байқалады.

2-Сурет – Онимнің лингвистикалық белгіленімі

Ақұштап Бақтыгереева "Өлең жазарда"

Автордың аты-жөні	Ақұштап Бақтыгереева
Стиль	көркем әдебиет
Мәтін аты	Өлең жазарда
Таралым типі	ғаламтор-ресурс
Мерзімі (жазылған уақыты)	
Дереккөзі	www.adebiportal.kz
Мәтін формасы	жазбаша
Мәтін тақырыбы	
Аудитория жасы	бейтарап
Субкорпус	көркем әдебиет
Мәтін типі	
Мәтін көзі	интернет-ресурс
Мәтін хронотопы	
Автордың жынысы	әйел
Автордың туған уақыты	1944
Сөзформа саны	75
Жарық көрген уақыты	
Корпусқа енгізілу уақыты	

ағындағы Тіл білімі институты

3-Сурет – Корпуста көрінетін метабелгіленім қойылған ақпарат

Ономастикалық ішкорпус төмендегідей белгіленім түрлерімен қамтамасыз етілді:

а) *Шартты белгілер*. Жалқы есімнің түрлеріне сәйкес шартты белгілер қойылды. Мысалы, топоним – t.topon, антропоним – a.antropon, фитоним – f.fiton, мифоним – m.mifon және т.б.

ә) *Орфографиялық норма және бейнорма*. Жалқы есімнің дұрыс жазылуы нормасы мен бейнормасы осы бөлікте қамтылды.

- Орфографиялық норма: Қобыланды, Райымбек, Қайсар.
- Бейнорма: Қобланды, Раимбек, Раймбек, Хайсар.

б) *Сөзжасамдық белгіленім*. Әр атаудың құрамы осы сөзжасамдық белгіленім бөлігінде беріледі. Географиялық атауларды жасауға ұйытқы болған грамматикалық формалар мен сөздер көрсетіледі. Жалқы есімдердің тарихилық сипаты да осы сөзжасам бөлігінде айқын көрінеді:

- Түбір сөз: Қаба, Обаған, Жем, Сағыз
- Туынды сөз: Тас+тақ, Мойын+ты, Шідер+ті, Қарқара + лы.
- Біріккен сөз: Есік+көл, Жаман+көл, Қоржын+көл, Арық+балық.
- Күрделі сөз (Сөз тіркесі): Тұлын Қаба өзені, Шалым-Тарым өзені.

в) *Геоақпарат*. ОІК-тағы ақпараттарды шартты түрде лингвистикалық және экстралингвистикалық деп бөлуге болады. Әсіресе топонимдердің түрлеріне келгенде тілдік емес географиялық ақпараттар берілді. Бірінші, географиялық нысанның орналасқан жері (геолокация) – геоақпарат бөлігіне орналастырылды. Мысалы, *Алтай – терістік жағы Ресей, шығысы Қытай, оңтүстік-шығысы Қазақстан Республикасындағы биік таулы аймақ. Ресейдің Алтай өлкесі Таулы Алтай автономиялық облысын қамтиды* (<https://qazcorpus.kz/>). Сондай-ақ, топонимдер үшін төмендегідей ақпараттарды беру маңызды болып саналады:

- Оронимдер мен гидронимдер үшін географиялық нысандардың биіктігін, ұзындығын, көлемі және т.б. сипаттамаларының берілуі;

- Ойконимдер мен урбанонимдер үшін маңызды саналатын хронологиялық ақпараттар (Мадиева, 2004: 170).

г) *Мәдени-танымдық ақпараттар*. Бұл белгіленім түрінде топонимнің шығу тарихы (этимологиясы), жер-су атауларының неліктен осылай деп аталғандығы жайында мәліметтер қамтылды. Мысалы, *Алтай оронимі (тау аты) туралы көптеген пікір бар. Мәселен, В.В. Радлов ал (биік) + тайга (орманды тау) деген сөздерден шығарса, Э.М. Мурзаев, В.А. Никоновтар «ала тау» мағынасымен байланыстырады. М. Ряснен ал+той «алып тау» десе, Е.Жанпейісов ал сөзі «үздіксіз, тізбектелген» мағынасын береді дейді. Ғалымдардың көпшілігі бұл атаудың негізінде алтын сөзінің жатқанын мойындайды (Б.Я. Владимирцов, Э.В. Севортян, Г. Дерфер, Г. Рамстедт, Г. Вамбери, А. Щербак, А. Әбдірахманов, Ғ. Қоңқашипаев, Т. Жанұзақов, т.б.). Оның ішінде Ғ. Қоңқашипаевтың келтірген дерегі дұрыс деп есептейміз. Моңғ. алт (алтын)+тай (туынды сын есім жұрнағы), сонда Алттай>Алтай «алтынды» болып шығады. Біздің қосарымыз академик В.В. Бартольдтың «Алтай оронимі кейін шыққан тәрізді» деген пікіріне сүйеніп, Алтай атауының көне түркі тілінен моңғол тіліне аударылған атау екенін ескерту. Өйткені, Орхон жазбаларында Алтын йыш (732-735 жж.), Қытайдың «Таншу» жылнамасында Алтынтаг (647 ж.), Моңғолдың «Құпия шежіресінде» Алтынтау (XIII ғасырдан бері) түрінде көрсетіледі де, Алтай атауы кездеспейді. Түркілік атауды моңғолдар жаулап алғаннан кейін, өз тілдеріне аударып атаған тәрізді. Ал мағынасы сақталып қалған. Жаулап, отарлап алған елдердің тілі келмегендіктен, немесе әдейі саясатпен жер-су аттарын өзгертуі тарихта бар оқиға (Бияров, 2002: 28).*

Аталған белгіленімдер пайдаланушыларға тілдің лексика-семантикалық, мәдени-семантикалық жүйесіне қатысты көлемді ақпарат алуға мүмкіндік береді. Мәдени-танымдық белгіленім қазақ тілін ана тілі немесе шет тілі ретінде үйренетіндерге, сондай-ақ дәстүрлі ұлттық білімдер жүйесіне қызығушылық танытқандарға, жастарға қазақ онимдерінің мәдени семантикасын, ұлттық кодты меңгеруге көмектеседі.

д) *Метабелгіленім*. Қазақ тілінің ұлттық корпусы сайтында қазақ тілінің мәтіндік қоры электронды түрде сақталады. Іздеген адам мәтіндер базасынан оним (оның түрлері) бойынша іздеу жүргізіп, іздеген атау қолданылған сөйлемдер тізімін және олардың дереккөзін таба алады. Табылған оним немесе оним түрі туралы мәтіндік қордағы барлық стиль бойынша ақпарат беріледі.

Жалпы, мәтін – сөздік қордың көрсеткіші. Қазақ тіліндегі мәтінді зерттеуші А.Шайкенова: «Мәтін – коммуникативтік бағыттағы, нақты бір ақпаратты өз бойына жиган, әрі өзгеге жеткізе отырып, белгілі бір уақыт пен кеңістіктен мәлімет (дерек) беретін, жеке тұлғаның рухани әлемін ашуға ықпалы зор күрделі тілдік таңба» (Шайкенова, 2012: 8) деген болатын.

Корпустың тірегі – мәтін. Мәтін сапалы, көп болған сайын корпус та жақсы болады. Корпусқа кез-келген мәтінді ала алмаймыз. Қазақ тілінің ұлттық корпусының негізін салған, филология ғылымдарының докторы, профессор Асқар Құдайбергенұлы Жұбанов: «Корпус термині әдетте

белгіленім жүргізілген көлемді мәтіндер жиыны дегенді білдіреді. Корпуста қамтылатын мәтіндер көбінесе жазба тілді негізге алады, яғни проза, поэзия т.б. стильдер қарастырылады. Алайда ұлттық корпустар белгілі бір тілдің болмысын түгел танытуы қажет. Корпусқа мәтіндер белгілі бір жүйе бойынша енгізіледі. Корпус мәтіндері жасалып отырған корпусстың мақсатына сай белгілі бір стильдер мен жанрлар немесе кезеңдер бойынша теңгерімді мөлшерге сай болуы қажет. Яғни мәтін де жасалып отырған корпус түріне қарай сәйкестендіріледі» (Жұбанов, 2017: 51) – деген болатын.

Осы орайда Ресей ғалымы В.В. Мамонтованың мына ойымен келіспеске болмайды: «В среде учёных появилось понимание того, что ряд корректных лингвистических исследований возможно провести только на большом речевом материале» (Мамонтова, 2010: 232) деген. Шынында да, корпус жұмысының нәтижесі үлкен көлемді мәтіндерден көрінеді.

Ономастикалық ішкорпус мәтініне алынатын мәтін де сараланып алынады. Бұл турасында зерттеуші, жас ғалым А.Қожахметова корпусстың зерттеу мақсатына қарай көлемді мәтіндер іріктеліп алынатындағын айтады. Корпус дегеніміз компьютер жадына сақталған көлемді мәтіндер жинағы, ал корпусстық лингвистикада ауқымды электронды мәтіндерді компьютердің көмегіне сүйене отырып зерттейді. Сол себептен жас ғалым корпусстық лингвистика мен мәтін лингвистикасының өзара тығыз байланысты екеніне назар аударады (Қожахметова, 2015: 303).

Енді ономастикалық ішкорпусстың мәтіндік базасына қандай еңбектер алынды деген мәселеге келсек, олар:

1. Қазақ ономастикасына қатысты жазылған ғылыми еңбектер, монографиялар, оқу құралдары (Т.Жанұзақ Қазақ ономастикасы. Бес томдық. Нұр-Сұлтан, 2021, т.б.)
2. Қазақ тілінің ономастика саласы бойынша қорғалған докторлық, кандидаттық диссертациялар (Б.Бияров, Ж.Исмаилова, А.Жартыбаев, Б.Тасболатов, т.б.)
3. ҚР әкімшілік аумақтарына қатысты анықтамалықтар, энциклопедиялық анықтамалықтар (Жеріңнің аты – Еліңнің хаты. Энциклопедиялық анықтамалық. Алматы: Аруна, 2006. – 808 б., т.б.)
4. Ономастикаға қатысты ақпараттар берілген мәдени-танымдық еңбектер (Қайдар, Ә. Қазақтар ана тілі әлемінде: этнолингвистикалық сөздік. Үш томдық. Адам. Қоғам. Табиғат. – Алматы: Сардар, 2013, т.б.).
5. Ономастика басылымдары мен хабаршылары (Ономастикалық хабаршы, т.б.)
6. М.Әуезов, Ғ.Мүсірепов, С.Сейфуллин, О.Бөкей, Қ.Жұмаділов және т.б. жазушылардың шығармалары («Абай жолы», «Тар жол тайғақ кешу», «Соңғы көш» және т.б.).
7. Қазақ тіліне, ономастика қатысты сайттар (atau.kz, tilalemi.kz, abai.kz, т.б.).

А.Қ.Жұбанов пен А.Ә.Жаңабекова аталған еңбегінде: «Корпустардың қай түрінде болмасын мәтіндер теңгерімділігі сақталады, яғни енгізілетін мәтіндер бейберекет алынбайды. Белгілі бір стиль, жанрдағы т.б. сипаттары бойынша мәтіндер көлем жағынан мүмкіндігінше біркелкі алынады. Өртүрлі жанрдағы мәтіндердің бірінің көлемі аз, бірінің көлемі үлкен болуы белгілі сандық, сапалық мәліметтер алуға кедергі келтіреді. Сондықтан әртүрлі сипаттағы мәтіндер теңгерімділігі сақталуы қажет. Демек, корпус жасауда енгізілетін мәтіндердің көлемі айқын болуы шарт» (Жұбанов, 2017: 52) дегенді айтты.

Бұл турасында «Коллекция текстов должна быть сбалансированной и иметь достаточную выборку по числу текстов и авторов, чтобы служить основой для статистический достоверных исследований лингвистических феноменов» (Шаров, 2003: 10) деп, орыс ғалымы да айтқан болатын.

Корпуста «метабелгіленім» термині қолданылады. Белгіленім корпустан іздеген сөзіңіз туралы толық ақпаратты көрсетеді. Мәтін метабелгіленімі іздеген атаудың қандай мәтіннен алынғанын, мәтіннің авторын, автордың жынысын, мәтіннің тақырыбын, жарияланған мерзімін, дереккөзін, стилін, таралу типін, т.б. көрсетеді. Корпус сайтына енгізілген мәтінге метабелгіленім қойылмаса, мәтін танылмайды. Ономастикалық ішкорпусстың метабелгіленімі төмендегі кестеге сәйкес қойылды (1-кесте).

1-Кесте – Метабелгіленімі қойылған еңбек

Мәтін аты:	Қазақ ономастикасы
Мәтін авторы:	Т.Жанұзақ
Автордың жынысы:	Ер
Автордың туған жылы:	1927
Шығарманың жарық көрген уақыты:	2021
Аудитория жасы:	Бейтарап
Мәтін формасы:	Жазбаша
Мәтін стилі:	Ғылыми-көпшілік
Мәтін типі (жанры):	Монография
Мәтін графикасы:	Кирил жазуы
Мәтін тақырыбы:	Ғылыми-танымдық
Мәтін көзі:	Электронды
Мәтіннің корпусқа енгізілу уақыты:	2023
Дереккөзі:	Т.Жанұзақ. Қазақ ономастикасы. Бес томдық. – Т.1. Нұр-Сұлтан. Халықаралық Түрік академиясы, 2021 ж.
Белгілеуші:	Пашан Д.М.
Ішкорпус:	Ономастикалық
Сөзформа саны:	117418

Біз жоғарыда көрсеткен, ономастикалық ішкорпусқа алып отырған мәтіндер де таңдалып, сараланған мәтіндер. Алынған мәтіндер ономастикалық ішкорпустың мақсаты мен іздеген адамның қажеттілігін толық қамтамасыз етуі тиіс деп білеміз.

Әзірленген әрбір ішкорпустың метамәтіндік белгіленімі мына ақпараттарды қамтиды: мәтін авторы, аудармашы, мәтіннің аударылған күні, дереккөзі, аудиожазбаның уақыты, тарихи мәтіннің жазылу уақыты, типі, стилі, енгізілген мерзімі және т.б. Пайдаланушы осы белгіленімдер арқылы қажетті ақпаратты ономастикалық ішкорпустан таба алады. Мәтіннің стилі, авторы, мәтіннің шыққан жылы және т.б. бойынша әзірленген параметрлер пайдаланушыға белгілі бір стиль бойынша зерттеу жүргізуге, белгілі бір автордың шығармашылығын, лингвистикалық мұрасын зерттеуге, сөздіктерге дәйектемелік мысалдар табуға мүмкіндік береді.

Қорытынды

Мақалада ономастикалық ішкорпусты құрастырудың негіздемесі мен тәжірибесі айтылды. Ішкорпусқа енгізілген оним түрлері, оларға жасалған белгіленім түрлері талқыланды. Жалқы есімдердің грамматикасы, орфографиясы, географиясы (топонимдерде) және мазмұны ашылып, сипатталады. Бүгінде ОІК-қа антропонимдер мен топонимдер енгізіліп, оларға тілдік белгіленім жасалды. Ономастикалық ішкорпусқа базасын қазақ ономастикасы бойынша қорғалған диссертациялар, жарық көрген монографиялар мен энциклопедиялар және көркем шығармалар құрайды. Мақалада осы дереккөздерге жасалған метабелгіленімнің үлгісі берілді. Бұл белгіленім түрі негізгі корпус пен барлық ішкорпустарға ортақ тақырыпшалардан тұрады.

ОІК-тағы сөзжасамдық белгіленімде атаудың құрамы, ондағы көне элементтер (өлі жұрнақтар) және басқа да қосымшалар көрсетілді. Келесі белгіленім түрі геоақпаратта атаудың нысанмен байланысы, яғни мекен-жайы берілді. Осы екі бөлімнен алынған мағлұмат мәдени ақпараттар бөлімінде тереңдей, толығы түседі. Бұл,

біріншіден, сөзжасам бөлігінен алынған ақпарат ары қарай этимологиясын ашуға негіз болады;

екіншіден, геоақпарат бөлігінде берілген түсіндірмелер атаудың бойында кодталған ақпараттармен байланысты болғандықтан, бұл онимнің мәдени семантикасын ашуға, ондағы жабық кодтарды анықтауда пайдалы болмақ. Бұл жалқы есімдерге жасалған белгіленімдер арасындағы сабақтастықты білдіреді;

үшіншіден, онимдердің дұрыс жазылу нормасы мен бейнормасы, қолданушыға атаудың қалыптану жайлы мағлұматтар береді.

Корпустың мүмкіндіктерін талдау, атап айтқанда, ономастикалық ішкорпус мысалында талдау, корпусты жасау тәжірибелі лингвистің сапалы талдауымен ақпараттарды берудің автоматты

жүйесін біріктірген кезде нәтижелі болатындығын көреміз. Сапалы талдау сенімді корпустың белгісі болып табылады.

Ономастикалық ішкорпусты құрастыру қазақ ономастикасының мүмкіндіктерін кеңейтеді. Өйткені ішкорпус ономастикалық материалдардың үлкен электронды қорына айналмақ. Бұған дейінгі кезеңдегі зерттеушілердің материал көлеміне қарағанда корпус әлдеқайда мол материал ұсыну мүмкіндігіне ие. Сондықтан Ономастикалық ішкорпусты құрастыруға отандық ономаст ғалымдар мен туған жер атауларына бей-жай қарамайтын барлық қоғам мүшелерінің де үлес қоса алатынын құлаққағыс етеміз. Ономастикалық ішкорпусты түзу барысында кем-кетік кетпеді дей алмаймыз. Алғаш рет жасалып отырған корпус болғандықтан, қиындықтары да бар. Алдағы уақытта жетілдіру мен дамыту жұмыстары тоқтамай жүргізілетін болады. Мәтіндер базасын түзудің бірқатар қиындықтары болғанымен, мәселелерді шешуге мүмкіндік бар.

Әдебиеттер

- Тілеубердиев Б.М. (2006) Қазақ ономастикасының лингвокогнитивтік аспектілері. – Алматы: Арыс, – 280 б.
 Подольская Н.В. (1988) Словарь русской ономастической терминологии. – Москва: Наука, – 192 с.
<https://qazcorpus.kz/>. Қазақ тілінің ұлттық корпусы
 Мадиева Г.Б. (2004) Имя собственное в контексте познания. – Алматы: Қазақ университеті, – 190 с.
 Бияров Б. (2002) Өр Алтайдың жер-су аттары. – Алматы, – 180 б.
 Шайкенова А.Ж. (2012) Көркем мәтін және дискурстық талдау. Филология студенттеріне арналған оқу құралы. – Петропавл.
 Жұбанов А.Қ., Жаңабекова А.Ә. (2017) Корпустық лингвистика. – Алматы: Қазақ тілі, – 336 б.
 Мамонтова В.В. (2010) Корпусная лингвистика в современной языковедческой парадигме Актуальные вопросы современной науки, № 12, с. 231-235.
 Қожахметова А. (2015) Көркем мәтін прагматикасын корпустық лингвистика кеңістігінде зерттеу ҚазҰУ хабаршысы Филология сериясы. №6 (158), – 300-305 бб.
 Шаров С.А. (2003) Представительный корпус русского языка в контексте мирового опыта Научно-техническая информация – Серия 2 Информационные процессы и системы, № 6, С. 10.

References

- Tileuberdiev B.M. (2006) Qazaq onomastikasynyñ lingvokognitivtik aspektileri. Almaty: Arys, – 280 b. [Linguocognitive aspects of Kazakh onomastics. – Almaty: Arys, – 280 p.] (in Kazakh)
 Podolskaya N.V. (1988) Slovar russkoj onomasticheskoy terminologii. M.: Nauka, – 192 s. [Dictionary of Russian onomastic terminology. – Moscow: Nauka, – 192 p.] (in Russian)
<https://qazcorpus.kz/>. Qazaq tiliniñ ulttyq korpussy [National Corpus of the Kazakh language] (in Kazakh)
 Madieva G.B. (2004) Imia sobstvennoe v kontekste poznaniya. – Almaty, Qazaq universiteti, – 190 s. [Proper name in the context of cognition. – Almaty: Kazakh University, – 190 p.] (in Russian)
 Biarov B. (2002) Ör Altaidyñ jer-su attary. – Almaty, – 180 b. [Toponyms of great Altai. – Almaty, – 180 p.] (in Kazakh)
 Shaikenova A.J. (2012) Körkem мәтін және diskurstyq taldaу. Filologiya studentterine arnalğan oqu quraly. – Petropavl. [Literary text and discursive analysis. Manual for philology students. – Petropavlovsk] (in Kazakh)
 Jubanov A.Q., Jañabekova A.Ä. (2017) Korpustyq lingvistika. – Almaty: Qazaq tili, – 336 b. [Corpus linguistics. – Almaty: Kazakh language – 336 p.] (in Kazakh)
 Mamontova V.V. (2010) Corpusnaya lingvistika v sovremennoj yazykovedcheskoj paradihme Aktualnie voprosi sovremennoj nauki, №12, s. 231-235. [Corpus linguistics in the modern linguistic paradigm Actual issues of modern science, No. 12, pp. 231-235.] (in Russian)
 Qojahmetova A. (2015) «Körkem мәтін pragmatikasyn korpustyq lingvistika keñistiginde zertteu» // QazÜU habarshysy. Filologiya seriasy. №6 (158), – 300-305 b. [The study of literary text pragmatics in the space of Corpus Linguistics KazNU bulletin Philology series. No. 6 (158), – pp. 300-305.] (in Kazakh)
 Sharov S.A. (2003) Predstavitelnyj korpus russkogo yazyka v kontekste mirovogo opyta. // Nauchno tekhnicheskaya informaciya. – Seriya 2 Informasionnye processy i sistemy, № 6, S.10. [Representative corpus of the Russian language in the context of world experience Scientific and Technical information - Series 2 Information Processes and Systems, No. 6, p. 10.] (in Russian)